UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

ESTUDIO DE MINERÍA DE DATOS PARA LA INFORMACIÓN DE MORTALIDAD EN MÉXICO

TESIS

Para obtener el título de:
Ingeniero en Computación

PRESENTA:

Paulina Galván Castro
Alejandra Meza Mendoza

Directores de Tesis:
Ing. Gabriela Betzabé Lizárraga Ramírez
Ing. José Enrique Larios Canale

Ciudad Universitaria, 2012

Nunca desistas de un sueño.
Sólo trata de ver las señales que te lleven a él.

Agradezco a dios por
poderme permitir llegar
a la culminación
de tanto trabajo y dedicación.

Agradezco a mis padres
todo su apoyo desde que
comencé mis estudios hasta
el día de hoy. Ya que sin ellos
no sería la persona que soy.

Agradezco a mi abuelo por su
cariño y apoyo incondicional.

Agradezco a mi compañero de toda
la vida, mi hermano, por ser mi amigo .

Agradezco a la Ing. Betzabé
por su paciencia y tiempo.

Agradezco a mi universidad, que dejo
grabado en mi el sello puma por siempre.

Por mi raza hablará el espíritu
P A U

A ti dios por darme la fuerza y el coraje necesario para hacer este sueño realidad, por estar conmigo en cada momento de mi vida. Por cada regalo de gracia que me has dado y que inmerecidamente eh recibido.

A mis padres a quienes amo profundamente, por su cariño, por estar ahí siempre al pie del cañón y por creer en mí.

Al ángel más hermoso que dios me pudo prestar, mi hermano gracias por siempre creer en mí por todo tu amor y porque con tu ejemplo llegue a este momento que juntos esperamos te amo.

A Mary mi amiga incondicional gracias por la paciencia y por siempre estar ahí para levantarme o para celebrar juntas lo triunfos sin ti no lo hubiera logrado. A Gaby mi hermanita gracias por tus buenos concejos por enseñarme a ver la vida de otra forma, gracias a las dos por su cariño.

A mis abuelitos por su cariño y principalmente a mi abuelito Salomón por enseñarme que la verdadera felicidad se encuentra en lo sencillo, te quiero mucho.

A la profesora Ing. Betzabe por su gran apoyo, paciencia y amistad. A la universidad que me forma como profesionista y a la que llevare siempre en el corazón.

Aprender sin pensar es inútil. Pensar sin aprender, peligroso. (Confucio)

ALE

ÍNDICE

OBJETIVO

Hacer uso de los diversos métodos y tareas de la minería de datos, con el fin de extraer información útil con la información de Mortalidad en México la cual fue proporcionada por el INEGI. Esta información será utilizada para realizar estudios de diferentes aspectos de las defunciones de la población en México, como identificar las instituciones de salud menos eficientes, ocupaciones peligrosas, la violencia familiar como causa de mortalidad y los embarazos de alto riesgo, con la finalidad de obtener reglas o patrones sobre las características de la defunción, con el objetivo detectar grupos más vulnerables.

INTRODUCCIÓN

La Minería de Datos es usada con el fin de extraer patrones, de describir tendencias, predecir conceptos y en general de obtener información útil, novedosa y principalmente desconocida, que generalmente se encuentra en la información que tenemos a nuestro alrededor y que esta almacenada de forma heterogénea.

Es de suma importancia que se entienda a la perfección que es la minería de datos, que nos ofrece, cuales son las herramientas que necesitamos, como se trabaja con ella y muchísimas cosas más por lo cual este trabajo consta de 3 capítulos teóricos para después en un cuarto capítulo apliquemos toda esta teoría en casos reales y prácticos.

Se tiene la situación de que día a día hay una gran cantidad de personas que fallecen por diferentes razones. Para atender esta situación se han realizado diferentes estudios estadísticos en donde tratamos de ver casos muy precisos, para ser concretos se manejan 4 grandes temas de la mortalidad en México.

La Minería de Datos no ayudara a darle un nuevo enfoque a esos estudios estadísticos con el fin de encontrar patrones de comportamiento de la ocupación más peligrosa, de la institución pública o privada de salud menos eficiente, de la violencia familiar ó del embarazo de alto riego.

Sera bueno comentar un poco más de cada una de los temas que se analizarán posteriormente. Por mencionar un tema para la ocupación peligrosa, en principio se busca en número de personas que hayan fallecido en su trabajo y esto para las diferentes ocupaciones, una vez obteniendo este resultado escogemos un algoritmo de minería de datos que nos arroje las posibles ocupaciones de dicho problema.

Para la institución menos eficiente se busca obtener el número de personas que murieron por institución en el año 2000 y el 2009, en este caso se escogerá otro

algoritmo de minería de datos que nos muestre que institución es la menos eficiente en cada año.

La violencia familiar es un tema muy importante ya que este es un tema que se ha tratado de evitar de varias formas a lo largo de los años, por lo cual para este estudio en principio se busca tener el número de personas que fallecen a causa de violencia familiar, tomando en cuenta aspectos constantes de dichas personas, una vez que se obtengan estos resultados, se aplicara otro algoritmo de minería de datos diferente al de los anteriores para que arroje el riesgo de que sufran violencia familiar ciertas personas en años posteriores.

Las mujeres que se embarazan a edades tempranas o avanzadas, estos casos son considerados como embarazos de alto riegos, por lo cual es un tema de interés, antes que nada se necesita saber cuáles son las edades críticas para estos casos, al igual que el número de personas que fallecen con esa edades en condición de embarazadas, obteniendo esto se escogerá el ultimo algoritmo de minería de datos que pueda mostrar a cierta edad que cantidad aproximada de mujeres van a morir en años posteriores.

Es importante que se escojan algoritmos diferentes en cada tema ya que con esto se pueda ver el comportamiento de cada uno de los algoritmos elegidos en este estudio y que nos ofrece la minería de datos, ya que como se sabe existe una gran cantidad de algoritmos que ayudan a resolver diferentes problemas.

Para llevar a cabo el estudio, se contó con las tablas de los historiales de mortalidad en México de la base de datos del INEGI.

Se sabe que existen muchos factores externos tales como los económicos, sociales, y geográficos, que intervienen en los temas descritos anteriormente, los cuales no se abordan en una totalidad en este trabajo el cual se centra principalmente en poner de relieve el gran potencial que tiene la minería de datos aplicados en diferentes áreas de la sociedad.

CAPÍTULO 1 INTRODUCCIÓN A LA MINERÍA DE DATOS

1.1 Definición de Minería de Datos

La "minería de datos" se crea, por la aparición de nuevas necesidades y especialmente, por el reconocimiento de un nuevo potencial: el valor que tiene la gran cantidad de datos almacenados en los sistemas de información de instituciones, empresas, gobiernos y particulares por ejemplo: nominas, inscripciones, prestamos por mencionar algunos, estos datos son usados diariamente pero rara vez se ocupan con fines analíticos.

Esta información representa transacciones o situaciones que se han producido en varios años, conformando así información histórica, descubriendo una necesidad de analizar los datos para la obtención de información útil para la organización, permitiendo obtener conocimiento novedoso para la toma de decisiones es decir, información que nos ayudara a explicar el pasado, entender el presente y predecir la información futura, es decir la minería de datos se presenta como una estrategia de análisis de la información, siendo la materia prima "los datos" que hay que explotar para obtener el verdadero "producto elaborado" el conocimiento.

Así esta primera parte de la tesis está enfocada a conocer qué es la "minería de datos" dentro del contexto de la "extracción del conocimiento", sus características y su relación con otras disciplinas.

Definiciones de minería de datos:

"Proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos" (Hernández, O., Ramirez, M. y Ferri, C., 2000).

"Un mecanismo de explotación, consistente en la búsqueda de información valiosa en grandes volúmenes de datos" (Hernández, O., Ramirez, M. y Ferri, C., 2000).

Se pueden identificar características importantes: trabaja con grandes volúmenes de datos, procedentes de sistemas de información, usando técnicas adecuadas para analizar los mismos sistemas y extraer conocimiento novedoso. Podríamos decir que el objetivo de la minería de datos es convertir datos en conocimiento. Se preocupa por el análisis de los datos y el uso de técnicas de software para encontrar patrones y regularidades.

1.2 Tareas

Pero, ¿Cómo es que se puede representar el conocimiento de la minería de datos? Puede ser en forma de relaciones y patrones previamente desconocidos, o bien en forma de una descripción más concisa es decir, un resumen de los mismos. Estas relaciones o

resúmenes constituyen la tarea de los datos analizados. Existen muchas formas de representar las tareas y cada una de ellas determina el tipo de técnica que puede usarse.

En la práctica, las tareas pueden ser de dos tipos: *predictivas* y *descriptivas.*

Tareas predictivas: pretenden estimar valores futuros o desconocidos de variables de interés, que denominamos *variables objetivo* o *dependientes,* usando otras variables o campos de la base de datos, a las que nos referiremos como *variables independientes* o *predictivas (Hernández, O., Ramirez, M. y Ferri, C., 2000).*

Tareas descriptivas: identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos (Hernández, O., Ramirez, M. y Ferri, C., 2000).

1.3 Tipos de Datos

Hablaremos ahora precisamente sobre los datos, que es la materia prima de la minería, conocer ¿a qué tipo de datos puede aplicarse la minería de datos?, conociendo en esta parte algunos tipos de datos.

Se pueden encontrar en diversas formas, dependiendo de la manera en que se encuentren almacenados y la forma de usarlos. Principalmente se dividen en 3 formas de datos:

Datos Relacionales

La obtención de información desde una base de datos relacional se ha resuelto tradicionalmente a través de lenguajes de consulta especialmente diseñados para ello, como lenguaje de consulta estructurado (SQL, Structured Query Language). La Figura 1.1 muestra una consulta típica sobre la tabla de mortalidad en México en el año 2004 que lista el promedio de edad de mortalidad en la entidad federativa que es Durango agrupados por sexo.

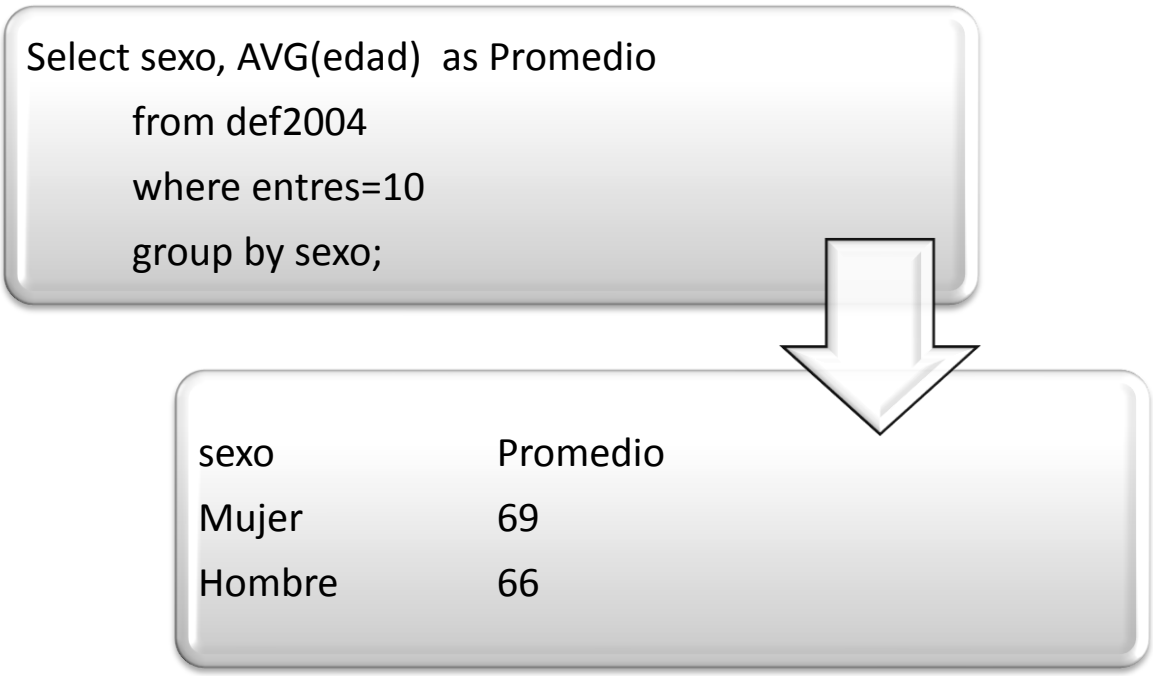


Figura 1.1 Consulta típica

Mediante una consulta podemos combinar en una sola tabla o *vista minable* aquella información de varias tablas que requiramos para cada tarea concreta de minería de datos.

Otros tipos de bases de datos

Bases de datos espaciales: contienen información relacionada con el espacio físico por ejemplo: una ciudad, una región montañosa. Incluyen datos geográficos, redes de transporte o información de tráfico, etc., donde las relaciones espaciales son muy relevantes. La minería de datos sobre estas bases es muy relevante permite encontrar patrones como por ejemplo las características de las casas en una zona montañosa.

Bases de datos temporales: almacenan datos relacionados con el tiempo, los atributos pueden referirse a distintos instantes o intervalos de tiempo, se pueden encontrar tendencias de cambio y características de evolución.

Bases de datos documentales: contienen descripciones de los objetos desde palabras claves hasta resúmenes de ese objeto, pueden contener documentos como una biblioteca digital o una base de datos de fichas bibliográficas.

Bases de datos multimedia: almacenan imágenes, audio y vídeo. Soportan objetos de gran tamaño ya que por ejemplo los videos pueden tener capacidades grandes y se necesita gran capacidad para su almacenamiento.

La World Wide Web

Repositorio de información más grande y diversa de los existentes en la actualidad, hay gran cantidad de datos en la web de donde se puede extraer información útil. Esta no es una tarea fácil ya que los datos multimedia como texto, imágenes, video y audio pueden estar en diferentes servidores.

1.4 El KDD y la minería de datos

En el momento que el usuario les atribuye algún significado a los datos se convierten en información. Cuando los especialistas elaboran o encuentran un modelo, haciendo que la interpretación de la información y ese modelo representen un valor agregado, entonces nos referimos al conocimiento.

El Descubrimiento de Conocimiento en Bases de Datos (KDD**,** Knowledge Discovery in Databases**)** apunta a procesar automáticamente grandes cantidades de datos para encontrar conocimiento útil en ellos, de esta manera permitirá al usuario el uso de esta información valiosa para su conveniencia.

El *KDD* se puede definir como "el proceso de identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de los datos" (Hernández, O., Ramirez, M. y Ferri, C., 2000). Teniendo como propiedades:

Válido: los patrones deben ser precisos para datos nuevos, y no sólo para aquellos que han sido usados en su obtención.

Novedoso: tienen que aportar algo desconocido tanto para el sistema como para el usuario.

Potencialmente útil: la información debe conducir a algún beneficio para el usuario.

Comprensible: fácil comprensión para el usuario.

Es un proceso complejo que incluye la selección, limpieza, transformación y proyección de los datos, la obtención de los modelos o patrones (el objetivo de la minería de datos), también la evaluación y posible interpretación de los mismos, tal y como se refleja en la Figura 1.2.

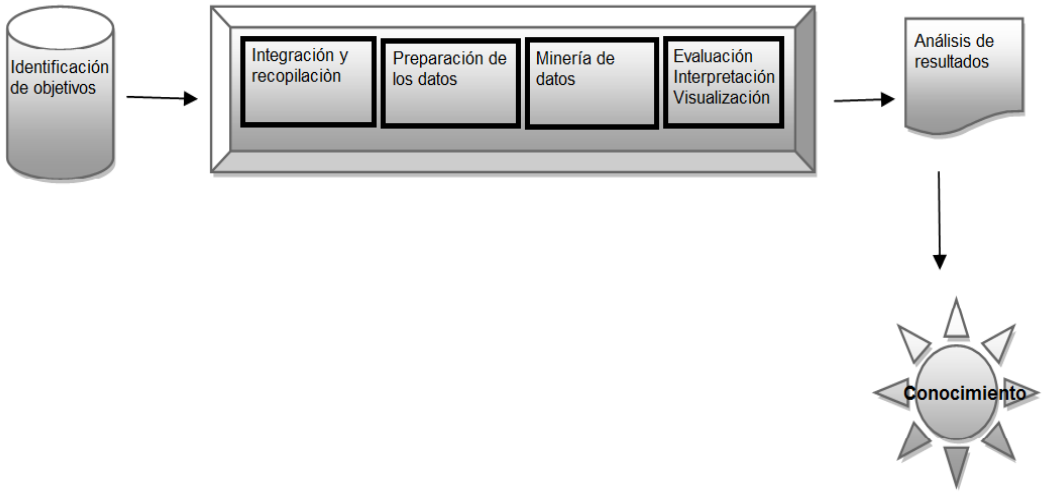


Figura 1.2 Proceso KDD

Por lo que podemos concluir que el KDD es el proceso global de descubrir conocimiento útil desde las bases de datos mientras que la minería de datos se refiere a la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos.

1.5 Relación con otras disciplinas

La minería de datos se ha desarrollado junto con otras áreas. Dentro de las disciplinas más influyentes se pueden destacar:

1.-**Bases de datos:** tecnologías de bases de datos y depósitos de datos, maneras eficientes de almacenar, accesar y manipular datos.

2.-**Visualización de datos**: interfaz entre humanos y datos, y entre humanos y patrones.

3.-**Estadística**: conceptos como la media, la varianza, las distribuciones, la regresión lineal y no lineal.

4.-**Computación paralela**: cómputo de alto desempeño, mejora de desempeño de algoritmos debido a su complejidad y a la cantidad de datos.

5.-**Recuperación de información**: obtener información sobre datos textuales por ejemplo los buscadores que existen en internet en los cuales con poner la palabra exacta se realiza un barrido de resultados.

6.-**Aprendizaje Automático**: área de la Inteligencia Artificial que se encarga de desarrollar algoritmos capaces de aprender, es decir la máquina aprende un modelo a partir de ejemplos.

7.-**Sistemas para la toma de decisiones**: herramientas que nos ayudan en la resolución de problemas, proporcionan información necesaria para una buena decisión, por ejemplo una herramienta es los árboles de decisión.

Figura 1.3 Relación con otras diciplinas

1.6 Aplicaciones

La minería de datos puede usarse en diversos ámbitos ya que al trabajar con cantidades grandes de datos y darles una utilidad benéfica para el negocio resulta atractivo para empresas, industria, medicina, administración entre otros. Se ha encontrado un enorme potencial para resolver problemas reales.

Algunos ejemplos reales de la aplicación de la minería de datos son:

Sector industria

- Optimización de Centrales Eléctricas
- Control de Trenes de Laminado en la Industria del Acero
- Optimización de Altos Hornos
- Gestión de Alarmas en Plantas Petroquímicas
- Control de Calidad en la Fabricación de Electrodomésticos
- Optimización del Proceso de Producción de Cemento
- Control de Calidad de Materiales Fabricados Industrialmente

Sector sanitario y farmacéutico

- Predicción de Ventas de Productos Farmacéuticos
- Diagnóstico de Accidentes Cerebro-Vasculares Agudos
- Supervisión de Calidad del precultivo en el Cultivo Industrial de Antibióticos
- Supervisión de la Evolución de Cultivos en la Fabricación de Antibióticos

Administración Pública y servicios

- Análisis y Control de Tráfico de Vehículos
- Predicción de Demanda de Tiempos de Trabajo para Reparto Postal
- Predicción de Flujos de Turismo

Sector financiero y seguros

- Estimación de Riesgos en la Concesión de Seguros de Crédito
- Detección y Control de Fraude en el Uso de Tarjetas de Crédito
- Segmentación de Clientes de Entidades Financieras

Podemos detallar más algunos ejemplos:

Predicción de Ventas de Productos Farmacéuticos

En este proyecto de minería de datos el objetivo fue desarrollar un modelo para predecir las ventas de un producto en un determinado mes, basándose en datos sobre las ventas en meses previos.

La empresa Bayer mantiene un registro histórico de diferentes datos, entre ellos las cifras de ventas. Basándose únicamente en los datos de ventas de uno de sus productos, sin indicadores adicionales, pretende desarrollar un modelo del comportamiento de dicho producto en el mercado que le permita predecir las ventas del mismo con cierta anticipación. En concreto, se dispone de las cifras de los últimos 56 meses (Hernández, O., Ramirez, M. y Ferri, C., 2000).

Predicción de Demanda Turística

El objetivo de este proyecto fue desarrollo de un modelo basado en redes neuronales que permita predecir el número de turistas que llegarán a Japón procedentes de EE.UU. Para ello se dispone de la evolución histórica de estas cifras desde enero de 1978 hasta septiembre de 1998, además de algunos indicadores nacionales de carácter económico como los ingresos o el volumen de importaciones y exportaciones.

Aplicando técnicas de minería de datos, se han desarrollado un conjunto de modelos, basados en redes neuronales, que permiten predecir las llegadas turísticas con un mes, o incluso un año, de anticipación a partir de las cifras de viajes de los meses anteriores.

CAPÍTULO 2 FASES DEL KDD (PROCESO DE EXTRACCIÓN DE CONOCIMIENTO)

Todo análisis requiere cierta metodología o pasos a seguir y, como ya se mencionó en el capítulo anterior, la Minería de Datos consta de fases que se realizan de una forma secuencial, hasta no haber concluido una, no se pasa a la siguiente; lo cual da cierta seguridad de que el resultado que se obtiene en este proceso sea auténtico y que la información obtenida en él sea verdaderamente útil para nosotros.

El Descubrimiento de Conocimiento en Bases de Datos (KDD Knowledge Discovery in Databases) es un proceso iterativo como se muestra en la figura 2.1, ya que la salida de alguna de las fases puede hacer volver a pasos anteriores, y porque a menudo son necesarias varias iteraciones para extraer el conocimiento de alta calidad. Es iterativo porque el usuario, y más generalmente un experto en el dominio del problema debe de ayudar en la preparación de los datos, validación del conocimiento extraído, etc. (Hernández, O., Ramirez, M. y Ferri, C., 2000).

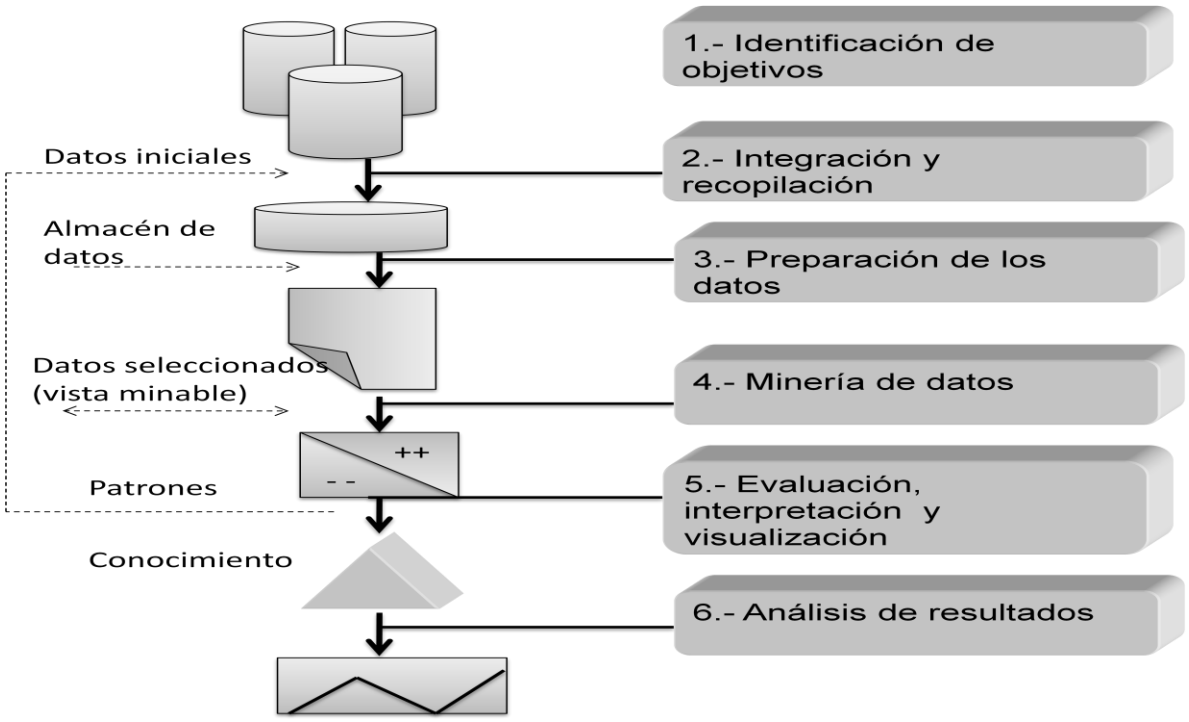


Figura 2.1 Fases de proceso del descubrimiento en bases de datos KDD

2.1 Fase de identificación de objetivos

Es una fase previa de análisis de las necesidades de la organización y definición del problema, en la que se establecen los objetivos de minería de datos.

El sistema de información es indispensable para empezar con el análisis, ya que de éste es donde se analizan los requisitos; define el ámbito del problema, define las métricas por las que se evaluará el modelo y define el objetivo final del proyecto de minería de datos. Estas tareas se traducen en preguntas como las siguientes:

¿Qué se está buscando?

¿Qué atributo del conjunto de datos se desea intentar predecir?

¿Qué tipos de relaciones se intenta buscar?

¿Desea realizar predicciones a partir del modelo de Minería de Datos o sólo buscar asociaciones y patrones interesantes?

¿Cómo se distribuyen los datos?

¿Se cuenta con un diagrama Entidad-Relación (DER) para saber cómo se relacionan las tablas?

La naturaleza de las respuestas será quien determine el tipo de Minería de Datos que se deba aplicar, así como la tecnología adecuada.

Por otra parte, para responder a estas preguntas, es probable que se deba dirigir un estudio de disponibilidad de datos para investigar las necesidades de los usuarios de la empresa con respecto a los datos disponibles. Si los datos no son compatibles con las necesidades de los usuarios, puede que se deba definir el proyecto nuevamente.

2.2 Fase de integración y recopilación

En la fase de integración y recopilación de datos se determina las fuentes de información que pueden ser útiles y dónde conseguirlas.

Recopilación de datos

Para analizar y extraer algo útil de los datos es necesario disponer de ellos. Esto en algunos casos puede parecer simple. Se parte de un archivo de datos a analizar. En otros, la diversidad y tamaño de las fuentes hace que el proceso de recopilación de datos sea una tarea compleja. En general, el problema de reunir un conjunto de datos que posibilite la extracción del conocimiento requiere decidir de qué fuentes, internas y externas, se van a obtener los datos; cómo se van a organizar y, finalmente, de qué forma se van a extraer.

El reconocimiento de los datos se refiere a la exploración de los mismos, pero en esta ocasión de manera más detallada.

Se debe conocer el negocio y el tipo de decisiones que toman los directivos, los aspectos que son cruciales en su negocio, las reglas y políticas que se utilizan, qué decisiones son críticas para la organización, qué conocimiento se requiere, entre otros aspectos.

Existen varias razones para las cuales es conveniente trabajar sobre una vista minable y no directamente sobre la base de datos, la primera de ellas es que una base de datos contiene varias tablas diferentes, por lo que la herramienta de minería de datos no tendría claro con que información trabajar y la segunda es que la mayoría de los métodos de minería de datos solo tratan con una única tabla.

Cuando tenemos todos los datos lo primero que podemos realizar es un resumen de las características de atributos. En este tipo de tablas se muestra las características generales de los atributos.

Al tener los datos integrados, una de las primeras actividades que se deben realizar es la descripción de los datos.

En esta etapa lo recomendable es: crear un diccionario de datos con los atributos de las tablas de la base que ha sido creada. Algunos puntos importantes son la tabla de donde provienen, el tipo de dato, el número total de registros, el número de valores nulos que contiene, entre otros.

En el caso de la visualización de los datos, se puede utilizar una herramienta bastante sencilla como los histogramas, que muestran la distribución de los valores de los atributos. Los histogramas ayudan a entender mejor la distribución de los datos.

Otra herramienta son las gráficas de dispersión que son especialmente útiles cuando los atributos a graficar son numéricos.

Por último, también se puede utilizar los diagramas de pareto, son gráficas en donde se organiza diversas clasificaciones de datos por orden descendente, de izquierda a derecha por medio de barras, de modo que se pueda signar un orden de prioridades. El análisis de pareto es una técnica que separa los pocos vitales de los muchos triviales, y se utiliza para separar gráficamente los aspectos significativos de un problema de los triviales para saber dónde dirigir los esfuerzos.

Integración

La integración es generalmente un proceso que se realiza durante la recopilación de datos y si se realiza un almacén de datos, durante el proceso de carga, mediante el sistema ETL (Extract, Transform and Load / Extraer, transformar y cargar).

Es importante mencionar que los datos se pueden obtener de distintas fuentes, véase la *figura 2.2* El primer problema a la hora de realizar una integración de distintas fuentes

de datos es identificar los objetos, es decir, conseguir que datos sobre el mismo objeto se unifiquen y datos de diferentes objetos permanezcan separados. Este problema se conoce como el problema del esclarecimiento de la identidad.

Existen dos tipos de errores que pueden ocurrir en esta integración:

Dos o más objetos diferentes se unifican: los datos resultantes mezclarán patrones de diferentes individuos y será un problema para extraer conocimiento. Esto será más grave cuanto más diferentes sean los objetos unificados.

Dos o más fuentes de objetos iguales se dejan separadas: los patrones del mismo individuo aparecerán repartidos entre varios individuos parciales. Este problema genera menos "ruido" que el anterior, aunque es especialmente problemático cuando se usan valores agregados.

En general el primer problema es menos frecuente que el segundo, ya que la unificación se realiza generalmente por identificadores externos en la base de datos, como por ejemplo: número de identidad, número de pólizas, matrículas, tarjeta de crédito o de fidelización, etc

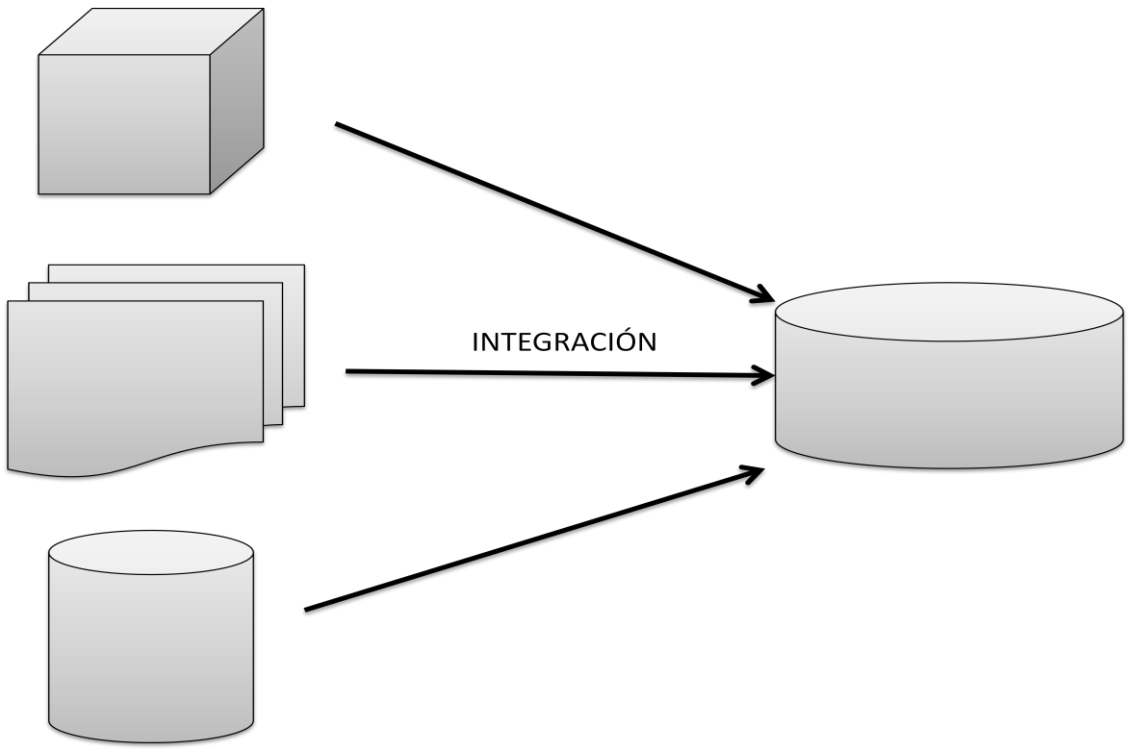


Figura 2.2 Integración de los datos

2.3 Fase de preparación de datos

Dado que los datos provienen de diferentes fuentes, pueden contener valores erróneos o faltantes. Estas situaciones se tratan en esta fase, en la que se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Además se proyectan para considerar únicamente aquellas variables o atributos que va a ser relevantes, con el objetivo de hacer más fácil la tarea propia de minería y para que los resultados de la misma sean más útiles. La selección incluye tanto una cifra o función horizontal (filas / registros), como vertical (columnas / atributos).

Este paso, del proceso de Minería de Datos consiste en recopilar, limpiar, transformar, explorar y seleccionar los datos que se pudieron identificar al definir el problema.

Desafortunadamente, la fase de selección, exploración y transformación de variables ha sido a la que menos importancia se le ha dado en la bibliografía, por ser una fase de enorme dificultad en la que los datos se analizan y exploran pero no se obtienen resultados definitivos. **La fase de preparación representa la clave del éxito de un proyecto de Minería de Datos**. Puede ser la diferencia entre el éxito y el fracaso, la diferencia entre resultados provechosos, la diferencia entre predicciones interesantes y averiguaciones absurdas.

Selección

La construcción de la vista minable es indispensable para que la herramienta de minería de datos pueda digerir un conjunto de datos y producir algo razonable. Cada herramienta recibe los datos de una manera diferente, por lo que es necesario prestar atención a este punto.

Un aspecto, de igual manera relevante en el proceso de extracción de conocimiento, es el conocimiento del dominio, en general este es un factor muy decisivo tanto en la aplicación del método de minería de datos como en su interpretación.

La siguiente figura esquematiza el proceso en el que se interrelacionan los datos, el conocimiento del dominio y de los usuarios.

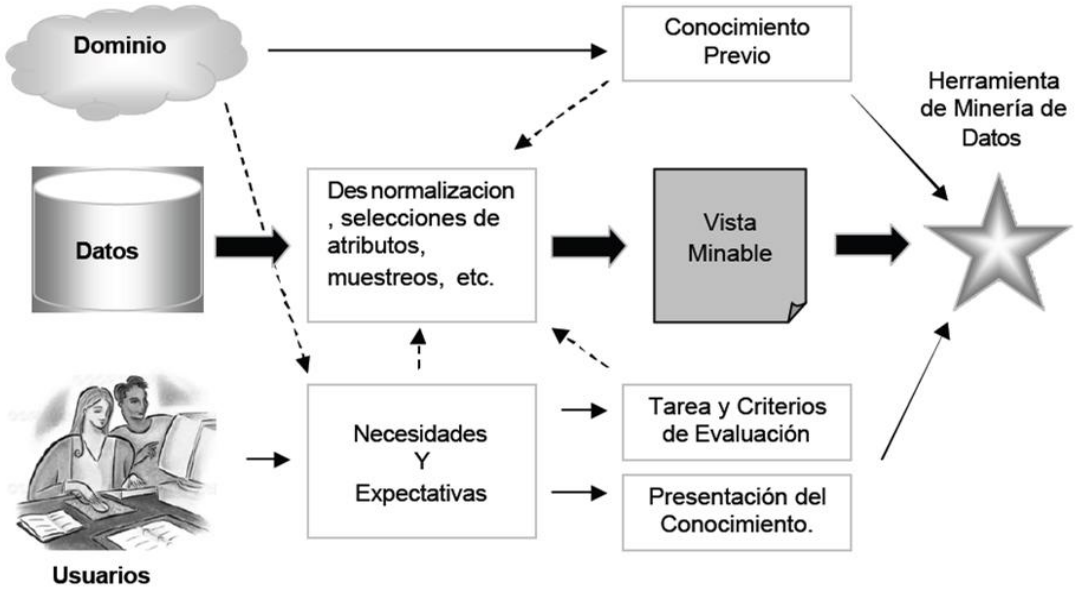


Figura 2.3 Construcción de una vista minable

Como se puede observar en la figura anterior, el obtener una vista minable implica que va acompañada de la tarea a realizar sobre ella y cómo evaluarla, así como la forma de presentar el resultado final, y en su caso, el conocimiento previo necesario.

Revisemos cada uno:

Vista Minable. Representa la parte de los datos que es pertinente analizar. Una vista minable consiste en una vista en el sentido más clásico de base de datos: una tabla. La mayoría de los métodos de minería de datos, son solo capaces de tratar con vistas minables. Por tanto una vista minable ha de recoger solamente la información necesaria para realizar la tarea de minería de datos.

Tarea, método y presentación. Para poder definir qué tipo de conocimiento se desea extraer y como se debe presentar, se bebe definir la tarea a utilizar, las entradas y salidas, con que método, y la manera en cómo se van a presentar los datos.

Calidad. El conocimiento extraído debe ser válido, novedoso e interesante, en otras palabras debe ser de calidad. En muchos casos hay que establecer ciertos criterios de comprensibilidad de los métodos, criterios de fiabilidad, criterios de utilidad y criterios de novedad o interés.

Conocimiento Previo. Cuando se emprende un programa de minería de datos es importante contar con el conocimiento de dominio, tanto a la hora de construir la vista minable como para ayudar al propio algoritmo de minería de datos.

Limpieza

La limpieza de datos (data cleaning / cleansing) puede, en muchos casos, detectar y solucionar problemas de datos no resueltos durante la integración, como los valores anómalos.

La inconsistencia y los valores nulos existen en casi todas las bases de datos. Los datos inconsistentes se ocasionan por distintas razones, como puede ser que los atributos de interés no están siempre disponibles o la información que se tiene es errónea.

Otros datos no se tienen almacenados porque al momento de introducir los datos se pensaba que no eran de interés.

Es por ello que la rutina de la limpieza de los datos se vuelve una pieza fundamental en dicho proceso, ya que ésta ayuda a rellenar valores nulos y va resolviendo las inconsistencias. Los datos sin limpiar pueden ocasionar confusiones para los procedimientos de análisis pudiendo entonces generar un modelo erróneo. En la *figura 2.4* se simula como se encuentran las bases de datos, llenas de basura y como deben de quedar tras la limpieza.

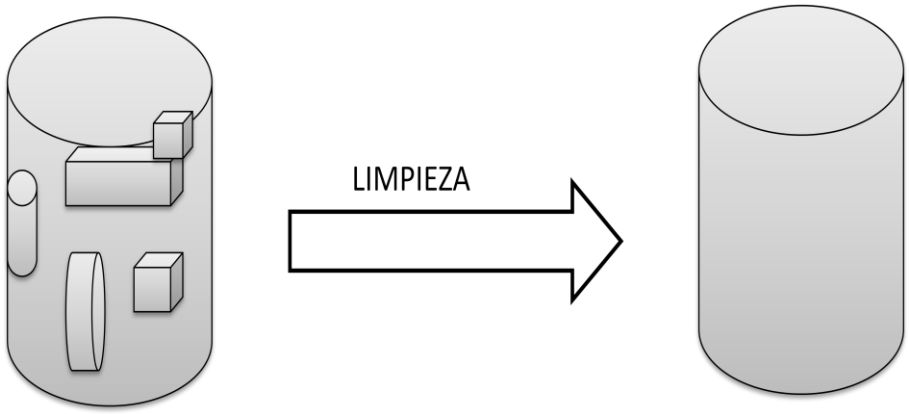


Figura 2.4 Limpieza de las bases de datos

2.4 Fase de minería de datos

En la fase minería de datos, se decide cual es la tarea a realizar (clasificar, agrupar, etc.) y se elige el método que se va a utilizar.

Esta fase es la más característica de KDD y, por esta razón, muchas veces se utiliza esta fase para nombrar todo el proceso. El objetivo de esta fase es producir nuevo conocimiento que puede utilizar el usuario. Esto se realiza construyendo un modelo el cual es una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos para explicar situaciones pasadas.

Para ello es necesario tomar una serie de decisiones antes de empezar el proceso:

Determina qué tipo de tarea de minería es el más apropiado

Elegir el tipo de modelo

Elegir el algoritmo de minería que resuelve la tarea y obtenga el tipo de modelo que estamos buscando. Esta elección es permitente porque existen muchos para construir los modelos.

Continuando con la explicación de la fase de Minería de Datos a nivel general, se dice comúnmente que este proceso de Minería de Datos convierte datos en conocimiento, tal cual alquimista pudiera convertir espigas de trigo en lingotes de oro, o como un minero puede obtener metales preciosos de un montón de rocas. También para algunos autores o estudiosos de dicha materia llegan a decir que el objetivo es extraer "verdad a partir de basura".

Si se pudiera referir al contexto de trabajo de las fases anteriores del *KDD*, en donde se prepararon los datos para al fin aplicar una técnica de minería de datos, se podrían representar en la siguiente figura 2.5.

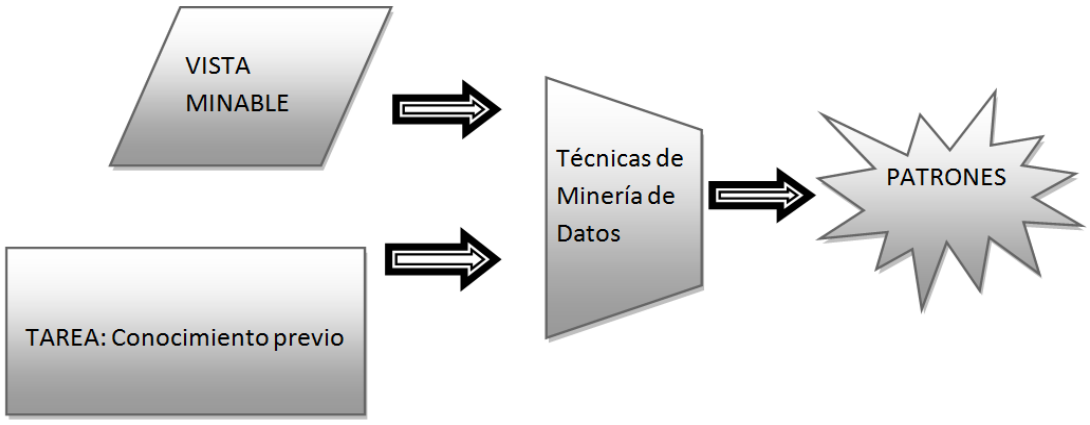


Figura 2.5 Proceso ideal de minería de datos

En la figura 2.5 las técnicas de Minería de Datos aparecen como una especie de colador que, al introducirle los datos junto con criterios asociados, descubre patrones de comportamiento o reglas.

Tareas de la minería de datos

Las distintas tareas pueden ser predictivas o descriptivas. Entre las tareas predictivas encontramos 1.1 la clasificación y 1.2 la regresión, mientras que 1.3 el agrupamiento (clustering), 1.4 las reglas de asociación, 1.5 las reglas de asociación secuenciales y1.6 las correlaciones son tareas descriptivas. Veamos en mayor detalle todas ellas.

La clasificación es quizá la tarea más utilizada. En ella cada instancia pertenece a una clase, la cual se indica mediante el valor de un atributo que llamamos la clase de la instancia. Este atributo puede tomar diferentes valores discretos, cada uno de los cuales corresponde a una clase. El resto de los atributos de la instancia se utilizan para predecir la clase. El objetivo del algoritmo es maximizar la razón de precisión de la clasificación de las nuevas instancias, la cual se calcula como el cociente entre las predicciones correctas y el número total de predicciones. Existen varias de las tareas de la clasificación, como son el aprendizaje de "rankings", el aprendizaje de preferencias, el aprendizaje de estimadores de probabilidad, etc.

La regresión es también una tarea predictiva que consiste en aprender una función real que se asigna a cada instancia a un valor real. La principal diferencia respecto a la clasificación; el valor a predecir es numérico. El objetivo es minimizar el error entre el valor predicho y el valor real.

El agrupamiento (clustering) es la tarea descriptiva por excelencia y consiste en obtener grupos "naturales" a partir de los datos (Hernández, O., Ramirez, M. y Ferri, C., 2000). Hablamos de grupos y no de clases por que a diferencia de la clasificación, en lugar de analizar datos etiquetados con una clase, los analiza para generar está etiqueta. Los datos son agrupados basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizado la similitud entre los distintos grupos. Es decir se forman grupos tales que los objetos de un mismo grupo son muy similares entre sí y, al mismo tiempo, son muy diferentes a los objetos de otro grupo. Al agrupamiento también se le suele llamar segmentación, ya que parte o segmenta los datos en grupos que pueden ser o no disjuntos. El agrupamiento está muy relacionado con la sumarización, que algunos autores consideran una tarea en sí misma, en la que cada grupo formado se considera como un resumen de los elementos que la forman.

Las correlaciones son una tarea descriptiva que se usa para examinar el grado de similitud de los valores de dos variables numéricas. Una fórmula estándar para medir la correlación lineal es el coeficiente de correlación r, el cual es un valor real comprendido entre -1 y 1. Si r es 1 (respectivamente -1). Las variables están perfectamente correlacionadas valores numéricos (perfectamente correlacionadas negativamente), mientras que si es 0 no hay correlación. Esto quiere decir que cuando r es positiva, las variables tienen un comportamiento similar (ambas crecen o decrecen al mismo tiempo) y cuando r es negativa si una variable crece la otra decrece. El análisis de correlaciones, sobre todo las negativas, puede ser muy útil para establecer reglas de ítems correlacionados.

Las reglas de asociación son también una tarea descriptiva, muy similar a las correlaciones, que tienen como objetivo identificar relaciones no explícitas entre atributos categóricos. Pueden ser de muchas formas, aunque la formulación más común es del estilo “si el atributo X toma el valor d entonces el atributo Y toma el valor de b”. Las reglas de asociación no implican una relación causa-efecto, es decir, puede no existir una causa para que los datos estén asociados.

Un caso especial de reglas de asociación, que recibe el nombre de reglas de asociación secundarias, se usa para determinar patrones secuenciales en los datos. Estos patrones se basan en secuencias temporales de acciones y difieren de las reglas de asociación en quelas relaciones entre los datos se basan en el tiempo.

2.5 Fase de evaluación, interpretación y visualización

Medir la calidad de los patrones descubiertos por un algoritmo de minería de datos no es un problema trivial, ya que esta medida puede atañer a varios criterios, algunos de ellos bastante subjetivo. Idealmente, los patrones descubiertos deben tener tres cualidades: ser precisos, comprensibles e interesantes. Según las aplicaciones pueden interesar mejorar algún criterio y sacrificar ligeramente otro, como en el caso del diagnóstico médico que prefiere patrones comprensibles aunque su precisión no sea muy buena

Evaluación

Los métodos de aprendizaje permiten construir modelos de hipótesis a partir de un conjunto de datos, o evidencias. En la mayoría de los casos es necesario evaluar la calidad de las hipótesis de la manera más exacta posible.

Por ejemplo, si en el ámbito de aplicación de un modelo surge un error en la predicción conlleva a importantes consecuencias (por ejemplo, la detección de células cancerígenas), es importante conocer la exactitud del nivel de precisión de los modelos aprendidos.

La etapa de evaluación de modelos es crucial para la aplicación real de las técnicas de minería de datos.

A continuación se mencionarán las técnicas de evaluación que se utilizarán para evaluar los modelos y reglas o patrones de comportamiento que se obtendrán en esta tesis.

Evaluación mediante validación cruzada

Un mecanismo que permite reducir la dependencia del resultado del experimento en el modo en el cual se realiza la partición, es utilizar validación cruzada (cross-validation). Este método, ver Figura 3.1.5, consiste en dividir el conjunto de evidencia en *k* subconjuntos disjuntos de similar tamaño. Entonces, se aprende una hipótesis utilizando el conjunto disjunto de similar tamaño. Entonces se aprende una hipótesis

utilizando el conjunto formado por la unión de $k-1$ subconjuntos y el conjunto restante se emplea para calcular un error de muestra parcial. Este procedimiento se repite k veces, utilizando siempre un subconjunto diferente para estimar el error de muestra parcial. El error de muestra final se calcula como la medida aritmética de los k errores de muestra parcial. De esta manera, el resultado final recoge la medida de los experimentos con k subconjuntos de prueba independientes.

Como hemos indicado, una de las ventajas de esta técnica es que los k subconjuntos de prueba son independientes. No obstante, esto no sucede en los conjuntos de entrenamiento.

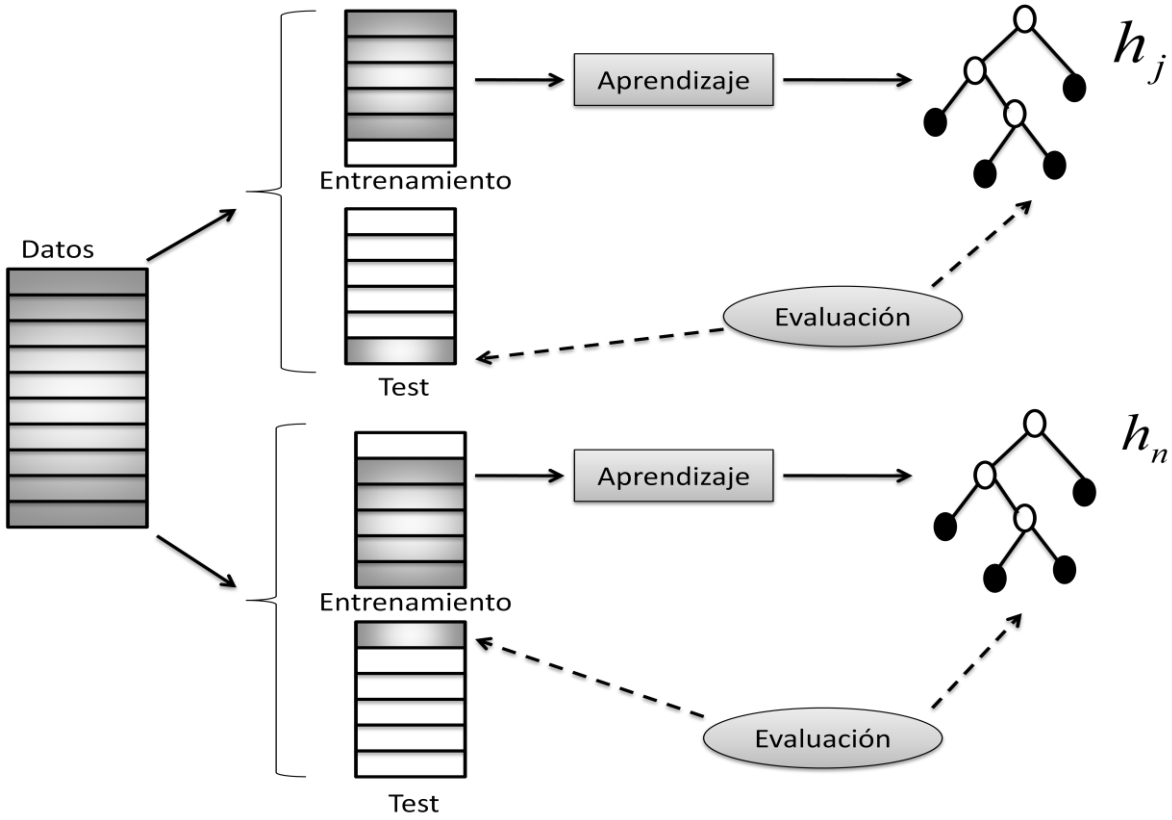


Figura 2.6 Evaluación mediante validación cruzada

Análisis ROC

Por desgracia, no siempre se dispone de una matriz de costes que permita estimar adaptar el aprendizaje a ese determinado contexto de costes. Muchas veces, la matriz de costes sólo conoce durante el tiempo de aplicación y no durante el aprendizaje, generalmente porque los costes varían frecuentemente o son dependientes del contexto.

En esta situación, lo que suele hacer es aprender un conjunto de clasificadores y seleccionar el que mejor se comporte para unas circunstancias o contextos de costes

determinados a posteriori. Para ello, la técnica denominada análisis de ROC (Receiver Operating Characteristc) provee herramientas que permiten seleccionar el subconjunto de clasificadores que tiene un comportamiento óptimo en general. Asimismo, el análisis ROC permite evaluar clasificadores de manera más independiente y completa a la clásica precisión.

El análisis ROC se utiliza normalmente para problemas de dos clases (se suelen denominar clase positiva y clase negativa), y este tipo de problemas se utiliza la siguiente notación para la matriz de confusión:

Estimado	Real	
	True Positives(TP)	False Positives (FP)
	False Negatives (FN)	True Negatives (TN)

Figura 2.7 Notación

Ejemplo de Análisis ROC:

Tenemos la matriz de de confusión/contingencia:

		Real	
Pred		Abrir	Cerrar
	Abrir	400	12000
	Cerrar	100	87500

		Real	
Pred		Abrir	Cerrar
	Abrir	0.8	0.12
	Cerrar	0.2	0.87

Figura 2.8 Ejemplo Análisis ROC

Se normaliza la matriz de confusión

True Positive Rate: TPR = TP / (TP + FN). (“*recall*” o ”*sensitivity*” o *“positive accuracy”)*

TPR=400 / (400+100) = 400/500 = 0.8

False Negative Rate: FNR = FN / (TP + FN). (“*positive error*”)

FNR = 100 / (400+100) = 100/500 = 0.2

True Negative Rate: TNR = TN / (TN + FP). (”*specificity*” o ”*negative accuracy*”).

TNR=87500 / (87500+12000) = 0.87

False Positive Rate: FPR = FP / (TN + FP). (“*negative error*”)

FNR=12000 / (87500+12000) = 0.12

Se grafica en el espacio ROC

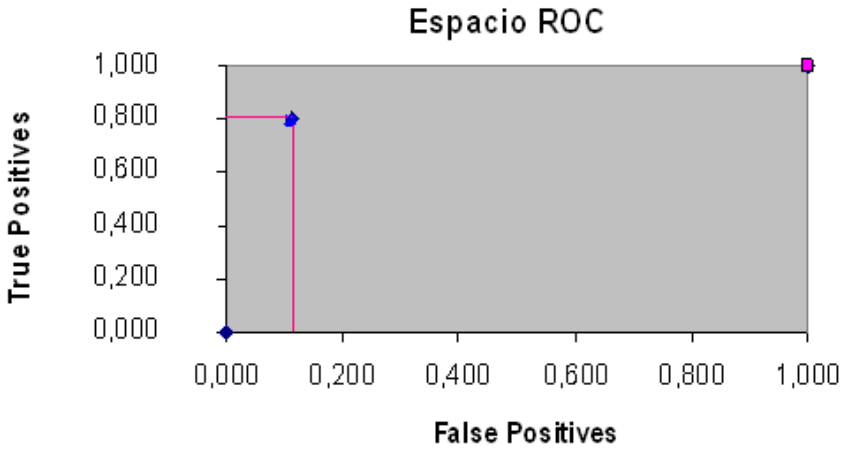


Figura 2.9 Gráfica en el espacio ROC

Este resultado se compara con nuestro rango de espacio ROC siguiente:

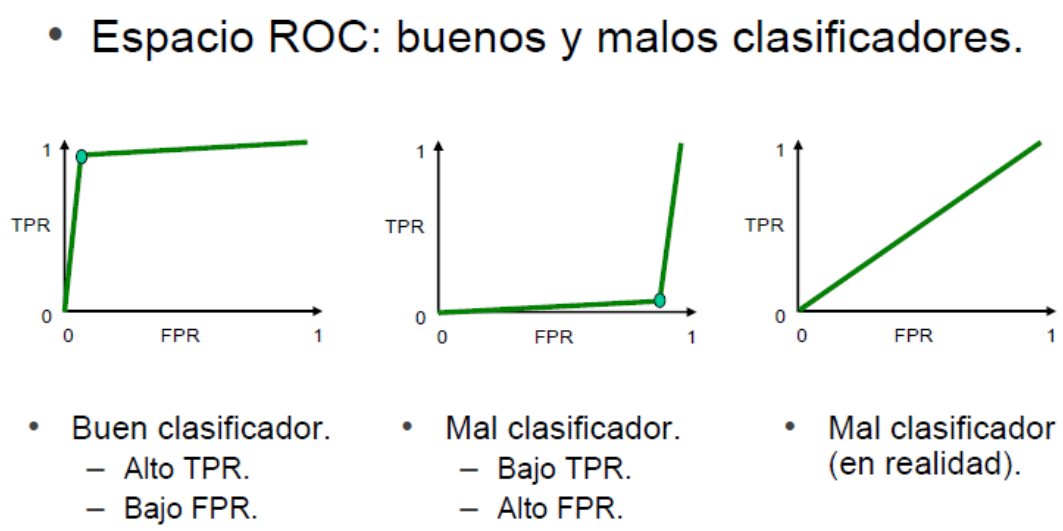


Figura 2.10 Gráficas de clasificadores

Criterios Subjetivos de Evaluación

Las dos técnicas anteriores apuntan a otros criterios que evalúan los modelos tales como:

Interés: es medir la capacidad de ese modelo para suscitar la atención del usuario al modelo.

Novedad: criterio relacionado con la capacidad de un modelo de sorprender al usuario con respecto al conocimiento previo que tenía sobre determinado problema.

Comprensibilidad: la comprensibilidad de un modelo es un factor muy importante y es una cuestión subjetiva desde que un modelo puede ser poco comprensible para un usuario y muy comprensible para otro.

Simplicidad: este criterio se basa en establecer el tamaño o complejidad del modelo. Este criterio está muy relacionado con el criterio de comprensibilidad.

Aplicabilidad: en este caso, la calidad de un modelo se basa en su capacidad de ser utilizado con éxito en el contexto real donde va a ser aplicado.

Interpretación

En esta fase interviene el sentido humano. Si aún no son muy claros los patrones generados anteriormente, se puede llegar a pensar que no se ha obtenido un resultado bueno ya que depende de la visión que tenga el analista para analizarlos y con ayuda de otras herramientas; por ejemplo, las estadísticas o incluso de las bases de datos se podrá tener una mejor visualización de los patrones o reglas generadas.

Visualización

La tarea de minería de datos produce una serie de modelos cuya fácil interpretación por parte del usuario, es clave para el éxito final del proceso de extracción de conocimientos desde base de datos. Con este fin, se han definido diferentes métodos y técnicas que permiten la visualización de los resultados de la etapa de aprendizaje. La visualización de los modelos permite que los usuarios puedan identificar fácilmente y de manera directa los patrones más significativos que ha descubierto el modelo. También permiten representar los modelos junto a los datos.

Asimismo, muchos de los métodos de visualización permiten que los propios usuarios modifiquen los modelos para referirlos o adaptarlos según su conocimiento o circunstancias del ámbito de aplicación.

Gran parte de los métodos de minería de datos producen modelos que pueden ser expresados con lenguaje natural o bien mediante expresiones matemáticas. Sin embargo, dado que una representación gráfica permite, por lo general, una mejor interpretación y comprensión por parte de los seres humanos, se han estudiado diferentes representaciones visuales de modelos resultantes de fase de aprendizaje. En concreto, se ha definido representaciones gráficas para árboles de decisión, reglas de asociación y redes bayesianas. En otro caso, en especial si sólo hay 2 o 3 dimensiones, se pueden mostrar los patrones sobre los mismos datos originales.

Un árbol de decisión puede verse como un grafo parcialmente ordenado donde los nodos sólo tienen un padre. Dado que un árbol puede ser de gran tamaño, muchas herramientas permiten mostrar segmentos parciales del árbol, empezando por las ramas superiores, y desplegar las partes que el usuario selecciones hasta llegar hasta las hojas.

Las redes bayesianas representan el conocimiento cualitativo de un modelo mediante un grafo dirigido acíclico, por lo que su representación gráfica es directa. El grafo expresa las relaciones de dependencia/independencia entre los diferentes atributos de un problema

Aunque las **reglas de asociación** no son directamente representables gráficamente, si es posible expresar mediante una representación visual llamada malla las relaciones entre los ítems o los conjuntos de ítems.

Por otra parte, también es posible representar gráficamente modelos de regresión o clasificación, siempre que los problemas tengan menos de 3 dimensiones (atributos), mostrando directamente el modelo en el espacio formado por los atributos. Incluyendo los datos en la representación, podremos observar la calidad de los modelos para ajustarse a los datos.

Una estrategia similar se puede adoptar en modelos de **agrupamiento**, pero en este caso en vez de los modelos, se visualizan los centros de los grupos aprendidos.

Por otra parte, dado que los métodos jerárquicos de agrupamiento se basan en la construcción de un árbol, se puede representar este árbol en un grafico llamado dendograma.

2.6 Fase de análisis de resultados

Una vez construido y validado el modelo puede usarse principalmente con dos finalidades: para que un analista recomiende acciones basándose en el modelo y en sus resultados o bien para aplicar el modelo a diferentes conjuntos de datos. También puede incorporarse a otras aplicaciones. Tanto en el caso de una aplicación manual o automática del modelo es necesario su difusión, es decir que se distribuya y se comunique a los posibles usuarios, ya sea por reuniones, intranet, etc. El nuevo conocimiento extraído debe integrar el know-how de la organización.

También es importante medirlo bien que el modelo evoluciona. Aun cuando el modelo funcione bien debemos continuamente comprobar las prestaciones del mismo. Esto se debe principalmente a que los patrones pueden cambiar.

La finalidad del proceso del *KDD* es obtener conocimiento para ayudar a entender mejor el entorno donde se desenvuelve la organización y, en definitiva, mejorar la toma de decisiones en dicho entorno.

CAPÍTULO 3 MÉTODOS O TÉCNICAS DE MÍNERIA DE DATOS

3.1 Extracción de patrones

La extracción de conocimiento a partir de los datos tiene como objetivo descubrir patrones que deben ser validos, interesantes y útiles para la toma de decisiones. Las técnicas de minería de datos para la extracción de patrones en especial las heredadas del aprendizaje automático y del reconocimiento de formas han querido emular la capacidad del hombre para ver patrones en nuestro alrededor, incluso donde no los hay.

Veremos en este capítulo técnicas de minería de muy diferentes tipos, pero antes de eso veremos de una manera muy general, que tienen todas estas técnicas en común, a que problemas se enfrentan y como se puede expresar los resultados de estas técnicas.

En la Figura 3.1 las técnicas de minería de datos parecen como una especie de colador que al introducir los datos produce una serie de patrones, aunque la idea es muy simple en realidad no es tan fácil, el proceso de extracción de patrones a partir de datos son computacionalmente costoso cuanto más expresivo, novedoso e interesante queremos que sean los patrones extraídos.

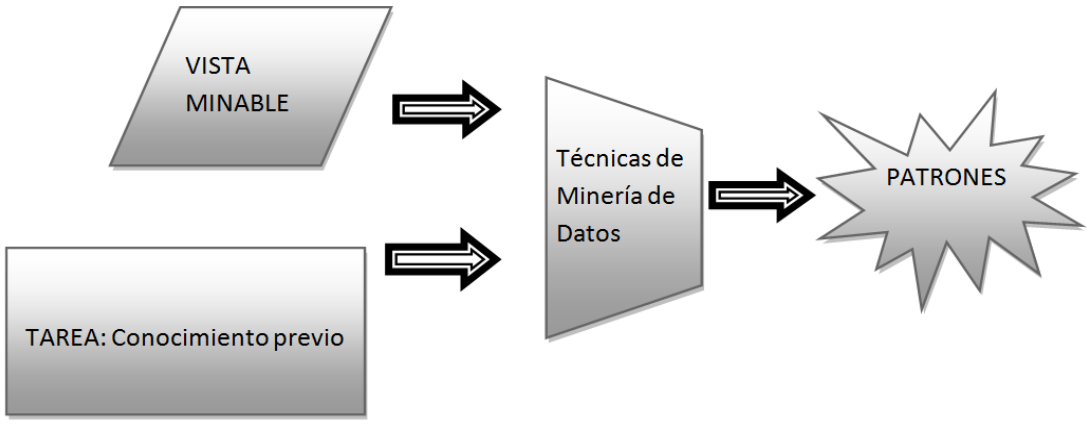


Figura 3.1 Proceso ideal de minería de datos

3.1.1 Tareas y Métodos

Debemos diferenciar una tarea de un método y destacar las tareas y métodos más relevantes. Un tipo de tarea de minería de datos es un tipo de problema de minería de datos por ejemplo “clasificar a las personas por su edad como preescolares, escolares, productivos o posproductivos” en este caso el tipo de tarea es Clasificación. Los métodos son las formas como se puede resolver esta tarea por ejemplo árboles de decisión o redes neuronales.

Cada una de las tareas, requiere métodos ó técnicas para resolverlas. Una tarea puede tener muchos métodos diferentes para resolverla y el mismo método puede resolver muchas tareas.

1) Técnicas algebraicas y estadísticas: expresan modelos y patrones mediante fórmulas algebraicas, funciones lineales, funciones no lineales, medias, varianzas, correlaciones, etc. Alguno de los algoritmos más conocidos es la regresión lineal.

2) Técnicas bayesianas: se basan en estimar la probabilidad de pertenencia (a una clase o grupo), utilizando para ello el teorema de Bayes.

3) Técnicas basadas en conteos de frecuencias y tablas de contingencia: estas técnicas se basan en contar la frecuencia en la que 2 o más sucesos se presenten conjuntamente. Ejemplo: algoritmo "Apriori".

4) Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas: Son técnicas que, además de su representación en forma de reglas, se basa en 2 tipos de algoritmos: los algoritmos denominados "divide y vencerás" como el ID3/C4.5 o el CART, y los algoritmos denominados "separa y vencerás", como el CN2.

5) Técnicas relacionales, declarativas y estructurales: representan los modelos mediante lenguajes declarativos, técnicas ILP (Programación lógica inductiva) han denominado minería de datos relacional.

6) Técnicas basadas en redes neuronales artificiales: se trata de técnicas que aprenden un modelo mediante el entrenamiento de los pesos que conectan un conjunto de nodos o neuronas.

7) Técnicas basadas en núcleo y máquinas de soporte vectorial: se trata de técnicas que intentan maximizar el margen entre los grupos o las clases formadas. Para ello se basan en unas transformaciones que pueden aumentar la dimensionalidad. Estas transformaciones se llaman núcleos (kernels). Existen muchísimas variantes, dependiendo del núcleo utilizado.
8 Técnicas estocásticas y difusas: forman lo que se denomina computación flexible (soft computing)

9) Técnicas basadas en casos, en densidad o distancia: son métodos que se basan en distancias al resto de elementos, ya sea directamente, como los vecinos más próximos (los casos más similares), de una manera más sofisticada, mediante la estimación de funciones de densidad.

3.1.2 El lenguaje de los patrones

El aprendizaje nos permite identificar regularidades en un conjunto de observaciones. Estas regularidades pueden ser representadas por patrones o modelos que los definan.

Estos patrones pueden ser utilizados para predecir observaciones futuras o explicar observaciones pasadas. Esta capacidad de predecir y explicar el entorno es fundamental para mejorar el comportamiento.

La característica más diferenciadora de los métodos de aprendizaje es la manera en la que se expresan los patrones aprendidos. En la Figura 3.2 podemos observar que tipos de patrones son capaces de expresar algunos de los métodos.

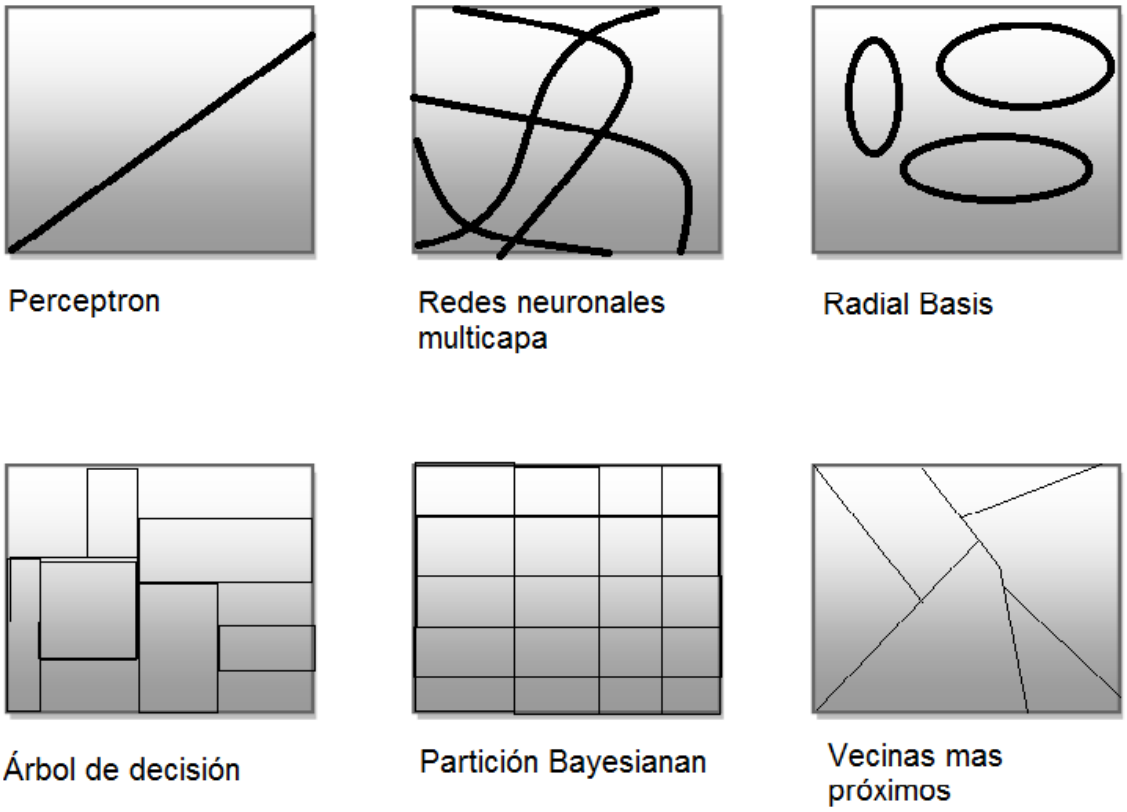


Figura 3.2 Tipos de patrones

3.2 Regresión

La regresión consiste en aprender una función que asigna a cada instancia un valor real. Ésta es la principal diferencia respecto a la clasificación; el valor a predecir es numérico. El objetivo en este caso es minimizar el error entre el valor predicho y el valor real

Por ejemplo, un empresario desea conocer cuál es el costo de un nuevo contrato basándose en los datos correspondientes a contratos anteriores. Para ello usa una fórmula de regresión de regresión lineal, ajustando con los datos pasados la función lineal y usándola para predecir el costo en el futuro. Otros ejemplos sería estimar las ventas del año 2010 o predecir el número de unidades defectuosas de una partida de productos.

La regresión se conoce con otros nombres: interpolación, cuando el valor predicho se encuentra en medio de otros, o también estimación cuando se trata de un valor futuro.

Modelo de regresión

Hablamos de modelo de regresión cuando la variable de respuesta y las variables explicativas son todas ellas cuantitativas. Si sólo disponemos de una variable explicativa hablamos de regresión simple, mientras que si disponemos de varias variables explicativas se trata de un problema de regresión múltiple.

La ecuación de regresión de la muestra que representa el modelo de regresión rectilíneo, sería:

$$Yi = b0 + b1Xi$$

Una vez obtenidos b0 (intercepción con el eje Y) y b1 (la pendiente) se conoce la línea recta y se puede trazar en el diagrama de dispersión y se puede ver si los datos originales están cerca de la línea o se desvían.

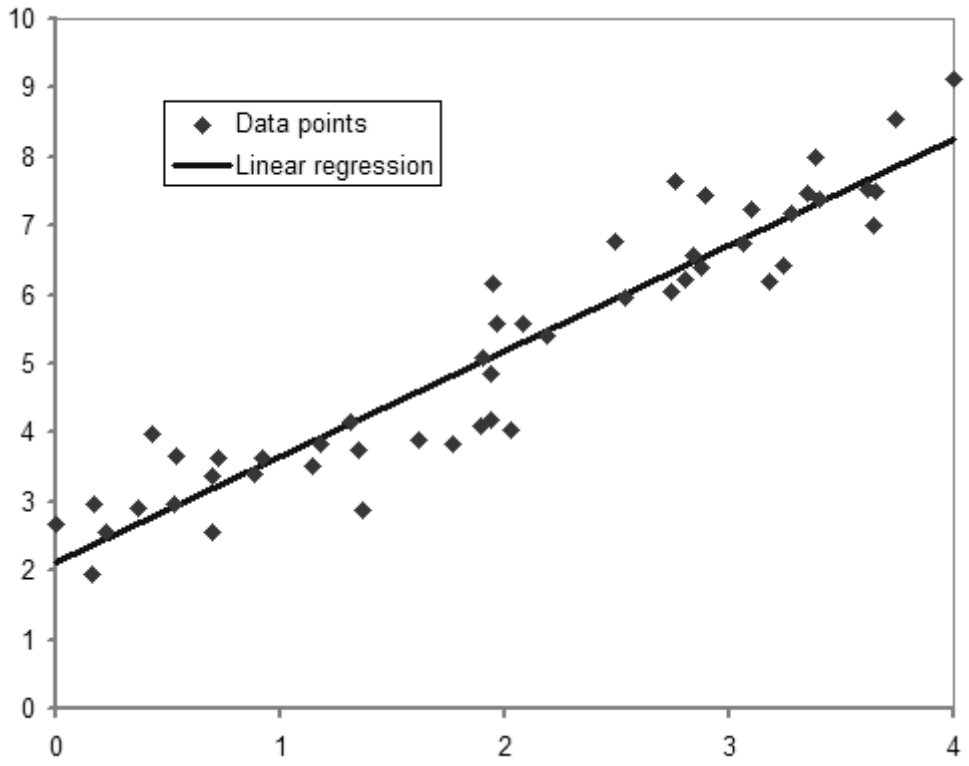


Figura 3.3 Regresion Lineal

Para poder crear un modelo de regresión lineal, es necesario que se cumpla con los siguientes supuestos:

- La relación entre las variables es lineal.
- Los errores son independientes.
- Los errores tienen varianza constante.
- Los errores tienen una esperanza matemática igual a cero.
- El error total es la suma de todos los errores.
- Aplicaciones de la regresión lineal

Líneas de tendencia

Una línea de tendencia representa una tendencia en una serie de datos obtenidos a través de un largo período. Este tipo de líneas puede decirnos si un conjunto de datos en particular (como por ejemplo, el PBI, el precio del petróleo o el valor de las acciones) han aumentado o decrementado en un determinado período. Se puede dibujar una línea de tendencia a simple vista fácilmente a partir de un grupo de puntos, pero su posición y pendiente se calcula de manera más precisa utilizando técnicas estadísticas como las regresiones lineales. Las líneas de tendencia son generalmente líneas rectas, aunque algunas variaciones utilizan polinomios de mayor grado dependiendo de la curvatura deseada en la línea.

Medicina

En medicina, las primeras evidencias relacionando la mortalidad con el fumar tabaco vinieron de estudios que utilizaban la regresión lineal. Los investigadores incluyen una gran cantidad de variables en su análisis de regresión en un esfuerzo por eliminar factores que pudieran producir correlaciones. En el caso del tabaquismo, los investigadores incluyeron el estado socio-económico para asegurarse que los efectos de mortalidad por tabaquismo no sean un efecto de su educación o posición económica. No obstante, es imposible incluir todas las variables posibles en un estudio de regresión. En el ejemplo del tabaquismo, un hipotético gen podría aumentar la mortalidad y aumentar la propensión a adquirir enfermedades relacionadas con el consumo de tabaco. Por esta razón, en la actualidad las pruebas controladas aleatorias son consideradas mucho más confiables que los análisis de regresión.

3.3 Reglas de Asociación y dependencia

Las reglas de asociación expresan patrones de comportamiento entre los datos en función de la aparición conjunta de valores de dos o más atributos. La característica principal de estas reglas es que tratan de atributos no nominales en concreto estas reglas expresan las combinaciones de valores de los atributos (items) que suceden más frecuentemente

Una regla de asociación es una proposición probabilística sobre la ocurrencia de ciertos estados en una base de datos. Pueden ser vistas como reglas de la forma SI α ENTONCES β , donde α y β son 2 conjuntos de items distintos. El conjunto α recibe el nombre de predecesor de la regla y a β se le denomina sucesor o consecuente.

En reglas de asociación existen dos conceptos importantes, la cobertura se llama soporte (*support*) y la precisión se llama confianza (*confidence*).

Se pueden leer como:

$$soporte(A \rightarrow B) = P(A \cup B)$$

$$confianza\ (A \rightarrow B) = P(B \mid A) = \frac{soporte\ (A \cup B)}{soporte\ (A)}$$

Ejemplo:

Supongamos que disponemos de información acerca de las transacciones que se realizan en un supermercado. Cada transacción estará identificada por un cliente y contendrá un conjunto de artículos.

Una muestra de la base de datos podría ser:

Cliente	Artículos
Juan	Papas, Huevos, Jamón
María	Pan, Leche, Huevos
Luis	Pan, Leche, Papas, Huevos
Ana	Pan, Leche

Tabla 3.1 Muestra de Datos

Si fijamos la relevancia mínima en un 50% estamos indicando que, para que una regla sea considerada, al menos debe cumplirse en el 50% de las transacciones de la base de datos, en nuestro caso 2 transacciones mínimo.

Dada la base de datos anterior, podemos obtener, entre otras, las siguientes reglas de asociación:

N	Regla	Soporte	Confianza
1	Pan Z Leche	75%	100%
2	Leche Z Pan	75%	100%
3	Papas Z Huevo	50%	100%

4	Huevos Z Papas	50%	66.60%
5	Huevos Z Papas	50%	66.60%
6	Huevos Z Papas	50%	66.60%
7	Huevos Z Papas	50%	66.60%
8	Huevos Z Papas	50%	66.60%
9	Huevos Z Papas	50%	66.60%
10	Huevos Z Papas	50%	66.60%

Tabla 3.2 Reglas de Asociación

De la simple base de datos de transacciones se puede extraer bastante información útil.

Por ejemplo, las dos primeras reglas nos muestran cómo es habitual comprar leche cuando se va a por el pan.

Las dos siguientes nos podrían indicar que los clientes del supermercado toman con relativa frecuencia huevos con Papas fritas. Estas reglas ponen de manifiesto que la fiabilidad de la regla de asociación A ->! B obviamente no tiene por qué coincidir con la fiabilidad de la regla inversa B -><>A.

Las seis reglas que vienen a continuación son resultado de combinar varios items (tres en este caso).

3.4 Métodos Bayesianos

Esta técnica se utiliza tanto en tareas descriptivas como predictivas. El objetivo consiste en estimar la probabilidad de pertenencia a una clase o grupo, utilizando el teorema de Bayes. Algunos algoritmos muy populares son el clasificador bayesiano naive, los métodos basados en máxima verosimilitud entre otros.

Hay 2 razones por las que los métodos bayesianos son relevantes al aprendizaje automático y a la minería de datos.

1) Son un método práctico para realizar inferencias a partir de los datos, induciendo modelos probabilísticos usados para razonar sobre nuevos valores observados.

2) Facilitan un marco de trabajo útil para la comprensión y análisis de numerosas técnicas de aprendizaje y minería de datos que no trabajan explícitamente con probabilidades.

Teorema de Bayes

El teorema de Bayes nos permite pasar de la probabilidad a priori P (suceso) a la probabilidad a posteriori P (suceso|observaciones) (Hernández, O., Ramirez, M. y Ferri, C., 2000). La probabilidad a priori puede verse como la probabilidad inicial, la que

fijamos sin saber nada más. Por ejemplo, si estamos en invierno, la probabilidad de que una persona padezca gripe en un momento determinado podría ser del 0.95 (P(Gripe=sí) = 0.95). La probabilidad a posteriori es la que obtendríamos tras conocer cierta información, por tanto, puede verse como un refinamiento de nuestro conocimiento. Por ejemplo, si sabemos que la persona en cuestión se ha vacunado contra la gripe, entonces la probabilidad a posteriori sería prácticamente cero P(Gripe=si|vacuna=si) = 0 .

El teorema de Bayes nos facilita un método sencillo y con una semántica clara para resolver esta tarea. Sin embargo, este método tiene un problema y es su altísima complejidad computacional, porque involucra muchas variables, haciéndolas en la mayoría de los casos inmanejables.

Naïve Bayes

El fundamento principal del clasificador Naïve Bayes, es la suposición de que todos los atributos son independientes conocido el valor de la variable clase.

El clasificador Naïve Bayes da lugar a un modelo gráfico probabilístico en el que existe un único nodo raíz (la clase), y en la que todos los atributos son nodos hoja que tienen como único padre a la variable clase. Gráficamente tendríamos la siguiente estructura:

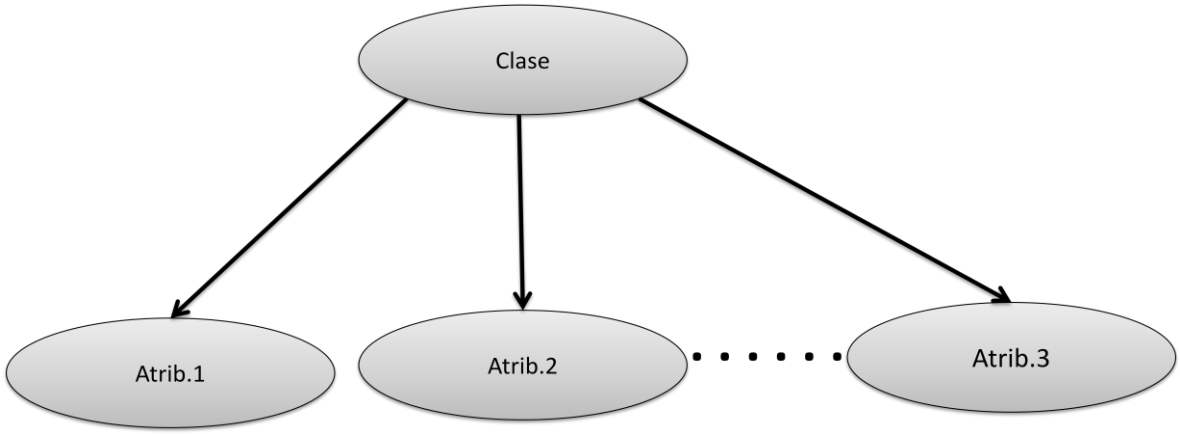


Figura 3.4 Topología de un Clasificador Naive Bayes

Redes Bayesianas

Las Redes Bayesianas RBs representan el conocimiento cualitativo del modelo mediante un grafo dirigido acíclico (Hernández, O., Ramirez, M. y Ferri, C., 2000). Este conocimiento se articula en la definición de relaciones de independencia/dependencia entre las variables que componen el modelo. El hecho de utilizar una representación gráfica para la especificación del modelo hace de las RBs una herramienta realmente muy atractiva en su uso como representación del conocimiento, aspecto muy importante de la minería de datos.

Las RBs no sólo modelan de forma cualitativa el conocimiento sino que además expresan de forma numérica la "fuerza" de las relaciones entre las variables. Esta parte cuantitativa del modelo suele especificarse mediante distribuciones de probabilidad.

Veremos ahora un ejemplo ilustrativo de Redes Bayesianas. Una red que refleja las relaciones entre las variables de un pequeño dominio para determinar si un determinado cliente comprará o no una computadora personal.

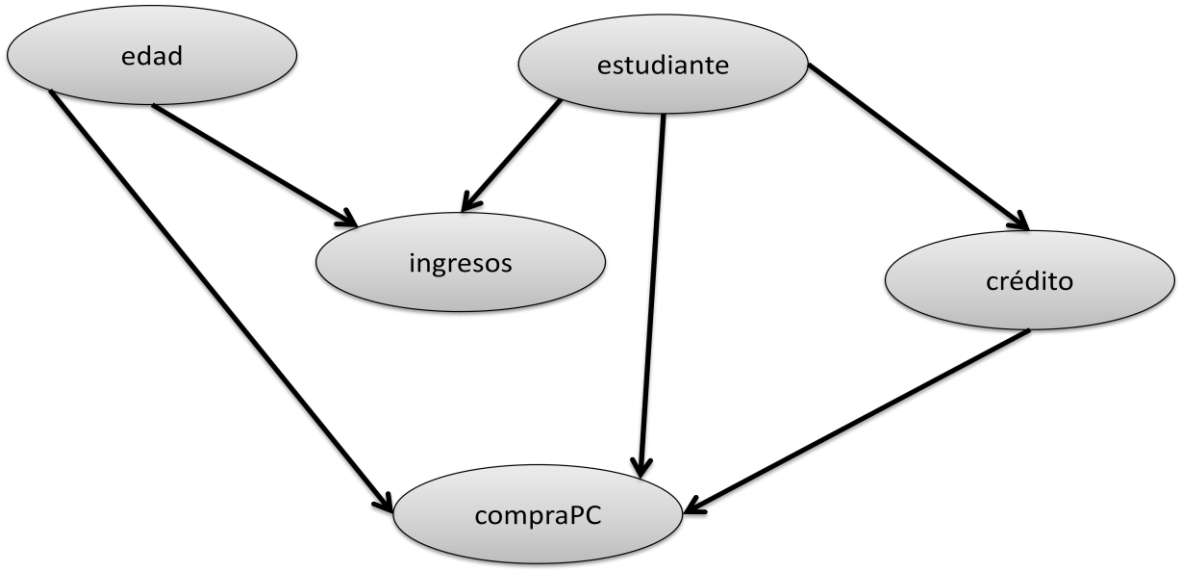


Figura 3.5 Red Bayesiana Compra PC

Existen numerosos paquetes software (gratuitos o no) que están dedicados al manejo de RBs.

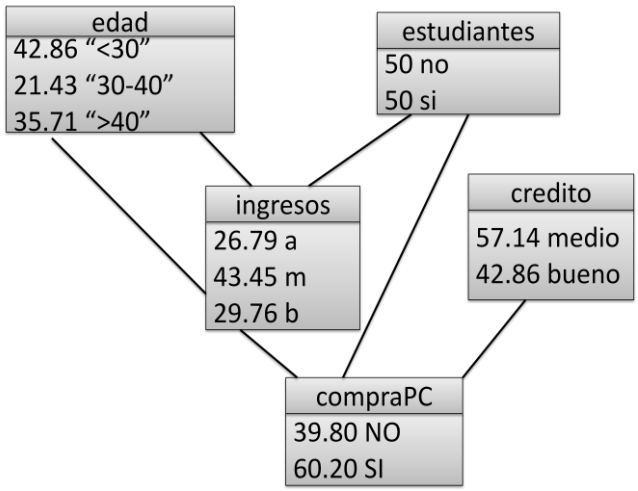


Figura 3.6 Ejemplo de compra de una PC: probabilidad sin evidencia

Podremos realizar inferencias a partir del modelo, como por ejemplo las probabilidades a priori, así vemos que el 60% de la población compraría una PC.

Sin embargo, podemos conocer alguna evidencia, como por ejemplo que la persona en cuestión es un estudiante, entonces la probabilidad de la variable estudiante cambiará al 100 %, y el resto de variables cambiarán en función de la observación.

Sistemas

Paquetes de software que por distintos motivos (experiencia en su uso, flexibilidad, etc.) son más familiares. Indicaremos con los que disponen de entorno gráfico de uso, con A los que facilitan un API (una interfaz de programación), con F los que facilitan el código fuente y con G los que son gratuitos.

WEKA (IAFG): se trata de un entorno genérico de minería de datos, que implementa una gran variedad de técnicas. Podemos encontrar los siguientes:

Naïve Bayes. Implementa el NB clásico y además permite usar kernels para la estimación en las variables numéricas.

Redes Bayesianas

Es de destacar que en todos ellos se puede realizar una combinación con los métodos de selección de variables.

Los métodos bayesianos son relevantes ya que es un método práctico para realizar inferencias a partir de los datos, induciendo modelos probabilísticos para formular hipótesis sobre nuevos valores observados. Son fáciles de usar, muy eficientes, pueden tratar muchos atributos y son muy robustos al ruido. Al igual que las técnicas anteriores no construyen modelos, sólo estiman una serie de probabilidades.

3.5 Árboles de Decisión

Los arboles de decisión con una técnica para el aprendizaje de modelos comprensibles de decisión. El término “modelo” indica que estas técnicas construyen un “modelo”, o “representación” de los datos y el término “comprensibles” hace referencia al hecho de que estos modelos se pueden expresar de una manera simbólica, pueden tener como resultado modelos inteligibles para los seres.

Los sistemas de aprendizaje basados en árboles de decisión son el método más fácil de utilizar y de entender. Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. Una de las grandes ventajas de los árboles de decisión es que, las opciones posibles a partir de una determinada condición son excluyentes. Esto permite llegar a una sola acción o decisión a tomar.

Ejemplo: en un hospital en el que se realizan operaciones de cirugía refractiva a aquellas personas miopes que lo soliciten, evidentemente no en todos los casos no están indicadas y hay casos en los que pueden estar excluidos con el objetivo de evitar riesgos potenciales o efectos secundarios, existen algunas condiciones claras por las cuales se puede determinar si, en principio, una persona está indicada para la operación.

En la Figura 3.7, se muestra un ejemplo del árbol de decisión para determinar recomendar o no cirugía ocular.

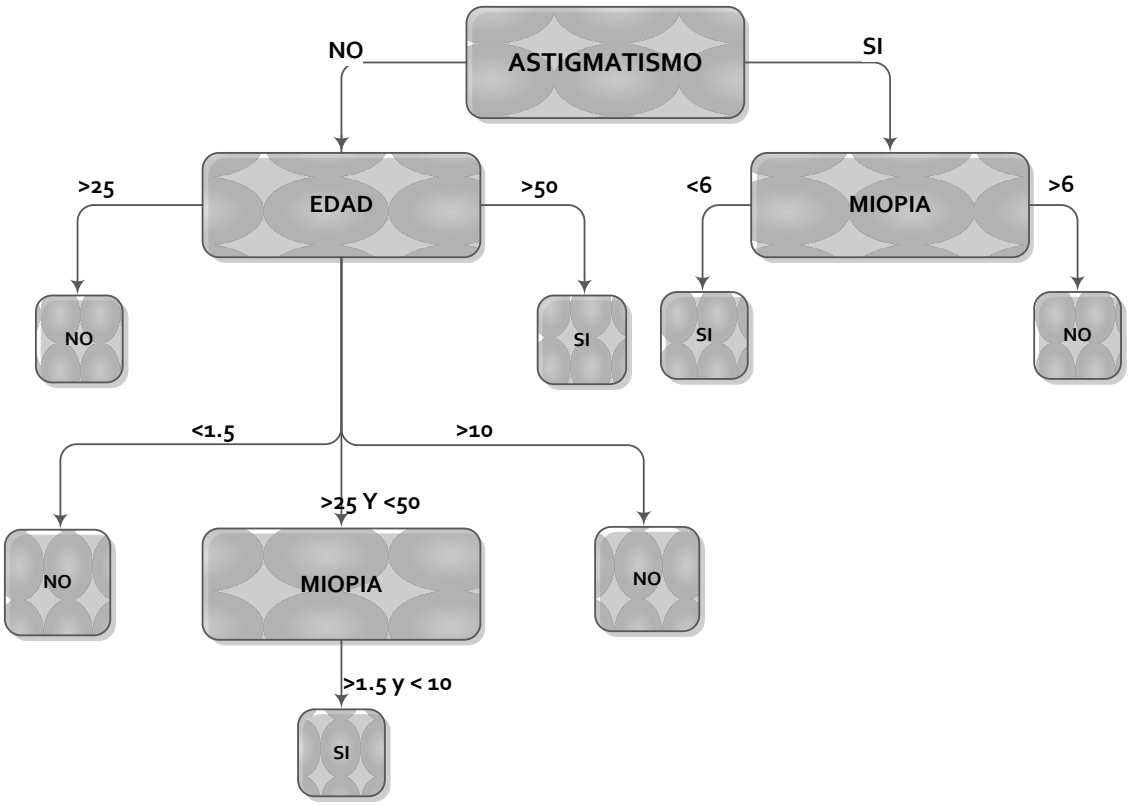


Figura 3.7 Árbol de decisión

Es sencillo aplicar el árbol de decisión a un nuevo paciente. Basta realizar las preguntas y seguir las respuestas hasta alguna de las hojas del árbol, catalogadas con un "no" o un "sí". Este árbol de decisión en concreto funciona como un "clasificador", es decir, dado un nuevo individuo nos lo clasifica en una de las dos clases posibles: "no" o "sí".

Respecto al ejemplo de la cirugía, se muestra el árbol de decisión expresado en forma de reglas.

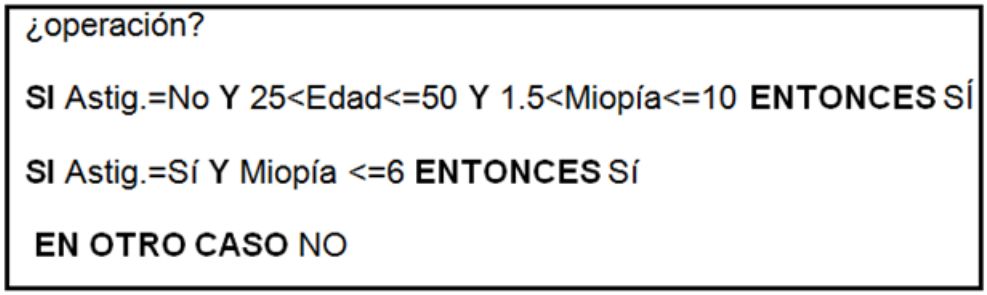


Figura 3.8 Reglas correspondientes al árbol de la Figura 3.7

Algoritmos o sistemas de aprendizaje más populares

CART y derivados: son métodos "divide y vencerás" que construyen árboles binarios tanto para clasificación como para regresión. La poda se basa en una estimación de la complejidad del error.

ID3, C4.5 y derivados: son métodos "divide y vencerás" y están basados en criterios de partición derivados de la ganancia. Tienen poda basada en reglas. Contiene métodos de colapsado de ramas.

AQ, CN2, y derivados: son métodos por "cobertura". Se pueden encontrar en paquetes de minería de datos con el nombre de "RULES" (por ejemplo en el Clementine).

SLIQ: modificaciones de árboles de decisión clásicos para conseguir escalabilidad para grandes volúmenes de datos, paralelización, etc.

3.6 Redes Neuronales Artificiales

Las redes neuronales artificiales son un método de aprendizaje cuya finalidad es de emular mediante computadoras el funcionamiento del cerebro. Las redes neuronales artificiales parten de la capacidad humana de procesar información debido a la naturaleza biológica de nuestro cerebro. Por tanto, para imitar esta característica debemos estudiar y basarnos en el uso de soportes artificiales semejantes a los existentes en nuestro cerebro.

Las redes neuronales artificiales (RNA) son sistemas conexionistas dentro del campo de la Inteligencia Artificial, las cuales, dependiendo del tipo de arquitectura neuronal, pueden tener diferentes aplicaciones. Pueden utilizarse para la minería de datos como en agrupamiento, clasificación.

3.6.1 Neuronas biológicas y artificiales

En una neurona real cada sinapsis representa la unión de un axón de una neurona con una dendrita de otra neurona. Una transmisión electro-química tiene lugar en la sinapsis. La información es entonces transmitida a lo largo de las dendritas hasta que alcanza el cuerpo de la célula. Allí tiene lugar el sumatorio de los impulsos eléctricos que lo alcanzan. La neurona se activará si el resultado es superior a un determinado límite o umbral.

Esto significa que enviará una señal a lo largo de su axón con la finalidad de comunicarse con otras neuronas. Ésta es la manera en la que la información pasa de una parte de la red de neuronas a otra. Normalmente modelamos una neurona biológica de la manera que se muestra en la parte derecha.

Cuando tenemos una red de neuronas, las salidas de unas se conectan con las entradas de otras. Si el peso entre dos neuronas conectadas es positivo, el efecto producido es de excitación. Por el contrario, si es negativo, este efecto es de inhibición. Intentando simular varios cientos de billones de neuronas. Imaginamos las neuronas actuando conjuntamente en capas como se muestra en la Figura 3.6.

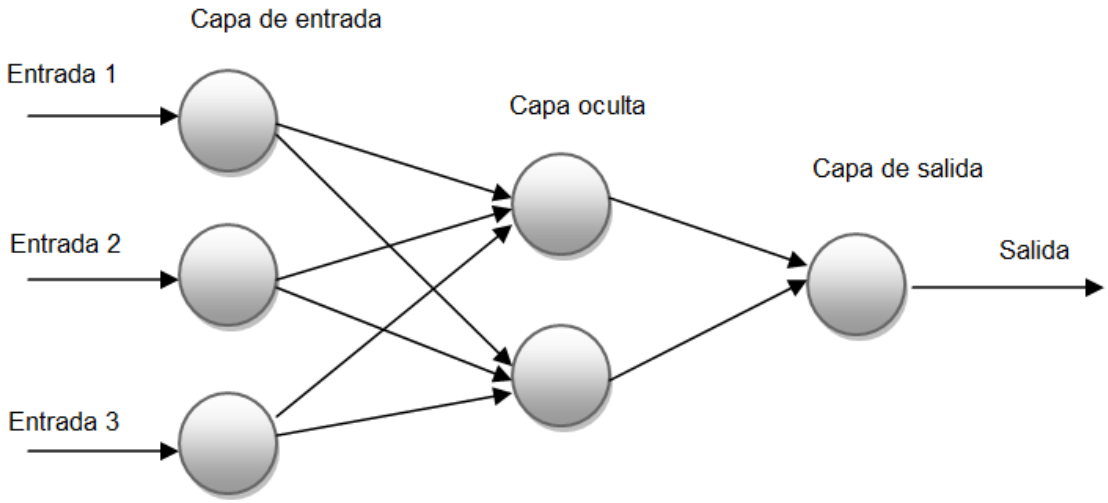


Figura 3.9 Ejemplo de red neuronal artificial formada por tres capas

Hay dos modos de trabajo en una RNA:

Modo de transferencia de la activación: cuando la activación es transmitida por toda la red y está asociado a la operación de propagación hacia adelante.

Modo de aprendizaje: cuando la red se organiza normalmente a partir de la transferencia de activación más reciente.

El aprendizaje en las redes neuronales artificiales

Los cambios son debidos a variaciones en los pesos de la red neuronal dando lugar al aprendizaje. Se cree que nuestro aprendizaje se debe a cambios en el rendimiento o eficiencia de las sinapsis, a través de las cuales se transmite la información entre neuronas.

Hay dos tipos principales de aprendizaje en RNA:

Aprendizaje supervisado. Con este tipo de aprendizaje, proporcionamos a la red un conjunto de datos de entrada y la respuesta correcta. Entonces podemos comparar la respuesta calculada por la red con aquella que se desea obtener. Entonces se ajustan los pesos para que la red produzca una respuesta correcta. Este tipo de aprendizaje es útil para las tareas de regresión y clasificación.

Aprendizaje no supervisado. Sólo se proporciona a la red un conjunto de datos de entrada. La red debe auto-organizarse. Este tipo de aprendizaje es útil para las tareas de agrupamiento y reducción de dimensionalidad.

3.7 Métodos basados en casos y en vecindad

Los métodos basados en vecindad reciben su nombre del hecho de que la predicción se basa fundamentalmente en la utilización del conjunto de ejemplos "vecinos" al dato que hay que procesar, o en el caso más general, porque la distancia entre cada ejemplo y el dato en cuestión es esencial en el proceso.

Existen tres maneras de procesar los ejemplos: la memorización de todos los ejemplos (dando lugar a métodos retardados), la memorización de parte (seleccionando ejemplos significativos, siendo, en cierto modo, un híbrido) y la creación de prototipo, es decir, representantes ficticios de un conjunto de datos (dando lugar a métodos no retardados).

Medidas de distancia

Las medidas de distancia más tradicionales son aquellas que se aplican sobre dos instancias o ejemplos, tales que todos los atributos son numéricos. Por ejemplo, se puede definir las siguiente distancia entre dos vectores, puntos o distancias x e y de dimensión n de muy distintas formas.

Distancia Euclídea. Es la distancia, como la longitud de línea recta que une dos puntos en el espacio euclídeo:

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

Figura 3.10 Distancia Euclidea

Métodos retardados frente a no retardados

Los métodos retardados, también llamados perezosos (del inglés lazy), se llama así porque retrasa la decisión de la generalización del conjunto de entrenamiento, hasta el instante en que se recibe un nuevo dato a procesar. Por el contrario. Los métodos anticipados o no retardados (del inglés eager) construyen un modelo de generalización, antes de tener que realizar dicha tarea de generalización, a partir del conjunto de ejemplos.

Una diferencia fundamental, es que los métodos retardados realizan una aproximación local al dato a generalizar, por lo tanto el modelo resultante podría verse como una combinación de aproximaciones locales, que dan mayor versatilidad a la solución. Por otra parte los métodos no retardados construyen una aproximación global utilizando la totalidad del conjunto de ejemplos. El problema de los métodos retardados es que se basan en la selección de un conjunto de ejemplos apropiados, o en algún tipo de ponderación de cada uno de los elementos del conjunto de ejemplos, para la construcción de la aproximación local.

Mapas de auto-organizativos de Kohonen

Fue modelizado en un principio como una red neuronal de dos capas, tal como se muestra en la figura.

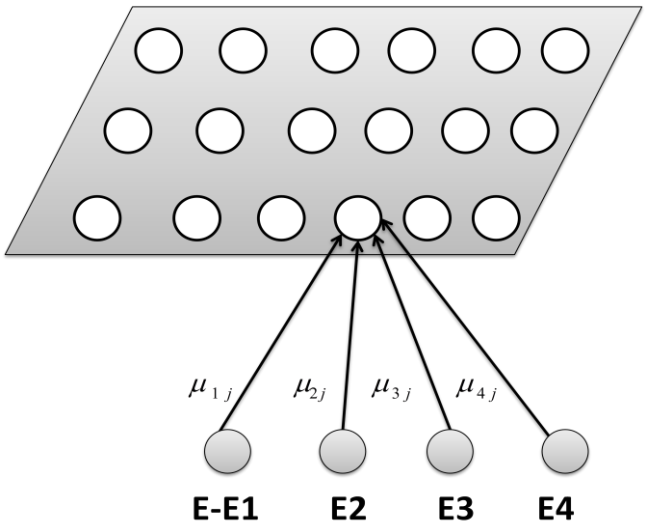


Figura 3.11 Arquitectura del Mapa de Kohonen

El modelo consta de dos capas, una de entrada, donde se introducen los ejemplos, y otra de competición, en el que cada célula representa a un prototipo. El objetivo de entrenamiento es que los prototipos capturasen ejemplos similares.

K medidas

El algoritmo K medidas se trata de un método de agrupamiento por vecindad en el que se parten de un número determinado de prototipos y de un conjunto de ejemplos a agrupar, sin etiquetar. La idea de K medidas es situar a los prototipos o centros en el espacio, de forma que los datos pertenecientes al mismo prototipo tengan características similares.

Todo ejemplo nuevo, una vez que los prototipos han sido correctamente situado, es comparado con éstos y asociados a aquél que sea el más próximo, en términos de una distancia previamente elegida.

El método tiene una fase de entrenamiento, que puede ser lenta, dependiendo del número de puntos a clasificar y de la dimensión del problema. Pero una vez terminado el entrenamiento, la clasificación de nuevos datos es muy rápida, gracias a que la comparación de distancias se realiza sólo con los prototipos.

Agrupamiento jerárquico

Los métodos jerárquicos se basan en la construcción de un árbol en el que sus hojas son los elementos del conjunto de ejemplos, y el resto de los nodos son subconjuntos de ejemplos que pueden ser utilizados como particionamiento del espacio. Este gráfico se denomina dendrograma y se muestra en la figura.

En la Figura 3.12 se muestra un ejemplo de un árbol de agrupamiento. En la raíz está el conjunto de ejemplos, en este caso 11 ejemplos etiquetados de *a* a la *k.* cada descendiente es una división del nodo de partida, de forma que van describiendo sucesivos subconjunto de ejemplos. Una particularidad de este tipo de árboles es que cada nodo está situado en un nivel diferente de todos los demás. De esta forma se genera una jerarquía de nodos, que da nombre al conjunto de métodos, que permite la obtención de diferentes soluciones.

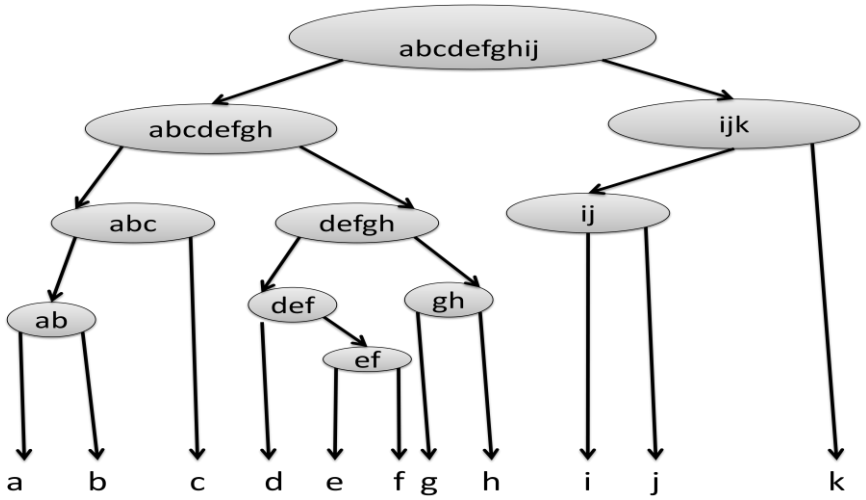


Figura 3.12 Ejemplo de un árbol de agrupamiento (dendrograma)

K vecinos

La regla del vecino más próximo simplemente asigna la clase del ejemplo más próximo, utilizando una función de distancias.

Una variante de este método son los *k* vecinos más próximos en el que se asigna la clase mayoritaria entre los *k* vecinos más próximos.

Una versión más elaborada de los vecinos más próximos se basa en determinar una región de cercanía, más que un valor constante de ejemplos a comparar.

Razonamiento basado en casos

El razonamiento basado en casos es un modelo computacional de razonamiento por analogía, basado en casos históricos existentes. Una de las premisas del razonamiento basado en casos en que gran parte de las soluciones a problemas encontradas por los especialistas no son originales, son variaciones sobre un problema tipo. A veces es mejor resolver un problema partiendo de la solución existente a un problema similar, que tratar de resolverlo desde el principio. Esto es lo que se llama razonamiento por analogía de casos previos. La forma de afrontar el problema es buscar un caso similar en el pasado, junto con la solución que funcionó para ese caso. Adaptar la solución para ajustar las diferencias existentes entre los dos casos y almacenar la solución sugerida, para que pueda ser utilizada posteriormente en nuevos casos.

Generalmente sistemas de razonamiento basado en casos consiste en:

- Una base de datos de ejemplos anteriores y su solución.
- Un conjunto de índices para poder recuperar casos pasados y almacenar los nuevos.
- Un conjunto de reglas para medir las similitudes entre casos

- Y una regla de modificación de soluciones existentes, en función de las diferencias entre los casos.

Los sistemas de razonamiento basados en casos tienen un gran número de ventajas: en los problemas en los que no existe un gran conocimiento del dominio, o donde la relación entre los atributos de los casos y la solución no está suficientemente clara como para que pueda ser representada en forma de reglas, o cuando la relación de casos que son "excepción a la regla" es alta, los sistemas de razonamiento basados en casos son buenos candidatos, ya que pueden hacer modelos de las excepciones y los casos nuevos.

Los sistemas de razonamiento basados en casos son también útiles a la hora de explicar o justificar una solución. Cuando la teoría del dominio es débil, es bastante difícil justificar o explicar una postura de forma razonable. Sin embargo el poder dar una analogía o encontrar un antecedente puede ser más eficaz que un argumento basado en un modelo.

Sistemas

Muchos sistemas genéricos de minería de datos incluyen varios de los métodos comentados, en especial los basados en modelo (los no retardados): k medidas, los mapas de Kohenen, LQV, etc. En cambio, los métodos retardados, como los vecinos más próximos o el agrupamiento jerárquico, no son tan habituales.

Quizás el sistema que dispone de más métodos basados en instancias es el WEKA. WEKA dispone de los métodos ya antes explicados, así como el IB1 y IBk (para vecinos más próximos), el kstar (que es el K*) o el cobweb.

Posiblemente el mayor problema de las implementaciones de estos métodos en los paquetes de minería de datos que no explotan su mejor característica, la probabilidad de variar la función de distancia. Muchas implementaciones de estos métodos incorporan una medida de distancia (generalmente la euclídea) y no permite variarla fácilmente.

CAPÍTULO 4 HACIENDO MINERÍA DE DATOS CON LA INFORMACIÓN DE MORTALIDAD EN MÉXICO

4.1 Introducción

Debido a la gran cantidad de información almacenada en las diferentes bases de datos del INEGI, se consideró un panorama muy interesante para realizar minería de datos, haciéndolo con el proceso KDD, proceso que se ha explicado a más detalles en capítulos anteriores.

Antes de profundizar en este capítulo, daremos una breve reseña de lo que es el INEGI.

El Instituto Nacional de Estadística y Geografía (INEGI) es una institución gubernamental de México, dedicada a la coordinación de los Sistemas Nacionales Estadísticos y Geográficos del país. Asimismo de la orientación y la promoción del desarrollo informático. Fue creado el 25 de enero de 1983 bajo decreto presidencial.

El INEGI trabaja en varias áreas, pero nosotros vamos a trabajar con la información que nos muestra todo lo referente a las defunciones generales en México.

A continuación se hablara de forma general sobre la bases de datos que se va a trabajar, se comentaran aspectos como, los antecedentes, objetivo general, periodicidad, diseño conceptual, cobertura temporal, cobertura geográfica, desglose geográfico.

Antecedentes

La estadística de defunciones generales, desde el inicio de su captación, en 1893, ha tenido una evolución similar a la de las estadísticas de nacimientos y matrimonios. En 1987, la Secretaría de Salud puso en operación a nivel nacional el certificado de defunción, que a partir de 1989 es el principal formato de captación.

Objetivo general

Generar las estadísticas sobre defunciones generales, que permitan caracterizar el fenómeno de la mortalidad en el país (INEGI, 1988).

Periodicidad

Anual.

Diseño conceptual

COBERTURA TEMÁTICA

Los conceptos básicos son: defunción general, definidos por la Organización de las Naciones Unidas (ONU) y la Organización Mundial de la Salud (OMS).

Las variables que capta la estadística de defunciones generales son las siguientes:

De la defunción:
Fecha de registro y fecha de ocurrencia, lugar geográfico de registro y domicilio de ocurrencia, lugar de ocurrencia, atención médica, condición de necropsia, causas de la defunción, condición y relación de embarazo y persona que certificó la defunción.

De muertes accidentales y violentas:
Presunción de accidente, homicidio, etc., ocurrencia en el desempeño del trabajo, lugar donde ocurrió (vivienda, área deportiva, etc.) y violencia familiar.

Del fallecido:
Sexo, edad, escolaridad, derechohabiencia, fecha de nacimiento, estado civil, nacionalidad, ocupación y lugar geográfico de residencia habitual.

CLASIFICADORES UTILIZADOS

- Catálogo de Municipios y Localidades.
- Clasificación Mexicana de Ocupaciones, 1998.
- Clasificación Internacional de Enfermedades, Décima Revisión, OMS.

RECOMENDACIONES INTERNACIONALES
Se siguen los lineamientos del Programa Internacional para Acelerar el Mejoramiento de los Sistemas del Registro Civil y Estadísticas Vitales, de la ONU, así como los del Manual de Sistemas y Métodos de Estadísticas Vitales, del mismo organismo.

Cobertura geográfica

Se captan las defunciones registradas en todo el país en el Sistema de Registro Civil, así como las defunciones accidentales y violentas en las que interviene el Ministerio Público.

Desglose geográfico

Entidad federativa, municipio y localidad.

FUENTES DE DATOS

Para la etapa de integración y recopilación de datos se obtuvo información de las siguientes fuentes de datos:

Sistema Nacional de Información de Salud SINAIS

Instituto Nacional de Estadística y Geografía (INEGI, 1988) Archivos en formato xBase contenidos en un archivo comprimido ZIP.

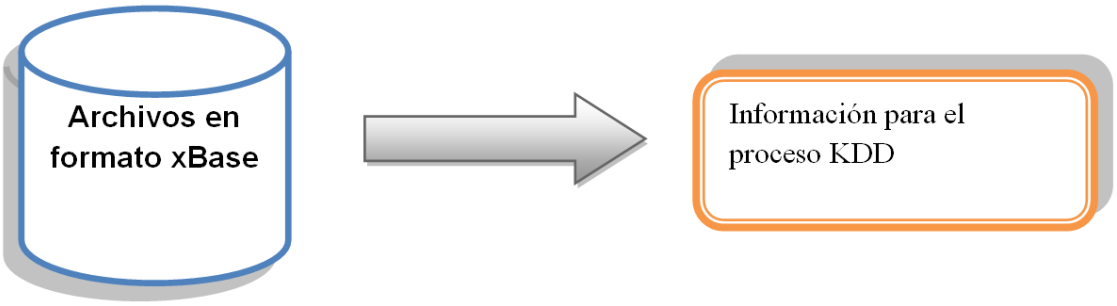


Figura 4.1 Fuente de datos

La adquisición de los datos fue a través de la página de internet del SINAIS, en la siguiente dirección:

http://sinais.salud.gob.mx/basesdedatos/index.html

Antes de profundizar un poco más en la cuestión práctica, hablaremos del contenido y la estructura de la información que se encuentra contenida en los archivos .ZIP con los siguientes nombres:

```
def2000
DEF2001
 def02
 def03
 DEF04
 DEF05
 DEF06
 DEF07
 DEF08
 def09
```

.

La Estadística de Mortalidad proporciona elementos para el análisis de los niveles de mortalidad general que se dan anualmente en el país, así como el crecimiento natural de la población al relacionar esta información con la natalidad.

Al presentar la frecuencia con que ocurren las defunciones, se abre la posibilidad para obtener indicadores que permitan evaluar la efectividad de los programas de salud pública (servicios médicos, sanitarios, de nutrición, de atención materno-infantil, entre otros), así como detectar las necesidades de servicios y recursos médicos.

Esta estadística, se genera a partir de la información asentada en las actas, certificados y cuadernos de defunciones inscritos en el Sistema Nacional del Registro Civil y Agencias del Ministerio Público clasificada para fines estadísticos en:

- Defunciones Generales
- Características de la defunción
- Características del fallecido

A través de la Consulta interactiva de datos de la Estadística de Mortalidad el Instituto Nacional de Estadística y Geografía ofrece una opción para un mejor aprovechamiento y análisis de la información anual de las defunciones generales, con una cobertura geográfica: nacional, estatal y municipal que considera la posibilidad de realizar consultas con diferentes niveles de desagregación, de acuerdo a las variables seleccionadas y diseño del tabulado, así como exportar la información obtenida a diversos formatos, como por ejemplo una hoja de cálculo.

Después de una breve descripción de cómo se encuentra la información en las bases de datos, de los años 2000 a 2009. Se detallara lo que se hizo en la práctica.

Una vez que se descargan los archivos .dbf, la forma en la cual los descomprimimos e integramos, directamente en el SQL Server[1] con la tarea de importar, esto se realizo de una forma muy sencilla seleccionando el formato de nuestro archivo de origen, el cual está en formato dBase.

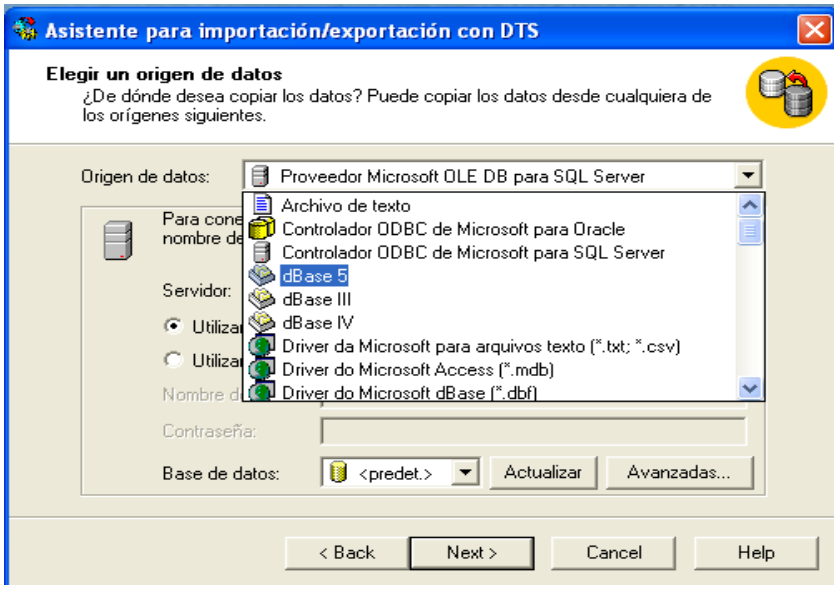


Figura 4.2 Asistente para importación / exportación

[1] SQL Server: Es un conjunto de objetos eficientemente almacenados. Los objetos donde se almacena la información se denominan tablas, y éstas a su vez están compuestas de filas y columnas.

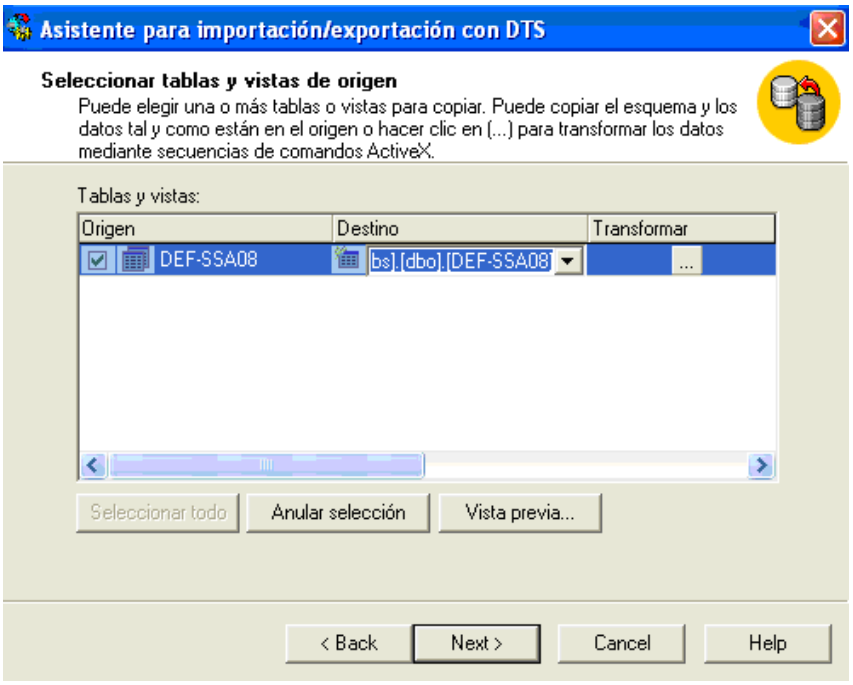
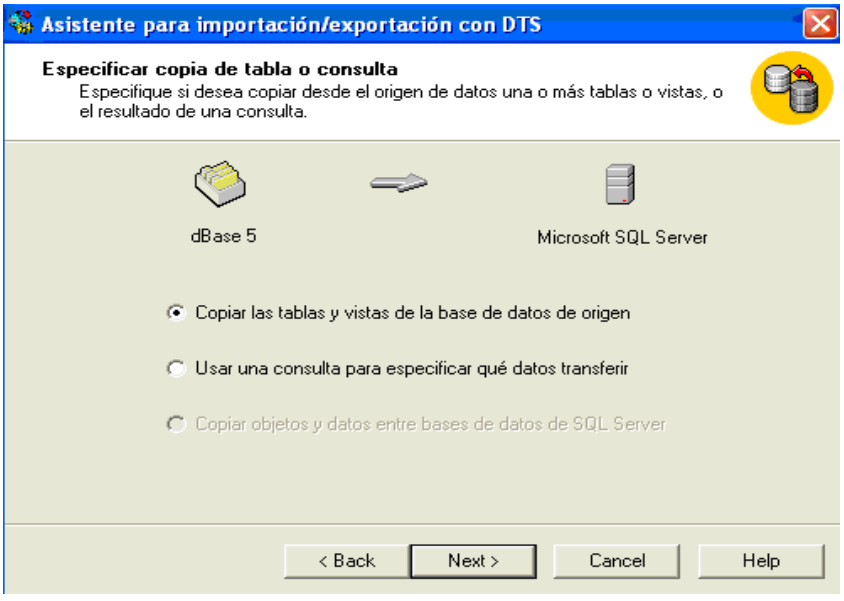


Figura 4.3 Seleccionar tablas y vistas de origen

Especificamos a que base de datos vamos a importar, es decir el destino de nuestros datos, se da siguiente (Next) y se realiza la importación, este paso se va a realizar con cada uno de los archivos que se quieran importar, ya que cada archivo será una tabla de nuestra base.



Figuar 4.4 Copiar las tablas y vistas archivo xbase a SQL Server

Una forma sencilla de pasar nuestra base ya integrada a otra máquina es con los siguiente comando, los cuales ya contienen la base en dos archivos principales con extensiones .mdf y .ldf . Ejecutando los siguientes comandos:

exec sp_attach_db @dbname = N'PAU', comando que nos dice que se va a exportar una base que lleva por nombre PAU.

@filename1 = N'c:\Program Files\Microsoft SQL Server\MSSQL.2\MSSQL\DATA\PAU_Data.mdf', con esta se hace la carga del primer archivo, se anota la dirección de donde se tiene el archivo.

@filename2 = N'c:\Program Files\Microsoft SQL Server\MSSQL.2\MSSQL\DATA\PAU_Log.ldf', con esta se hace la carga del segundo archivo, se anota la dirección de donde se tiene el archivo.

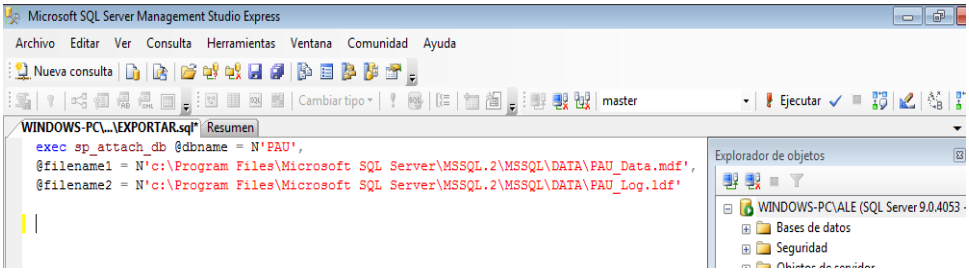


Figura 4.5 exec sp_attach_db @dbname

DESCRIPCIÓN DE LOS DATOS

Los datos al ser descomprimidos y exportados al manejador son presentados de la siguiente forma:

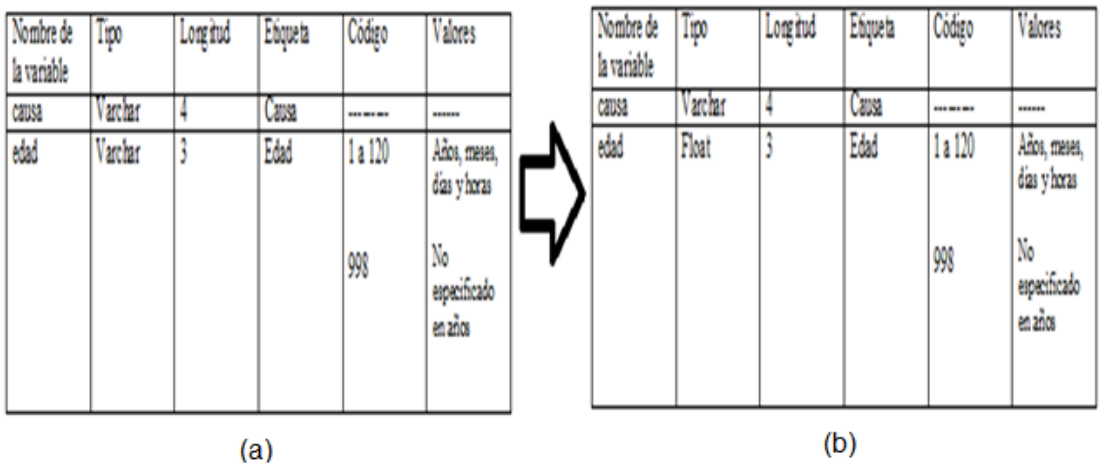
	CONTROL	ENTRES	MPORES	TLOC	ENTDEF	MPODEF	CAUSA	SEXO	CVEEDAD	EDAD	ANODEF	MESDEF	DIADEF	ANOREG
1	15220140011	22	14	14	22	14	A010	1	A	78	2000	1	12	2000
2	15220120001	22	12	5	22	12	A010	2	A	83	2000	4	9	2000
3	15220140001	22	14	14	22	14	A010	2	A	68	2000	4	6	2000
4	15220140001	22	14	14	22	14	A010	1	A	42	2000	5	4	2000
5	15080370013	8	37	15	8	37	A010	1	A	25	2000	7	14	2000
6	15070470001	7	47	1	7	47	A010	2	A	37	2000	1	16	2000

Figura 4.6 Muestra de datos

Cada año está representado por una tabla. Para visualizar la información que se encuentra almacenada en cada una de las tablas, solo se da clic derecho en la tabla que queremos ver y se abre la tabla.

Para tener más organizada la información, y que pueda ser manipulada por cualquier persona que quiera trabajar con esta base, se creó un diccionario de datos, en donde se describen cada uno de los atributos que forman las tablas, esta descripción se realizó por año, ya que en cada año se muestran los mismos atributos, pero con nombres diferentes, y en otros casos hay atributos diferentes es decir no todas las tablas tienen la misma cantidad de atributos, pero todos tienen de 45 atributos en adelante. Este diccionario de dato se encuentra al final de esta tesis en el Anexo (Diccionario de datos).

Algo muy importante de mencionar es que varios datos sufrieron cambios para que la manipulación ó el trabajo con ellos fuera más eficiente, esto sucede porque es difícil manipular un **varchar**[2] y esto nos llevo a modificarlos a un **float**[3], esta modificación se realizo con todos los atributos a acepción de CAUSA, debido a que este atributo almacena datos que tienen tanto letras como números. Esto cambio de tipo de dato se dio en los años 2000 al hasta 2005, ya que en los años posteriores este cambio ya estaba realizado.



Nombre de la variable	Tipo	Longitud	Etiqueta	Código	Valores
causa	Varchar	4	Causa	.....	.....
edad	Varchar	3	Edad	1 a 120 998	Años, meses, días y horas No especificado en años

(a)

Nombre de la variable	Tipo	Longitud	Etiqueta	Código	Valores
causa	Varchar	4	Causa	.....	.....
edad	Float	3	Edad	1 a 120 998	Años, meses, días y horas No especificado en años

(b)

Figura 4.7 (a) Datos originales y (b) Datos modificados

Hablando de cómo fue la modificación, esto se realizó de una manera muy sencilla ya que el SQL Server es un manejador muy amigable y nos permite hacerlo de esta manera, primero seleccionamos un atributo y con clic derecho elegimos la opción de modificar y nos despliega todos los atributos de cada tabla y escogemos el tipo de dato al que lo vamos a cambiar o a cambiarlos, ya que se pueden cambiar todos los atributos a la vez.

[2] **Varchar**: Se utiliza cuando los tamaños de las entradas de datos de columna varíen de forma considerable

[3] **float**: se usa para almacenar valores numéricos con decimales. Se utiliza como separador el punto (.). Definimos campos de este tipo para precios, por ejemplo.

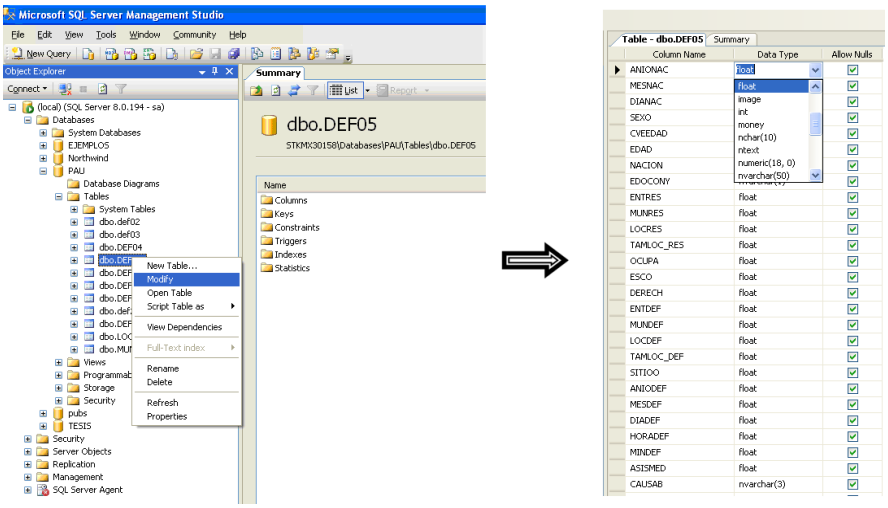


Figura 4.8 Modificación de los datos en SQL Server

La selección de los atributos se realizó con los atributos que nos podrían decir información interesante como la edad de las defunciones, o la causa de la defunción, dejando atributos que no consideramos importantes como día de la defunción o mes de registro.

Una vez ya realizada la selección de los atributos que nos servirán para hacer las vistas minables, pasamos a realizar la limpieza sobre dichos datos, es decir detectar datos anómalos, la presencia de datos faltantes o perdidos en general valores que no se ajustan al comportamiento general de los datos. Esto lo hacemos para saber cómo se van a tratar dichos datos, ya sean eliminarlos, remplazarlos, ignorarlos, entre otras soluciones, en esta ocasión tal vez no sea necesario el uso de estas posibles soluciones, ya casi en su totalidad la base de datos se encuentra limpia.

Para detectar los datos antes mencionados lo que hacemos son queries. A continuación se presenta una tabla que nos muestra el atributo y el número de errores que presenta en la base de datos, al igual que el querie con el que se obtuvo dicho resultado.

Atributo	**Consulta para encontrar errores**	**No. de errores**
Edad	SELECT EDAD FROM DEF03 WHERE (edad<1 OR edad>120) AND edad <> 998	0
Causa	N/A	0
Derechohabiente	select distinct(derech) from DEF03 where derech not in (1,2,3,4,5,6,0)	1
Relación con el embarazo	select relemb from DEF03 where relemb not in (1,2,0)	0
Atención medica	select asismed from DEF03 where asismed not in (1,2,0)	0
Estado civil	select edociv from DEF03 where edociv not in (1,2,3,4,5,6,8,0)	0
Violencia familiar	N/A	0
Edad	SELECT EDAD FROM DEF03 WHERE (edad<1 OR edad>120) AND edad <> 998	0
Complicación embarazo	select CONEMB from DEF03 where CONEMB not in (1,2,3,0)	0
Trabajo	select traba from DEF03 where traba not in (1,2,8,0)	0
Escolaridad	select esco from DEF03 where esco not in (1,2,3,4,5,6,7,8,0)	0
Ocupación	select ocupa from DEF03 where ocupa not in(2,11,12,13,14,21,41,51,52,53,54,55,61,62,71,72,81,82,83,98,99,0)	0

Tabla 4.1 Consultas para encontrar errores

Como pudimos ver la base de datos se encuentra limpia, a excepción del atributo DERECHOHABIENTE lo que sucede con este atributo es que hay un 46 que no tiene ningún significado, es decir hay una inconsistencia en este dato. Lo que se va a ser con este atributo como solo es un error, será reemplazar con el valor de 0 que significa 'no especificado'.

La transformación es una tarea muy importante ya que reduce los datos y esto facilita el manejo de la base de datos. Los datos son transformados o consolidados de forma apropiada para la extracción de información. Existen varias vías para la reducción de datos, pero la que se utilizo más para la base que se está trabajando es la discretización.

Para mostrar cómo fue que se realizó esta desratización, se presenta la tabla siguiente con los datos que nos servirán para nuestra vistas minables:

No.	NOMBRE LOGICO EN BASE DE DATOS	NOMBRE DE VARIABLE	CODIGO	DESCRIPCIÓN
1	sexo	SEXO	0	No especificado
			1	Hombre
			2	Mujer
2	edadvalor	EDAD VALOR	001 A 120	Edad en años
			998	No especificado en años
			001 A 011	Edad en meses
			98	No especificado en meses
			001 A 029	Edad en días
			98	No especificado en días
			001 A 023	Edad en horas
			98	No especificado en horas
			97	No especificado en minutos
3	edociv	ESTADO CIVIL	0	Se ignora
			1	Soltero
			2	Viudo
			3	Divorciado
			4	Unión libre
			5	Casado
			8	No aplica a menores de 12 años
4	entres	ENTIDAD RESIDENCIA HABITUAL.	1	Aguascalientes
			…	….
			32	Zacatecas
			33	Estados unidos de Norteamérica
			34	Otros países de Latinoamérica
			35	Otros países
5	ocupa	OCUPACION HABITUAL	0	No especificado
			2	Inactivos

No.	NOMBRE LOGICO EN BASE DE DATOS	NOMBRE DE VARIABLE	CODIGO	DESCRIPCIÓN
			11	Profesionistas.
			12	Técnicos.
			13	Trabajadores de la educación.
			14	Trab. Arte, espec y deportes.
			21	Fun. Y direc. Del sector pub., priv. Y social.
			41	Trabajadores agropecuarios.
			51	Sup. Y per. De control en act. Industriales.
			52	Obreros y artesanos en la prod. Industrial.
			53	Oper. De maqui. En la producción industrial.
			54	Ayud., auxi. Y peones en la prod. Industrial.
			55	Operadores de transporte.
			61	Jefes y coordi. En activi. Admvas. Y servicio.
			62	Trabajadores de apoyo en actividades admvas.
			71	Comerciantes y agentes de ventas.
			72	Vendedores ambulantes.
			81	Trabajadores en servicios personales.
			82	Trab. En servicios domesticos.
			83	Trab. De fza. Armada, protección y vigilancia.
			98	No aplica a menores de 12 años
			99	Ocupación insuficientemente especificada.
6	esco	ESCOLARIDAD	0	No especificado
			1	Sin escolaridad
			2	Primaria incompleta (de uno a cinco años)
			3	Primaria completa
			4	Secundaria incompleta
			5	Secundaria completa
			6	Bachillerato o preparatoria
			7	Profesional
			8	No aplica a menor de 6 años
7	derech	DERECHOHABIENTE	0	Derechohabiencia no especificada
			1	Ninguna
			2	Imss
			3	Issste
			4	Pemex
			5	Secretaria de la defensa nacional
			6	Secretaria de marina
			7	Seguro popular
			8	Otra
8	asismed	ASISTENCIA MEDICA	1	Con atención medica
			2	Sin atención medica

No.	NOMBRE LOGICO EN BASE DE DATOS	NOMBRE DE VARIABLE	CODIGO	DESCRIPCIÓN
			0	No especificado
9	causa	CLAVE CAUSA	A00	Equivalencia del catalogo de la 10a. Revisión
			…..	
			Y89	
10	traba	OCURRIO TRABAJO	1	Si
			2	No
			0	No especificado
			8	NO APLICA A MUERTE NATURAL (A00 a R99)
11	violfam	VIOLENCIA FAMILIAR	0	No especificado
			1	Hubo violencia familiar
			2	Hubo violencia no familiar
			8	NO APLICA PARA MUERTE NATURAL (A00 – R99) o VIOLENTAS, EXCEPTO HOMICIDIOS
12	conemb	CONDICION EMBARAZO	1	El embarazo
			2	El parto
			3	El puerperio
			4	43 días a 11 meses después del parto o aborto
			5	No estuvo embarazada durante los once meses previos a la muerte
			6	Estuvo embarazada un año o más antes de la muerte
			8	No aplica
13	relemb	FUERON COMPLICACIONES DEL EMBARAZO	1	Si tuvieron relación las causas
			2	No tuvieron relación las causas
			8	No aplica

Tabla 4.2 Discretización de los atributos

Algo importante de mencionar es que esta discretización o la descripción que se le asigno a cada código ya venía incluida desde que se descargaron los archivos en formato excel.

Procedimiento para llevar a cabo la minería de datos

A continuación encontraremos un esquema que nos muestra de forma más ilustrativa, el proceso que se llevara para cada una de las vistas, primero tenemos que definir bien cuál es el problema para así saber hacia dónde dirigirnos y obtener objetivos, "Definición del problema", una vez hecho esto si los datos con los que vamos a trabajar los obtenemos de una fuente externa lo que se hace es extraerlos "base de datos del INEGI" y los almacenamos en el manejador que vamos a ocupar, "Almacenamiento de los datos a SQL Server", para hacer un mejor estudio de los datos, con una herramienta externa realizamos la exploración de los mismos con PASW, "Exploración de los datos", una vez que se hizo esta exploración lo que procede es una "Preparación y limpieza de los datos" que se tienen cargados en el manejador, posteriormente procedemos a realizar la "Minería de datos", trabajaremos con WEKA y Rapid Miner estas herramientas trabajan con las vistas que armamos en SQL Server, con los resultados obtenidos de estos, que son "Patrones y modelos", pasamos a la "Evaluación e interpretación" de dichos modelos y patrones si llegara a ser el caso en el que los resultados del análisis no fueran coherentes o estuvieran muy erróneos lo que se hace es regresar al paso de "Preparación y limpieza de los datos" y repetir todo el proceso desde ese punto, una vez que los resultados sean correctos, pasamos a la "Difusión del conocimiento", es decir saber hacia qué sector de la población va ir dirigido y saber cómo corregir ó mejorar el problema que se plateo en la "Definición del problema".

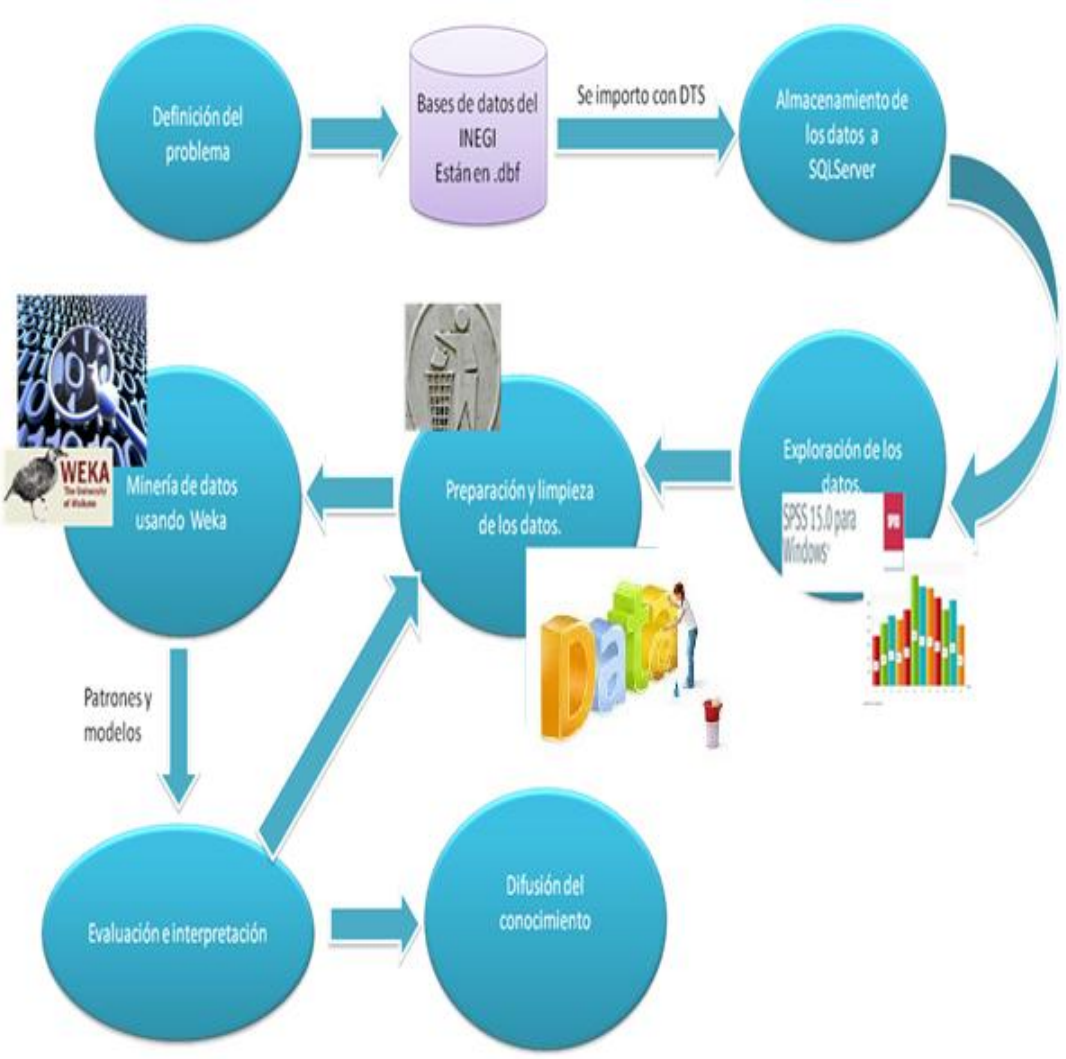


Figura 4.9 Procedimiento para llevar a cabo la minería de datos

Si realizamos un buen preprosesamiento de los datos es decir toda la preparación de dichos datos nos permiten aplicar modelos de aprendizaje o lo que es lo mismo realizar la minería de datos de una forma rápida y sencilla, esto nos dará como resultado modelos o patrones de mayor calidad para que así los resultados sean precisos.

4.2 Eficiencia de Instituciones Médicas

En esta fase analizaremos los objetivos, y esto lo haremos con las siguientes preguntas:

- **¿Qué parte de los datos es pertinente analizar? (Eficiencia de Instituciones Médicas)**

  Analizaremos datos como son, su ocupación, entidad de residencia, estado civil, si recibió o no asistencia médica, y uno de los datos con más peso a analizar será el de derechohabiencia, el cual funge como estructura principal de toda nuestra vista.

- **¿Qué tipo de conocimiento se desea extraer y cómo se debe presentar? (Eficiencia de Instituciones Médicas)**

  Se desea extraer información que nos indique personas fallecidas de los organismos médicos públicos o privados a los cuales se encuentran afiliadas las personas (derechohabiencia), también saber cuál es su ocupación para saber porque fueron asignados a dichos organismos, será de suma importancia saber su estado civil ya que con esto nos mostrara sí el trabajador afilio, a su esposa(o), hijos, padres al mismo organismo, algo que es indispensable es saber si la persona recibió o no atención médica. También se vería en que entidad se encuentran más personas afiliadas a cada uno de los organismos existentes.

- **¿Qué conocimiento puede ser válido, novedoso e interesante? (Eficiencia de Instituciones Médicas)**

  Se requiere conocer que organismos médicos públicos o privados con mayor número de personas fallecidas, afiliadas a dichos organismos, para saber qué tipo de gente se canaliza a cada uno de los organismos existentes es decir que ocupación desempeñan. Todo esto para saber que organismos hay que mejorar y así las personas afiliadas a cada organismo público o privado, puedan recibir una mejor atención en todos los aspectos.

- **¿Qué reglas o modelos de decisión están utilizando? (Eficiencia de Instituciones Médicas)**

  Reglas de asociación

- **¿Qué decisiones son críticas? (Eficiencia de Instituciones Médicas)**

  Identificar cuáles son los organismos médicos más ineficientes y el porqué de su ineficacia, ya que estén bien identificados, sin errores ni dudas entonces mejorarlos para un bienestar de la sociedad.

- **¿Cómo se distribuyen los datos? (Eficiencia de Instituciones Médicas)**

  En una sola base de datos.

Análisis estadístico previo a la minería

Exploración de los datos para saber qué datos pueden ser los más adecuados para poder formar la vista minable.

Primeramente**,** se tiene el objetivo de obtener las estadísticas para todos los casos posibles empezando por el panorama general hasta los casos particulares. Esto mismo se usará para la Minería de Datos ya que resulta mucho mejor analizar por casos concretos que por un caso general; cuando se tiene un caso muy general, pueden existir varios casos extremos que probablemente provoquen una cierta inclinación o ponderación de más a los modelos haciendo que éstos varíen más allá de lo que sería en realidad, es decir, analizando un caso específico.

Se realizaran una serie de histogramas que nos dirán como es que se encuentran las datos, con un software llamado **PASW statictics 18**, describiremos el procedimiento para realizar el primer histograma que relaciona los atributos 'Derechohabiencia' con 'Estado de residencia'.

Lo primero que se tiene que hacer es seleccionar los atributos a graficar de nuestra base de datos con la siguiente consulta:

```
SELECT  ENTRES, DERECH FROM dbo.DEF2000
union all
SELECT  ENTRES, DERECH FROM dbo.DEF2001
union all
SELECT  ENTRES, DERECH FROM dbo.def02
UNION all
SELECT  ENTRES, DERECH FROM dbo.def03
union all
SELECT  ENTRES, DERECH FROM dbo.DEF04
UNION ALL
SELECT  ENTRES, DERECH FROM dbo.DEF05
union all
SELECT  ENTRH, DERHAB FROM dbo.DEF06
UNION ALl
SELECT   ENTRH, DERHAB FROM dbo.DEF07
union all
SELECT   ENTRH, DERHAB FROM dbo.DEF08
UNION ALl
SELECT   ENTRH, DERHAB FROM dbo.DEF_09
```

Realizando una union con todos los 10 años, guardar en un archivo excel o .csv , porteriormente abrir **PASW statictics 18,** abrir un origen de datos existente y elegir el archivo previamente guardado.

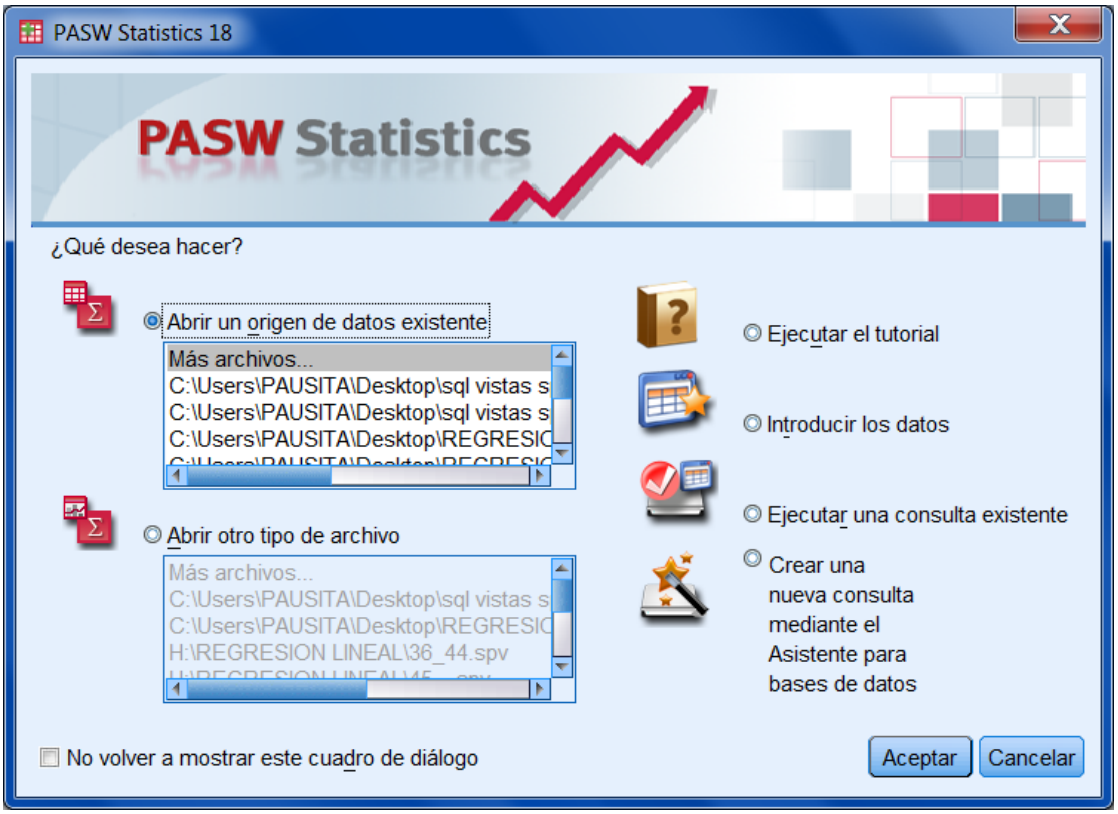


Figura 4.10 Abrir origen de datos existente (PASW)

Posteriormente se abriran los datos, se observa que los datos aún no estan discretizados, traen los valores originales de la base de datos, esto se cambia directamete con la herramienta en la pestaña de vista de variables, en esta pestaña en la vista de valores asignamos el significado de cada número.

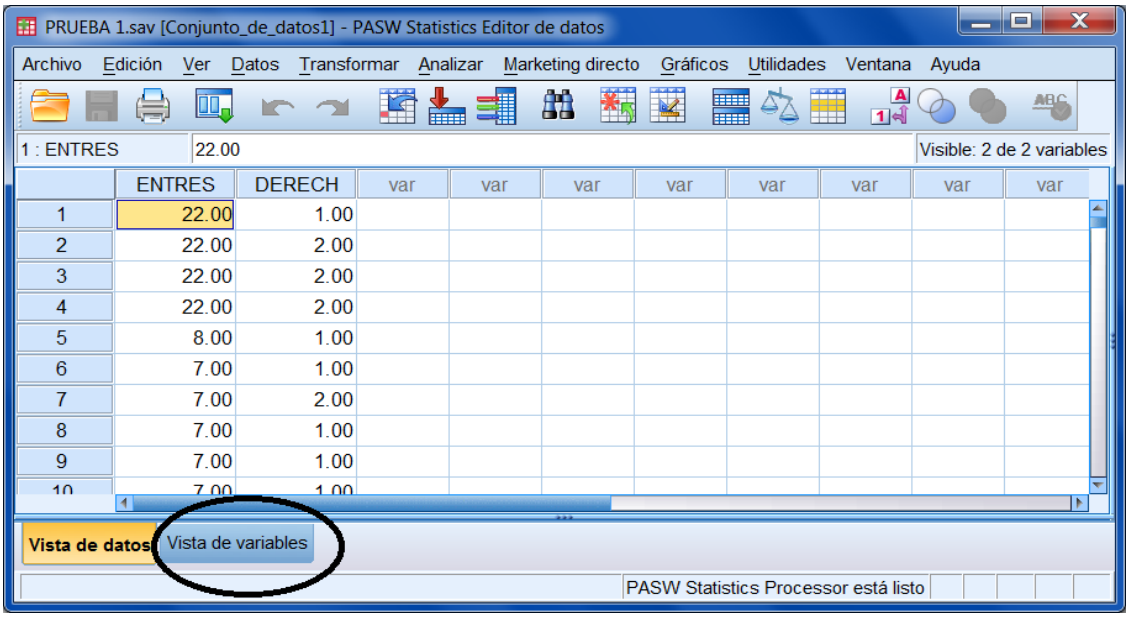


Figura 4.11 PASW statictics 18 editor de datos

En este caso asignaremos el nombre de cada estado de la republica, por ejemplo 1 = "Aguascalientes" en el caso del atributo ENTRES(entidad de residencia) igualmente para el otro atributo del cual queremos ver su comportamiento en una gráfica el cual es DERECHOHABIENCIA.

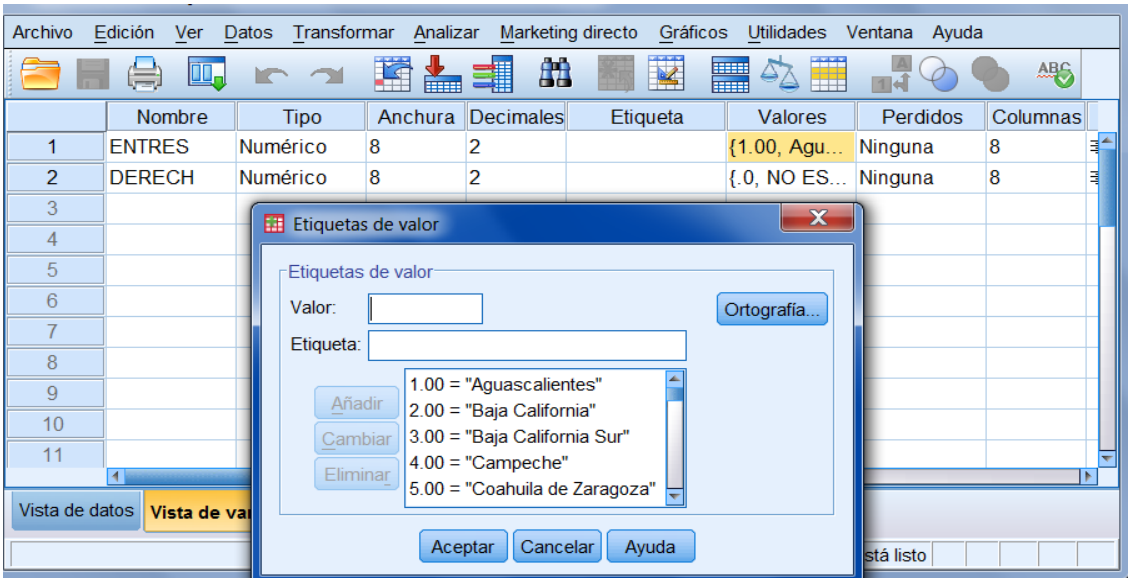


Figura 4.12 PASW statictics 18 etiquetas de valor

Después de haber realizado esto en la pestaña Graficos, se presentara una pantalla para la generación de graficos, escojemos Histograma y arrastramos la gráfica que querramos, poteriormente arrastramos los atributos, dependiendo cual estará en el eje X y cual será la pila de nuestro histograma.

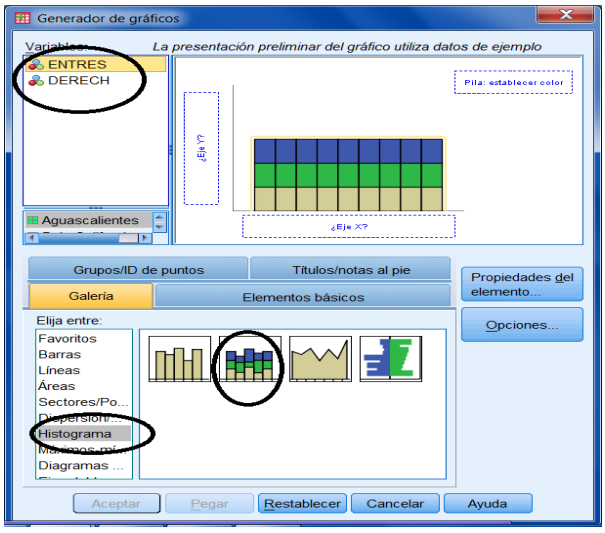


Figura 4.13 Generador de graficos

Posteriormente le damos aceptar y se generara nuestro grafico en una ventana de resultado, la cual podemos guardar.

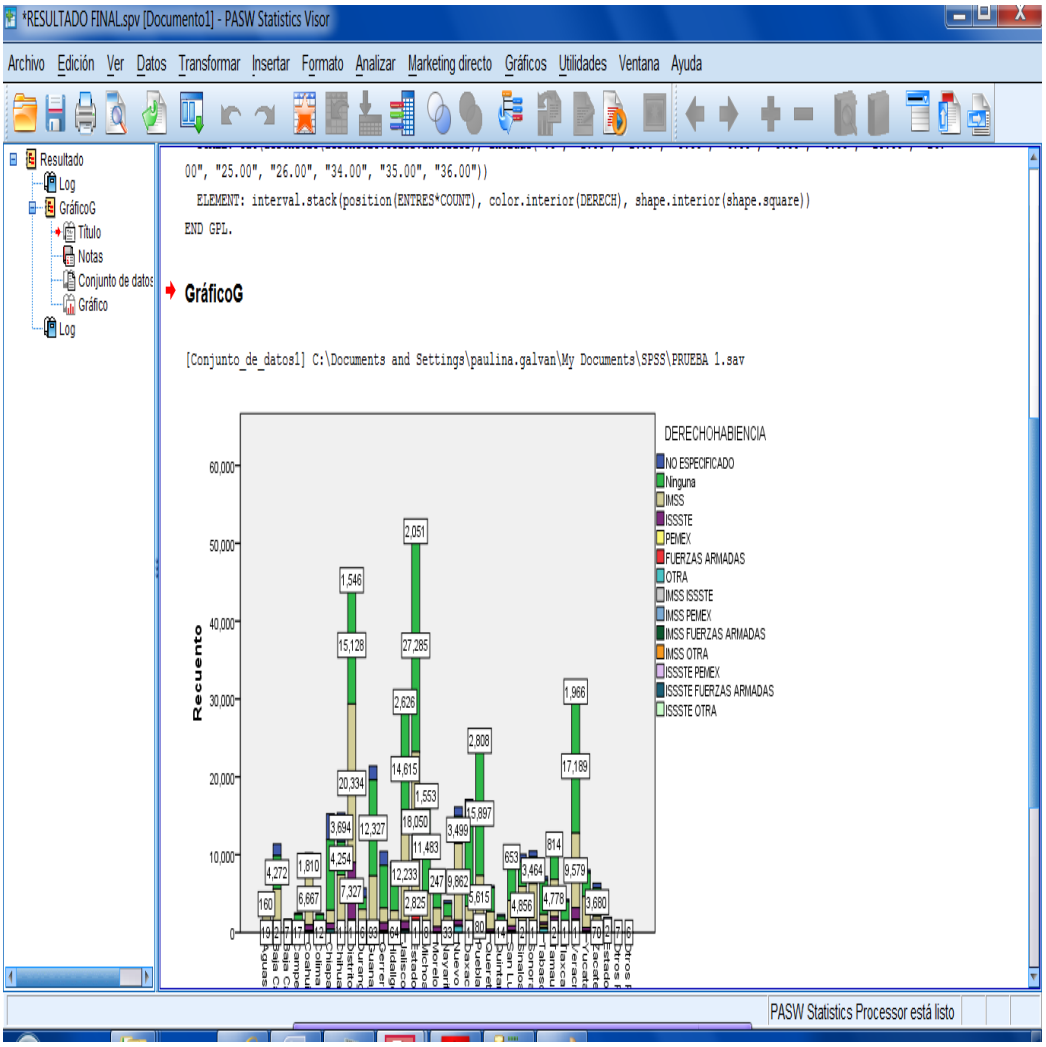


Figura 4.14 PASW statictics 18 Visor

Este procedimiento se repite en todas las gráficas de nuestro análisis estadístico previo a la minería de datos. Cabe aclarar que las gráficas que a continuación se presentan son de los 10 años que estudiaremos 2000 al 2009.

La gráfica que a continuación se muestra nos indica que organizaciones tienen más personas afiliadas, esto se observa por cada estado, algunos de los resultados más generales que nos arrojo dicho análisis, es que en la mayoría de los estados el IMSS es la organización que más afiliados tiene.

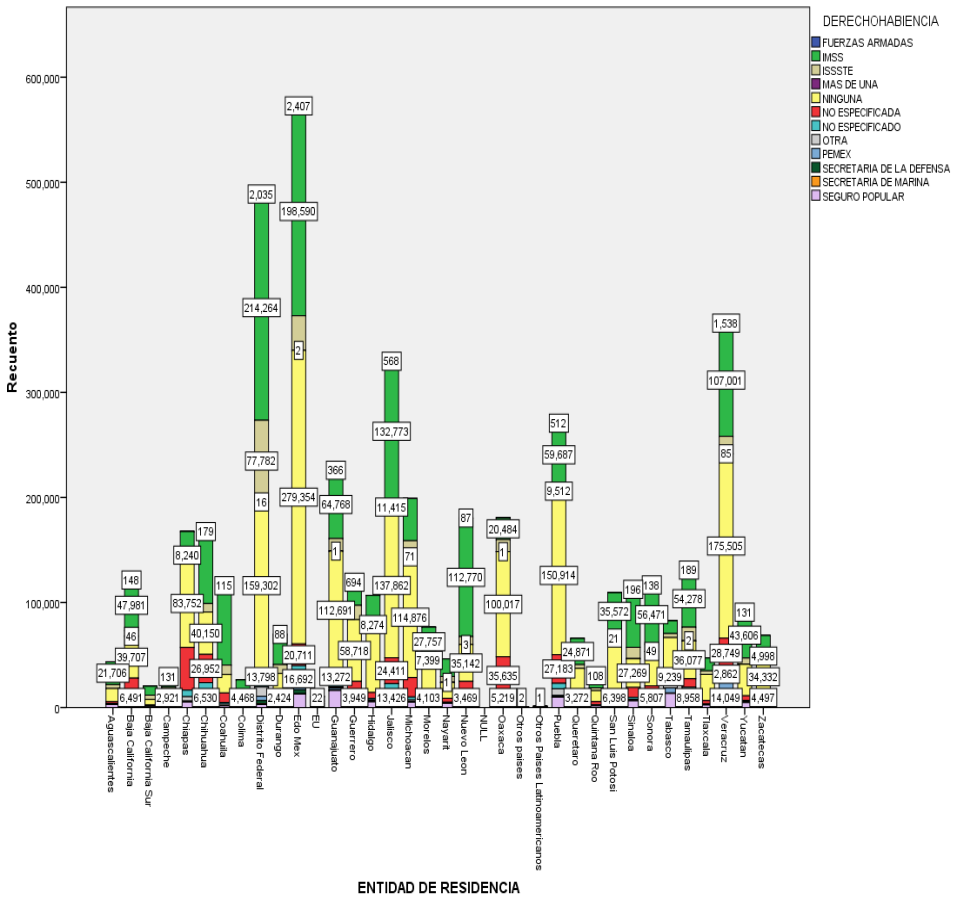


Figura 4.15 Histograma Derechohabiencia -Entidad Residencia

La siguiente gráfica, muestra el estado civil de las personas y la ocupación que desempeñan, los resultados que nos arroja, es que hay una gran cantidad de personas inactivas de las cuales una gran parte son viudos y otra cantidad considerable son casados, una de las ocupaciones que más destaca es la de trabajadores agropecuarios que en su mayoría son casados, otra de las ocupaciones que también son más desempeñadas son las de obreros y artesanos y al igual que en la anterior la mayoría de dichos empleados son casados. Las ocupaciones que menos se desempeñan son aquellas que tienes niveles administrativos altos.

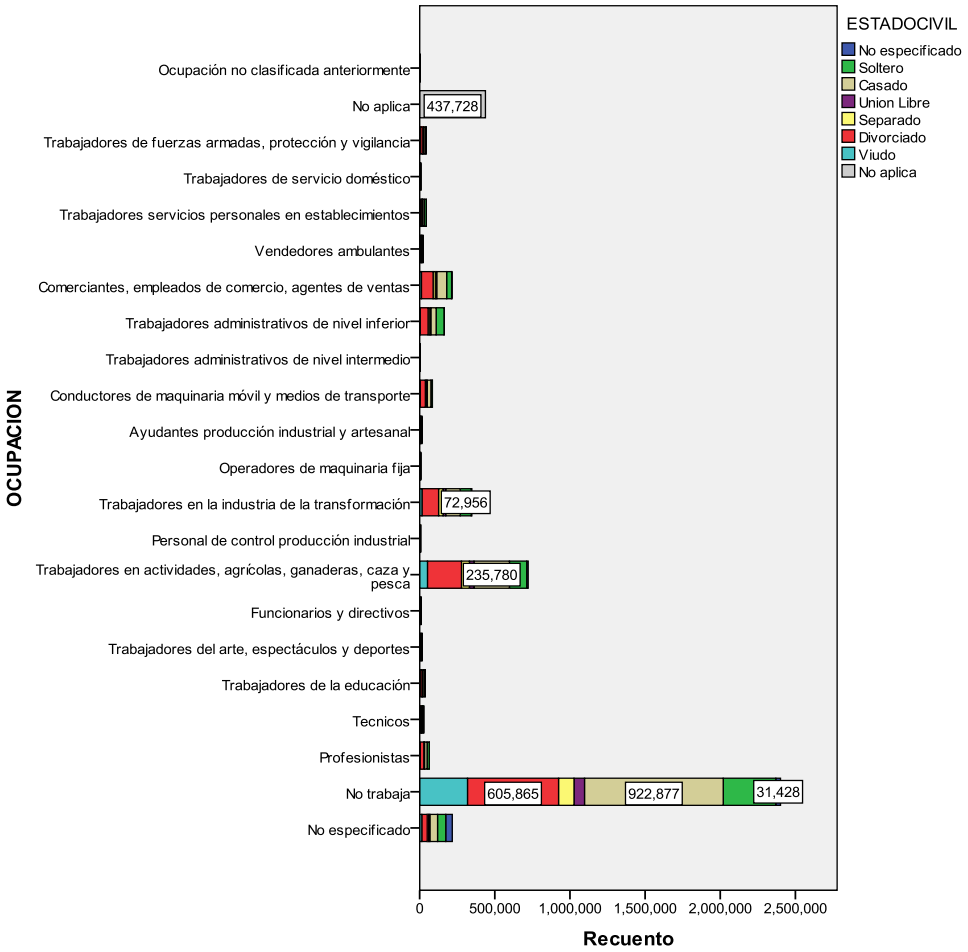


Figura 4.16 Histograma Estado Civil – Ocupación

La gráfica siguiente nos muestra si recibieron o no asistencia médica en todos los estados de la republica, la mayoría de los estados indican que las mayoría de sus habitantes reciben asistencia medica. El Estado de México tiene el mayor número de personas y se observa que el número de personas sin asistencia fue de 72,170 en estos 10 años.

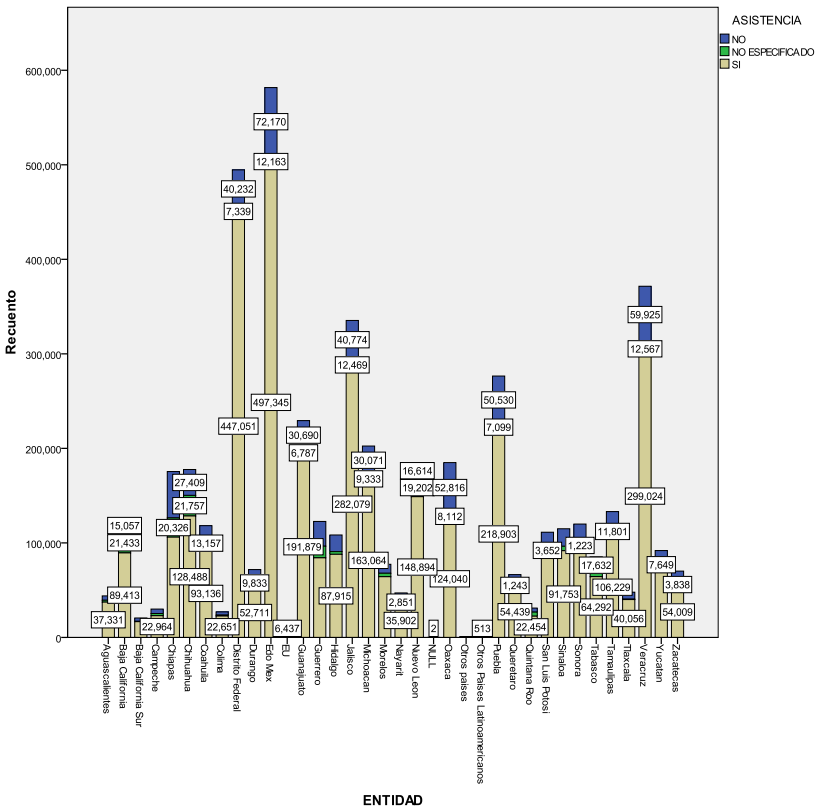


Figura 4.17 Histograma Asistencia Medica – Entidad Residencia

Después del análisis estadístico previo que se realizó anteriormente decidimos que atributos son los más importantes para poder formar la vista minable. Y ha explicar con más detalle cada uno de los atributos que se va a utilizar.

ATRIBUTO	DESCRIPCIÓN
DERECHOHABIENCIA	Columna vertebral de una vista minable, muestra cuantas y cuáles son los organismos médicos públicos o privados en los que se encuentran afiliadas las personas, el nombre del atributo puede cambiar en cada año, los nombres completos se encuentran en el ANEXO (diccionario de datos), pero lo que no cambia es su contenido.
OCUPACIÓN	Ocupación que desempeña la persona afiliada. Para saber si influye mucho o poco la ocupación de las personas en la asignación del organismo médico y si influye en su

ATRIBUTO	DESCRIPCIÓN
	defunción.
ASISTENCIA	Si la persona recibió o no "atención médica", por parte del organismo en el que estuvieron afiliados, en el caso de que haya estado afiliados a alguno.
ESTADO CIVIL	Nos va a mostrar si la persona afiliada a algún organismo puede tener afiliados a sus parientes como son, esposa(o), hijos ó padres, es decir aunque la persona afectada o fallecida no hubiera trabajado pudo pertenecer a algún organismo ya que fue afiliado por algún pariente cercano.
ENTIDAD	Nos indica que organismos están más presentes en cada entidad.

Tabla 4.3 Descripción de atributos

Los valores que puede tomar a lo largo de los 10 años se pueden observar en la Tabla 4.2 Discretización de los atributos (página 56).

Ahora bien ¿Cómo fue que llegamos a la conclusión que estos atributos serían los indicados para esta vista?, esto se logro con las siguientes consultas:

Primero hicimos un conteo de todas las personas que tenían el atributo "asistencia médica" en 1, es decir que si recibieron atención médica y revisamos a que organismo médico está afiliado, haciendo un conteo agrupado por derechohabiencia. Esto se realizó con los 10 años.

```
SELECT COUNT(DERECH)AS CONTEO, ( CASE
WHEN ASISMED= 1   THEN   'SI'
WHEN ASISMED= 2 THEN 'NO'
WHEN ASISMED= 0 THEN 'No especificado'
END) AS ASISTENCIA, DERECH
FROM dbo.def2000
GROUP BY  DERECH, ASISMED
ORDER BY COUNT(DERECH) ASC
```

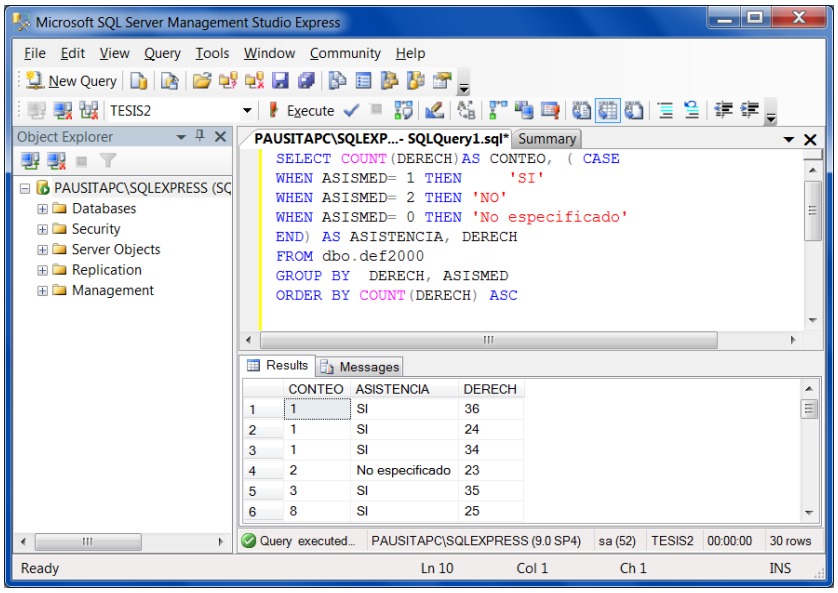


Figura 4.18 Seleccionando SI en asistencia medica

Un atributo que también se utilizó fue el de "entidad de residencia" que nos muestra cuales son las entidades en donde hay más personas que si reciben atención médica.

```
SELECT COUNT(ENTRES)AS CONTEO, ( CASE
WHEN ASISMED= 1 THEN 'SI'
WHEN ASISMED= 2 THEN 'NO'
WHEN ASISMED= 0 THEN 'NO Especificado'
END) AS ASISTENCIA, ENTRES
FROM dbo.def2000
GROUP BY ENTRES, ASISMED
ORDER BY COUNT(ENTRES)ASCASC
```

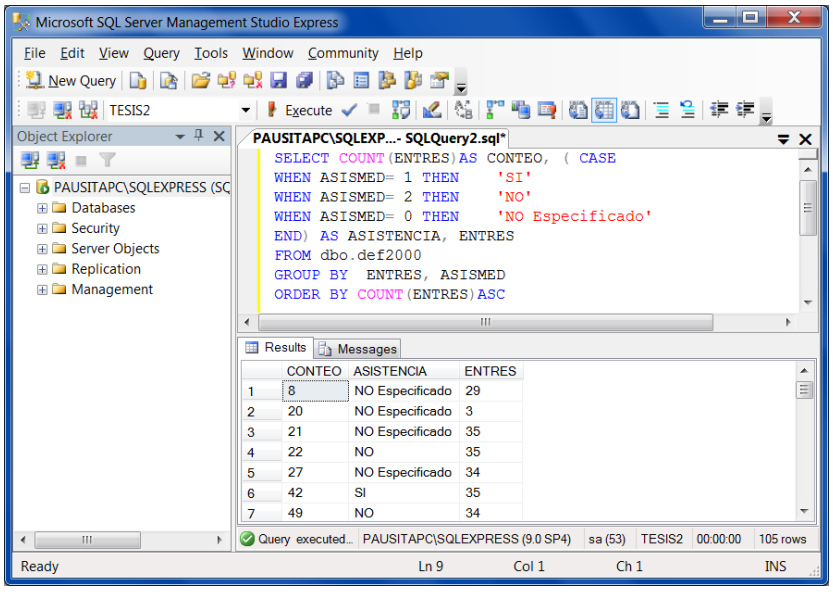


Figura 4.19 Seleccionando Entidad de Recidencia

Otro atributo que también se estudio fue el de “estado civil” el cual nos indica cual era su estado civil en el momento de su fallecimiento, esto nos sirve para saber si la persona afiliada a alguna de las organizaciones fue afiliada directamente por su trabajo o fue afiliado por algun familiar cercano, como pueden ser un esposo o esposa, padres e hijos.

```
SELECT COUNT(EDOCIV)AS CONTEO, ( CASE
WHEN ASISMED= 1   THEN  'SI'
WHEN ASISMED= 2 THEN    'NO'
WHEN ASISMED= 0 THEN    'No especificado'
END) AS ASISTENCIA, EDOCIV
FROM dbo.def2000
GROUP BY  EDOCIV, ASISMED ORDER BY COUNT(EDOCIV)ASC
```

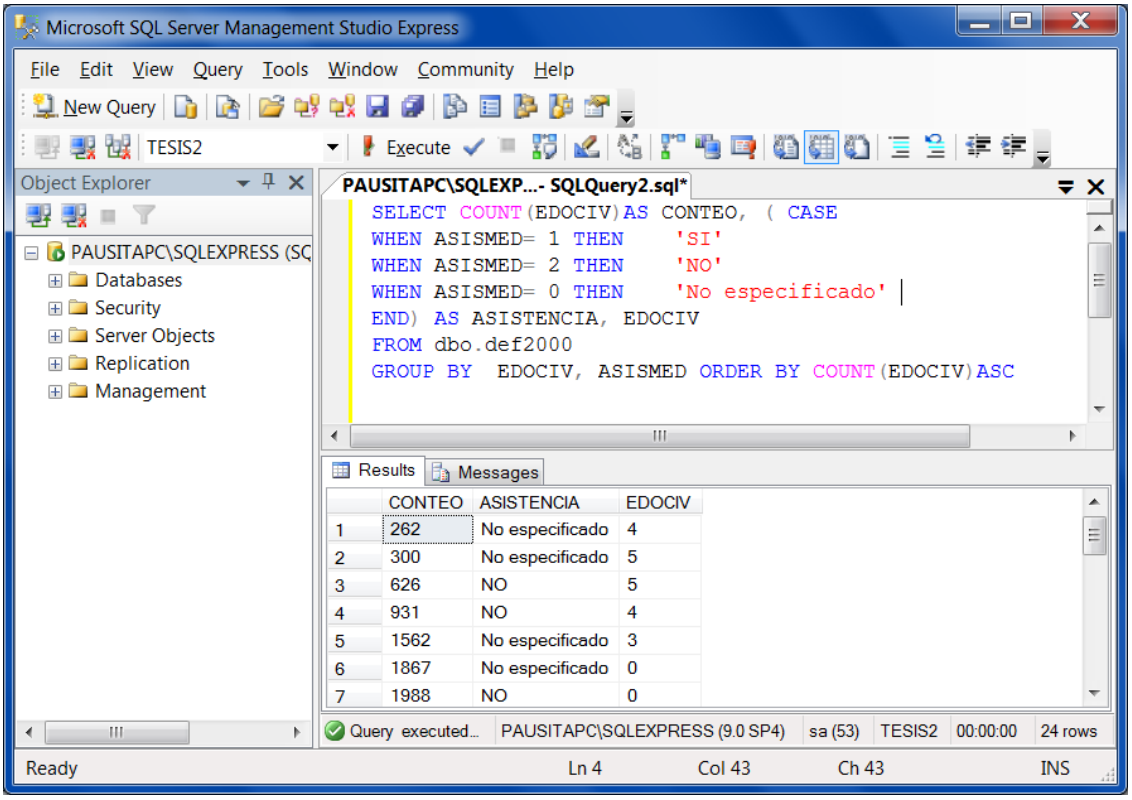


Figura 4.20 Seleccionando Estado Civil

Un atributo que nos muestra mucha información es “ocupación”, con dicho atributo podremos saber que ocupación desempeñaba la persona fallecida y si juntamos esta información con saber a que derechohabiencia pertenecia podremos saber que organismos se asigna dependiendo de cual es tu ocupación.

```
SELECT COUNT(OCUPA)AS CONTEO, ( CASE
WHEN ASISMED= 1   THEN  'SI'
WHEN ASISMED= 2 THEN    'NO'
WHEN ASISMED= 0 THEN    'No especificado'
END) AS ASISTENCIA, OCUPA
FROM dbo.def2000
GROUP BY  OCUPA, ASISMED
```

```
ORDER BY COUNT(OCUPA)ASC
```

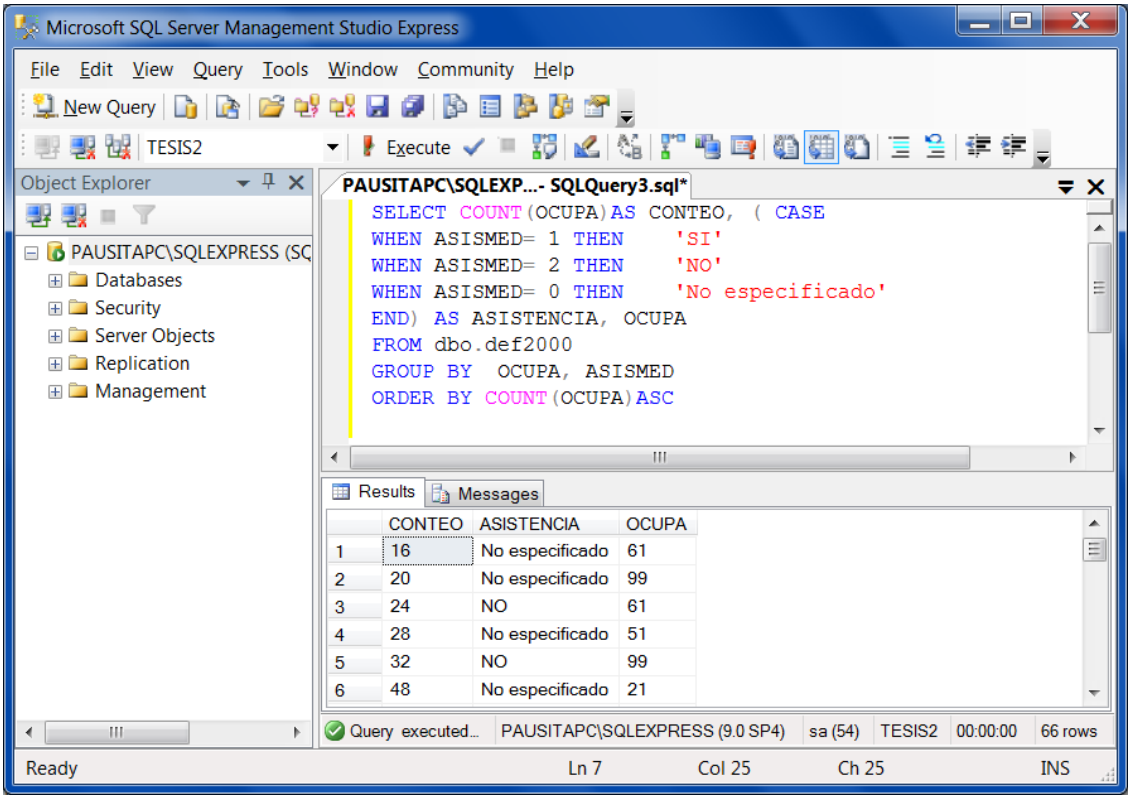


Figura 4.21 Seleccionando Ocupación

Después de todas las consultas realizadas ya podemos formar la vista minable sin discretizar aun, el query completo para la vista de 'derechohabientes' está en el apartado de ANEXO (scrips) al final de esta tesis, al igual que se hizo para este año se hará para los 10 años, que se vayan a analizar.

```
SELECT  DISTINCT DERECH, OCUPA, ENTRES, (CAS
WHEN ASISMED= 1THEN'SI'
WHEN ASISMED= 2 THEN 'NO'
END) AS ASISTENCIA
FROM dbo.def2000
WHERE (ASISMED='1' OR ASISMED='2') AND
DERECH IN (1,2,3,4,5) and edad between 0 and 120
AND EDOCIV IN (1,2,3,4,5,6)
```

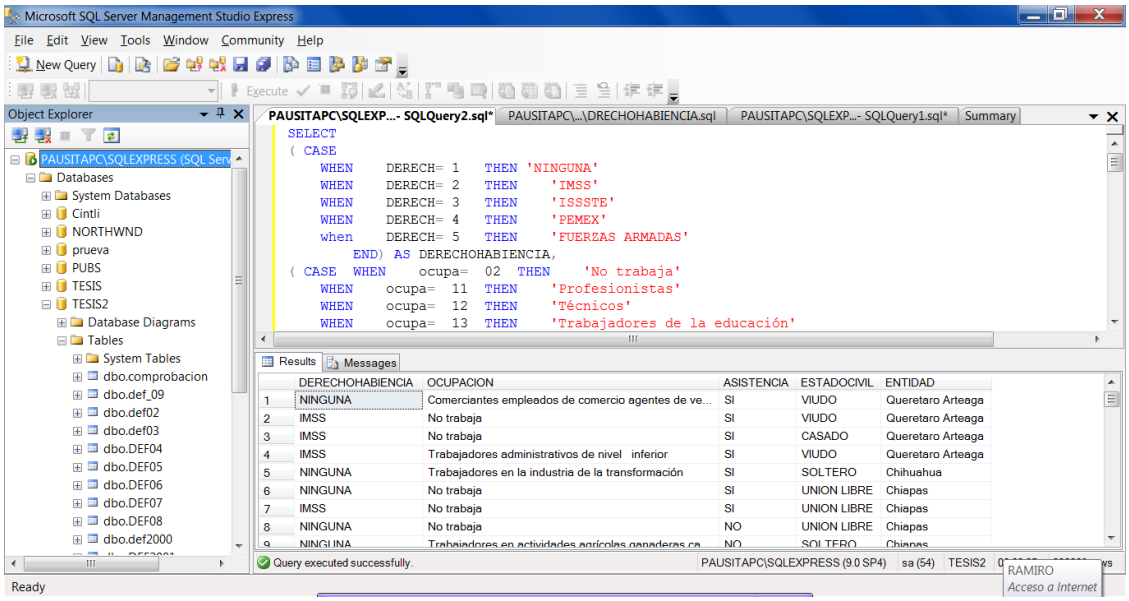


Figura 4.22 Vista Minable sin Discretizar

Ya teniendo nuestra vista minable empezaremos a realizar minería de datos usando el software "WEKA"[4] con esta herramienta podremos hacer minería de datos.

A continuación se muestra un ejemplo de la vista minable que obtuvimos:

DERECHOHABIENCIA	OCUPACION	ASISTENCIA	ESTADOCIVIL	ENTIDAD
IMSS	No trabaja	SI	CASADO	Aguascalientes
NINGUNA	Trabajadores en la industria de la transformaci¢n	SI	CASADO	Aguascalientes
NINGUNA	No trabaja	SI	VIUDO	Aguascalientes
NINGUNA	Trabajadores en actividades, agr¡colas, ganaderas, caza y pesca	NO	SOLTERO	Aguascalientes
NINGUNA	No trabaja	NO	VIUDO	Aguascalientes
NINGUNA	Trabajadores en actividades, agr¡colas, ganaderas, caza y pesca	NO	SOLTERO	Aguascalientes
NINGUNA	No trabaja	SI	CASADO	Aguascalientes
IMSS	Trabajadores en actividades, agr¡colas, ganaderas, caza y pesca	SI	CASADO	Aguascalientes
NINGUNA	Comerciantes, empleados de comercio, agentes de ventas	SI	CASADO	Aguascalientes
NINGUNA	No trabaja	SI	CASADO	Aguascalientes
NINGUNA	Trabajadores en actividades, agr¡colas, ganaderas, caza y pesca	SI	VIUDO	Aguascalientes
NINGUNA	Comerciantes, empleados de comercio, agentes de ventas	SI	VIUDO	Aguascalientes
IMSS	Trabajadores en la industria de la transformaci¢n	NO	SOLTERO	Aguascalientes
IMSS	No trabaja	SI	CASADO	Aguascalientes
NINGUNA	Conductores de maquinaria m¢vil y medios de transporte	SI	CASADO	Aguascalientes

Figura 4.23 Muestra de la Vista Minable

Como ya se habia mencionado, la herramienta que se ocupo para la minería de datos fue WEKA-3-7-2, esta herramienta presenta problemas de capacidad de memoria y esto ocasionaba que no nos abriera la vista minable que obtuvimos de SQLServer, por lo cual se decidió trabajar con WEKA desde MS-DOS para así poder extender la capasidad de la memoria en WEKA. En este caso tampoco fue posible trabajar con los 10 años en conjunto ya que aun extendiendo la memoria no hay suficiente espacio como se observa en la figura 4.24, por lo que se trabajara con el año 2000 y con el año 2009 solamente.

[4] WEKA(Waikato Environment for Knowledge Analysis - *Entorno para Análisis del Conocimiento de la Universidad de Waikato*): Es un software programado en Java que está orientado a la extracción de conocimientos desde bases de datos

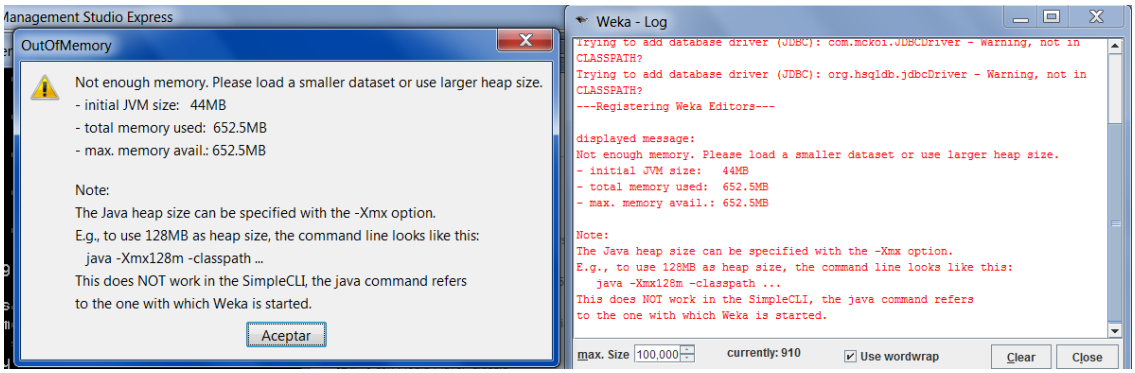


Figura 4.24 Error de memoria WEKA

WEKA se distribuye como un fichero ejecutable comprimido de java (fichero"jar"), que se invoca directamente sobre la máquina virtual JVM. Y lo ejecutaremos de la siguiente manera:

Figura 4.25 Ejecución WEKA

Una vez invocada, aparece la ventana de entrada a la interfaz gráfica (*GUIChooser*), que nos ofrece cuatro opciones posibles de trabajo:

- **Explorer**
- **Experimenter**
- **KnowledgeFlow**
- **Simple CLI**

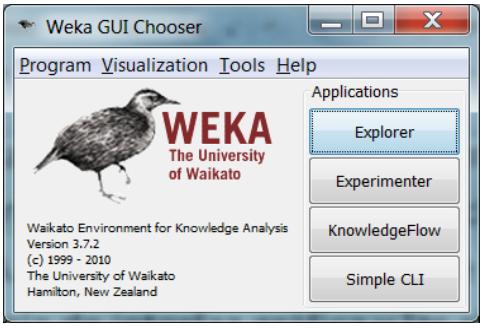


Figura 4.26 Pantalla Inicial WEKA

Nos centraremos únicamente en la primera opción (Explorer). Una vez seleccionada, se crea una ventana con 6 pestañas en la parte superior que se corresponden con diferentes tipos de operaciones, en etapas independientes, que se pueden realizar sobre los datos:

- Preprocess: selección de la fuente de datos y preparación (filtrado).
- Clasify: Facilidades para aplicar esquemas de clasificación, entrenar modelos y evaluar su precisión
- Cluster: Algoritmos de agrupamiento
- Associate: Algoritmos de búsqueda de reglas de asociación
- Select Attributes: Búsqueda supervisada de subconjuntos de atributo representativos
- Visualize: Herramienta interactiva de presentación gráfica en 2D.

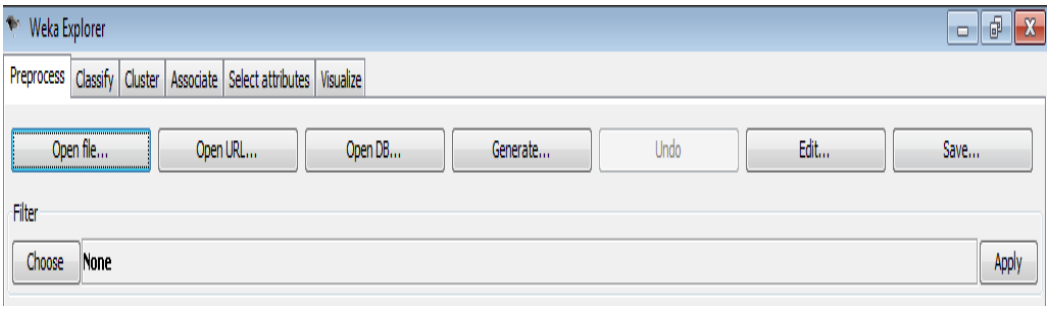


Figura 4.27 WEKA Explorer

Una vez ya modificada la memoria y permitiéndonos el acceso a WEKA, empezamos a trabajar con el primer conjunto de datos que abarca el años 2000 y así igualmente con el año 2009, ya que tenemos la vista la cargamos en WEKA y empezamos a hacer minería de datos.

La herramienta de WEKA acepta archivos .arff data file y archivos .csv el cual fue el que usamos ya que para nosotros es más fácil la manipulación de los archivos .csv.

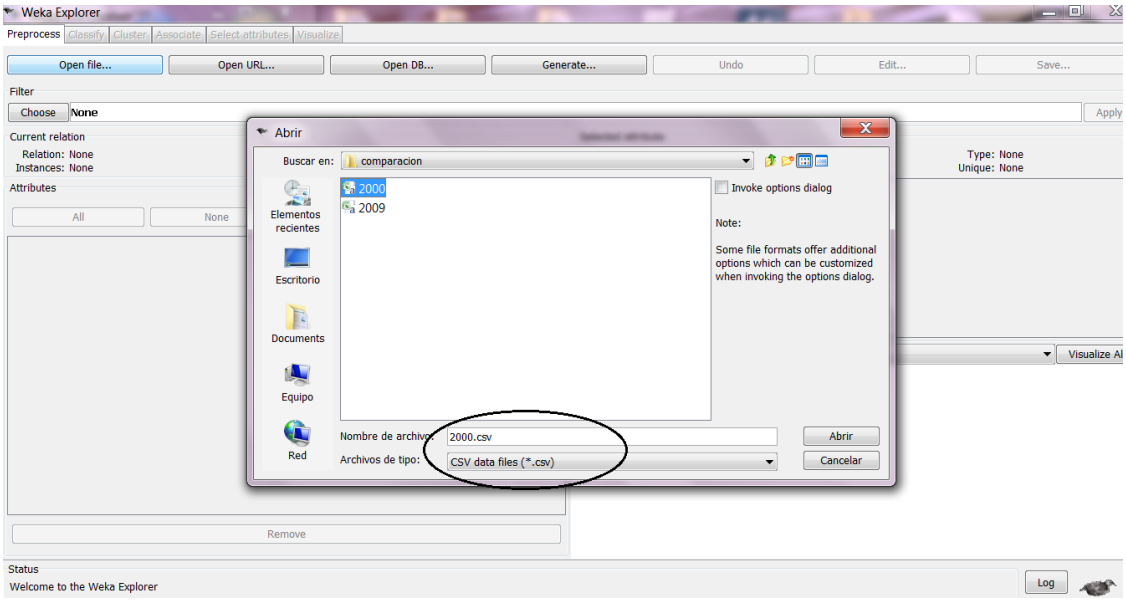


Figura 4.28 Archivo 2000.csv

Vamos a meter nuestra vista a WEKA teniendo una interface de entrada bastante amigable donde seleccionamos la fuente de los datos y hacemos la preparación:

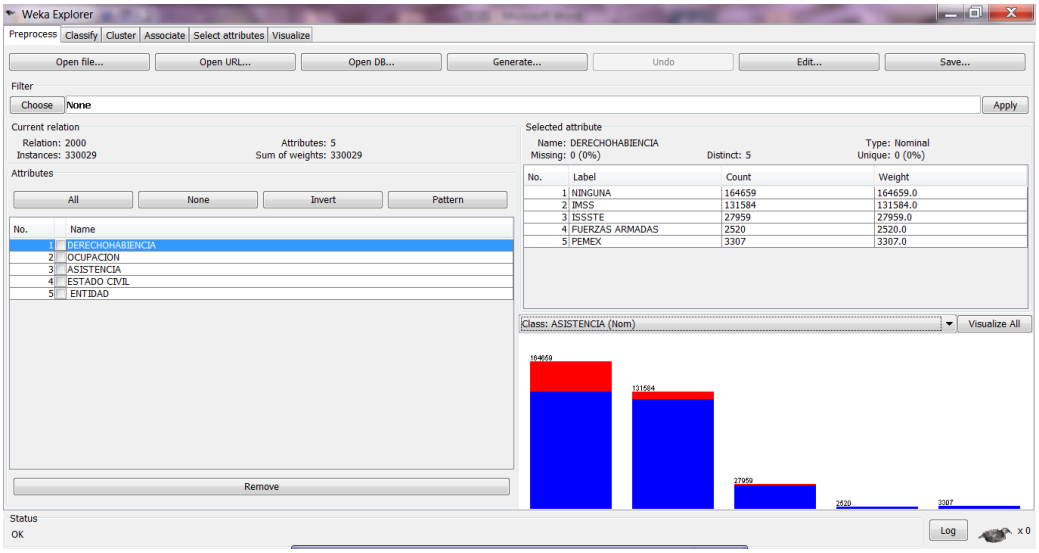
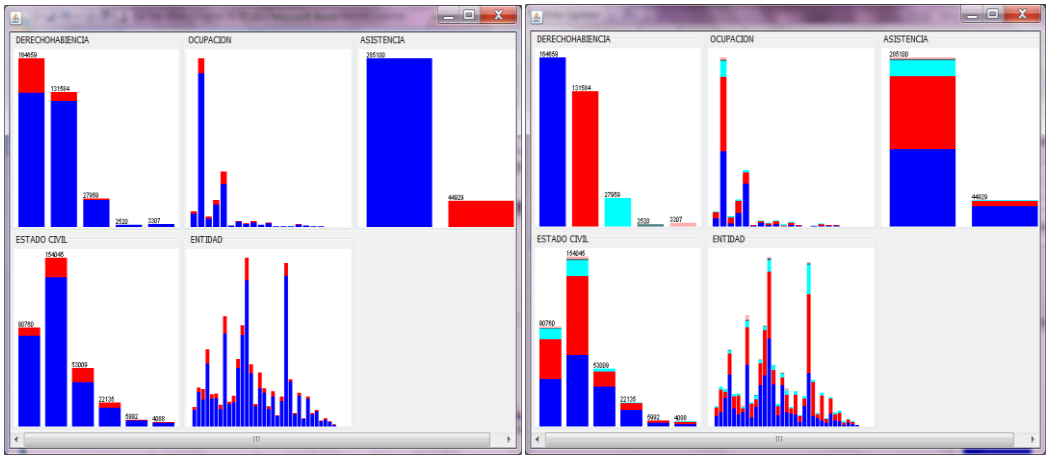


Figura 4.29 WEKA Explorer Preprocess

Una vez cargados los datos, aparece un cuadro resumen, donde se observa el número de instancias y el número de atributos. Con un listado de todos los atributos disponibles, con los nombres especificados en el fichero, de modo que se pueden seleccionar para ver sus detalles y propiedades.

Donde podemos observar atributos como un conteo de registros y una ventana donde podemos hacer histogramas con nuestros atributos, en este caso aremos un histograma

con el atributo de asistencia como base, ya que es uno de los importantes para nuestra minería. Siendo 'si' en asistencia y 'no' en asistencia como se observa en la Figura a).



a) Asistencia: si (azul) , no (rojo) b) Derechohabiencia

Figura 4.30 Histogramas Respecto a Asistencia

Se observa en nuestro histograma a) la asistencia en si fue de 285100 para 2001 y el total de registros con no asistencia es de 44929 registros, se observa que en el caso de no asistencia la mayoría de registros esta en 'ninguna' derechohabiencia y estado civil 'casado'.

En el histograma b) se observa que la mayoría de incidencias en derechohabiencia es 'ninguna' esto nos dice que en derechohabiencia la mayoría de las defunciones no tenían ningún servicio de salud, en las entidades de Edo. De México y DF.

En este caso todos nuestros atributos son nominales esto hace que podamos ocupar las reglas de asociación.

En la parte de Visualización de WEKA se observan las instancias en 2D que relacionan atributos. Al seleccionar la opción **Visualize** del *Explorer* aparecen todos los pares posibles de atributos en las coordenadas horizontal y vertical. La idea es que se seleccione la gráfica deseada para verla en detalle en una ventana nueva.

Vamos a visualizar variables que sean de nuestro interés.

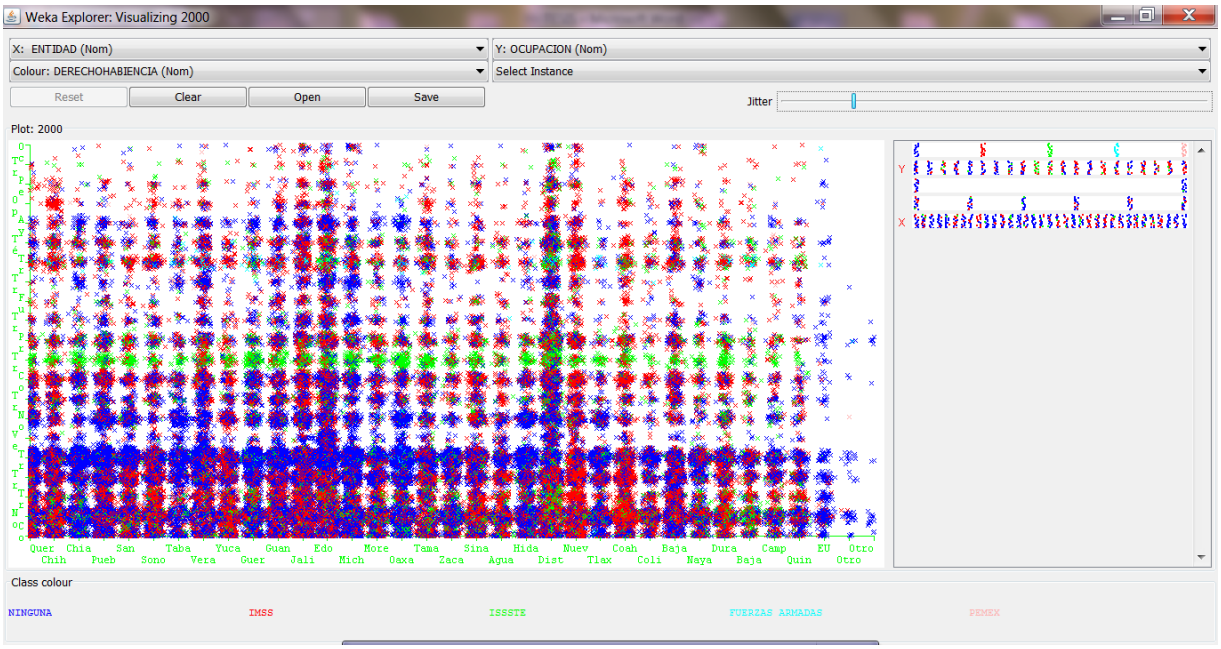


Figura 4.31 Visualize de WEKA

Podemos observar como en una ocupación en especial la institución que es mayoría es el ISSSTE (Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado) en color verde, siendo la ocupación “Trabajadores de la educación”, esto es en todos los estados de la republica. Cabe mencionar que esto se mantuvo en todos los años.

Al seleccionar ocupación en X y derechohabiencia en el eje Y se observa a las ocupaciones y la institución a la que pertenecían y nuestro indicador es asistencia en ‘SI’ azul y ‘NO’ rojo, en este grafico podemos observar varias cosas, donde se concentran nuestros puntos rojos mayormente es en NINGUNA derechohabiencia y también observamos una acumulación en la intersección de fuerzas armadas y ocupación “trabajadores de las fuerzas armadas”. Lo vemos mejor en la siguiente pantalla donde ya esta filtrado.

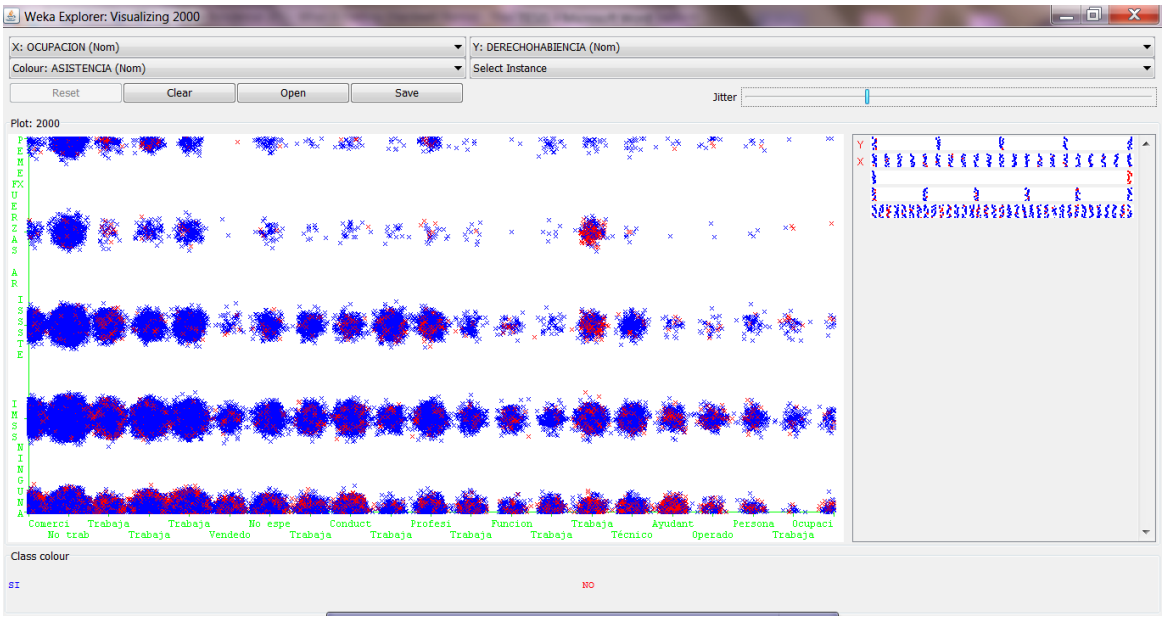


Figura 4.32 Visualización Asistencia

Ahora en esta gráfica podemos observar dos grupos que se distinguen de los demás, el de color rojo que nos hace referencia a estado civil ‘viudo’ en la ocupación ‘no trabaja’ con una relación muy estrecha en todos los estados. Y la segunda es en color verde estado civil ‘soltero’ y la ocupación ‘ayudante producción industrial y artesanal’.

La visualización nos ayuda mucho para ver antes te meter los datos a una técnica o tarea de minería de datos.

En la siguiente pantalla podemos ver la asistencia en ‘si’ y en ‘no’ y los colores representan a las instituciones de salud, se puede observar que para la asistencia en ‘si’ domina mas el color rojo que representa al IMSS(Instituto Mexicano del Seguro Social) y en asistencia ‘no’ el color azul que representa ninguna derechohabiencia.

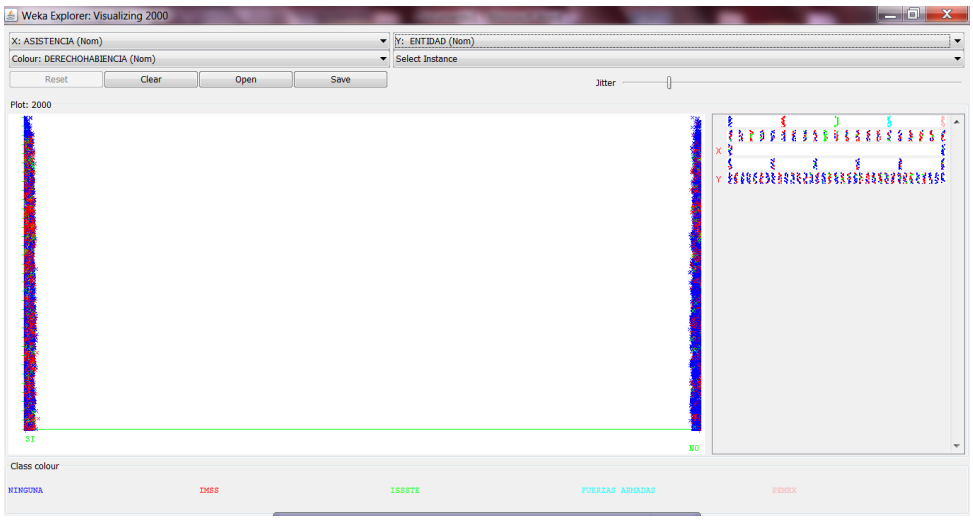


Figura 4.33 Visualización Derechohabiencia-Entidad

4.2.1 Reglas de asociación

Los algoritmos de asociación permiten la búsqueda automática de reglas que relacionan conjuntos de atributos entre sí. Son algoritmos no supervisados, en el sentido de que no existen relaciones conocidas a priori con las que contrastar la validez de los resultados, sino que se evalúa si esas reglas son estadísticamente significativas.
Utilizaremos el algoritmo a priori en las reglas de asociación ya que es el más común y primero no meteremos ningún filtro a nuestra vista de "derechohabientes".

La ventana de Asociación (**Associate** en el Explorer), tiene los siguientes elementos:

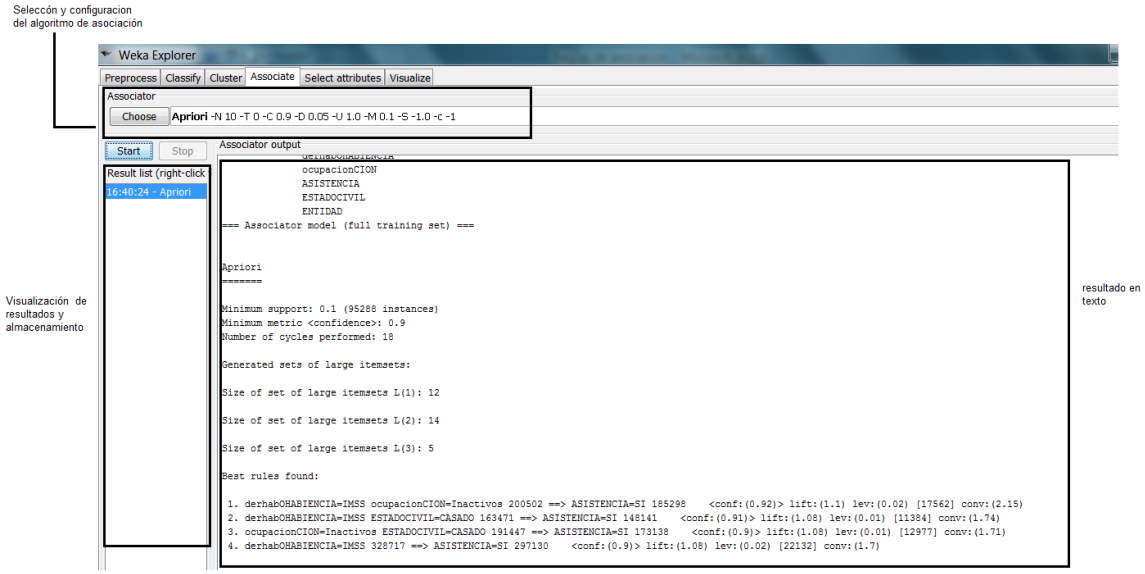


Figura 4.34 WEKA Explorer Associate

Con cada una de las vistas de los años 2000 y 2009 que son los años que se estará trabajando, se encuentran las reglas más generales y algunos casos específicas que se van a extraer por cada organización (IMSS, ISSSTE, etc.).

AÑO 2000

Se empezara a estudiar y a obtener las reglas de asociación del año 2000, para esto como se comento antes se dividió por organización, para el año 2000 hubo tres divisiones que fueron:

- IMSS
- ISSSTE
- FUERZAS ARMADAS , PEMEX(Petróleos Mexicanos)

Es importante mencionar que para hacer nuestras reglas de asociación se utilizo un criterio muy importante lo ilustramos con la figura 4.35.

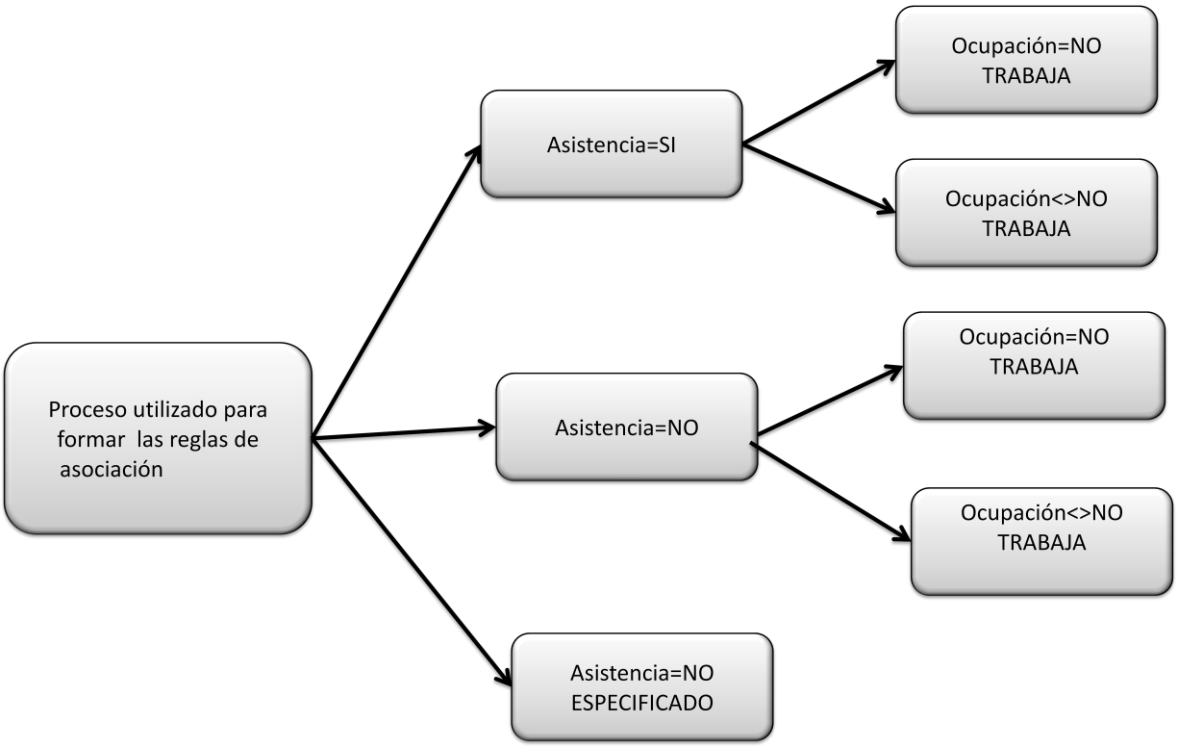


Figura 4.35 Criterio de filtros

Con base en el criterio antes mencionado, se filtrarán ahora los atributos para así hacer una búsqueda de reglas de asociación más detallada.

La forma en cómo se puede filtrar la información en el WEKA es en la pestaña de preprocess y en la parte de filtros utilizamos el filtro no supervisado de RemoveWithValues.

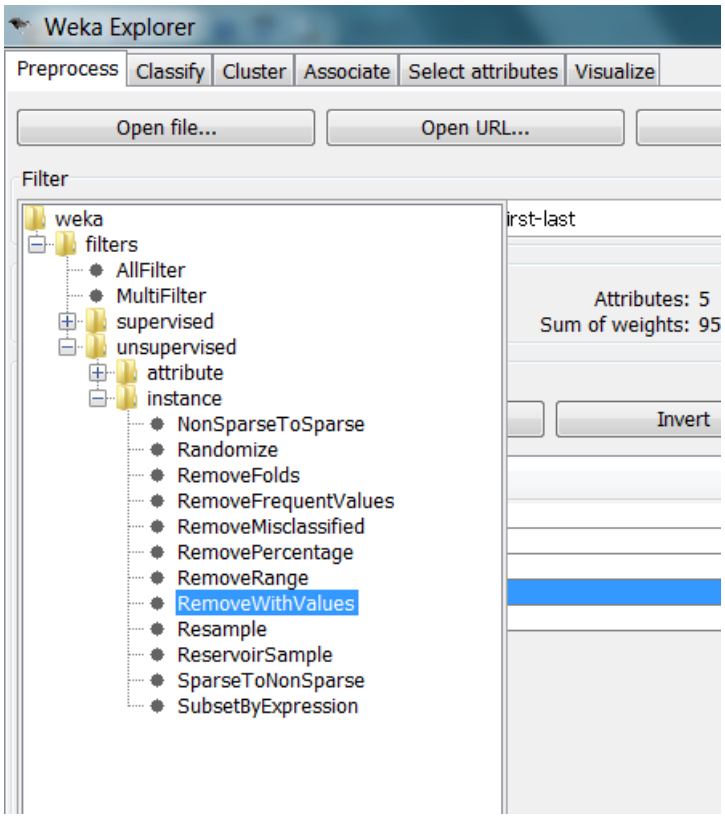
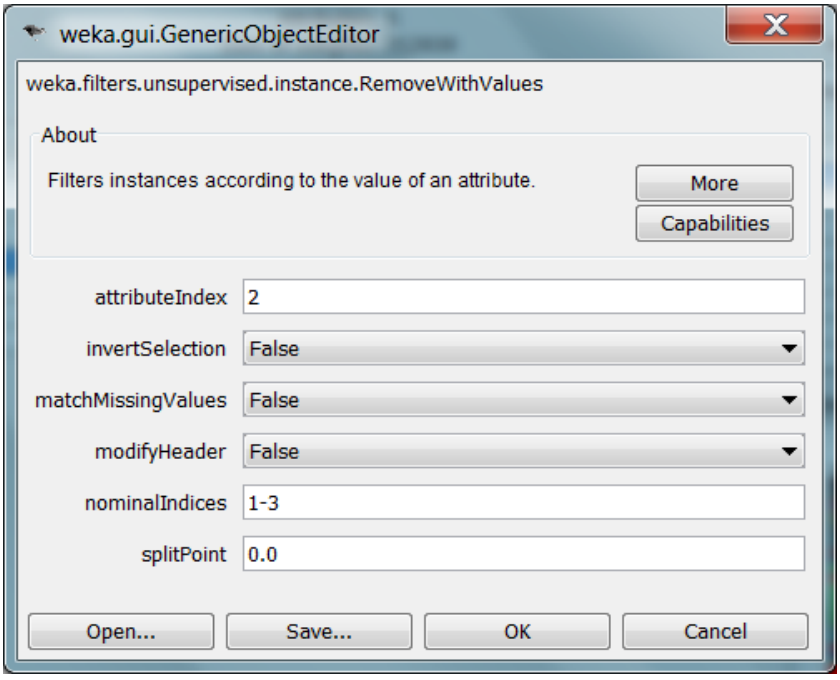


Figura 4.36 Filtros en WEKA



4.37 Configuración de la instacia RemoveWhitValues

Una vez hechos los filtros correspondientes vamos a empezar a revisar y estudiar las primeras reglas de asociación del IMSS, las reglas que se obtuvieron fueron las siguientes:

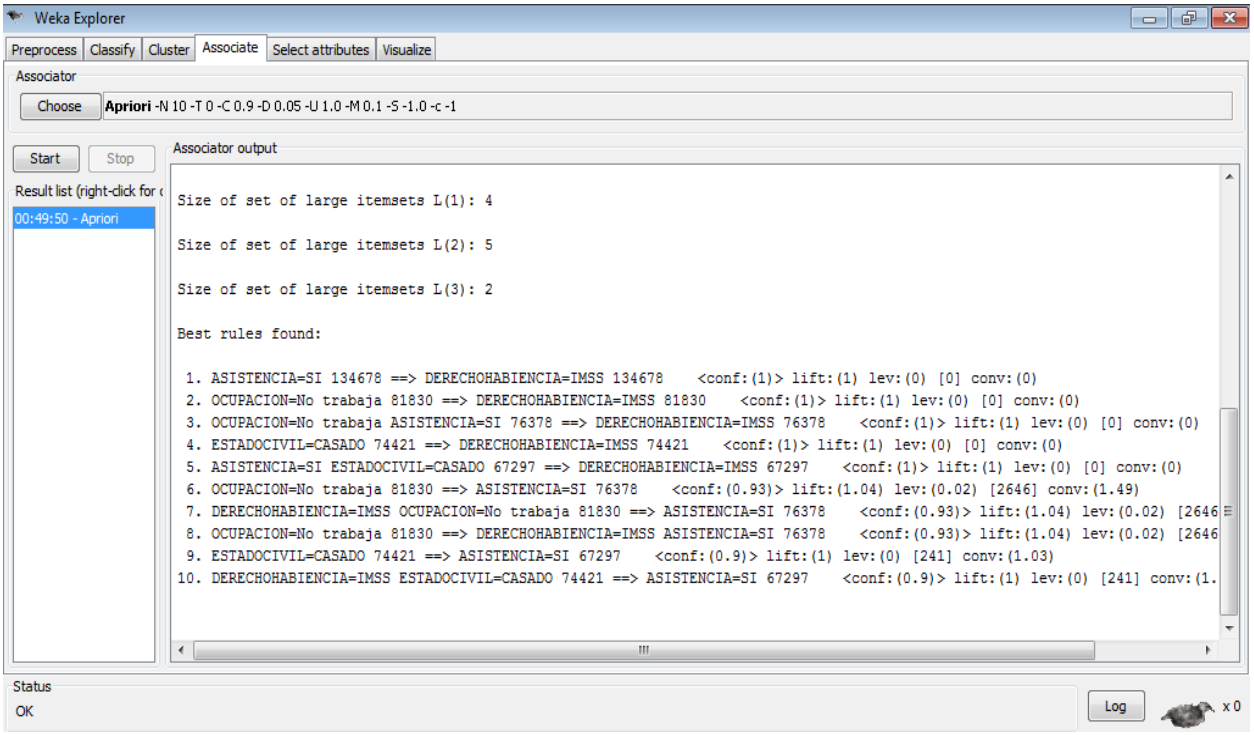


Figura 4.38 1er resultado de reglas de asociación

En la pantalla anterior nos muestra todas las reglas que obtiene WEKA sin hacer ningún tipo de filtro, de estas reglas tenemos que quedarnos con las que realmente sean importantes, valiosas y únicas, para esto utilizamos los filtros ya antes mencionados.

La pantalla que veremos a continuación ya se muestra con los filtros es decir ya nos muestra reglas con más detalles, de las siguientes reglas que se muestran solo agarramos reglas que sean únicas, para que la información no sea tan repetitiva y de cada una de estas reglas que se escogen se dará una explicación a detalle.

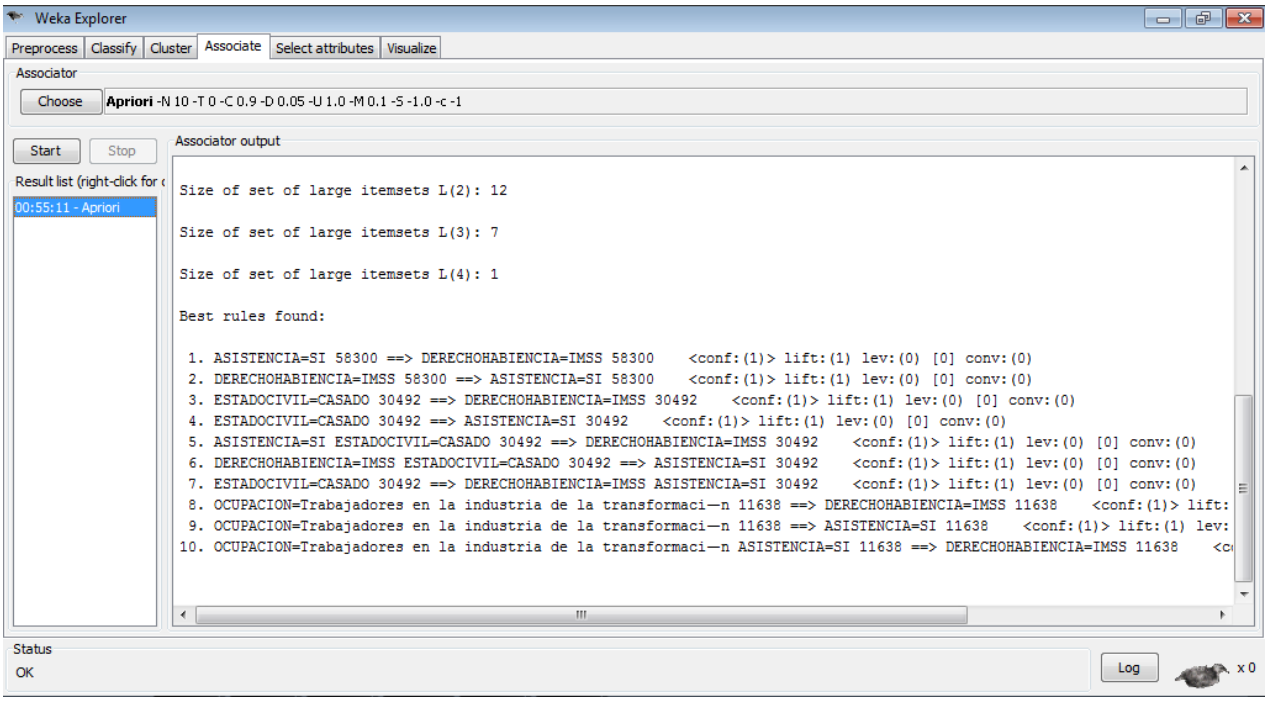


Figura 4.39 Reglas con filtros

Como se menciono antes etas son las reglas que se obtuvieron y que fueron escogidas como valiosas:

- ASISTENCIA=SI ==> DERECHOHABIENCIA=IMSS==> OCUPACION=NO TRABAJA76378

5. ASISTENCIA=SI ESTADOCIVIL=CASADO 67297 ==> DERECHOHABIENCIA=IMSS 67297 <conf:(1)> lift:(1)
lev:(0) [0] conv:(0)

- ASISTENCIA=SI ==> DERECHOHABIENCIA=IMSS==> OCUPACION=TRABAJA 58300

10. OCUPACION=Trabajadores en la industria de la transformación ASISTENCIA=SI 11638 ==> DERECHOHABIENCIA=IMSS 11638 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

De las siguientes reglas de asociación lo que más podemos destacar es que, hay más personas que recibe atención médica que la que no, es importante mencionar que hay un pequeño porcentaje que se queda indefinido es decir no está especificado si recibió o no atención médica, otros datos interesantes, es que de las personas afiliadas al IMSS y que fallecieron su ocupación esta como no trabajan, esto puede ser porque suelen ser familiares cercanos de los trabajadores (hijos, esposa(o), padres, etc.). Las ocupaciones que más resaltan es la de "Trabajadores en la industria de la transformación", por ultimo algo que también sobresalió en estas reglas fue el estado civil que la mayoría son personas casadas, todo esto fue resultado de la institución del IMSS.

Nuestra siguiente institución a estudiar y analizar es el ISSSTE, comenzaremos con mostrar sus reglas de asociación resultantes

- ASISTENCIA=SI ==> DERECHOHABIENCIA=ISSSTE==> OCUPACION=NO TRABAJA 15811

5. ASISTENCIA=SI ESTADOCIVIL=CASADO 13326 ==> DERECHOHABIENCIA=ISSSTE 13326 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

- ASISTENCIA=SI ==> DERECHOHABIENCIA=ISSTES==> OCUPACION= NO TRABAJA 15811

10. ASISTENCIA=SI ENTIDAD=Distrito Federal 2685 ==> DERECHOHABIENCIA=ISSSTE 2685 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

- ASISTENCIA=SI ==> DERECHOHABIENCIA=ISSSTE==> OCUPACION=NO TRABAJA 15811

5. ASISTENCIA=SI ESTADOCIVIL=VIUDO 1494 ==> DERECHOHABIENCIA=ISSSTE 1494 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

- DERECHOHABIENCIA=ISSSTE==> OCUPACION=TRABAJA 11704

8. OCUPACION=Trabajadores administrativos de nivel inferior 257 ==> DERECHOHABIENCIA=ISSSTE 257 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

Lo que se observo de las reglas de asociación anteriores fue que son mucho más las personas que reciben atención médica en la institución del ISSSTE, la gran mayoría de las personas afiliadas a esta institución no trabajan, y hablando de personas que si trabajan la ocupación que más destaca es la de "Trabajadores administrativos de nivel inferior". Su estado civil son más los casados siguiéndolos con una cantidad no despreciable los que son viudos. El estado en donde se hace más presente el ISSSTE es el Distrito Federal.

Por último estudiaremos el tercer grupo que son las Fuerzas Armadas y Pemex, como se hizo en las anteriores primero mostraremos sus reglas de asociación resultantes, para después pasar a estudiarlas y analizarlas.

- ASISTENCIA=SI ==> DERECHOHABIENCIA=PEMEX==> OCUPACION=NO TRABAJA 4100

3. DERECHOHABIENCIA=PEMEX OCUPACION=No trabaja ESTADOCIVIL=CASADO 1356 ==> ASISTENCIA=SI 1287 <conf:(0.95)> lift:(1.05) lev:(0.01) [57] conv:(1.8)

- ASISTENCIA=SI ==> DERECHOHABIENCIA=FUERZAS ARMADAS==> OCUPACION=NO TRABAJA 4100

7. DERECHOHABIENCIA=FUERZAS ARMADAS OCUPACION=No trabaja 1707 ==> ASISTENCIA=SI 1606 <conf:(0.94)> lift:(1.04) lev:(0.01) [57] conv:(1.56)

- ASISTENCIA=NO ==> DERECHOHABIENCIA=FUERZAS ARMADAS==> OCUPACION=NO TRABAJA 1728

1. DERECHOHABIENCIA=FUERZAS ARMADAS 257 ==> ASISTENCIA=NO 257 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

- ASISTENCIA=NO ==> DERECHOHABIENCIA=FUERZAS ARMADAS==> OCUPACION=TRABAJA 1728

9. DERECHOHABIENCIA=FUERZAS ARMADAS OCUPACION=Trabajadores de fuerzas armadas, protecci—n y vigilancia 110 ==> ASISTENCIA=NO 110 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

- ASISTENCIA=NULL ==> DERECHOHABIENCIA=FUERZAS ARMADAS==> OCUPACION=TRABAJA 261

8. DERECHOHABIENCIA=FUERZAS ARMADAS ESTADOCIVIL=CASADO 38 ==> ASISTENCIA=NULL 38 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

Las conclusiones que obtuvimos de estas reglas, fue que hay más gente afiliada a Pemex que a las Fuerzas Arma, en ambas hay más personas que si reciben atención médica que las que no. También en esta vemos que las personas que se afilian a estas instituciones no trabajan y de las que si trabajan la mayoría resulto que desempeñan la ocupación de "trabajadores de fuerzas armadas, protección y vigilancia. El estado civil que presentan las victimas es de casados.

Una vez que es analizada cada una de las instituciones procedemos a ilustrar con un cuadro que nos muestra de manera más general y numérica como se están moviendo los registros, es decir qué cantidad de registros pose las instituciones, dependiendo de las posibles opciones o caminos que tomamos para obtener las reglas ya antes mencionadas.

INSTITUCIÓN	ASISTENCIA=SI		ASISTENCIA=NO		ASISTENCIA=NO ESPECIFICADA	TOTAL
	TRABAJA	NO TRABAJA	TRABAJA	NO TRABAJA		
IMSS	76378	58300	6425	3187	5181	149471
ISSSTE	15811	11704	1168	710	744	30137
OTRAS	4100	1728	261	163	174	6426
NINGUNA						204713

Tabla 4.4 Tabla de Derechohabiencia 2000

*OTRAS=PEMEX Y FUERZAS ARMADAS

Como observamos un dato que es muy importante y que es un tanto alarmante ya que lo que nos muestra el cuadro es que hay muchas personas que en el 2000 no estaban afiliadas a ninguna institución, después lo que podemos observar es que el IMSS es el que siempre arroja los mayores números es decir el IMSS es el que tiene a más personas afiliadas y por tal motivo es el que va a presentar más incidencias en las diferentes situaciones siguiéndolo ISSSTE y final mente terminando con Pemex y las Fuerzas Armadas

Para concluir con el año 2000 hacemos una comparación entre las instituciones más potenciales es decir IMSS e ISSSTE en este caso al igual que analizaremos a los que no estaban afiliados a ninguna institución.

Del total de derechohabientes cual es el porcentaje de defunciones de cada institución, tenemos que la institución con mas defunciones respecto a su 100% de derechohabientes es el IMSS, con 0.33%. Este porcentaje es bajo con considerando que son más de 45 millones de personas afiliadas a esta institución.

INSTITUCIÓN	Tota de derechohabientes	Total de defunciones	Porcentaje
IMSS	45 053 710	149 471	0.3317%
ISSSTE	10 065 861	30 137	0.2993%

Tabla 4.5 Tabla IMSS-ISSSTE 2000

Ahora bien ahora empezaremos a estudiaremos a los defunciones sin asistencia, en donde se tomara como 100% al total de defunciones por cada institución, lo que nos da como resultado al IMSS.

INSTITUCIÓN	Tota de defunciones	Total de defunciones sin asistencia	Porcentaje
IMSS	149 471	9612	6.43%
ISSSTE	30 137	1878	6.23%

Tabla 4.6 Tabla IMSS-ISSSTE sin asistencia 2000

Tomando en cuenta las defunciones con asistencia la institución con mayor porcentaje es el ISSSTE, como se observa en la tabla siguiente.

INSTITUCIÓN	Tota de defunciones	Total de defunciones con asistencia	Porcentaje
IMSS	149 471	134678	90.1%
ISSSTE	30 137	27515	91.2%

Tabla 4.7 Tabla IMSS-ISSSTE con asistencia 2000

Para las defunciones que no tuvieron ninguna afiliación a alguna institución, podemos hacer una comparación con la población de nuestro país en el año 2000. En México en el 2009 había 100.3 millones de mexicanos.

INSTITUCIÓN	Población total en el 2000	Población sin derechohabiencia	Porcentaje
Ninguna	100,349,766	204713	0.2039%

Tabla 4.8 Tabla IMSS-ISSSTE sin derechohabiencia 2000

AÑO 2009

Para el año 2009, se realizó un estudio de reglas de asociación, por cada institución de la siguiente manera:

- IMSS
- SEGURO POPULAR
- ISSSTE
- SECRETARÍA DE MARINA, PEMEX Y SECRETARIA DE LA DEFENSA NACIONAL.
- NINGUNA

A continuación se dará una explicación de las reglas de asociación más relevantes por cada sección mencionada anteriormente.

IMSS

- ASISTENCIA=NO ==> DERECHOHABIENCIA=IMSS ==> OCUPACION=TRABAJA 9015

5. ASISTENCIA=NO ESTADO CIVIL=CASADO 4766 ==> DERECHOHABIENCIA=IMSS 4766 conf:(1)

10. OCUPACION=Trabajadores administrativos de nivel inferior ASISTENCIA=NO 1913 ==> DERECHOHABIENCIA=IMSS 1913 conf:(1)

Se observa que las reglas relevantes para el IMSS son personas que trabajaban, casadas y no recibieron asistencia esto con 4766 registros con una confianza de 1 y trabajadores administrativos de nivel inferior derechohabientes del IMSS que trabajaban y que no recibieron asistencia.

SEGURO POPULAR

- ASISTENCIA=NO ==> DERECHOHABIENCIA=SEGURO POPULAR ==> OCUPACION=TRABAJA 6056

5. OCUPACION=Trabajadores en actividades agrícolas ganaderas caza y pesca ASISTENCIA=NO 3189 ==> DERECHOHABIENCIA=SEGURO POPULAR 3189 conf:(1)

10. ASISTENCIA=NO ESTADO CIVIL=CASADO 2560 ==> DERECHOHABIENCIA=SEGURO POPULAR 2560 conf:(1)

- ASISTENCIA=SI ==> DERECHOHABIENCIA=SEGURO POPULAR ==> OCUPACION=TRABAJA 29866

7. OCUPACION=Trabajadores en actividades agrícolas ganaderas caza y pesca 11075 ==> DERECHOHABIENCIA=SEGURO POPULAR ASISTENCIA=SI 11075 conf:(1)

10. ASISTENCIA=SI ESTADO CIVIL=CASADO 10538 ==> DERECHOHABIENCIA=SEGURO POPULAR 10538 conf:(1)

Para el Seguro Popular tenemos trabajadores agrícolas que hayan o no hayan recibido asistencia médica, aunque eran derechohabientes de seguro popular. Personas casadas que no recibieron asistencia médica con 2560 registros. De igualmente las personas casadas que aunque si recibieron asistencia fallecieron con 10538 reg.

ISSSTE

- ASISTENCIA=NO ==> DERECHOHABIENCIA=ISSSTE ==> OCUPACION=TRABAJA 1847

5. ASISTENCIA=NO ESTADO CIVIL=CASADO 1068 ==> DERECHOHABIENCIA=ISSSTE 1068 conf:(1)

10. OCUPACION=Trabajadores administrativos de nivel inferior ASISTENCIA=NO 353 ==> DERECHOHABIENCIA=ISSSTE 353 conf:(1)

- ASISTENCIA=SI ==> DERECHOHABIENCIA=ISSSTE ==> OCUPACION=TRABAJA 13217

5. ASISTENCIA=SI ESTADO CIVIL=CASADO 7708 ==> DERECHOHABIENCIA=ISSSTE 7708 conf:(1)

10. ASISTENCIA=SI ENTIDAD=Distrito Federal 2597 ==> DERECHOHABIENCIA=ISSSTE 2597 conf:(1)

Para el ISSSTE las personas que son casadas, trabajan y no tubieron asistencia con 1068 registros. Igualmente para trabajadores de nivel inferior que no tuvieron asistencia aunque pertenecía la ISSSTE con 353 registros. Aunque si recibieron asistencia y son casados con 7708 registros. Se formo una regla con asistencia en sí, DF e ISSSTE con 2527 registros con personas que trabajaban.

NINGUNA

- ASISTENCIA=NO ==> DERECHOHABIENCIA=NINGUNA ==> OCUPACION=TRABAJA 27573

5. OCUPACION=Trabajadores en actividades agrícolas ganaderas caza y pesca ASISTENCIA=NO 12124 ==> DERECHOHABIENCIA=NINGUNA 12124 conf:(1)

- ASISTENCIA=SI ==> DERECHOHABIENCIA=NINGUNA ==> OCUPACION=TRABAJA 69393

7. OCUPACION=Trabajadores en actividades agrícolas ganaderas caza y pesca 26410 ==> DERECHOHABIENCIA=NINGUNA ASISTENCIA=SI 26410 conf:(1)

Se observa en las reglas de asociación que se formaron que los trabajadores agrícolas no tienen ninguna institución de salud, con 12124 registros. Para los trabajadores agrícolas que si recibieron asistencia sin ser afiliados a ninguna institución con 26410 registros.

INSTITUCIÓN	ASISTENCIA=SI		ASISTENCIA=NO		ASISTENCIA=NO ESPEFICICADA	TOTAL
	TRABAJA	NO TRABAJA	TRABAJA	NO TRABAJA		
IMSS	59192	103459	9015	9491	4132	185289
SEGURO POPULAR	29866	24640	6056	4232	1555	66349
ISSSTE	13217	22156	1847	1869	699	39788
OTRAS	1951	5513	395	633	185	8677
NINGUNA	69393	65250	27573	14520	5859	182595

Tabla 4.9 Tabla IMSS-ISSSTE 2009

*OTRAS= PEMEX, MARINA, DEFENSA

Ahora bien en nuestro país hay 3 instituciones que en el 2009 tienen el mayor número de defunciones y son IMSS, ISSSTE y SEGURO POPULAR, esto también porque son las instituciones más grandes del país, realizaremos un estudio para evaluar cuál de estas instituciones es la más deficiente, esto de la siguiente forma.
Del total de derechohabientes cual es el porcentaje de defunciones de cada institución, tenemos que la institución con mas defunciones respecto a su 100% de derechohabientes es el IMSS, con 0.37%. Este porcentaje es bajo con considerando que son casi 50 millones de personas afiliadas a esta institución.

INSTITUCIÓN	Tota de derechohabientes	Total de defunciones	Porcentaje
IMSS	49,134,310	185,289	0.37%
SEGURO POPULAR	31,100,000	66,349	0.21%
ISSSTE	11,589,483	39,788	0.34%

Tabla 4.10 Tabla IMSS-ISSSTE-SEGURO POPULAR 2009

Ahora bien, empezaremos a estudiaremos a los defunciones sin asistencia, en donde se tomara como 100% al total de defunciones por cada institución, lo que nos da como resultado el SEGURO POPULAR.

INSTITUCIÓN	Tota de defunciones	Total de defunciones sin asistencia	Porcentaje
IMSS	185,289	18,606	10.04%
SEGURO POPULAR	66,349	10,288	15.50%
ISSSTE	39,788	3,716	9.33%

Tabla 4.11 Tabla sin asistencia 2009

Tomando en cuenta las defunciones sin asistencia la institución con mayor porcentaje es el ISSSTE, como se observa en la tabla siguiente.

INSTITUCIÓN	Tota de defunciones	Total de defunciones con asistencia	Porcentaje
IMSS	185,289	162,651	87.78%
SEGURO POPULAR	66,349	54,506	82.15%
ISSSTE	39 788	35,373	88.90%

Tabla 4.12 Tabla con asistencia 2009

En nuestro resultado tenemos dos situaciones:
El ISSSTE es la institución con más defunciones con asistencia, esto nos indica que la calidad de la atención a pacientes en esta institución es deficiente, ya que hay en el año 2009 hubo 35373 defunciones con asistencia.

La institución con más defunciones que no recibieron asistencia médica es el seguro popular, es decir no fueron atendidos aunque eran derechohabientes del SEGURO POPULAR. Esto nos indica que no hay suficiente personal e instalaciones para todos los afiliados, ya que ni siquiera se les brindo la asistencia.

Para las defunciones que no tuvieron ninguna afiliación a alguna institución, las cuales son comparadas a las defunciones del IMSS muy similares con 182595 registros en el año 2009, podemos hacer una comparación con la población de nuestro país en el año 2009. En México en el 2009 había 107.6 millones de mexicanos.

Población total en México en 2009	Defunciones sin ninguna institución afiliada	Porcentaje
107.6 Millones	182,595	0.169%

Tabla 4.13 Tabla sin ninguna institución afiliada 2009

Haciendo el análisis de los 2 años 2000 y 2009 se observa que sus resultados son constantes ya que la institución con mayor porcentaje de fallecimientos es la misma en ambos años (IMSS). La institución con el porcentaje mayor en defunciones con asistencia médica es el ISSSTE. Por último encontramos una variante en el porcentaje de las defunciones sin asistencia ya que en el 2009 la institución con el mayor porcentaje es el IMSS y en el 2009 es el SEGURO POPULAR, quedando en 2º lugar el IMSS. Se puede destacar que el SEGURO POPULAR a pesar de ser una institución relativamente joven tiene el mayor porcentaje en defunciones que no recibieron asistencia médica, no olvidando que esta institución no es la que tiene mayor número de derechohabientes.

Conclusión "Eficiencia de Instituciones Medicas"

Nuestras posibles soluciones al estudio realizado anteriormente se basan en 2 aspectos:
Con asistencia médica siendo el ISSSTE la institución menos eficiente propondríamos una mejora en la atención de los pacientes, mejorar las instalaciones y que el personal (médicos, enfermeras, administrativos, etc.) estén mejor capacitados, con esto se lograra un mejor servicio lo cual nos dara como resultado una disminución en el número de defunciones con asistencia médica.
Sin asistencia médica, resaltan 2 instituciones (SEGURO POPULAR, IMSS). Empezaremos hablando del IMSS ya que esta institución aparece en ambos años aclarando que para el año 2009 ocupa el 2º lugar, una posible mejora es que esta institución aumente su infraestructura, es decir que cuente con la capacidad suficiente para atender a todos sus derechohabientes afiliados. En cuanto al seguro popular que tiene el mayor porcentaje en defunciones sin asistencia el problema es que esta institución tiene demasiados afiliados y no tiene ingresos de los trabajadores lo que hace que su calidad baje y no atienda a todos los afiliados. Por lo cual proponemos que los derechohabientes aporten conforme a sus ingresos una ayuda económica a la institución para que la institución no decaiga y atienda a todos sus afiliados.

4.3 Ocupaciones Peligrosas

En esta fase analizaremos cuales son los objetivos con las siguientes preguntas:

- **¿Qué parte de los datos es pertinente analizar? (Ocupaciones peligrosas)**

  Analizaremos datos como son, el saber si la persona falleció desempeñando una actividad relacionada con su trabajo, y si fue así entonces saber cuál era la ocupación que desempeñaba dicha persona, para tener una información más completa también se requiere saber la escolaridad que tenían, algo importante es la edad, esta la vamos a tomar de un rango que es de 12 años hacia arriba ya que es la edad en la que las personas pueden ser activas, un atributo que nos va a decir con más precisión la causa de la muerte es la causa, en esta solo vamos a tomar 13 que son las más frecuentes, es decir la demás enfermedades no son muy comunes o ni se hacen presentes por tal motivo no las nombraremos en esta vista.

- **¿Qué tipo de conocimiento se desea extraer y cómo se debe presentar? (Ocupaciones peligrosas)**
  Se desea extraer información que nos indique, cuales son las ocupaciones en donde hubieron más personas fallecidas a la hora de desempeñar sus labores, es decir detectar cuales son las ocupaciones más peligrosas y así poder saber cómo ayudar a las personas que desempeñan dichas actividades.
- **¿Qué conocimiento puede ser válido, novedoso e interesante? (Ocupaciones peligrosas)**
  Se requiere detectar cuales son las ocupaciones más peligrosas, una vez que se hayan detectado, investigar el porqué son peligrosas, si es que afecta su nivel de estudios ya que muchas veces las personas que tienen un nivel de estudios bajo son las más desprotegidas en sus ambientes de trabajo, otro dato interesante es el de la edad ya que en muchos trabajos tienen a muchas personas con edad avanzada trabajando y les dejan actividades pesadas.
- **¿Qué reglas o modelos de decisión están utilizando? (Ocupaciones peligrosas)**
  Mostraremos un modelo que refleje que ocupaciones son las más peligrosas, para esto utilizaremos clusters.

- **¿Qué decisiones son críticas? (Ocupaciones peligrosas)**
  Saber identificar bien cuáles son las ocupaciones más peligrosas y el porqué son peligrosas, es decir sino trabajan con el material adecuado para su seguridad, sino reciben una instrucción antes para la realización de sus actividades o sino tienen la edad y habilidades necesarias para poder desempeñar dicho cargo.
  ¿Cómo se distribuyen los datos? (Ocupaciones peligrosas)
  En una sola base de datos.

Una vez estudiado los objetivos pasamos a realizar el análisis estadístico previo.

Es decir hacer la exploración de los datos para saber qué datos pueden ser los más adecuados para poder realizar la siguiente vista minable.

La gráfica que a continuación se muestra nos indica que la mayoría de las personas que fallecieron fueron por causa natural y una cantidad mínima de personas murieron en sus trabajos, otro dato importante es que una gran cantidad de personas eran inactivas, una de las ocupaciones que más se desempeñaban son las de agropecuarios, y la ocupación en la que menos se desempeñaron las personas son en las de directivos ya sea de sector publico ó privado.

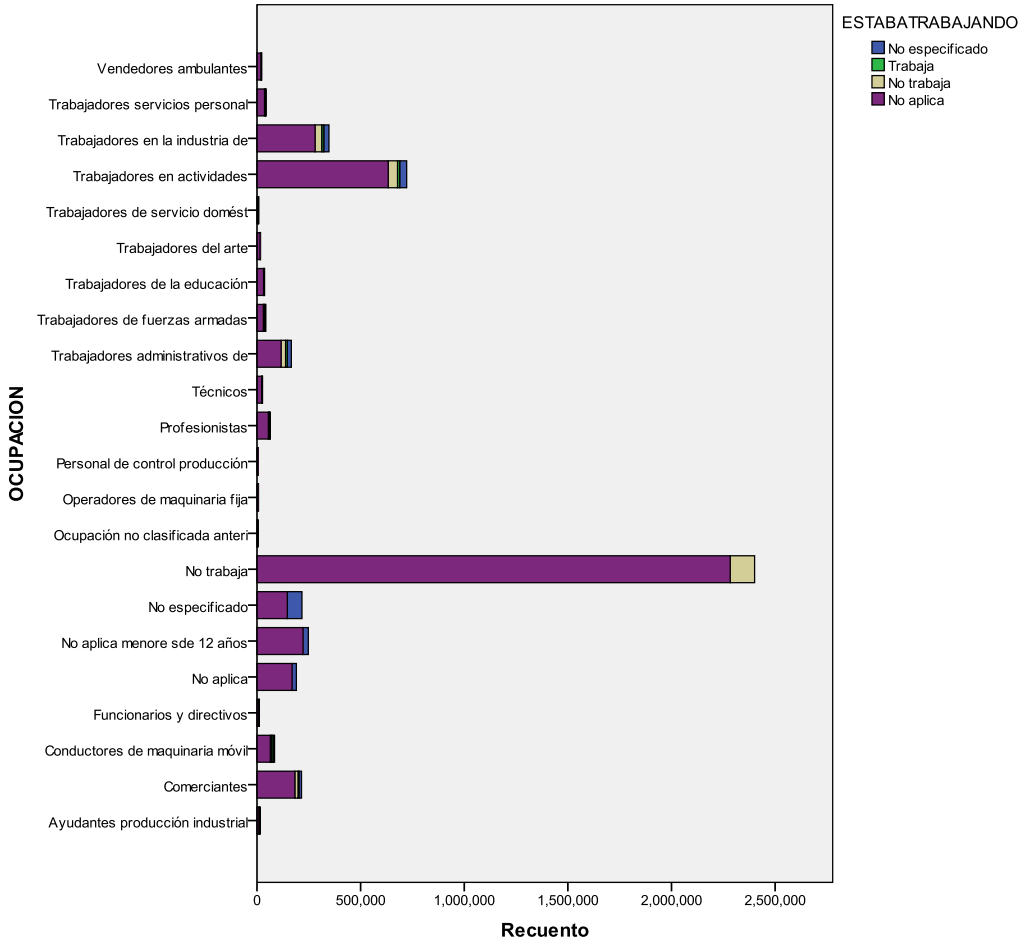


Figura 4.40 Histograma Ocupación- Estaba Trabajando

Como en este caso solo nos interesa analizar a las personas que fallecieron desempeñando alguna actividad relacionada con su trabajo (OCURRIO TRABAJO=1), una vez filtrando todos los registros que cumplan con estas condiciones se analizará cual era la ocupación que desempeñaban, esto nos lo mostrara la gráfica siguiente. Una vez analizada podemos observar que los que están en continuo riesgo son, los trabajadores agropecuarios, los obreros y artesanos en la producción industrial, operadores de

transporte y trabajadores de apoyo en actividades administrativas, las anteriores son las que más sobresalen del estudio estadístico previo que se realizó.

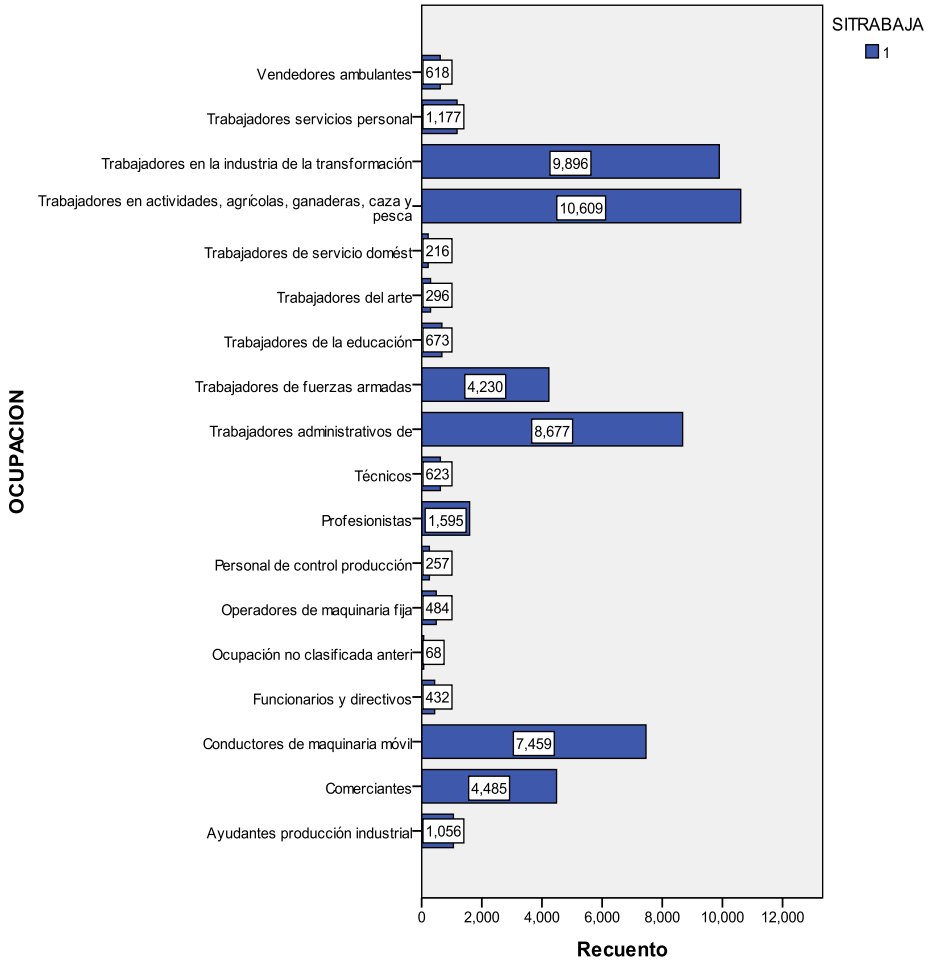


Figura 4.41 Histograma Ocupación- Si Trabajando

Uno de los atributos que también son de suma importancia es la escolaridad de los individuos, como el filtro que ya se había hecho antes (OCURRIO TRABAJO=1), pasaremos a estudiar el atributo de escolaridad el cual se refleja en la gráfica que a continuación se presenta, dicha gráfica nos muestra que la mayoría de las personas poseían un nivel de estudios muy bajo, por ejemplo el nivel máximo que llegaron a obtener dichos individuos, eran secundaria incompleta y completa, esto es el nivel que la mayoría de las personas alcanzaron.

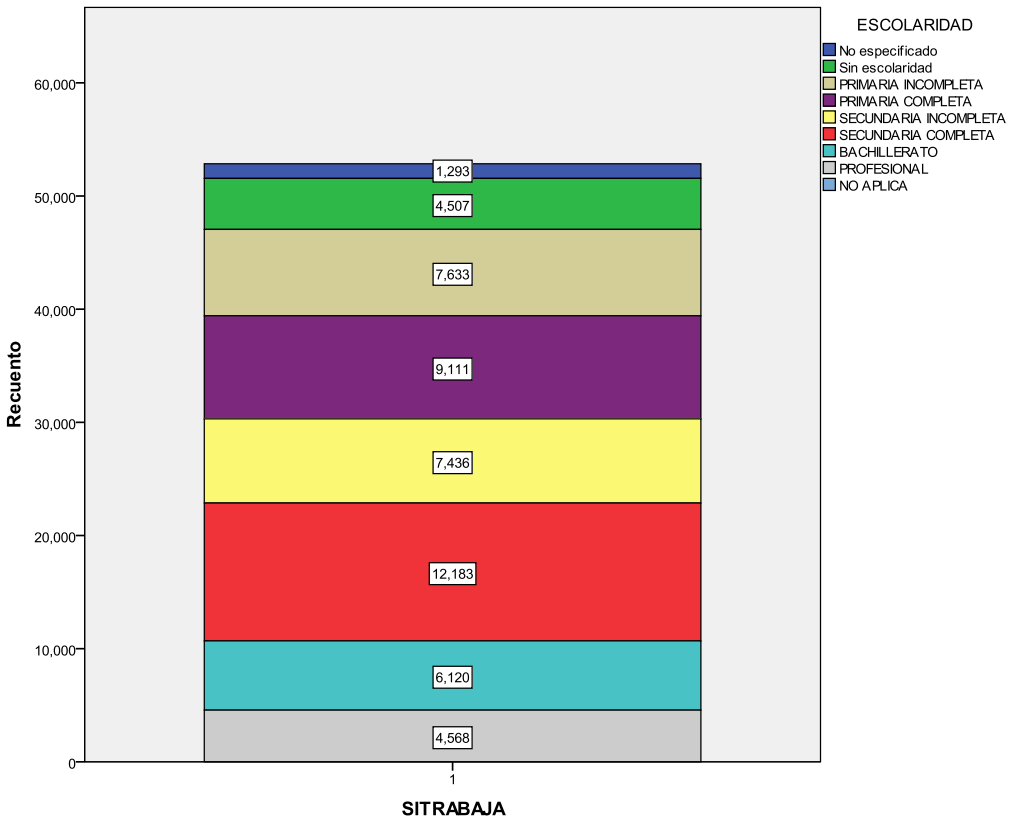


Figura 4.42 Histograma Escolaridad - Si Trabajando

Después del análisis estadístico previo que se realizó anteriormente pasamos a decidir que atributos son los más importantes para poder formar las vistas minables. Y ha explicar con más detalle cada uno de los atributos que se va a utilizar.

ATRIBUTO	DESCRIPCIÓN
CAUSA	Atributo de causa donde solo tomaremos las 13 causas más comunes.
EDAD	Solo ocuparemos del rango de 12 a 120 años ya que es la edad en la que una persona puede empezar a ser productivas, es importante conocer este atributo para saber que personas de determinadas edades no pueden desempeñar dichos puestos o tal vez saber si se tienen que jubilar antes, ya que entre más viejos las personas son menos hábiles o si son muy jovenes para ocupaciones que requieren mayor experiencia.
OCUPACIÓN	Ocupación que desempeñaba la persona, esto es para saber cuáles son las ocupaciones en donde favorecen menos a los empleados, como ya vimos anteriormente. Podríamos decir que este es el

ATRIBUTO	DESCRIPCIÓN
	atributo más importante de la vista.
ESCOLARIDAD	Es indispensable ya que con esto sabremos porque fueron asignados a dichos trabajos, ya que la mayoría de las veces desde aquí está el problema ya que como pasa siempre no hay tanto apoyo a la educación y la mayoría de la población se queda con un nivel de estudios muy bajo, y por lo cual se le asignan trabajos donde se benefician poco a los empleados.
TRABAJO	'ocurrió desempeño trabajo' es el que filtra un número de registros considerable, ya que solo vamos a tomar a las personas que fallecieron el trabajo (Ocurrió desempeño trabajo=1), y los demás los ignoramos.

Tabla 4.14 Tabla Atributo Descripción

El atributo causa tiene una descripción sobre la causa real, esto se muestra en la tabla siguiente:

CAUSA	DESCRIPCIÓN
V496	Ocupante no especificado de automóvil lesionado por colisión con otros vehículos de motor, y con los no especificados, en accidente de tránsito.
V092	Peatón lesionado en accidente de tránsito que involucra otros vehículos de motor, y los no especificados.
V878	Persona lesionada en otros accidentes especificados de transporte de vehículo de motor sin colisión (tránsito).
V899	Persona lesionada en accidente de vehículo no especificado.
V892	Persona lesionada en accidente de tránsito, de vehículo de motor no especificado.
V499	Ocupante (cualquiera) de automóvil lesionado en accidente de tránsito no especificado.
V099	Peatón lesionado en accidente de transporte no especificado.
V093	Peatón lesionado en accidente de tránsito no especificado.
V489	Ocupante de automóvil lesionado en accidente de transporte sin colisión, ocupante no especificado de automóvil, lesionado en accidente de tránsito.
V89	Accidente de vehículo de motor o sin motor, tipo de vehículo no especificado.

CAUSA	DESCRIPCIÓN
V09	Peatón lesionado en otros accidentes de transporte, y en los no especificados.
V49	Ocupante de automóvil lesionado en otros accidentes de transporte, y en los no especificados.
V87	Accidente de tránsito de tipo especificado pero donde se desconoce el modo de transporte de la víctima.
V48	Ocupante de automóvil lesionado en accidente de transporte sin colisión.
X99	Agresión con objeto cortante.
X598	Exposición a factores no especificados, otro lugar especificado.
X599	Exposición a factores no especificados, lugar no especificado.
X958	Agresión con disparo de otras armas de fuego, y las no especificadas, otro lugar especificado.
X594	Exposición a factores no especificados, calles y carreteras.
X954	Agresión con disparo de otras armas de fuego, y las no especificadas, calles y carreteras.
X959	Agresión con disparo de otras armas de fuego, y las no especificadas, lugar no especificado.
X59	Exposición a factores no especificados.
X95	Agresión con disparo de otras armas de fuego, y las no especificadas.
X09	Exposición a humos, fuegos o llamas no especificados.
X954	Agresión con disparo de otras armas de fuego, y las no especificadas, calles y carreteras
W878	Exposición a corriente eléctrica no especificada, otro lugar especificado.
W178	Otras caídas de un nivel a otro, otro lugar especificado.
W87	Exposición a corriente eléctrica no especificada.
W17	Otras caídas de un nivel a otro.
W19	Caída no especificada.
W13	Caída desde, fuera o a través de un edificio u otra construcción.
W20	Golpe por objeto arrojado.
Y094	Agresión por medios no especificados, calles y carreteras.
Y09	Agresión por medios no especificados.

Tabla 4.15 Causas y su descripción

Podemos encontrar las características de estos atributos en la Tabla 4.2 Discretización de los atributos (página 56).

Ahora bien ¿Cómo fue que llegamos a la conclusión que estos atributos serían los indicados para esta vista?, esto se logro con las siguientes consultas. Primero hicimos un conteo del atributo "causa" es decir cuál fue la causa de su fallecimiento, donde encontramos que de las múltiples enfermedades registradas en el catálogo, las más comunes son 13.

```
select count(causa),causa from def2000
where traba=1
group by causa
order by  1  desc
```

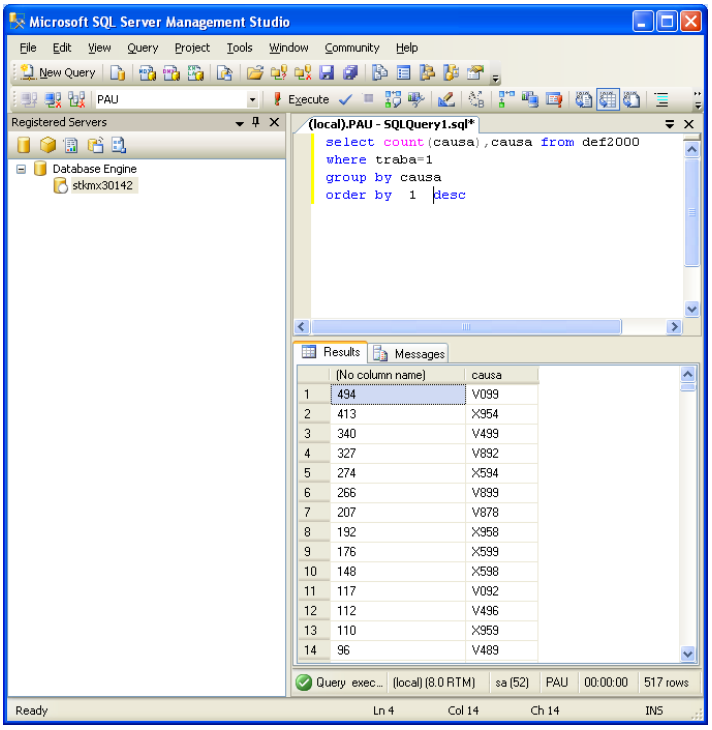


Figura 4.43 Consulta que cuenta causa

Existieron 1320 personas que fueron 'Trabajadores en actividades, agrícolas, ganaderas, caza y pesca, es la ocupación que más se desempeña, otras de las ocupaciones que más resultaron fueron las de 'Trabajadores en la industria de la transformación', 'Trabajadores administrativos de nivel inferior', 'Conductores de maquinaria móvil y medios de transporte', 'Trabajadores de fuerzas armadas, protección y vigilancia', 'Comerciantes, empleados de comercio, agentes de ventas' como lo muestra la consulta siguiente.

```
select count(OCUPA),OCUPA from def2000
where traba=1
group by OCUPA
order by  1  desc
```

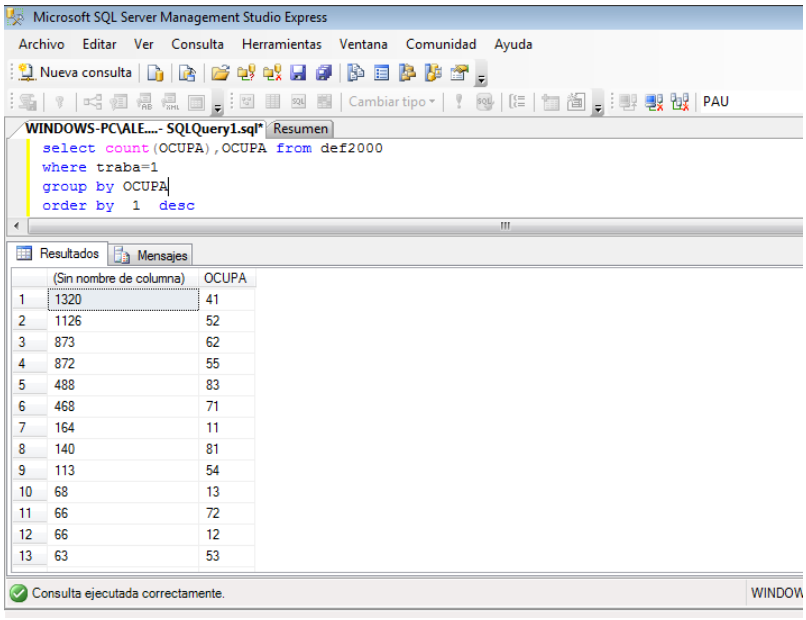


Figura 4.44 Consulta que cuenta ocupación

Después de las consultas encontramos que, la mayoría de las personas solo terminaron la primaria como vemos es un nivel muy bajo de estudios el que las personas poseían, otro de los niveles hasta donde también, y como era de esperarse el nivel que menos se alcanza es el de profesional. Como lo nuestra la siguiente consulta.

```
select count(ESCO),ESCO from def2000
where traba=1
group by ESCO
order by  1  desc
```

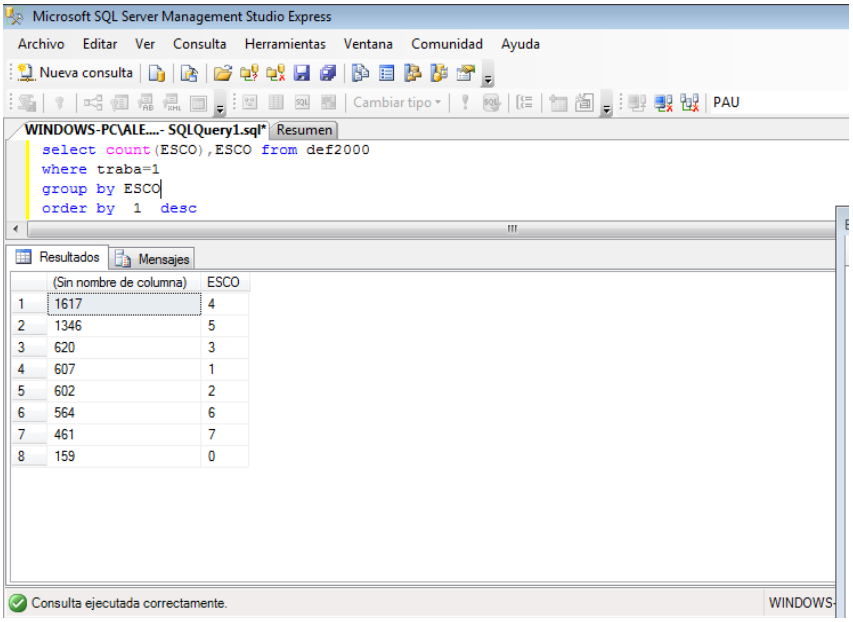


Figura 4.45 Consulta que cuenta escolaridad

Para la edad como se tomo un rango de 12 hacia arriba, el resultado que nos arrojo, fue que las personas que fallecieron tenían una edad de 27, 30, 25 años entre los más comunes, como se puede ver son edades jovenes, de lo que tal vez se puede sacar una hipótesis, que entre más joven sea el individuo, tienen mas riesgo en su salud y en su vida.

```
select count(EDAD),EDAD from def2000
where traba=1
AND EDAD>='12'
group by EDAD
order by  1  desc
```

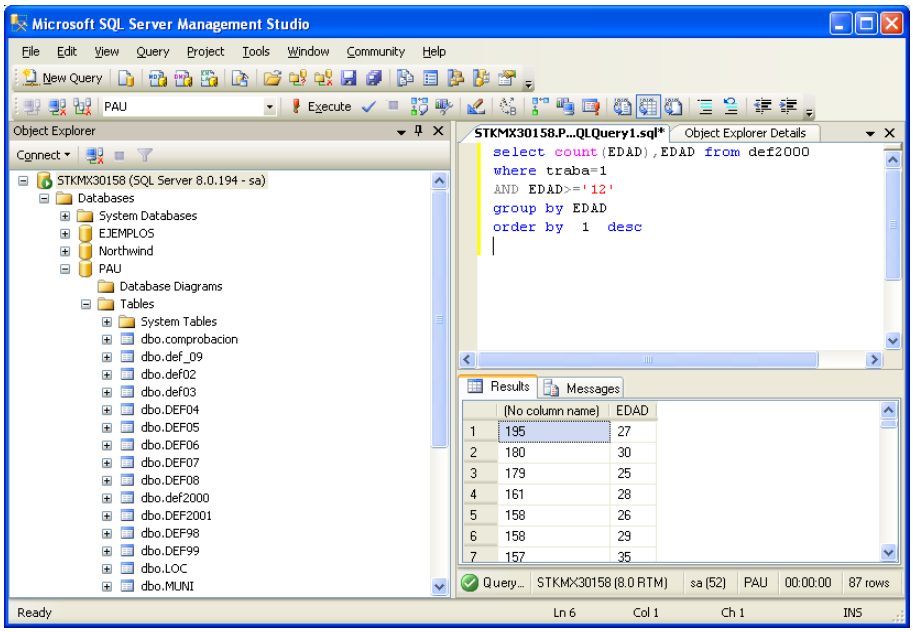


Figura 4.46 Consulta que cuenta edad

Una vez que se probaron los atributos por separado entonces, armamos la vista, es decir juntar todos los atributos antes estudiados en una solo consulta, las cual será la vista minable que utilizaremos en WEKA.

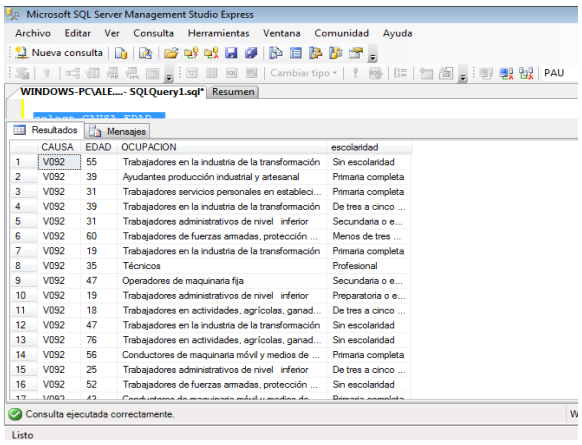


Figura 4.47 Vista Minable ‘Ocupación Peligrosa’

Ahora cargamos el archivo .csv que contiene los 10 años para analizar. En este caso la herramienta WEKA si soporto los 10 años con un total de 35695 registros en total.

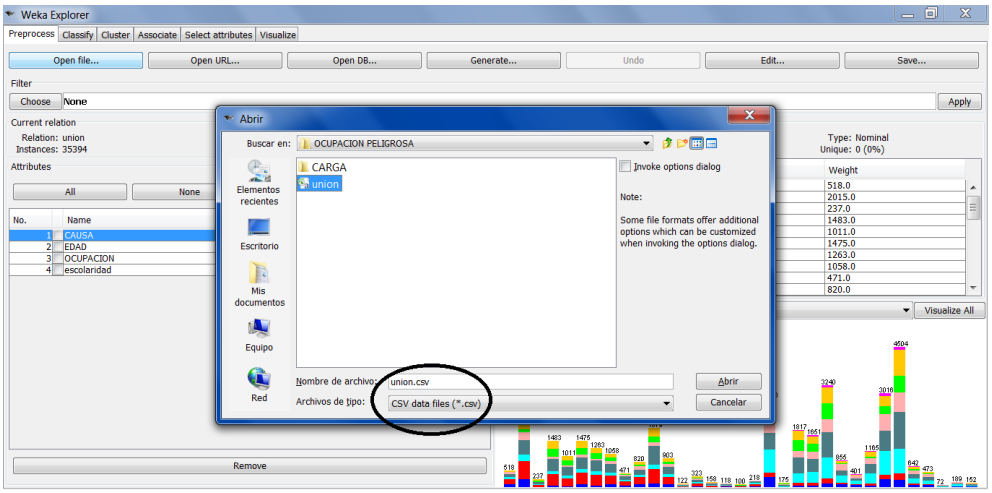


Figura 4.48 Escoger archivo .csv

Después de meter nuestra vista a la herramienta con la que estamos trabajando (WEKA) empezaremos por ver nuestros histogramas que nos genera el WEKA.

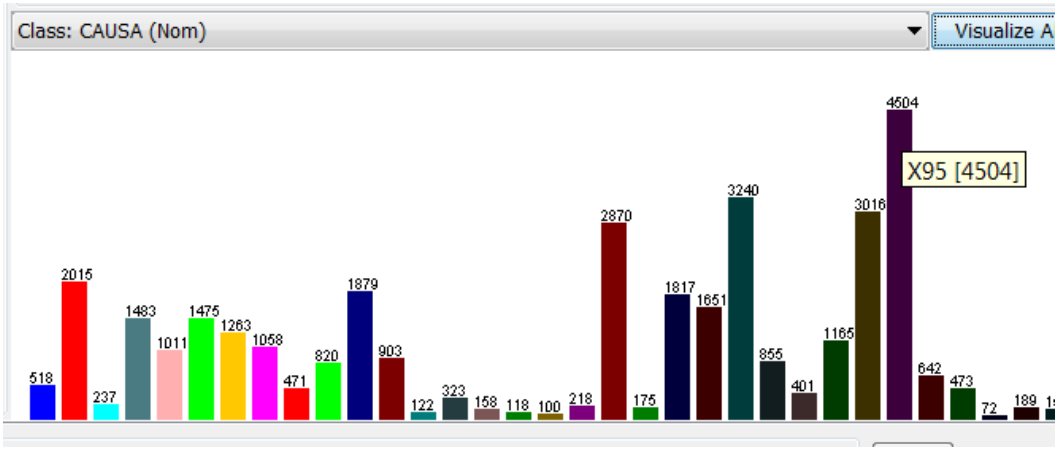


Figura 4.49 Histograma generado por WEKA de Causa

Donde podemos observar que la causa principal es la ‘X95’ Agresión con disparo de otras armas de fuego, y las no especificadas en color morado con 4504 registros.

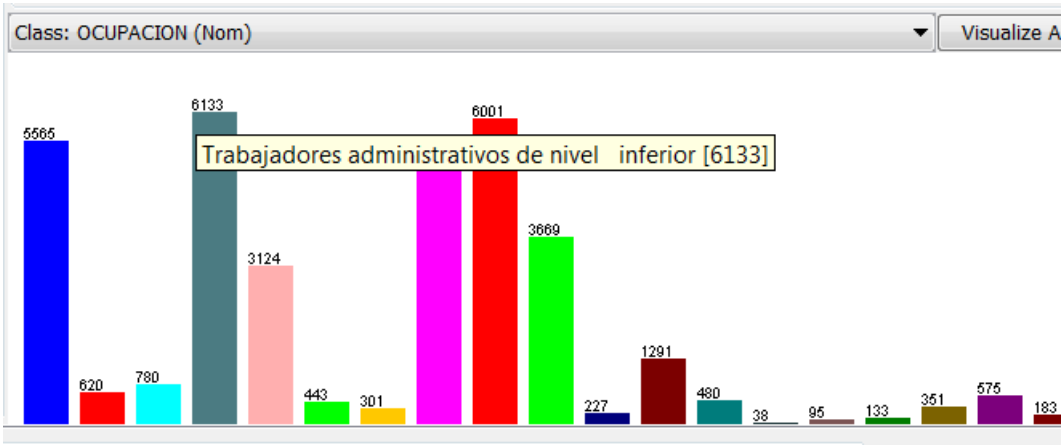


Figura 4.50 Histograma generado por WEKA de Ocupación

La ocupación principal es conductores de maquinaria movil y medios de transporte.

En la parte de la visualización en WEKA se observa que hay ocupaciones que predominan en nuestro análisis las cuales son:

- Trabajadores en la industria de la transformación
- Trabajadores de fuerzas armadas, protección y vigilancia
- Trabajadores administrativos de nivel inferior
- Trabajadores en actividades, agrícolas, ganaderas, caza y pesca
- Comerciantes, empleados de comercio, agentes de ventas

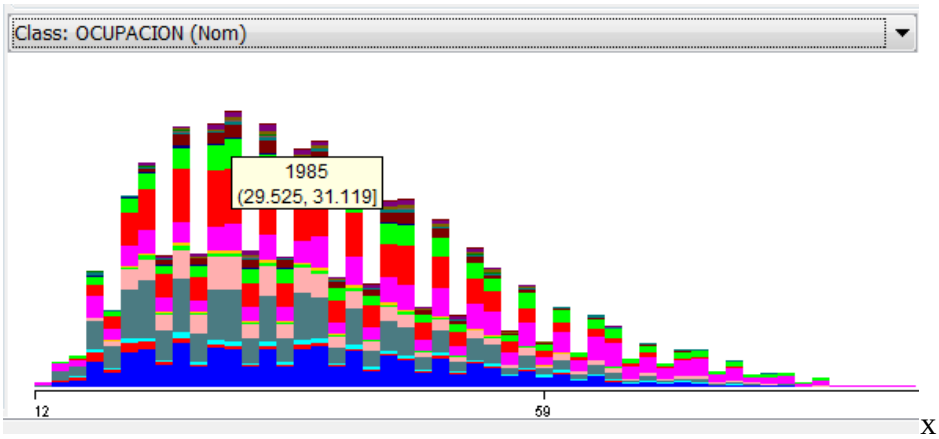


Figura 4.51 Histograma generado por WEKA de Ocupación-Edad

En este histograma se observa la ocurrencia de las ocupaciones a lo largo de las edades, teniendo el punto más alto en el intervalo de edad de 29 a 31 años con 1985 registros.

Dado que todas estas personas trabajaban, surge una pregunta: ¿Por qué hay ocupaciones en donde no hay tantas incidencias como en otras? Y se observa que las edades de estas profesiones en donde si hay accidentes son tan variadas que van desde los 12 años hasta después de los 80 años. Se observa que los niños de 12 años que realizan ocupaciones de "Trabajadores en actividades, agrícolas, ganaderas, caza y

pesca" y "trabajadores administrativos de nivel inferior" tienen más incidencias que en cualquier otra profesión. Se debe aclarar que se tomaran los 10 años para la vista minable con todos los años del 2000 al 2009 tenido en total 35395 registros ya que la herramienta nos lo permite.

4.3.1 Clusters [5]

Para esta vista vamos a hacer minería usando clusters, realizaremos la minería de los 10 años en un archivo .CVS los cuales iremos introduciendo a WEKA.

Los algoritmos de clustering permiten clasificar un conjunto de elementos de muestra en un determinado número de grupos basándose en las semejanzas y diferencias existentes entre los componentes de la muestra.

Ahora que ya hemos cargado los datos nos vamos directamente a la pestaña Cluster, selecionamos en *Choose* y escogemos **SimpleKMeans** como algoritmo de clustering.

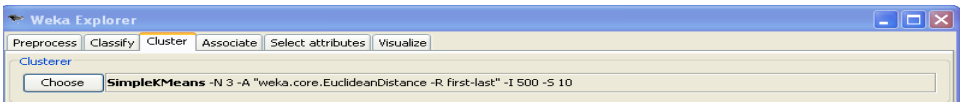


Figura 4.52 Pestaña de Clusters

A continuación seleccionamos el nombre del algoritmo para configurar sus propiedades. En este ejemplo vamos a querer obtener 4 **clusters**, así que configuramos el atributo *numClusters* con valor 4 y pulsamos en *OK*.

[5] Clusters: Es un algoritmo de agrupamiento.

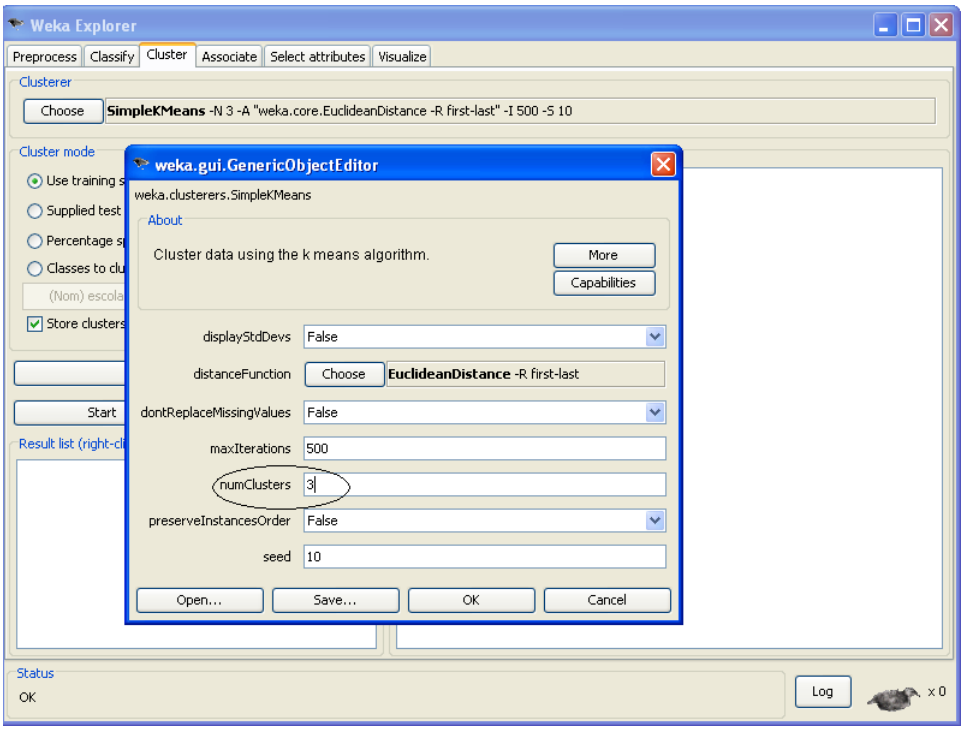


Figura 4.53 Pestaña de Propiedades para Clusters

Ahora ya sí que estamos preparados para ejecutar el **clustering** así que seleccionamos *Start* y en un momento estaremos viendo los resultados.

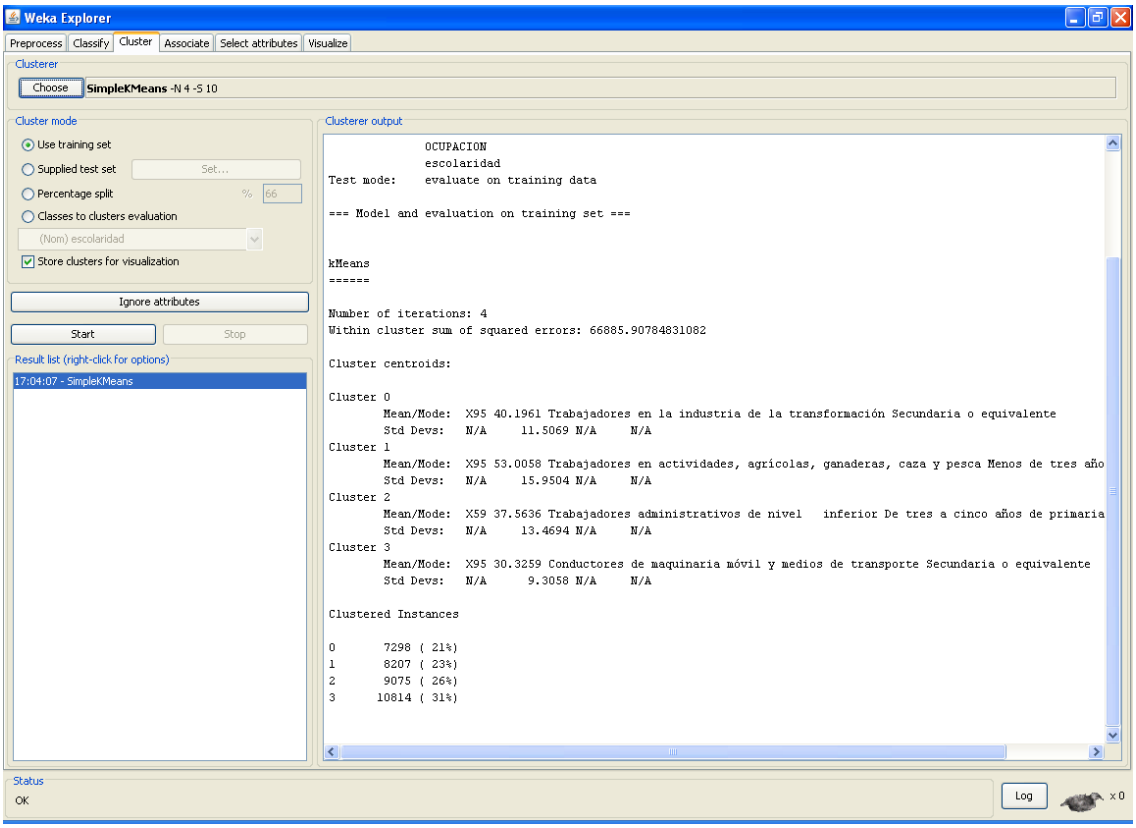


Figura 4.54 Resultado de clusters para los 10 años

En la pantalla resultante tenemos los grupos que se han formado para los10 años, con los siguientes resultados:

kMeans

Number of iterations: 4

Within cluster sum of squared errors: 66885.90784831082

Cluster centroids:

Cluster 0

Mean/Mode: X95 40.1961 Trabajadores en la industria de la transformación Secundaria o equivalente

Cluster 1

Mean/Mode : X95 53.0058 Trabajadores en actividades, agrícolas, ganaderas, caza y pesca Menos de tres años de primaria

Cluster 2

Mean/Mode : X59 37.5636 Trabajadores administrativos de nivel inferior De tres a cinco años de primaria

Cluster 3

Mean/Mode : X95 30.3259 Conductores de maquinaria móvil y medios de transporte Secundaria o equivalente

Clustered Instances

0	**7298 (21%)**
1	**8207 (23%)**
2	**9075 (26%)**
3	**10814 (31%)**

Podemos ver la distribución de una manera más gráfica pinchando con el botón derecho sobre la entrada correspondiente del listado de la derecha y después en *Visualize cluster assignments*.

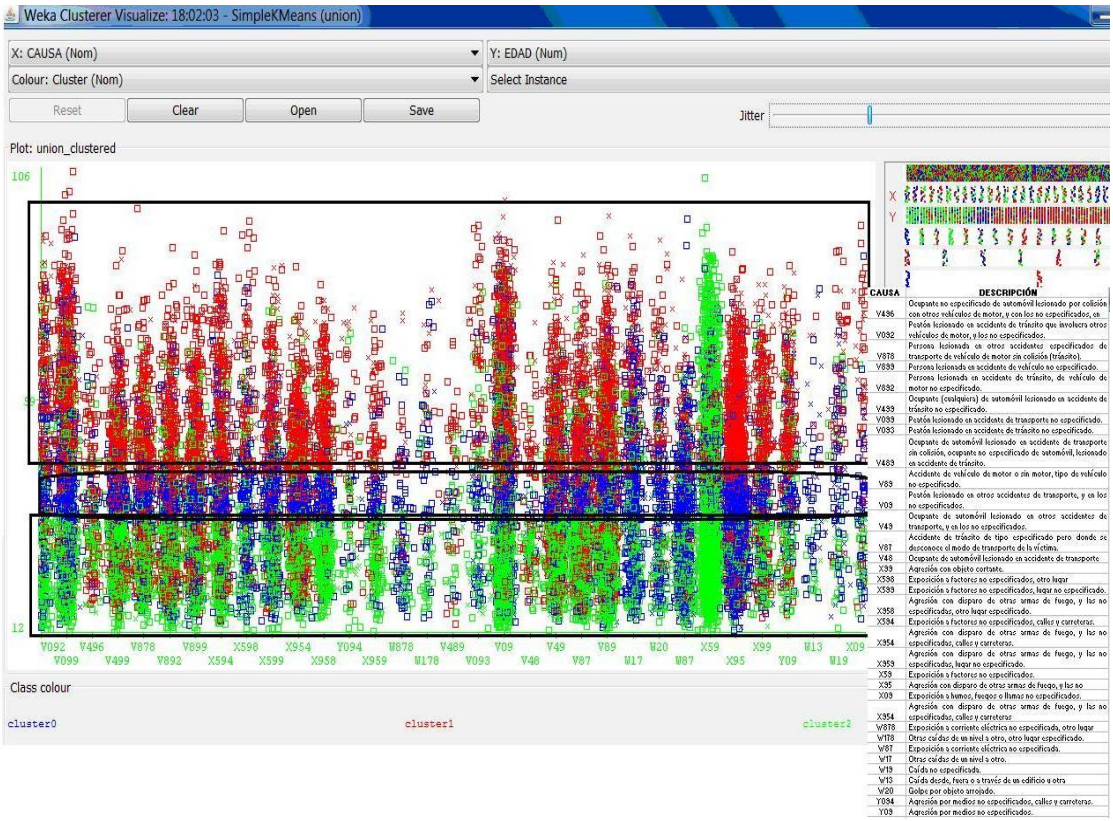


Figura 4.55 Visualización de nuestro cluster

Se observar como los grupos se mantienen constantes a lo largo de las causas, pero en cuestión de edades se marca el rango de edades para cada cluster, para el primer cluster de color verde se marca el rango de edades de 12 a 35 años, para el segundo cluster la edades van de 35 a 48 años y por último el tercero de color rojo se marca de 48 a 83 años. Con respecto a las ocupaciones tenemos que en los grupos las 3 ocupaciones que formaron nuestros clusters fueron:

1. Conductores de maquinaria móvil y medios de transporte
2. Trabajadores administrativos de nivel inferior
3. Trabajadores en la industria de la transformación
4. Trabajadores agrícolas, caza y pesca.

Se observo que en la formación de los 3 clusters la escolaridad máxima fue secundaria o equivalente, no se obtuvo agrupación en escolaridades como profesionistas o bachillerato.

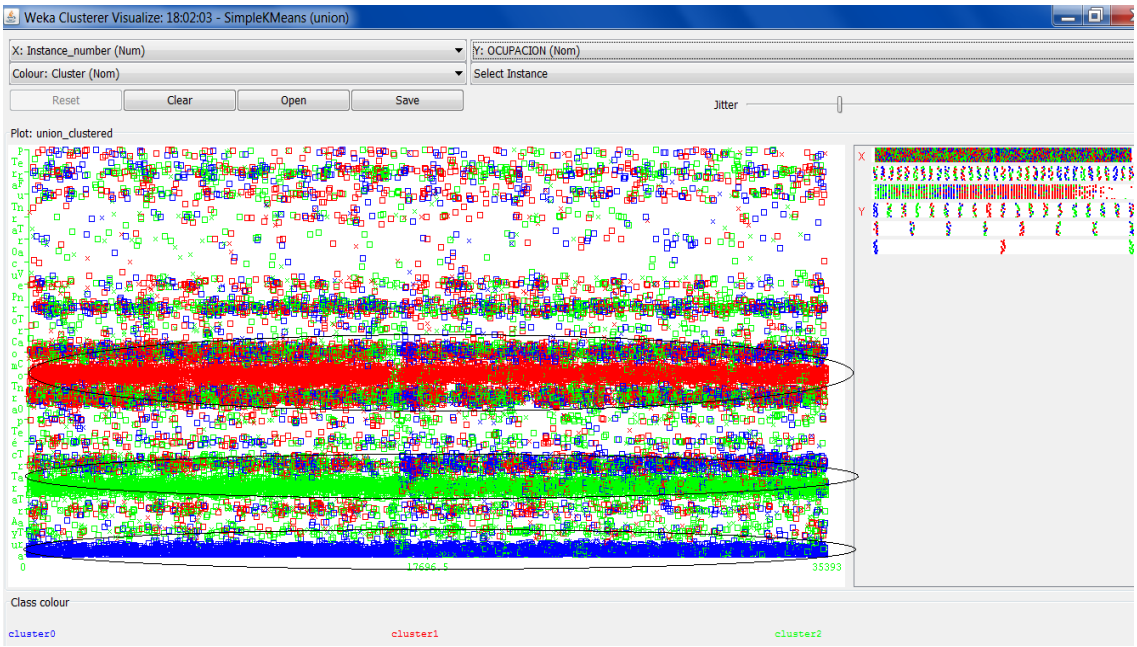


Figura 4.56 Visualización de nuestros clusters respecto a ocupación

Ahora bien, cuáles fueron los hallazgos que se encontraron al realizar clusters de los 10 años, algo que se observo fue que la edad se mantuvo entre 30 a 53 años. También se observa que la escolaridad es en su mayoría fue primaria y secundaria, en algunos casos incompleta.

Se observa fue que las ocupaciones de "Trabajadores en actividades, agrícolas, ganaderas, caza y pesca", "Conductores de maquinaria móvil y medios de transporte" y "trabajadores de la industria de la transformación" la causa es la misma en estas tres ocupaciones (X95) "Agresión con disparo de otras armas de fuego, y las no especificadas" por lo tanto se considera esta la causa principal, siendo esta causa no un accidente si no un problema de inseguridad que ataca principalmente a las personas que tienen un escolaridad máxima de secundaria y como consecuencia su trabajo no tiene mucha calidad.

Podemos observar que este tipo de causas se agrupan solamente en estas 4 ocupaciones, dejando a otras como Trabajadores del arte, espectáculos y deportes o Trabajadores de la educación en las cuales no hay ningún riesgo de este tipo.

Ahora vamos hacer una comparación de nuestro primer año de estudio (2000) y nuestro último año de estudio (2009) cuales fueron los cambios que se presentaron y que fue lo que se mantuvo constante.

En el 2000 se observan 2989 registros donde las causas principales fueron "Peatón lesionado en accidente de transporte" y "Agresión con disparo de otras armas de fuego, y las no especificadas, calles y carreteras" con edades entre los 31 y 48 años con profesiones como Trabajadores en la industria de la transformación, conductores de transporte y trabajadores administrativos de nivel inferior. En este año la causa con la

que se formaron 2 clusters fue “peatón lesionado en accidente de transporte no especificado”.

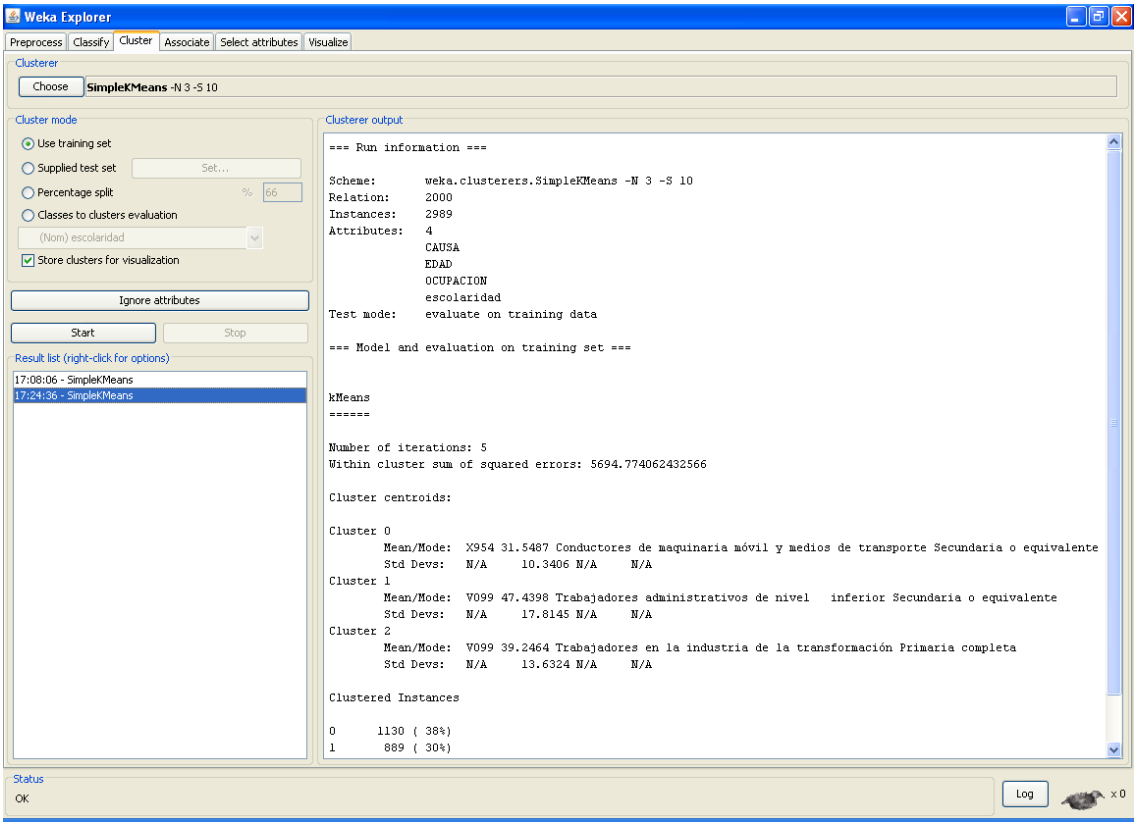


Figura 4.57 Resultados año 2000

kMeans

Number of iterations: 5

Within cluster sum of squared errors: 5694.774062432566

Cluster centroids:

Cluster 0

Mean/Mode: X954 31.5487 Conductores de maquinaria móvil y medios de transporte Secundaria o equivalente

Cluster 1

Mean/Mode: V099 47.4398 Trabajadores administrativos de nivel inferior Secundaria o equivalente

Cluster 2

Mean/Mode: V099 39.2464 Trabajadores en la industria de la transformación Primaria completa

Clustered Instances

0 1130 (38%)

1 889 (30%)

2 970 (32%)

Para nuestro año 2009 los registros aumentaron a 3776 casi 900 registros más que en el 2000 y se observo que la causa principal y la cual predomino fue la X95 “agresión con arma de fuego” con otra profesión muy distinta a las que resaltaron en el 2000 “Comerciantes“ con solo estudios de secundaria y con edad de 37 años. Y otra ocupación que no se había agrupado en nuestros clusters general o en el cluster de el año 200 es “ Trabajadores de actividades agrícolas, ganaderas, pesca” con la causa de peatón lesionado en accidente de transporte.

kMeans

Number of iterations: 5

Within cluster sum of squared errors: 7067.253699100283

Cluster 0

Mean/Mode: X95 33.7999 Trabajadores en la industria de la transformación De tres a cinco años de primaria

Std Devs: N/A 11.9065 N/A N/A

Cluster 1

Mean/Mode: X95 37.4917 Comerciantes, empleados de comercio, agentes de ventas Secundaria o equivalente

Std Devs: N/A 10.862 N/A N/A

Cluster 2

Mean/Mode: V09 52.0085 Trabajadores en actividades, agrícolas, ganaderas, caza y pesca Menos de tres años de primaria

Std Devs: N/A 15.3742 N/A N/A

```
Clustered Instances

0       1459 ( 39%)
1       1381 ( 37%)
2        936 ( 25%)
```

Figura 4.58 Clusters año 2009

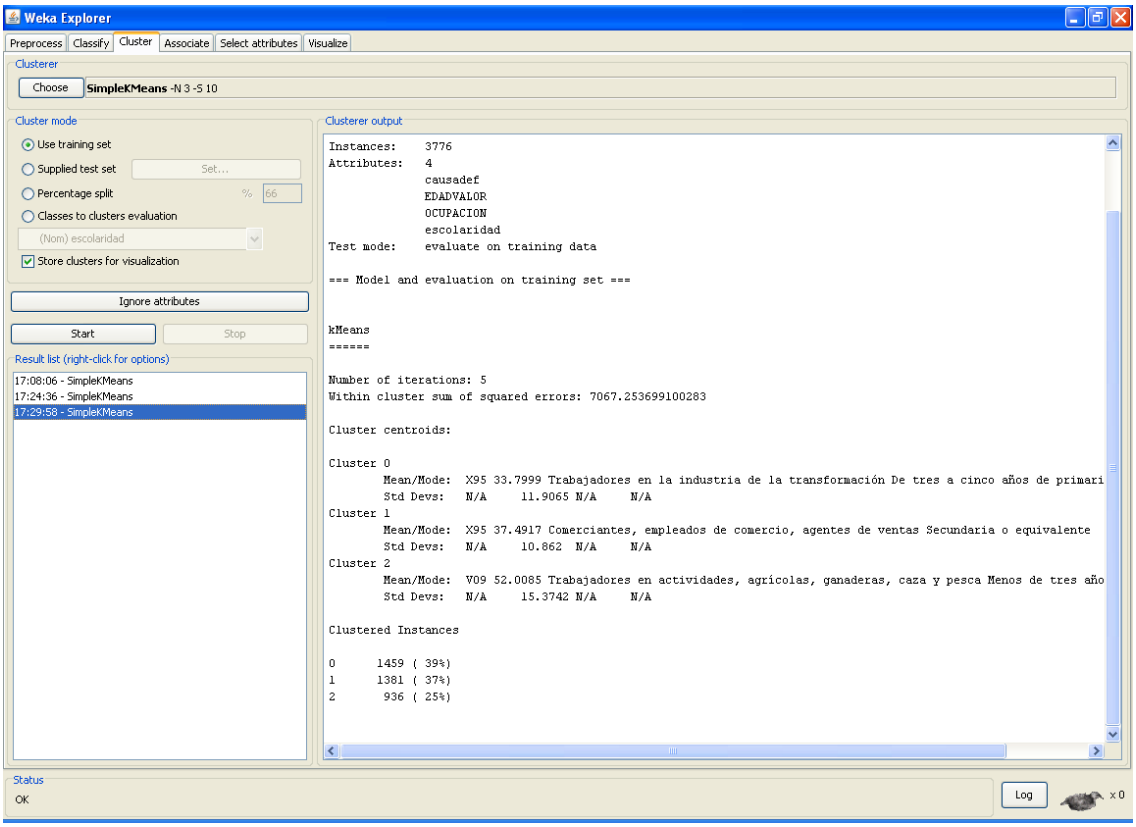


Figura 4.59 Resultados año 2009

Una ocupación que también salió en un cluster en el 2009 fue agricultores con la causa de “Peatón lesionado en otros accidentes de transporte, y en los no especificados”.

Conclusión “Ocupaciones Peligrosas”

Algo que confírmanos es que su nivel de escolaridad si influye en su ocupación y en el riesgo que se tiene de tener algún accidente en el trabajo, en la formación de nuestros clusters la escolaridad más alta fue secundaria, la mayoría de incidencias fueron con escolaridades bajas esto nos hace concluir que entre menos estudios tengas, tendrás un trabajo con menos seguridad y más riesgo de sufrir algún accidente fatal.

Las profesiones principales han cambiado conforme los años, en el 2009 la causa principal se refiere mas a un problema de seguridad hacia los comerciantes y trabajadores de fabricas en especial, ya no un accidente como tal, esta causa no se considera como un accidente como podría ser un “Peatón lesionado en accidente de transporte”.

Si indagamos más sobre esta causa “Agresión con disparo de otras armas de fuego, y las no especificadas, calles y carreteras” nos damos cuenta de que hubo más incidencias en el año 2009, la escolaridad pasó de secundaria es su mayoría a ser primara:

AÑO 2000				AÑO 2009			
	CAUSA	**EDAD**	**OCUPACIÓN**		**CAUSA**	**EDAD**	**OCUPACIÓN**
1130 **38%**	X954 **Agresión con disparo de otras armas de fuego, y las no especificadas, calles y carreteras**	31.549	Conductores de maquinaria móvil y medios de transporte Secundaria o equivalente	**1459** **39%**	X95 **Agresión con disparo de otras armas de fuego, y las no especificadas.**	33.8	Trabajadores en la industria de la transformación De tres a cinco años de primaria
889 **30%**	V099 **Peatón lesionado en accidente de transporte no especificado.**	47.44	Trabajadores administrativos de nivel inferior Secundaria o equivalente	**1381** **37%**	X95 **Agresión con disparo de otras armas de fuego, y las no especificadas.**	37.492	Comerciantes, empleados de comercio, agentes de ventas Secundaria o equivalente
970 **32%**	V099 **Peatón lesionado en accidente de transporte no especificado**	39.246	Trabajadores en la industria de la transformación Primaria completa	**936** **25%**	V09 **Peatón lesionado en otros accidentes de transporte, y en los no especificados.**	52.009	Trabajadores en actividades, agrícolas, ganaderas, caza y pesca Menos de tres años de primaria

Tabla 4.16 Comparación 2000 - 2009

La solución propuesta es poner cámaras de seguridad en los comercios, que esto sea un requisito por negocio establecido, en cuento a los comercios informales como tianguis y plazas, que allá más vigilancia.

4.4 Violencia Familiar

En esta fase analizaremos cuales son los objetivos con las siguientes preguntas:

Antes es importante mencionar que esta vista se clasifico en dos, en violencia general (niño, jóvenes, adultos, ancianos) y violencia contra las mujeres.

- **¿Qué parte de los datos es pertinente analizar? (Violencia Familiar)**

Para ambas clasificaciones, vamos a analizar si existió violencia familiar y la edad, para la violencia general vamos a analizar la edad desde 0 años hasta 120 años. Y para la violencia contra las mujeres también analizaremos el estado civil, su escolaridad, trabajaremos con las mujeres que estaban inactivas y que entren en un rango de edad de 15 a 120 años.

- **¿Qué tipo de conocimiento se desea extraer y cómo se debe presentar? (Violencia Familiar)**

Se extraerá información como su estado civil (soltero, casado, divorciado, unión libre), que sean mujeres que no trabajen es decir inactivas, su escolaridad (sin nivel, primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, superior), en este caso estudiaremos solo a mujeres (sexo=femenino), todo esto para la violencia contra las mujeres, y para la de violencia general vamos a extraer la edad de los individuo (niños, jóvenes, adultos, ancianos), algo que vamos a obtener para ambas clasificaciones de violencia es, si hubo violencia familiar (violencia familiar=si). Con todo esto vamos a crear un nuevo atributo que se va a llamar "RIESGO" con el cual podremos hacer una predicción sobre el nivel de riesgo que puede sufrir cada persona en los próximos años.

- **¿Qué conocimiento puede ser válido, novedoso e interesante? (Violencia Familiar)**

Se quiere predecir el riesgo que corre una persona de sufrir violencia familiar y a causa de esto, morir, en los próximos años, basándose en los datos correspondientes a defunciones anteriores a causa de la violencia familiar. Como ya se dijo antes va a existir dos tipos de riego, hablando únicamente de mujeres o hablando de la población en general.

- **¿Qué reglas o modelos de decisión están utilizando? (Violencia Familiar)**

Mostraremos un modelo que refleje en qué grado aumenta la violencia familiar y saber si ira en decremento, o en aumentó, utilizaremos árboles de decisión (j48) para resolver esta tarea.

- **¿Qué decisiones son críticas? (Violencia Familiar)**

Interpretar los resultados de manera correcta, saber qué hacer con los resultados y saber hacia que parte de la sociedad dirigir los resultados.

- **¿Cómo se distribuyen los datos? (Violencia Familiar)**

En una sola base de datos.

- **¿Qué atributo del conjunto de datos se desea intentar predecir? (Violencia Familiar)**

  El atributo a predecir es 'Riesgo' con bases en los datos de nuestra base, de defunciones anteriores a causa de la violencia familiar.

Antes de pasar la construcción, de las vistas minables hablaremos un poco de cómo fue que se armo el atributo de Riego, para las dos clasificaciones, se realizo un estudio antes sobre quiénes son los que más sufren de violencia familiar, este estudio estadístico se hizo por atributo, se obtuvo de diversas fuentes de internet, a continuación mostraremos los resultados de dicho estudio.

Violencia General (niños, jóvenes, adultos, ancianos):

Rangos de edades	Porcentaje (%)
0-5	82
6--9	81
10--19	72
20-44	51
45-64	42
65-en adelante	36

Tabla 4.17 Violencia general

Como podemos observar la edad que sufre más de violencia familiar, son de 0 a 9 años, que vienen siendo los niños, después serían los jóvenes y los que menos sufren de violencia familiar son entre adultos y ancianos, que entran en un rango de edad de 45 años en adelante.

Violencia Contra las Mujeres:

Escolaridad	Porcentaje (%)
Sin nivel	74
Primaria incompleta	4
Primaria completa	25
Secundaria incompleta	21
Secundaria completa	30
Superior	18

Tabla 4.18 Violencia Mujeres

El resultado obtenido fue que las mujeres que no tienen ningún nivel de estudios son las que más sufre de violencia familiar y las mujeres que tienen un porcentaje menor de sufrir violencia familiar son las que tienen un nivel de estudio a nivel universitario.

Estado civil	Porcentaje (%)
Casadas	50
Unión libre	35
Solteras	19
Divorciadas	5
Viudas	2

Tabla 4.19 Estado civil

El estado civil es uno de los atributos que nos arroja más información sobre la violencia familiar, como se nos podemos dar cuenta la mujeres casadas y que viven en unión libre son las que tienen un riego, más alto de sufrir violencia familiar las solteras también tienen un riesgo considerable, las que tienen un riego bajo o muy bajo de sufrir violencia, son las mujeres divorciadas y viudas.

Una vez analizado los objetivos y los estudios realizados anteriormente pasamos a realizar el análisis estadístico previo.

Es decir hacer la exploración de los datos para saber qué datos pueden ser los más adecuados para poder realizar la siguientes vistas minables.

La gráfica generada en PASW que a continuación se muestra se confirma que efectivamente el estado civil que sufren más de violencia familiar son las personas casadas, en uniones libres y solteras, las que menos sufren de violencia familiar las personas divorciadas y viudas.

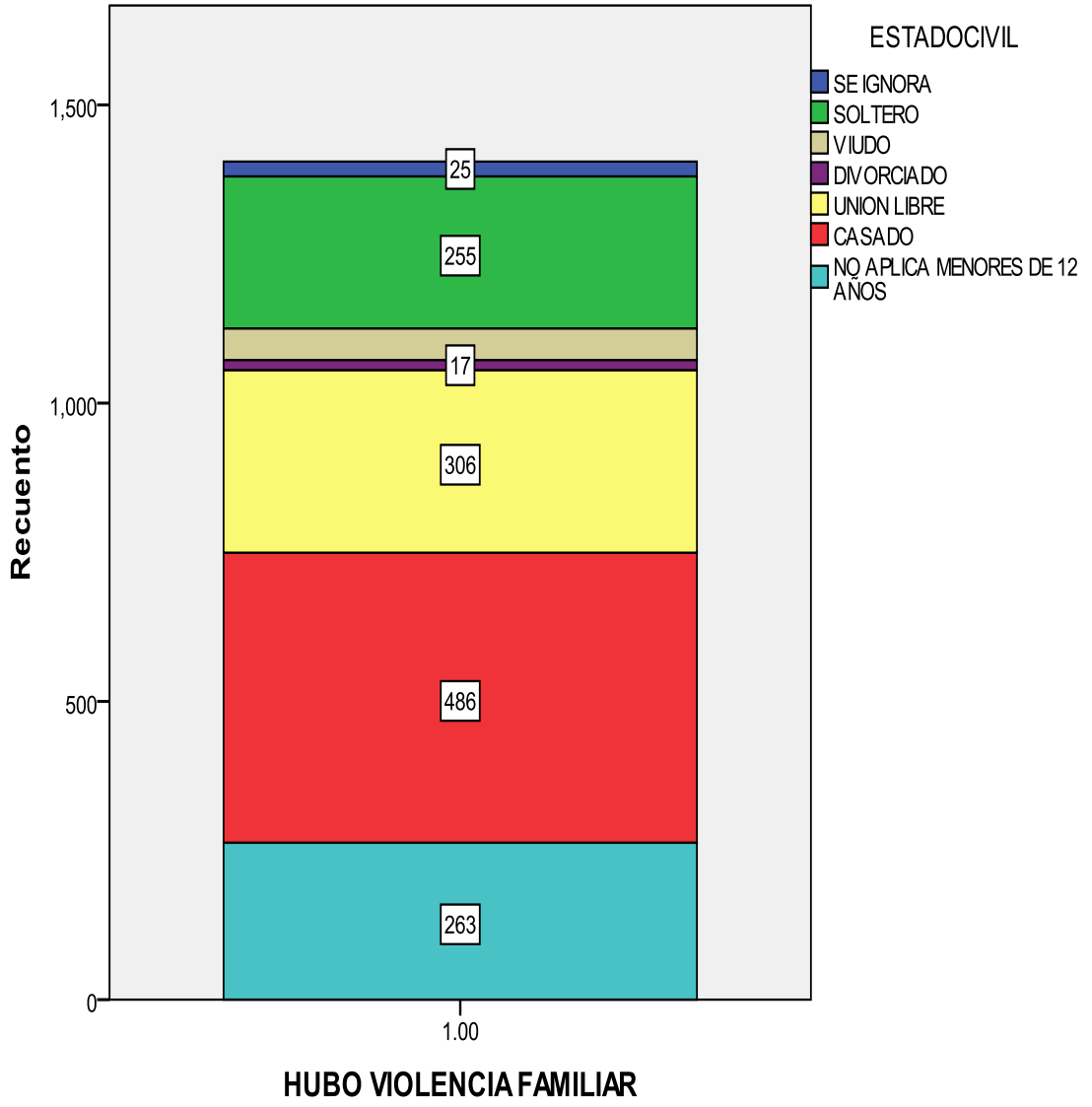


Figura 4.60 Histograma Estado Civil-Hubo violencia

La siguiente gráfica lo que nos muestra es la relación que existe entre la ocupación y si es que sufrieron de violencia familiar, como podemos observar las ocupaciones más afectadas son las personas inactivas, trabajadores de servicio domestico y también los trabajadores de actividades agrícolas, ganadería, caza y pesca, y los menos afectados son los profesionistas.

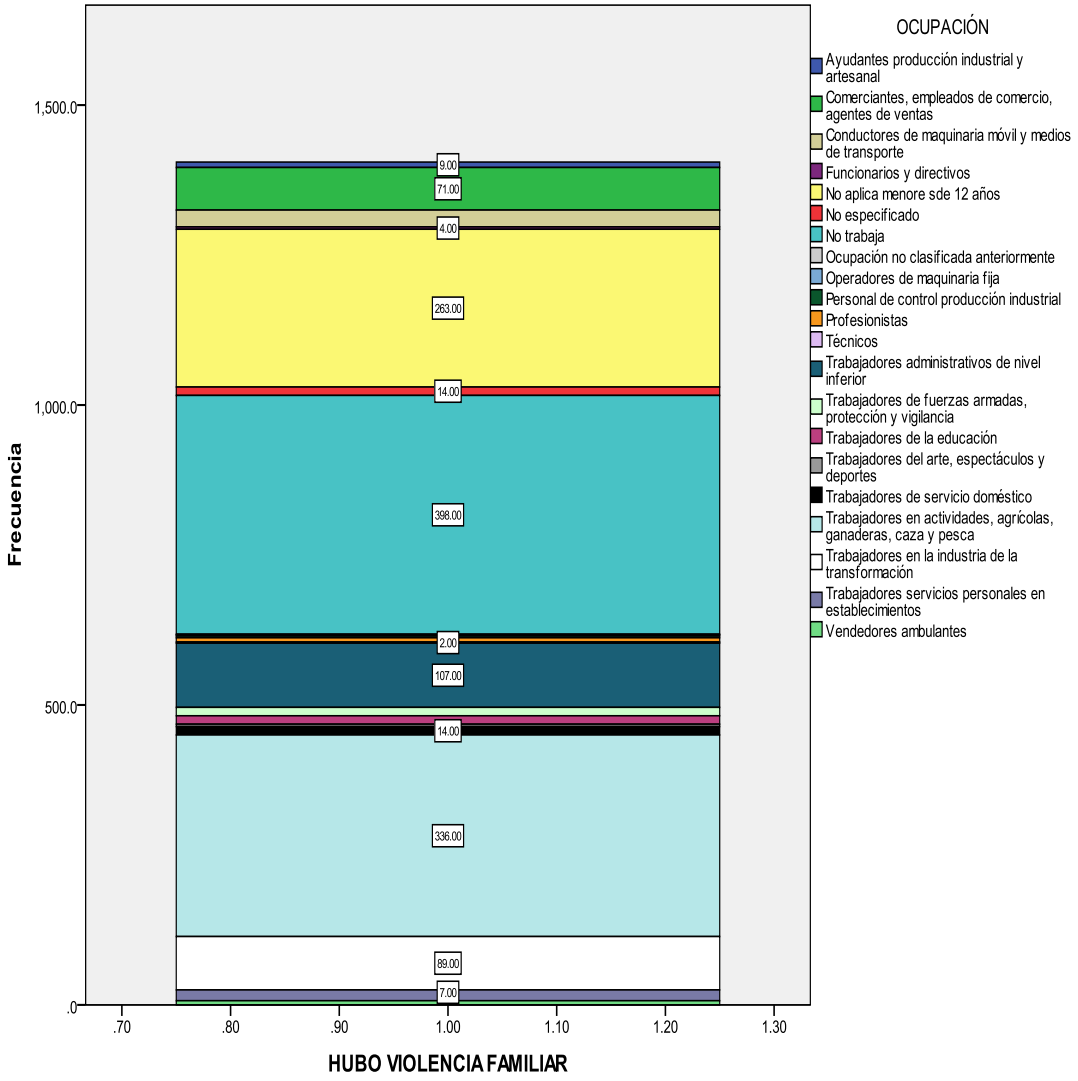


Figura 4.61 Histograma Ocupación - Hubo Violencia

Esta gráfica nos muestra un resultado similar al de los estudios anteriormente hechos, pero con ciertas variaciones como lo son que las personas sin escolaridad no son tantas como en las estadísticas de los estudios anteriores. En este análisis los niveles de estudios menos más afectados son, primaria completa e incompleta y después las personas que tienen como nivel d estudios la secundaria, los menos afectados son profesional, cabe aclarar que en estas estadísticas estamos tomando a la población en general.

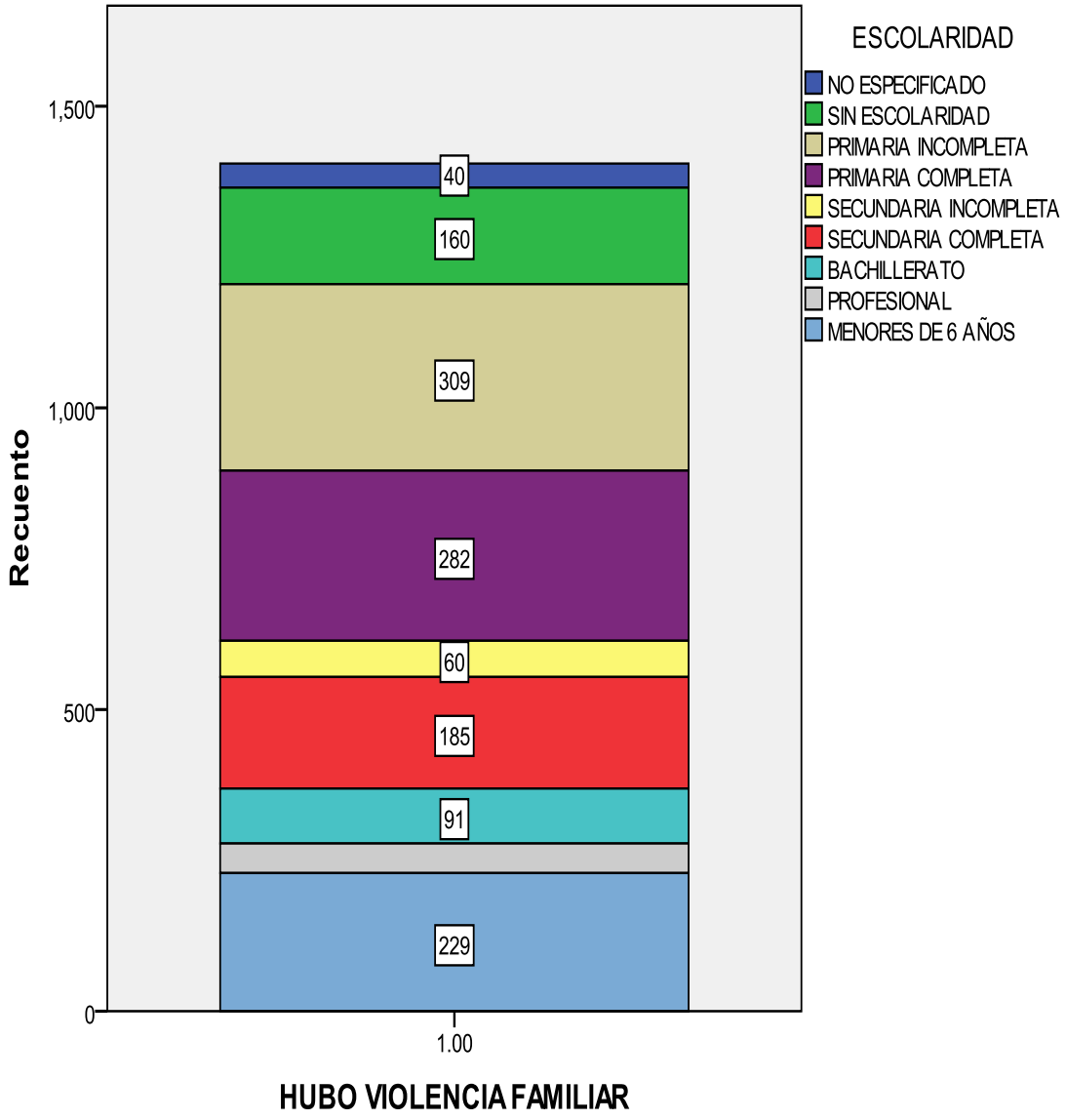


Figura 4.62 Histograma Escolaridad - Hubo Violencia

Hablando ahora sí, únicamente de la población femenina, veremos su relación con otros atributos como es el de la escolaridad, nos muestra un resultado que nos indica que la mayoría de la población femenina no tiene escolaridad o tienen la primaria incompleta ò completa y la menor parte de la población femenina tienen un nivel de estudio profesional o la secundaria incompleta.

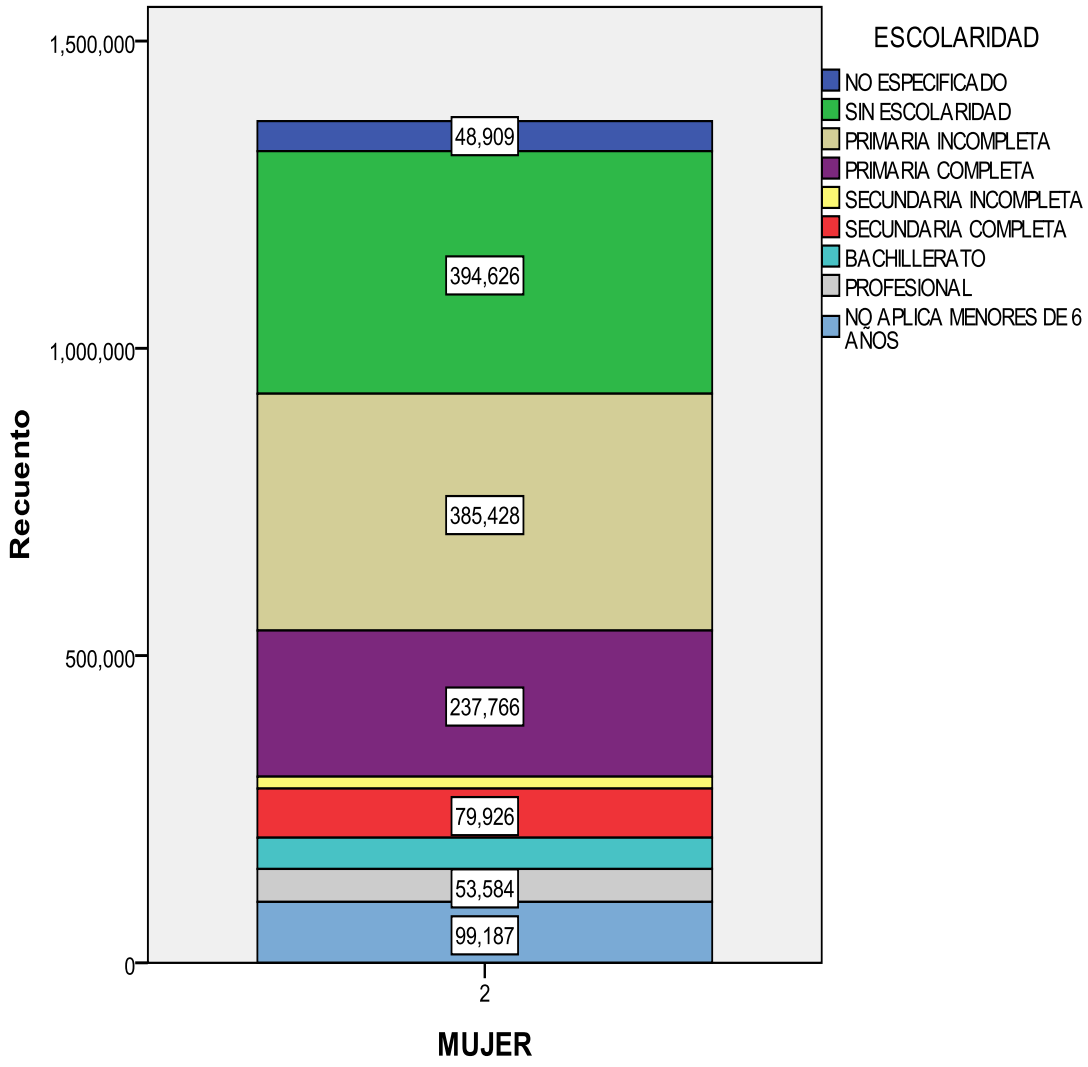


Figura 4.63 Histograma Mujer-Escolaridad

Otro de los atributos que también son de suma importancia, es el de ocupación el cual nos muestra que la gran mayoría casi una totalidad d la población femenina no trabaja y la minoría es profesionista, esto es algo alarmante y puede ser un factor de suma importancia para la violencia contra las mujeres.

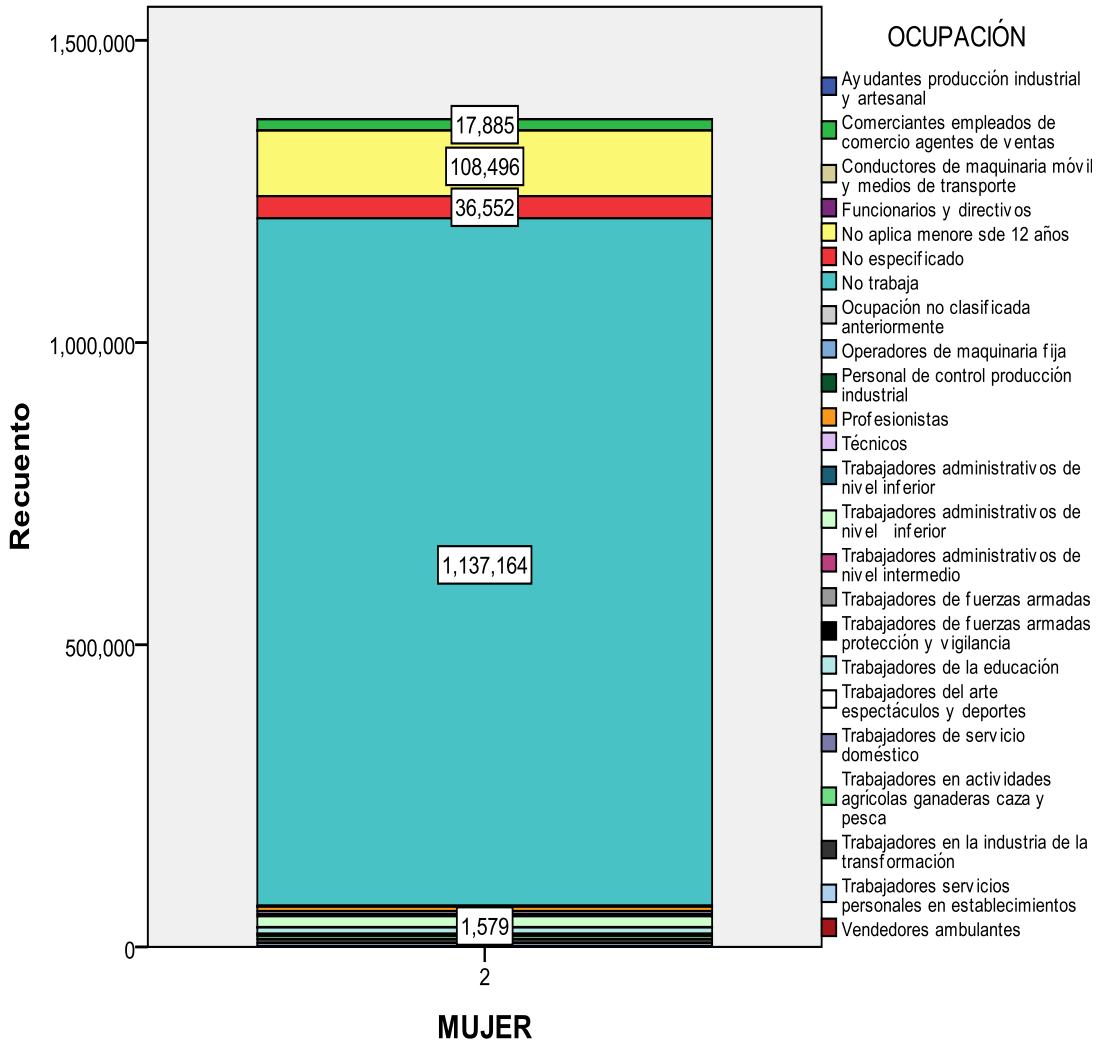


Figura 4.64 Histograma Mujer-Ocupación

Ahora analizaremos los resultados de la relación de mujeres con su estado civil, y lo que se obtuvo fue que existen más mujeres casadas, viudas y solteras, existiendo una población mínima de mujeres divorciadas y en unión libre.

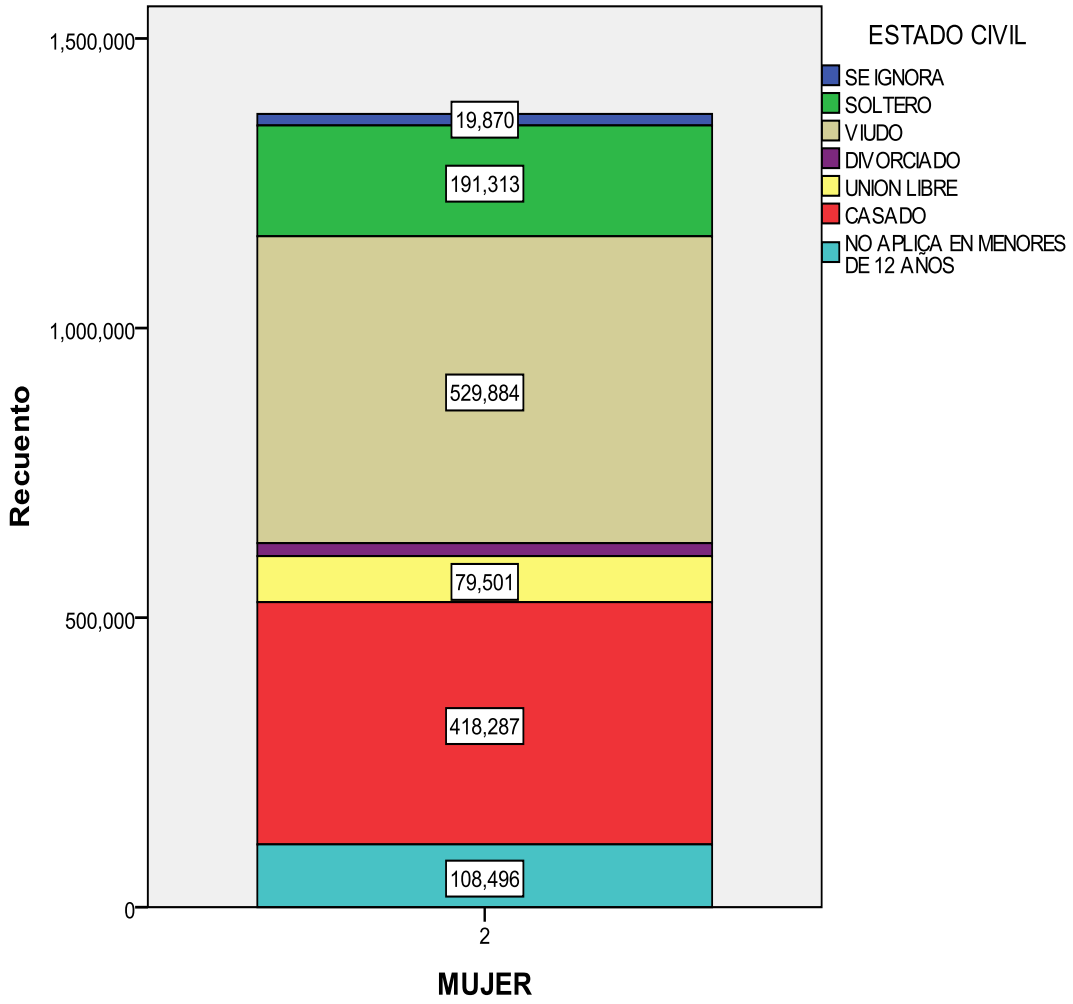


Figura 4.65 Histograma Mujer-Estado Civil

La gráfica no muestra el resultado del estudio que nos indica que personas sufren más de violencia familiar con respecto a su edad, hablando de la violencia general, lo que obtuvimos fue que los que los niños de 1 a 3 años de edad son los que más sufren de violencia, y después son los vemos otro pico en los de 27 años, los que menos sufre de violencia son los que tienen edades de 12, 53 ó de 80 a más años, existiendo una excepción en las personas de 97 años de edad ya que en este vuelve a existir un pico no tan grande, pero si significativo.

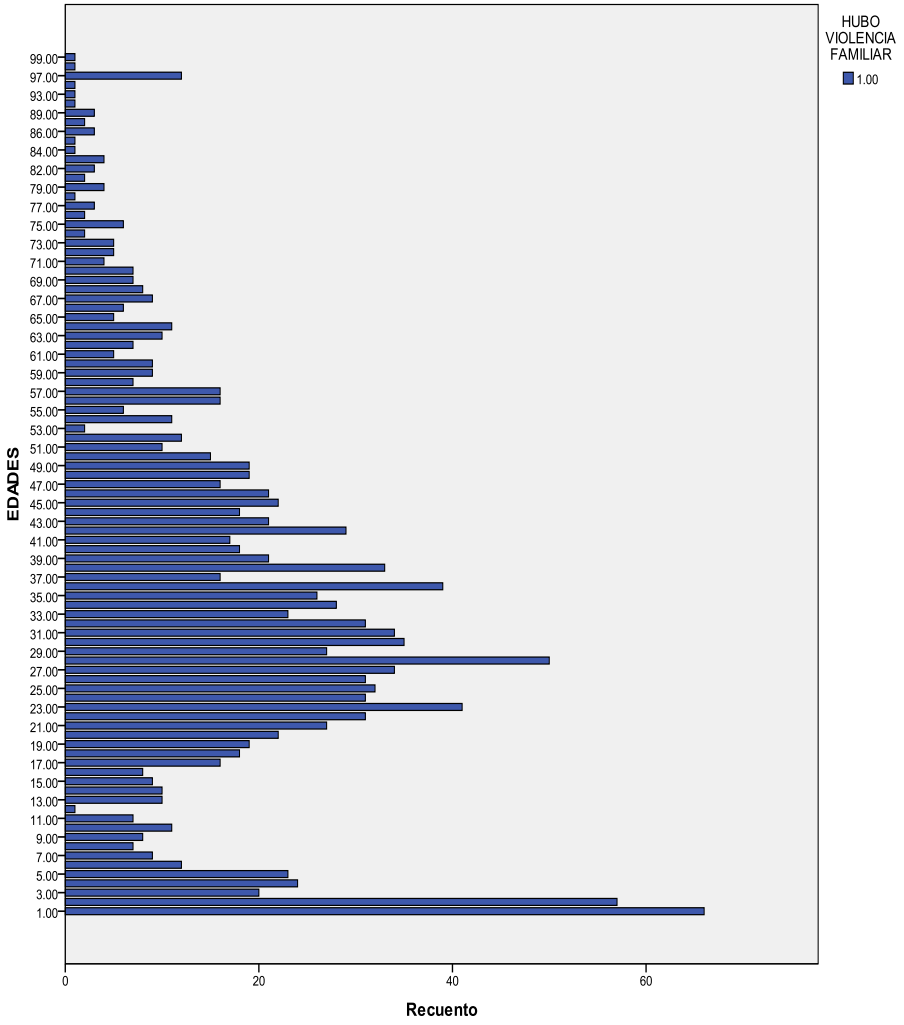


Figura 4.66 Histograma Edad-Hubo violencia

Se debe aclarar que los años que se van a analizar son del 2004 al 2008 ya que en nuestra base de datos solo en esos años existe nuestro atributo ‘violencia familiar’ que es el que nos marca la pauta para nuestra predicción.

Y el año del 2009 lo vamos a ocupar como datos de prueba para verificar si es correcta nuestra predicción que tenemos con nuestros datos de entrenamiento.

Después del análisis estadístico previo que se realizó anteriormente pasamos a decidir que atributos son los más importantes para poder formar las vistas minables. Y ha explicar con más detalle cada uno de los atributos que se va a utilizar.

ATRIBUTO	DESCRIPCIÓN
VIOLENCIA FAMILIAR	Fue el atributo que tomamos como base para realizar nuestras vistas, ahora bien este atributo nos dice si hubo violencia familiar al momento del fallecimiento.
ESTADO CIVIL	Podremos saber si la persona estaba casada, soltera, viudo, en unión libre, divorciado, menor de 12 años o se ignora.
EDAD	Para saber la edad tenemos que tomar en cuenta 2 cosas, la edad como tal, con un valor numérico y un indicador que nos dice si la edad está en años (A), días (D), meses (M), horas (H). Cabe aclarar que para la violencia contra las mujeres solo vamos a utilizar el indicador de años (A) ya que solo estamos tomando en cuenta un rango de 5 a 120 años, para las mujeres que sufren d violencia.
OCUPACIÓN	Nos indicara que ocupaciones son en las que las personas sufren más de violencia familiar, por ejemplo en estudios realizados anteriormente, para lo de la violencia contra las mujeres nos indica que las mujeres que están inactivas son las que más sufren de violencia.
ESCOLARIDAD	Mostrara el grado máximo de estudios que alcanzaron las personas fallecidas, y ver su relación con la violencia familiar, es decir cuál es el nivel de estudios, en donde se hace más presente la violencia familiar.
SEXO	Solo lo usaremos para la violencia contra las mujeres ya que este atributo junto con el de violencia familiar son la base para armar dicha vista minable.

Tabla 4.20 Atributo Descripción

Ahora bien ¿Cómo fue que llegamos a la conclusión que estos atributos serían los indicados para esta vista?, esto se logro con las siguientes consultas:

Primero hicimos un conteo de todas las personas que tenían el atributo "violencia familiar" en 1, es decir "Hubo violencia familiar" y revisamos su estado civil, haciendo un conteo agrupado por estado civil.

```
select count(edocivil) as conteo, edocivil from DEF06
WHERE violfam = 1
```

```
group by edocivil
order by count(edocivil) asc
```

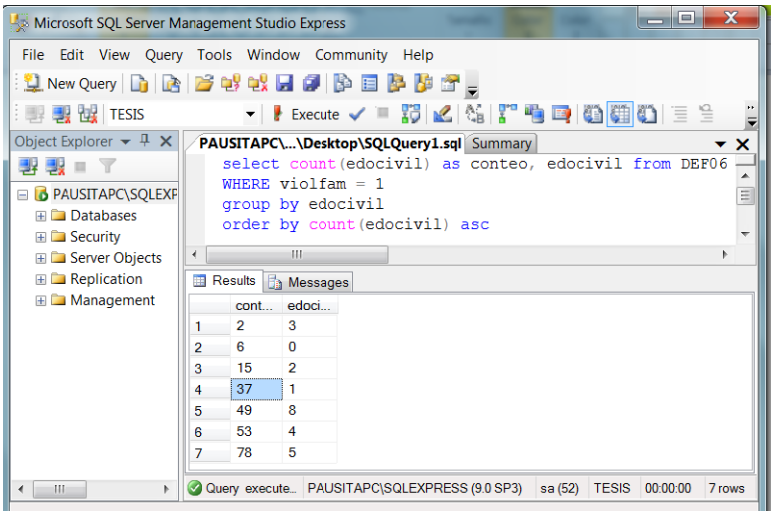


Figura 4.67 Conteo ‘estado civil’ donde hubo violencia familiar

Se observa que casado, union libre y menores de 12 años encabezan nuestro conteo para el año 2006, siendo esto asi en los demas años. Esto nos indica que las personas que fallecienron por violencia familiar estaban en su mayoria con un conyuge o eran menores de 12 años.

Otro atributo en el que notamos una gran variedad de cambio y del cual podríamos sacar información importante fue la edad, esto lo hicimos con la siguiente consulta:

```
select count(edadvalor) as conteo, edadvalor from DEF06
WHERE violfam = 1
group by edadvalor
order by edadvalor asc
```

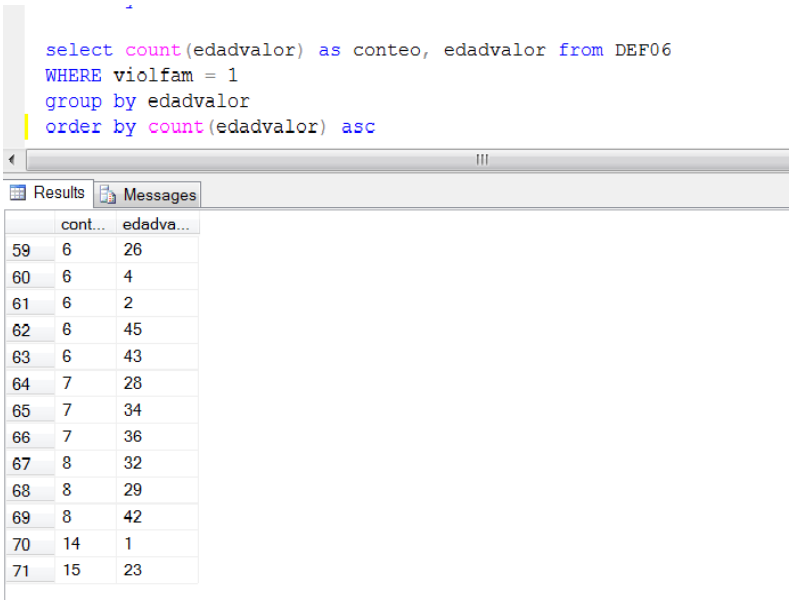


Figura 4.68 Conteo de ‘edad’ donde hubo violencia familiar

Se puede observar que para el año 2006 se tiene bastante variación en cuento a las edades.

Ahora bien empezaremos a establecer si una persona tiene riesgo o no de sufrir violencia familiar. Con base en los resultados de las consultas se establecieron los siguientes criterios:

Hablando solo para la población femenina y de un rango de edad de 15 a 120 años.

Riesgo / Atributo	Muy Alto	Alto	Bajo	Muy Bajo
Ocupación	Inactiva	Cualquiera ó Inactivo	Cualquiera ó Inactivo	Cualquiera ó Inactivo
Escolaridad	Sin Nivel	1. Secundaria Completa	Secundaria incompleta	1. Primaria incompleta
		2. Primaria completa		2. Superior
Estado Civil	1. Casadas	Solteras	Divorciadas	Viudas
	2. Unión Libre			

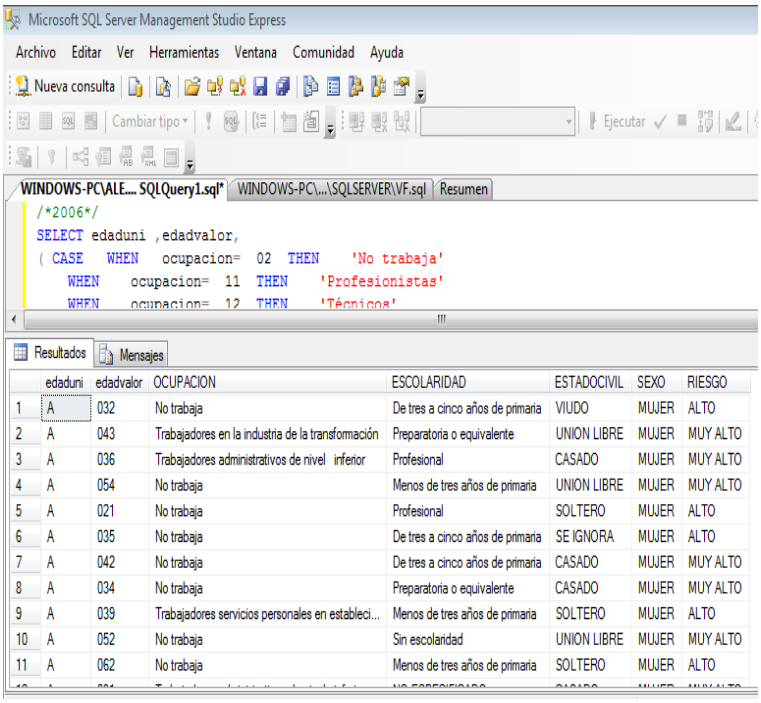
Tabla 4.21 Riesgo/Atributo (mujeres)

Para la clasificación de violencia general vamos a tomar en cuenta a toda la población.

Riego / Atributo	Muy Alto	Alto	Bajo
Edad	0 a 9	10 a 44	45 en adelante

Tabla 4.22 Riesgo/Atributo (general)

Después se procede a programar nuestra vista ‘Violencia Familiar’ ya con el nuevo atributo llamada “RIESGO”.



	edaduni	edadvalor	OCUPACION	ESCOLARIDAD	ESTADOCIVIL	SEXO	RIESGO
1	A	032	No trabaja	De tres a cinco años de primaria	VIUDO	MUJER	ALTO
2	A	043	Trabajadores en la industria de la transformación	Preparatoria o equivalente	UNION LIBRE	MUJER	MUY ALTO
3	A	036	Trabajadores administrativos de nivel inferior	Profesional	CASADO	MUJER	MUY ALTO
4	A	054	No trabaja	Menos de tres años de primaria	UNION LIBRE	MUJER	MUY ALTO
5	A	021	No trabaja	Profesional	SOLTERO	MUJER	ALTO
6	A	035	No trabaja	De tres a cinco años de primaria	SE IGNORA	MUJER	ALTO
7	A	042	No trabaja	De tres a cinco años de primaria	CASADO	MUJER	MUY ALTO
8	A	034	No trabaja	Preparatoria o equivalente	CASADO	MUJER	MUY ALTO
9	A	039	Trabajadores servicios personales en estableci...	Menos de tres años de primaria	SOLTERO	MUJER	ALTO
10	A	052	No trabaja	Sin escolaridad	UNION LIBRE	MUJER	MUY ALTO
11	A	062	No trabaja	Menos de tres años de primaria	SOLTERO	MUJER	ALTO

Figura 4.69 Vista Minable en Sql Server 2006

Se debe aclara que todo este proceso se ejecuto para los años 2004, 2005, 2006, 2007 y 2008, haciendo uniones para que salga un resultado general de todos los años desde el 2004 al 2008, ya que estos van a ser tomados como los datos de entrenamiento (el query completo lo encontramos en el ANEXO (scrips) y los datos del 2009 serán tomados de prueba.

4.4.1 J48 y Predicción con Rapid Miner

Ahora pasaremos realizar la minería de datos para obtener nuestra predicción con el año de prueba 2009.

Se hará una predecir por medio de reglas encontradas si las personas tienen RIESGO de sufrir violencia familiar y por consecuencia fallecer. Se generaron reglas en base en los años anteriores, 5 años anteriores en donde se tomaran como datos de entrenamiento posteriormente se creara un modelo con ayuda del algoritmo de árboles de decisión J48.

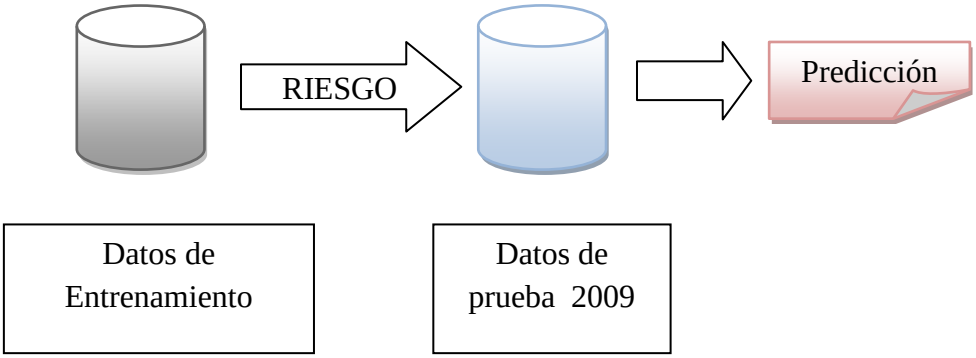


Figura 4.70 Proceso para la predicción

Primero nos enfocaremos en violencia general, se tomara en cuanta solo los años del 2004 al 2009, tomando el 2009 como nuestro año de **prueba**.

Esto se realizara con la herramienta Rapid Miner de la siguiente forma:

Iniciamos Rapid Miner con esta pantalla de bienvenida.

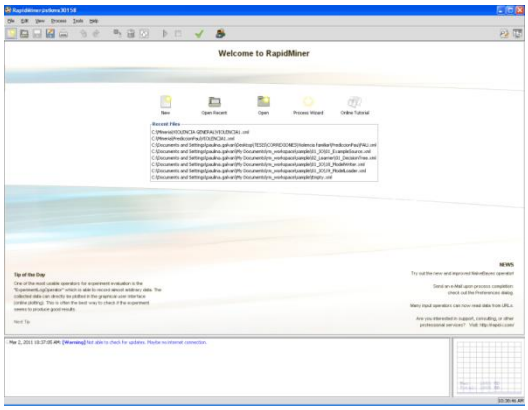


Figura 4.71 Pantalla Inicial Rapid Miner

En la parte de arriba, hay que dar clic en la parte de modo principiante / experto de tal forma que quede en modo experto. Esto es con el fin de que se puedan mostrar todos los parámetros. Dar clic en nuevo (o new). Se mostrara la pantalla siguiente con un root inicial (Figura 4.71)

Se procede a crear el árbol de procesos en Rapid Miner. Hacer clic con el botón derecho del mouse. Del menú contextual seleccionar New Operator, IO, Examples, CSVExampleSource o ExelExampleSource. Figura 4.72

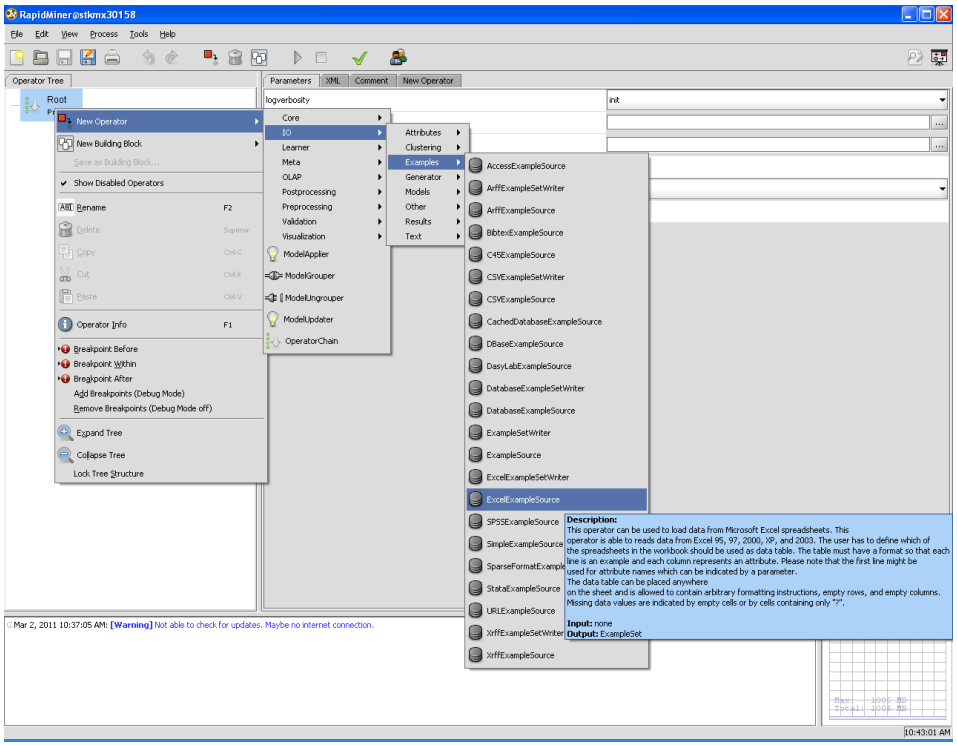


Figura 4.72 Seleccionar un archivo de fuente CSVExampleSource o ExelExampleSource.

De la misma manera haciendo clic con el botón derecho del mouse sobre el operador root, seleccionar New Operator, Visualization, ExampleVisualizer. Esto con el fin de poder visualizar los datos con datos estadísticos y poder verificar que los datos que se quieren cargar sean precisamente esos y que estén completos.

Se vuelve a teclear con el botón derecho del mouse sobre el operador root para seleccionar un tipo de validación con el fin de obtener la confiabilidad del modelo (es decir, qué tan acertado es y qué tan bien aprende de los datos). En este caso se usará la validación cruzada o XValidation. Del menú seleccionar New Operator, Validation y XValidation, para su configuración la casilla **create_complete_model** debe estar seleccionada. Véase la Figura 4.73

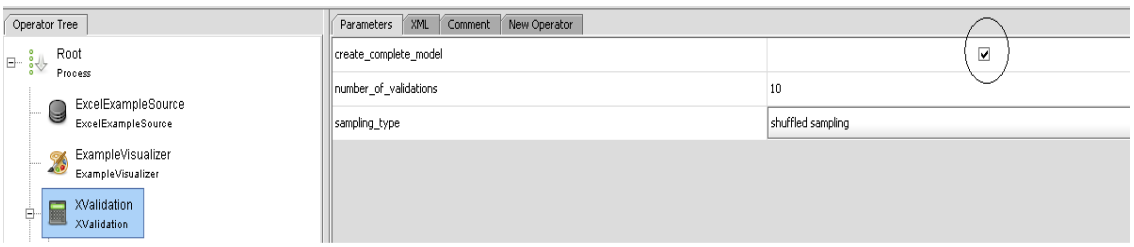


Figura 4.73 Seleccionar XValidation y crear modelo.

Sobre el operador XValidation dar clic con el botón derecho del mouse para agregar un operador cadena, el cual sirve para unir varios operadores a un operador. Se empieza con el operador cadena u OperatorChain del menú contextual: New Operator, OperatorChain (Figura 4.74).

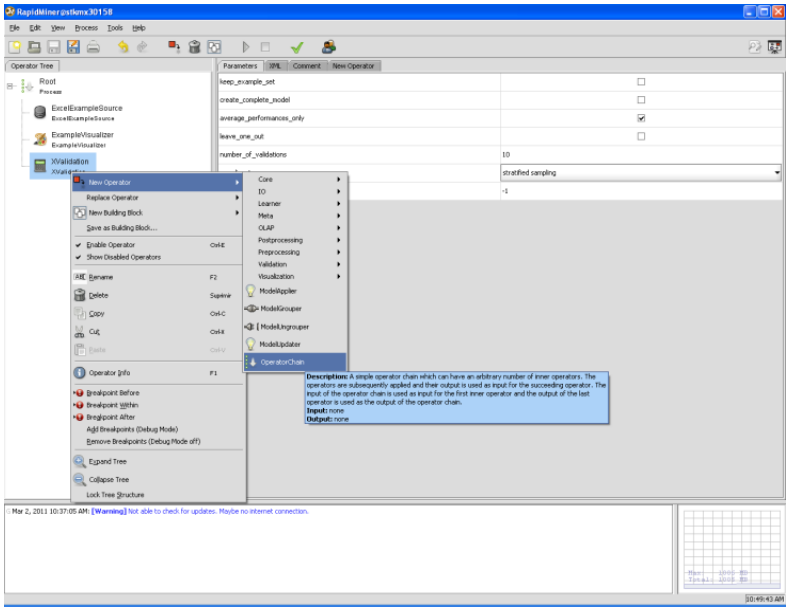


Figura 4.74 Seleccionar el Operador Cadena

Dar clic sobre este nuevo operador (j48) para poder visualizar los parámetros del lado derecho. Sólo seleccionar la casilla **keep_example_set** y los demás valores no se mueven ya que son los que nos marcan por default Véase la Figura 4.75

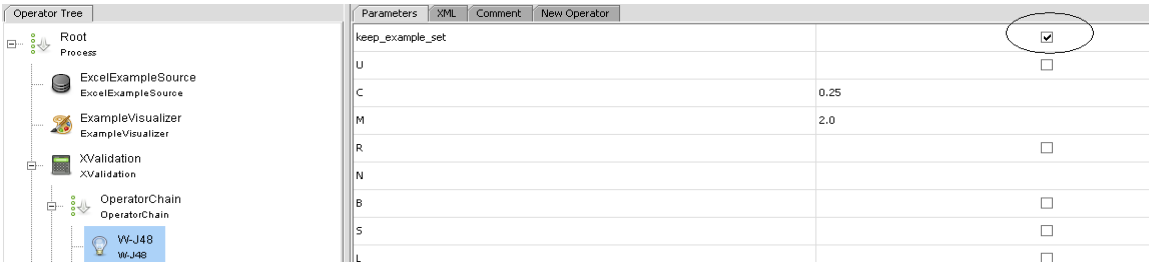


Figura 4.75 Seleccionar árbol W-J48

Se puede también agregar un operador que guarde el modelo generado. Esto es útil cuando se quiere volver a aplicar el modelo a otros datos (por ejemplo, de prueba). Dar clic con el botón derecho del mouse sobre el operador cadena u OperatorChain y en el menú contextual seleccionar New Operator, IO, Models, ModelWriter. Véase la Figura 4.76

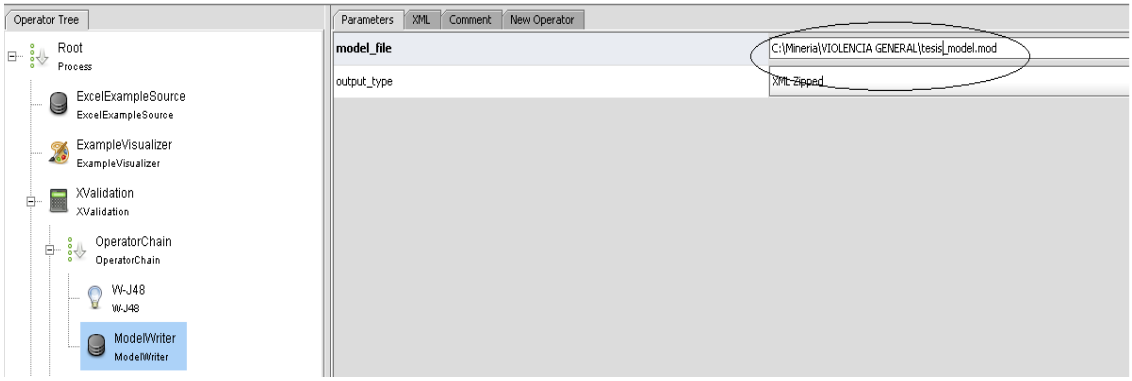


Figura 4.76 Seleccionar ModelWriter

En el ModelWriter se guardara nuestro modelo en un archivo .mod y una vez seleccionado el nombre y la ruta en dónde se va a guardar, dar clic en abrir. No hay problema si el archivo no existe, ya que éste se creará automáticamente. Si el archivo a guardar ya existe, se sobrescribirá.

Sobre el operador XValidation dar clic con el botón derecho del mouse para seleccionar otro operador cadena u OperatorChain con el fin de unir los procesos. Esto se hace porque en Rapid Miner los operadores están limitados a tener pocos nodos hijos, entonces con este operador cadena se pueden unir más operadores. Véase la Figura 4.77

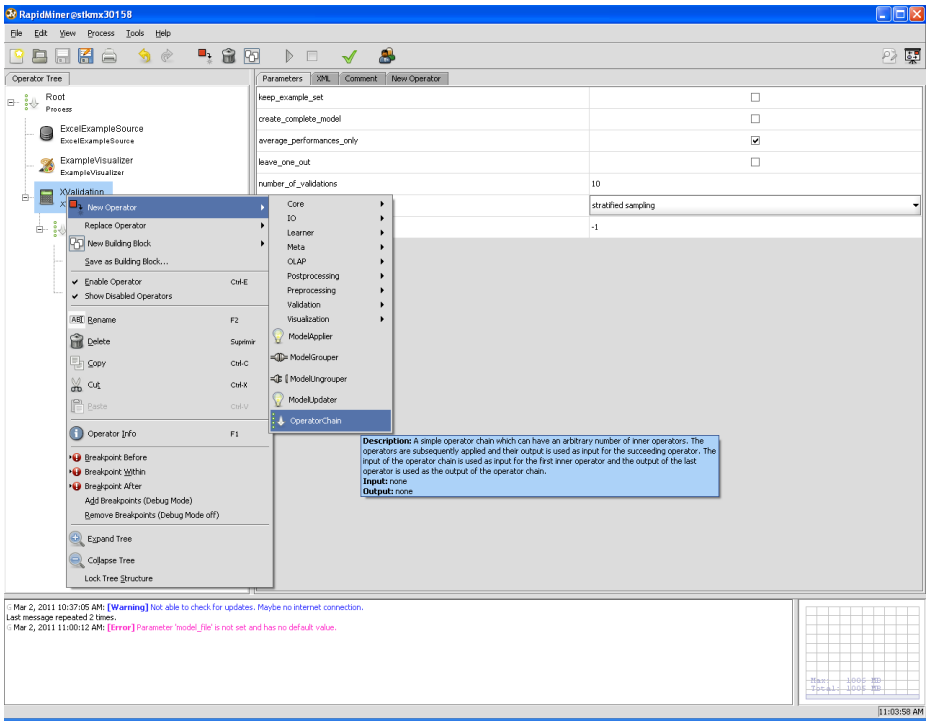


Figura 4.77 Seleccionar OperatorChain con el fin de unir los procesos

Ahora dar clic con el botón derecho del mouse sobre el nuevo operador OperatorChain para seleccionar un operador que aplique el modelo o ModelApplier. En el menú seleccionar New Operator, ModelApplier.

Dar clic sobre ModelApplier para poder visualizar los parámetros del lado derecho. En este caso, seleccionar las casillas keep_model y create_view. Véase la Figura 4.78

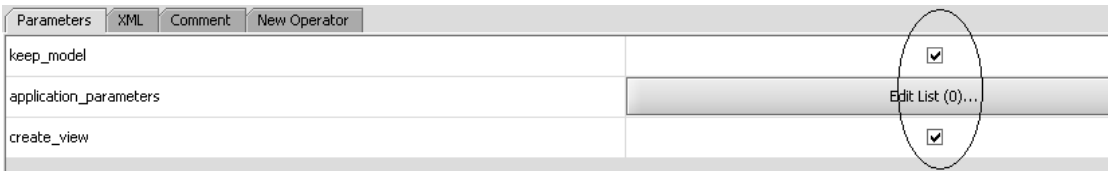


Figura 4.78 Seleccionar ModelApplier

Ahora se va a agregar un operador que evalúe el desempeño del modelo o performance. Se da clic sobre el operador cadena con el botón derecho del mouse y se selecciona New Operator, Validation, ClassificationPerformance. Después dar clic sobre el operador performance para poder visualizar los parámetros del lado derecho y marcar las casillas keep_example_set para poder seguir viendo los datos con los que se está trabajando; y otros parámetros de evaluación que se consideren pertinentes tales como accuracy (exactitud), classification error, root_mean_squared_error y correlation. En el parámetro main_criterion se elige **accuracy**. Véase la Figura 4.79

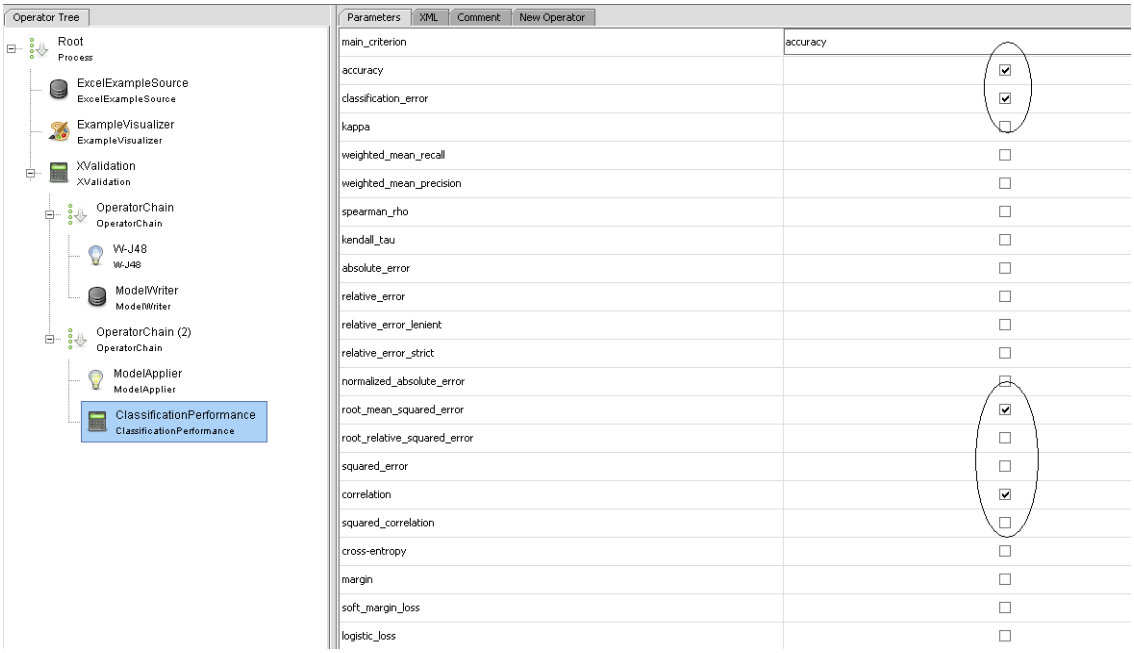


Figura 4.79 Seleccionar ClassificationPerformance

Se procede a crear el árbol de procesos en Rapid Miner. Hacer clic con el botón derecho del mouse. Del menú contextual seleccionar New Operator, IO, Examples,

CSVExampleSource o ExelExampleSource, donde se pondrá la ruta del archivo el cual no tiene la nueva columna RIESGO, la cual será calculada por el modelo. Figura 4.80.

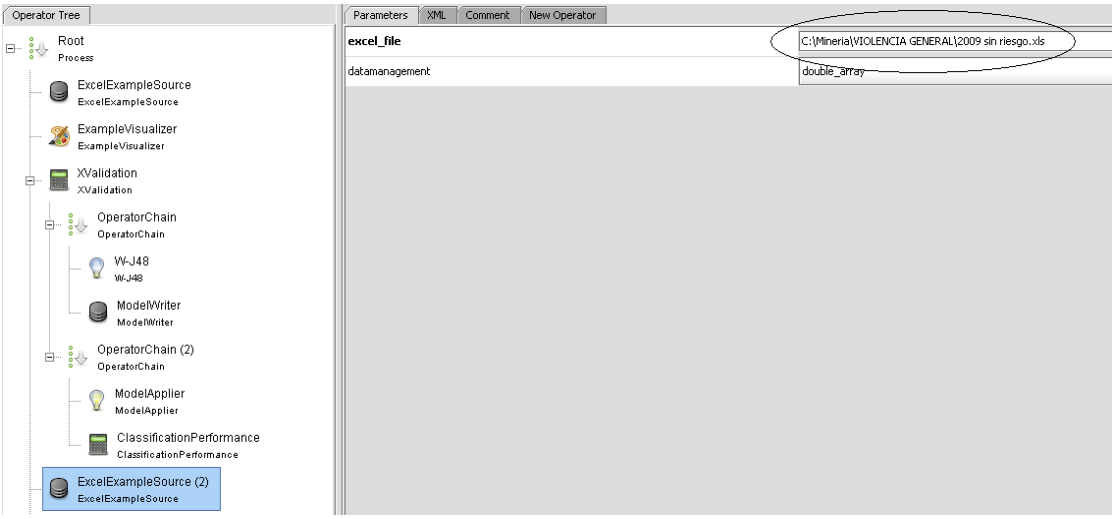


Figura 4.80 Seleccionar CSVExampleSource o ExelExampleSource

De la misma manera haciendo clic con el botón derecho del mouse sobre el operador root, seleccionar New Operator, Visualization, ExampleVisualizer. Esto con el fin de poder visualizar los datos con datos estadísticos y poder verificar que los datos que se quieren cargar sean precisamente esos y que estén completos. Figura 4.81

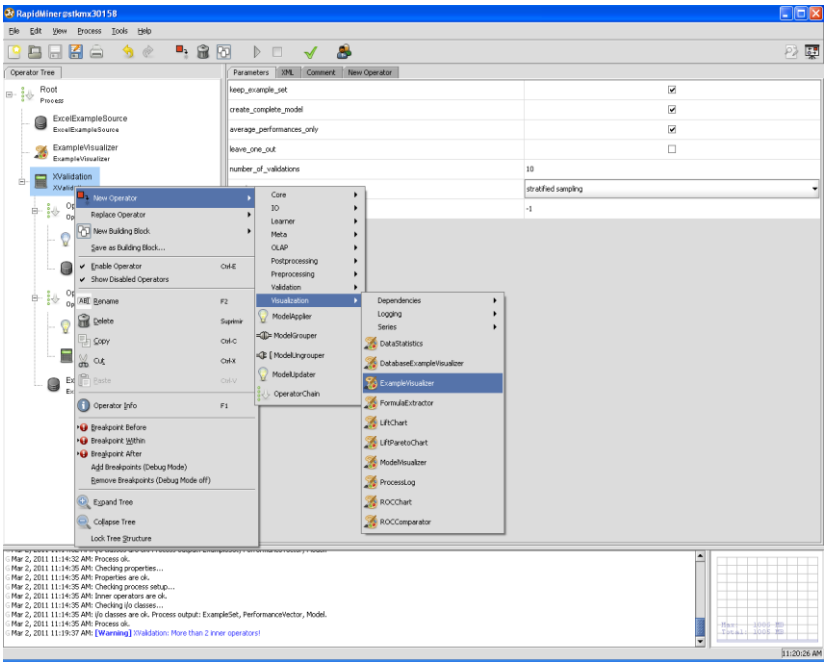


Figura 4.81

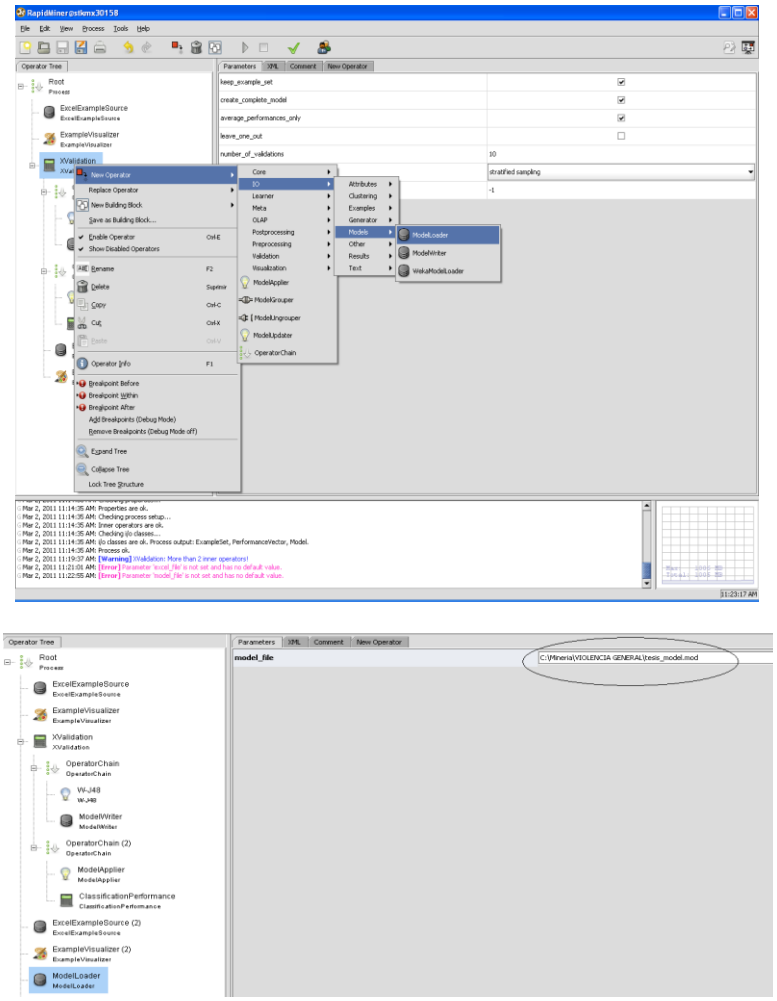


Figura 4.82

Dar clic con el botón derecho del mouse sobre el operador cadena u OperatorChain y en el menú contextual seleccionar New Operator, IO, Models, ModelLoader. Véase Figura 4.82 el cual lee un modelo a partir de un archivo que se genero por un operador en un proceso anterior, una vez al ir modelo generado, se puede aplicar varias veces a los nuevos datos adquiridos.

Y finalmente compilamos y corremos el proceso. Véase Figura 4.83

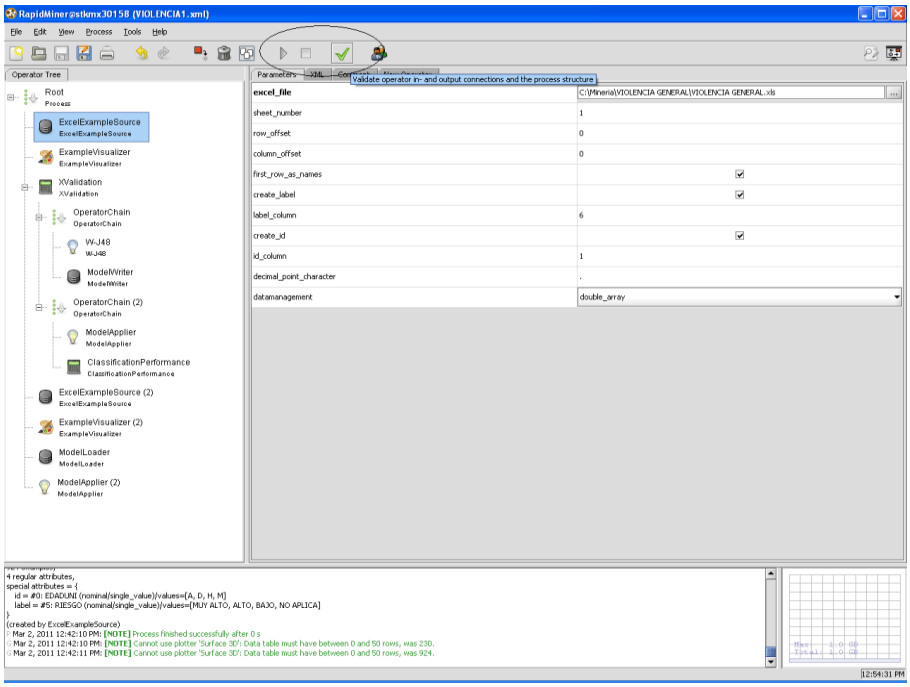


Figura 4.83 Compilar y correr proceso.

Una vez que termina el proceso se muestran los resultados de las predicciones, es decir, Rapid Miner automáticamente agrega tres nuevas columnas: una con el resultado de la predicción y las otras tres con el grado de confianza de ser, en este caso, un “ MUY ALTO”, un “ALTO” o un “BAJO”. Estas dos últimas columnas no son tan contundentes como para afirmar de la veracidad. Lo mejor es comprobar estos resultados con los reales para observar la efectividad del modelo.

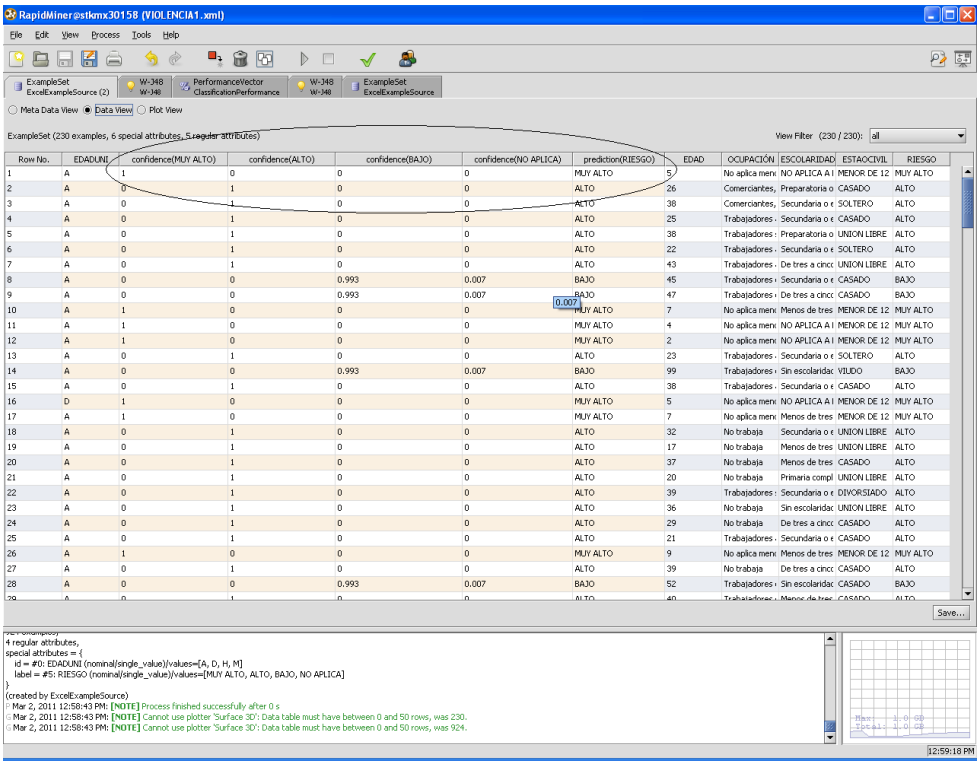
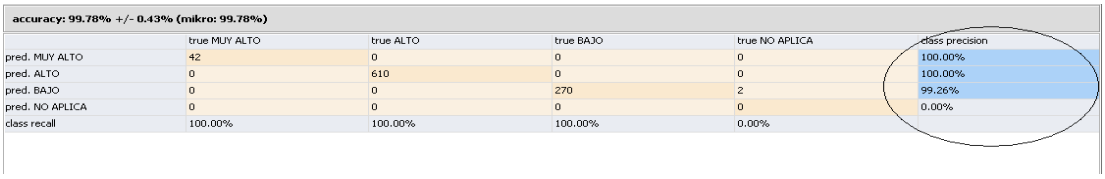


Figura 4.84 Vista de los datos resultantes

Dando clic sobre Meta Data View se podrá ver datos de las variables que se usan, algunas estadísticas, el rango de sus valores además de saber si hay datos desconocidos (Figura 4.84).

Los resultados de precisión (precision), exactitud (accuracy) y la correlación son los siguientes:



accuracy: 99.78% +/- 0.43% (mikro: 99.78%)

	true MUY ALTO	true ALTO	true BAJO	true NO APLICA	class precision
pred. MUY ALTO	42	0	0	0	100.00%
pred. ALTO	0	610	0	0	100.00%
pred. BAJO	0	0	270	2	99.26%
pred. NO APLICA	0	0	0	0	0.00%
class recall	100.00%	100.00%	100.00%	0.00%	

Figura 4.85 Resultados precisión y exactitud

Con una precisión de 100% para MUY ALTO y ALTO y 99.26% para bajo. Ahora que lo mejor es comprobar estos resultados con los reales para observar la efectividad del modelo.

En nuestro árbol j48 se genero el siguiente resultado:

EDAD <= 44

| EDAD <= 9: MUY ALTO (42.0)

| EDAD > 9: ALTO (610.0)

EDAD > 44: BAJO (272.0/2.0)

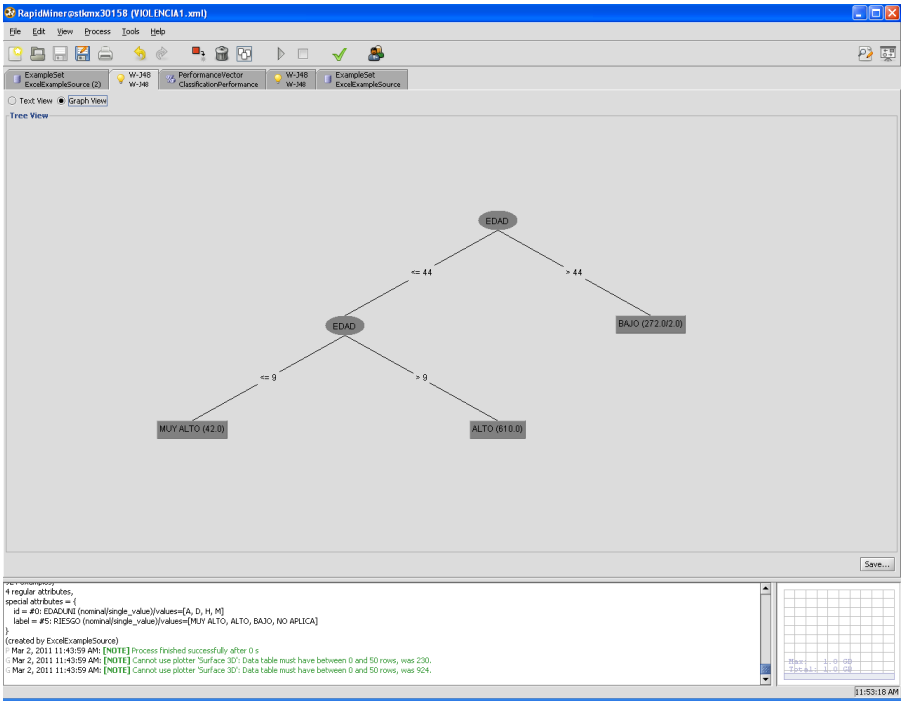


Figura 4.86 W-J48 Resultante

Se puede observar que si la edad es menor o igual a 44 y además es menor o igual años el riesgo es muy alto y si la edad es mayor a 9 años el riesgo es alto, además que se observa que el riesgo es bajo para personas mayores a 44 años.

Ahora bien vamos a realizar la comprobación de nuestros datos con predicción creando dos columnas una llamada "predicción" en la cual pondremos los resultados de "Riesgo" de nuestro modelo, la otra columna llamada "comprobación", llevara nuestro atributo "RIESGO" calculado con sql como especifico en nuestros datos de entrenamiento. Haciendo una comparación donde diremos que si los resultados son iguales poner un "1" y si no poner "0", siendo 230 reg. en total.

No	**Predicción RIESGO WEKA**	**Comprobación RIESGO SQL**	EDAD	OCUPACIÓN	ESCOLARIDAD	ESTADO CIVIL	EDADUNI
1	**MUY ALTO**	**MUY ALTO**	5	No aplica menores de 12 años	NO APLICA A MENOR DE 6 AÑOS	MENOR DE 12 AÑOS	A

No	**Predicción RIESGO WEKA**	**Comprobación RIESGO SQL**	EDAD	OCUPACIÓN	ESCOLARIDAD	ESTADO CIVIL	EDADUNI
2	**ALTO**	**ALTO**	26	Comerciantes empleados de comercio agentes de ventas	Preparatoria o equivalente	CASADO	A
3	**ALTO**	**ALTO**	38	Comerciantes empleados de comercio agentes de ventas	Secundaria o equivalente	SOLTERO	A
4	**ALTO**	**ALTO**	25	Trabajadores administrativos de nivel inferior	Secundaria o equivalente	CASADO	A

Figura 4.87 Ejemplo de comparación

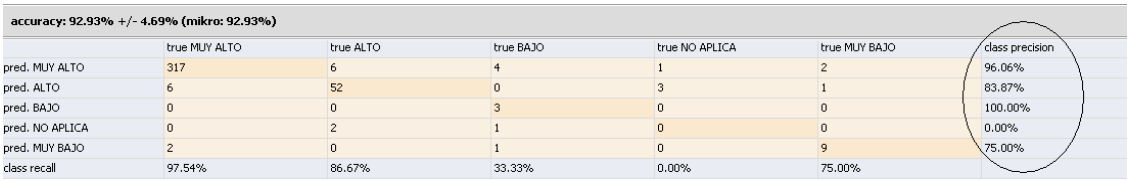
Entonces obteniendo el desempeño para cada tabla se tiene:

Acertó 230 --> 100%

No acertó: 0 --> 0 %

Violencia Familiar Mujeres

El proceso que se describió anteriormente se repite para violencia familiar para mujeres tenido los siguientes resultados:



accuracy: 92.93% +/- 4.69% (mikro: 92.93%)

	true MUY ALTO	true ALTO	true BAJO	true NO APLICA	true MUY BAJO	class precision
pred. MUY ALTO	317	6	4	1	2	96.06%
pred. ALTO	6	52	0	3	1	83.87%
pred. BAJO	0	0	3	0	0	100.00%
pred. NO APLICA	0	2	1	0	0	0.00%
pred. MUY BAJO	2	0	1	0	9	75.00%
class recall	97.54%	86.67%	33.33%	0.00%	75.00%	

Figura 4.88 Resultado precisión y exactitud (mujeres)

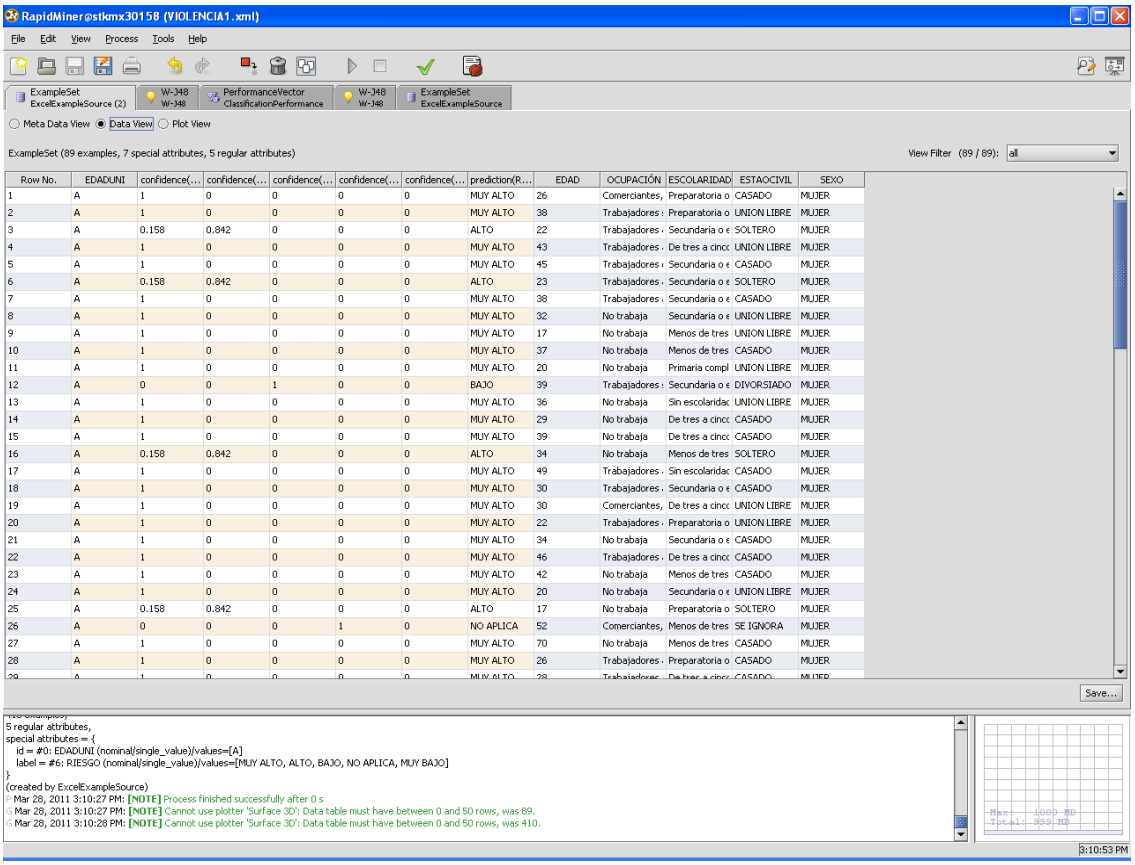


Figura 4.89 Resultado predicción mujeres

Se observa una columna con el resultado de la predicción y las otras tres con el grado de confianza de cero, en este caso, “MUY ALTO”, “ALTO”,“BAJO” o “MUY BAJO”. Estas dos últimas columnas no son tan contundentes como para afirmar la veracidad. Lo mejor es comprobar estos resultados con los reales para observar la efectividad del modelo.

Entonces obteniendo el desempeño para cada tabla se tiene:

Acertó 87 --> 97.7%

No acertó 2 --> 2.3 %

La forma en cómo se comprobó fue creando dos columnas una llamada “predicción” en la cual pondremos los resultados de “Riesgo” de nuestro modelo, la otra columna llamada “comprobación”, llevara nuestro atributo “RIESGO” calculado con sql como especifico en nuestros datos de entrenamiento. Haciendo una comparación donde diremos que si los resultados son iguales poner un “1” y si no poner “0”, siendo 89 reg. en total, en el caso de mujeres.

Al realizar esta tarea se observo que el modelo ha aprendido bien con el número de datos que entrenamiento no más de 500 registros desde el 2004, podemos utilizar estos datos para realizar predicciones en el futuro basándonos en datos como escolaridad, ocupación, estado civil, tenido una estrecha relación las mujeres que no tienen estudios tienen más riesgo de sufrir violencia familiar y en violencia general los niños con más riesgo de sufrir maltrato.

Como vemos el modelo al ser bien entrenado nos puede arrojar datos que se apegan a la realidad arrojándonos una predicción casi al 100% acertado, pero siempre con su grado de error, ya que estos también son parte de un modelo de predicción pero como nos podemos dar cuenta los errores que pertenecen a los datos no acertados son casi nulos ya que nos marca un porcentaje bajo, y también es un indicador que nos muestra que en verdad el modelo aprendió muy bien.

Conclusión "Violencia Familiar"

Una vez que hemos pasado por todo el proceso, revisamos a que sector de la población se va dirigir dicho conocimiento extraído, como se ha visto este conocimiento va más dirigido a Derechos Humanos y también a la secretaria de salud. Tomando un poco el papel de estas dos instituciones mencionaremos algunas de las posibles soluciones que podrían mejorar la situación para que baje el índice de muertes a causa de la violencia familiar, para empezar hablaremos de la violencia contra los niños ya que son los que más sufre de ella algunas de las posibles soluciones serían:

- Visitas de trabajadora social a las casas para supervisar las condiciones de vida de los infantes.
- En las guarderías que existan educadoras que estén capacitadas para identificar las conductas de los niños que sufren violencia familiar, y reportarlas con las autoridades correspondientes.

Hablando a hora de la violencia contra la población femenina las posibles soluciones que se podrían tomar son:

- Educar a la población pequeña (niños), sobre el respeto hacia a las mujeres y cualquier ser vivo es decir infundir bien los valores en la casa y escuelas.
- Poner penas más severas a los agresores.
- Proteger tanto física como psicológicamente a las agredidas

4.5 Mujeres embarazadas

En esta fase analizaremos los objetivos, y esto lo haremos con las siguientes preguntas:

- **¿Qué parte de los datos es pertinente analizar?**

Se analizara la información de las mujeres que hayan fallecido a causa de alguna complicación con su embarazo, debido a su edad (hablamos de mujeres de 18 años o menores ó de mujeres de 38 hacia arriba).

- **¿Qué tipo de conocimiento se desea extraer y cómo se debe presentar?**

Se extraerá el número de mujeres que fallecieron embarazadas o en periodo de lactancia, es importante aclarar que hablamos de mujeres que se embarazan a una edad avanzada ó a una edad temprana.

- **¿Qué conocimiento puede ser válido, novedoso e interesante?**

Conocer el número de mujeres que mueren a causa de que se embarazan a una edad temprana ó a una edad avanzada y estas complicaciones en su embarazo la llevan a la muerte, ya sean antes o después del parto.

También podremos identificar cual es el rango de edad con más índice de mortalidad para poner más atención en ese sector de la población.

- **¿Qué reglas o modelos de decisión están utilizando?**

Utilizaremos la regresión lineal simple, siendo esta una tarea sencilla. Tomando en cuenta que la regresión es solo un modelo de aproximación donde nuestras variables serán edad y conteo defunción de mujeres.

- **¿Qué decisiones son críticas?**

Interpretar los resultados de manera correcta, saber qué hacer con los resultados y hacia qué sector de la sociedad dirigir los resultados.

Antes de empezar a realizar la fase de la minería de datos, es importante realizar un análisis estadístico previo. Es decir hacer una exploración de los datos para saber qué datos pueden ser los más adecuados para poder realizar la vista minable. En este caso sabemos que nuestros atributos deben ser atributos numéricos para poder realizar adecuadamente la regresión lineal.

Nuestro atributo principal ó más relevante será la edad, entre otras cosas porque es un atributo numérico. Pero no se debe de perder de vista que este atributo debe de ir en conjunto con el atributo “relación con el embarazo”, ya que es la parte de población que nos interesa estudiar.

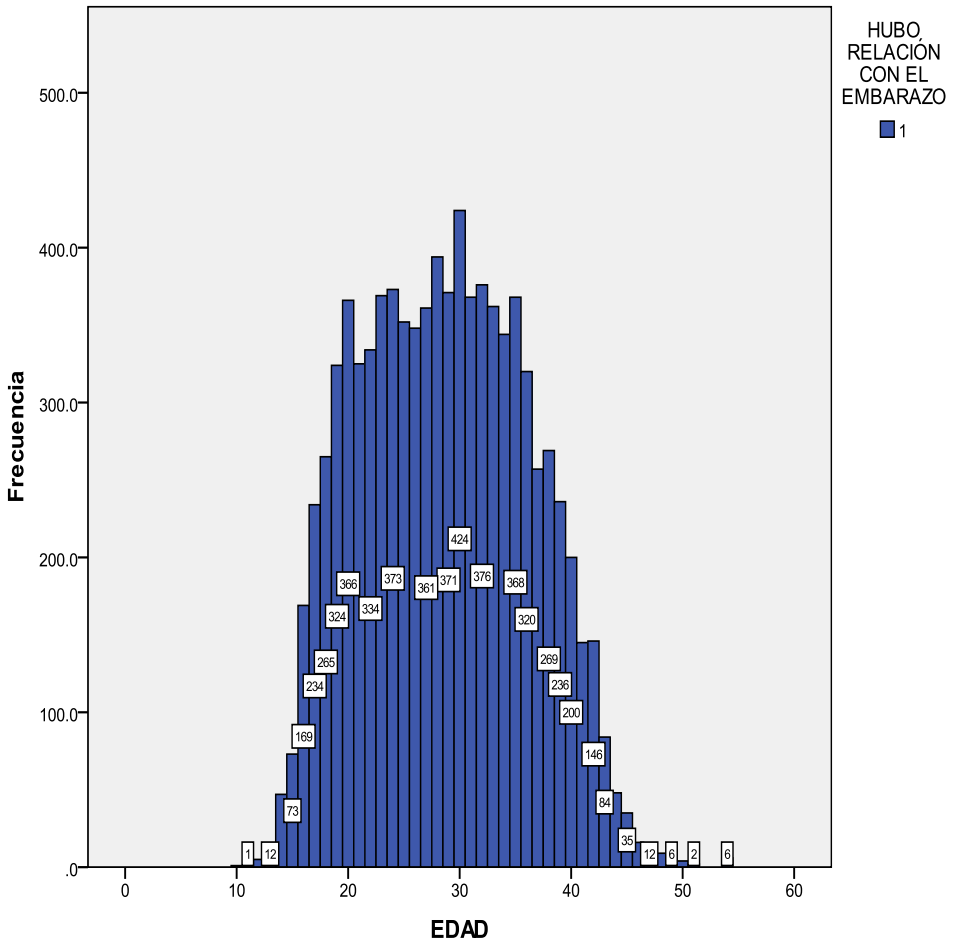


Figura 4.90 Frecuencia de Edad

En esta gráfica podemos ver cómo se comportan las edades con este atributo, teniendo más incidencias en edades de los 20 a los 40 años.

Pero qué pasa con las mujeres que no están en una edad fértil, tomaremos las edades menores ó iguales a 18 años y mayores ó igual a 38 años. Esto resultado lo

obtendremos realizando gráficas del comportamiento de las causas de la defunción con edades.

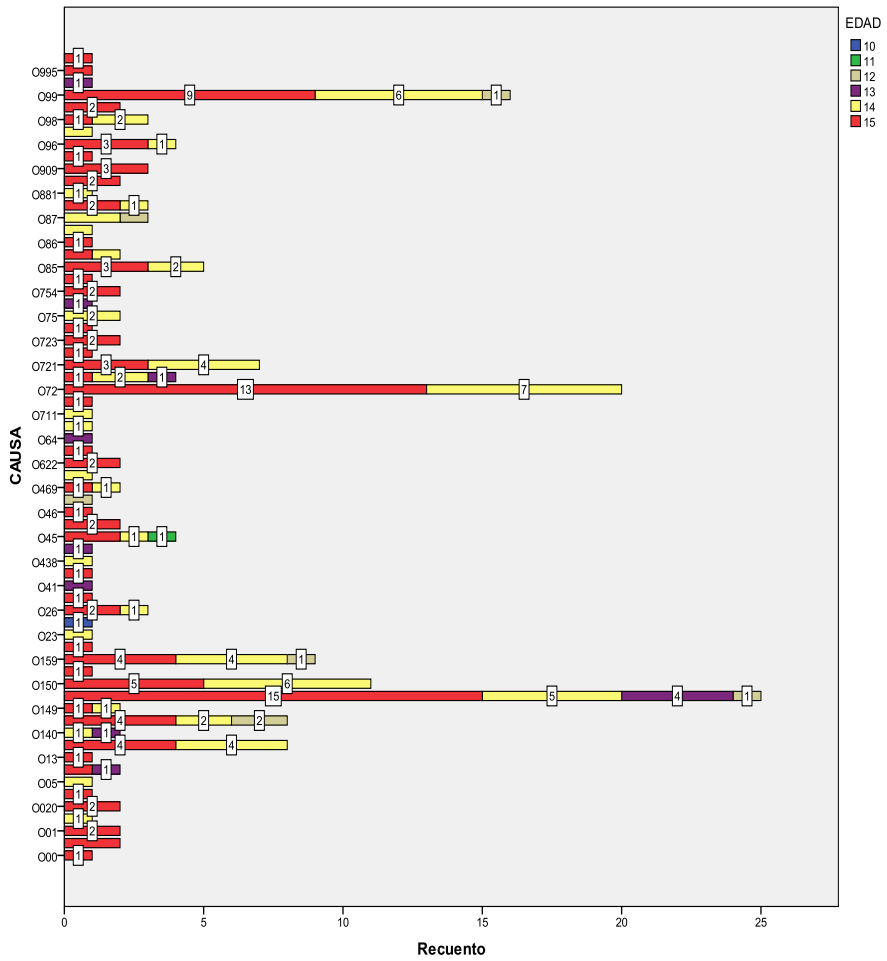


Figura 4.91 Causa-Edad de 10-15

Se puede observar en la gráfica siguiente que en su mayoría son defunciones de mujeres de entre 14 y 15 años, y sus causas principales son:

- Otras enfermedades maternas clasificables en otra parte, pero que complican el embarazo, el parto y el puerperio
- Hemorragia postparto
- Eclampsia

Siendo la cantidad de incidencias no mayor a 25 por causa.

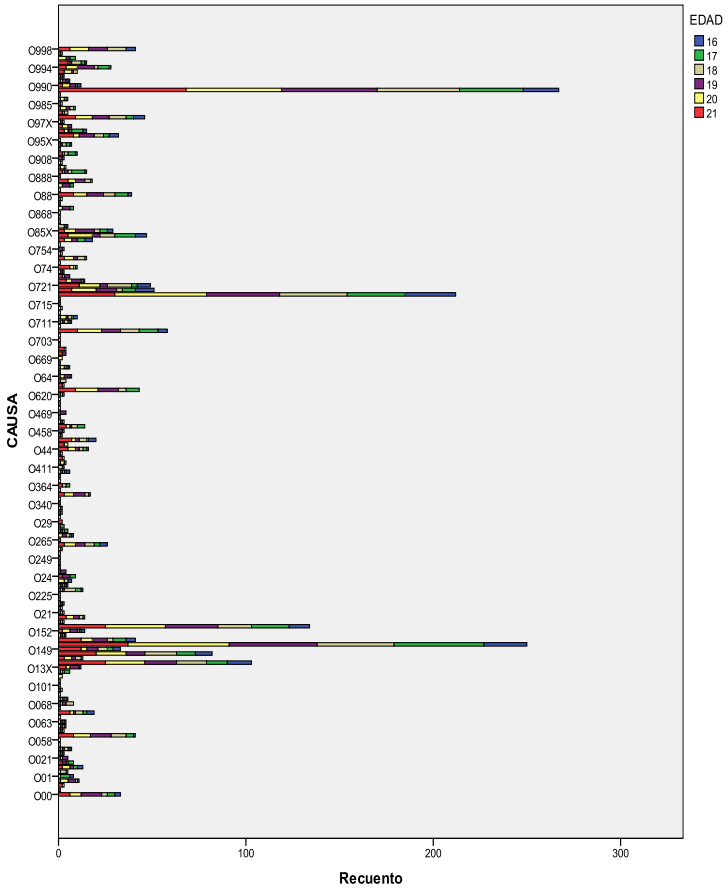


Figura 4.92 Causa-Edad de 16-21

En esta gráfica se tomara el rango de edades de 16 a 21 años y se observa que las incidencias aumentan llegando a 300 por causa, en donde las principales causas son:

- Eclampsia
- Otras enfermedades maternas clasificables en otra parte, pero que complican el embarazo, el parto y el puerperio
- Hemorragia postparto

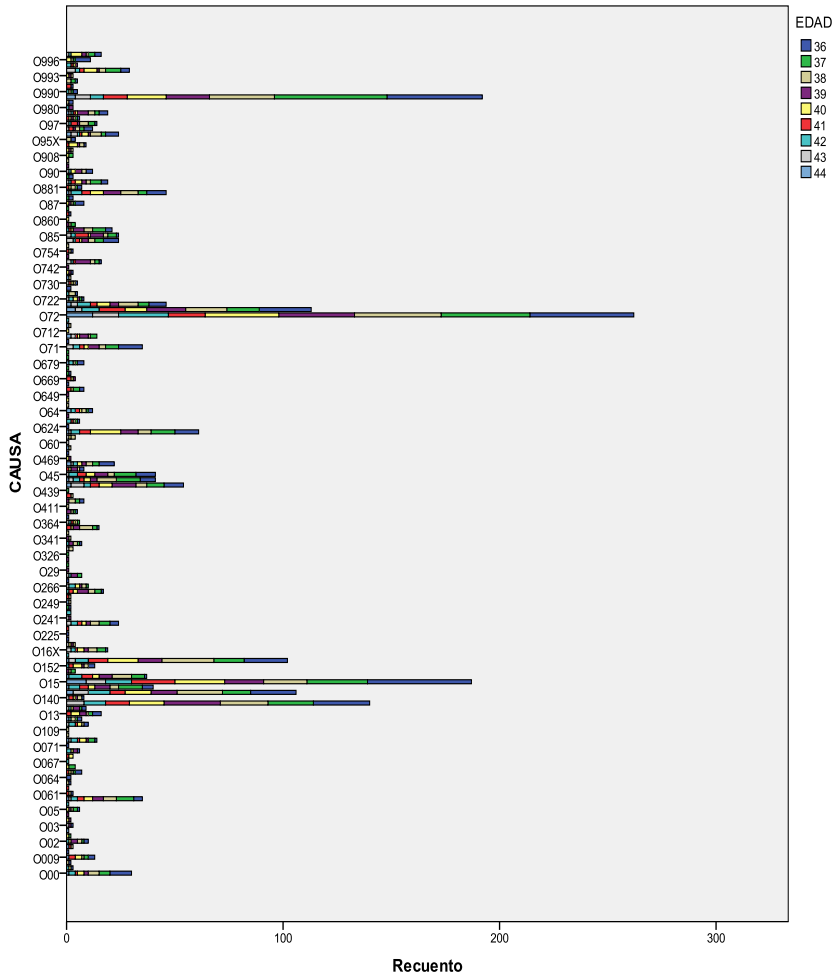


Figura 4.93 Causa-Edad de 36-44

Tomando edades de 36 a 44 años las incidencias se mantienen constantes en 300 incidencias, siendo que en estas edades ya hay más riesgo de tener algún problema en el embarazo. Las principales causas son:

- Hipertensión gestacional [inducida por el embarazo] con proteinuria significativa
- Eclampsia
- Otras enfermedades maternas clasificables en otra parte, pero que complican el embarazo, el parto y el puerperio

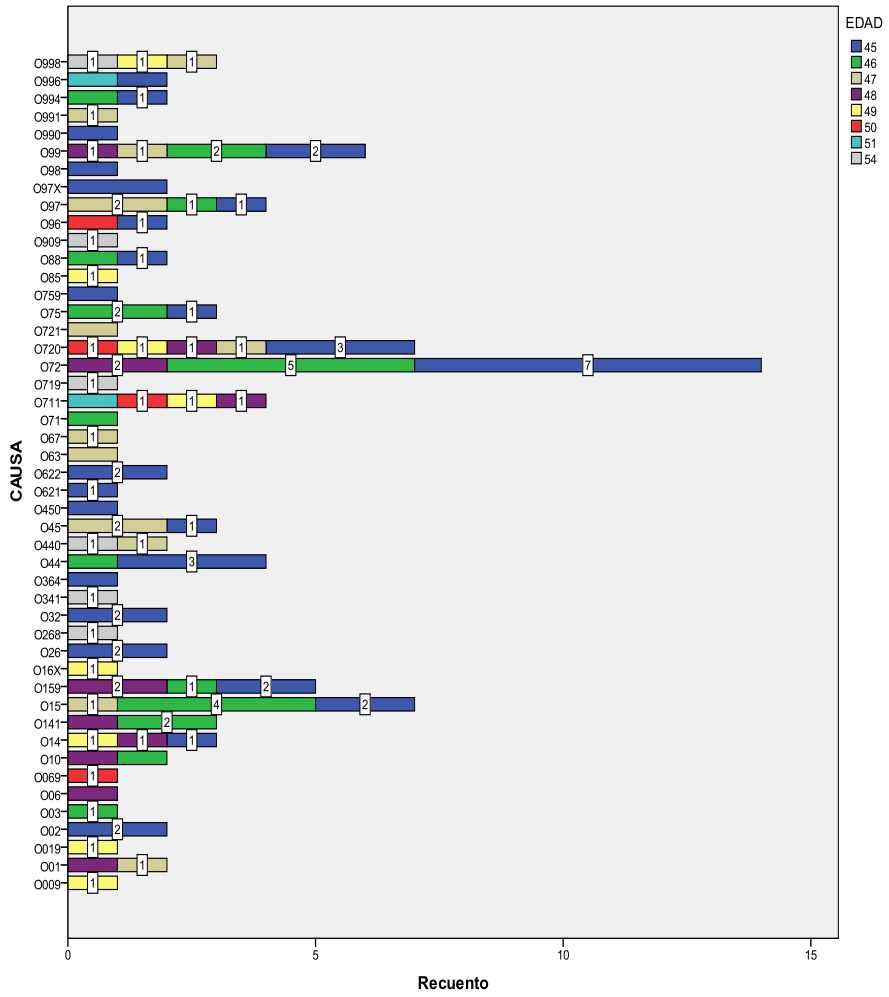


Figura 4.94 Causa-Edad de 45 – 54

Tomando las edades más avanzadas de 45 a 54, se observa que los casos son menores, máximo 15 por causa, teniendo como principales causas:

- Hemorragia postparto
- Eclampsia, en período no especificado
- Hemorragia del tercer período del parto

Podemos tener a todas las mujeres las cuales su defunción tuvo alguna relación con el embarazo. Para la siguiente consulta, utilizamos el atributo 'relación con embarazo' en 1=SI y traemos la edad de cada defunción donde hubo relación con el embarazo.

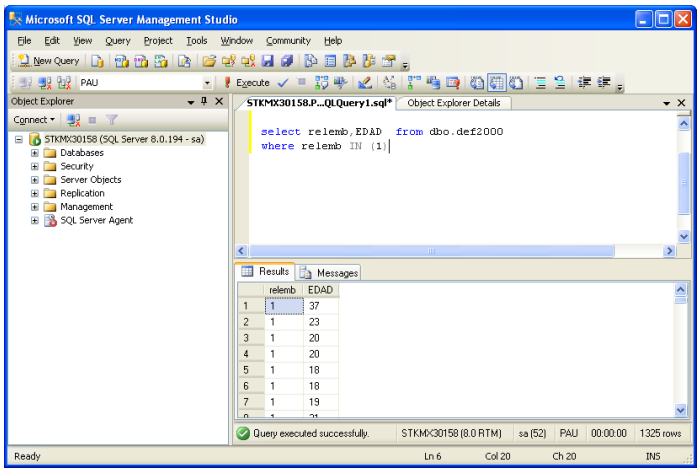


Figura 4.95 Consulta

Las causas que se van a utilizar en esta vista son todas aquellas que empiecen en 'O' ya que en nuestro catalogo de causas todas las que tienen 'O' al comienzo son causas relacionadas con el embarazo. Las causas completas relacionadas con el embarazo se encuentran en nuestro Anexo (Causas).
Por ejemplo:

O009		Embarazo ectópico, no especificado
O068		Aborto no especificado completo o no especificado, con otras complicaciones especificadas y las no especificadas

Tabla 4.23 Ejemplo causas relacionadas con el embarazo

En este caso serán 2 consultas y por lo tanto dos vistas minables una para menor o igual a 18 años y otra para mayores o iguales a 35 años.

```
Select edad as EDAD, count(edad)as NoMujeres from dbo.def2000
WHERE CAUSA LIKE 'O%' AND EDAD <> 998 and
edad <= 18
Group by edad
ORDER BY count(edad) DESC

Select edad as EDAD, count(edad)as NoMujeres from dbo.def2000
WHERE CAUSA LIKE 'O%' AND EDAD <> 998 and
edad > 38
Group by edad
ORDER BY count(edad) DESC
```

Se hace la unión de todos los años (1998 – 2008) para formar nuestra única vista minable, que usaremos como nuestros datos de entrenamiento, dejando al 2009 como un año de prueba. Ya que recordemos esta es una vista predictiva

4.5.1 Regresión Lineal

Antes de comenzar debemos tener bien claro que es lo que vamos a obtener de la regresión. Obtendremos una ecuación que nos ayudara a saber de manera aproximada el de número de mujeres fallecidas que habrá en años posteriores.

La ecuación que utilizamos es la siguiente:

$$\mathbf{y = a + b \cdot x}$$

Donde "y" sería la variable dependiente (No. de mujeres), es decir, aquella que viene definida a partir de la variable independiente "x" (año). Para definir la recta hay que determinar los valores de los parámetros "a" y "b":

El parámetro "a" es el valor que toma la variable dependiente "y", cuando la variable independiente "x" vale 0, y es el punto donde la recta cruza el eje vertical.

El parámetro "b" determina la pendiente de la recta, su grado de inclinación.

Realizaremos la minería de datos con WEKA, lo primero que veremos es una regresión tomando solo encuenta el año y el número de mujere que fallecieron en cada uno de estos años (del 1998 al 2008), esto se hizo para tener una visualización más general de lo que sucedió en cada año y de cómo seguira la tendecia es decir, ira decrementando o incrementando el indice de mujeres. Al usar la herramienta WEKA y ver su visualización el resultado que nos arrojo, lo vemos en la siguiente imagen, en la cual vemos que en los años de 1998 y 1999 existio un número muy signficativo de mujeres fallecidas, para los años sigientes la cifra disminuyo y la gráfica se ha mantenido estable hasta el 2008, con algunos aumentos y decremetos minimos, pero en general con una tendencia de decremento.

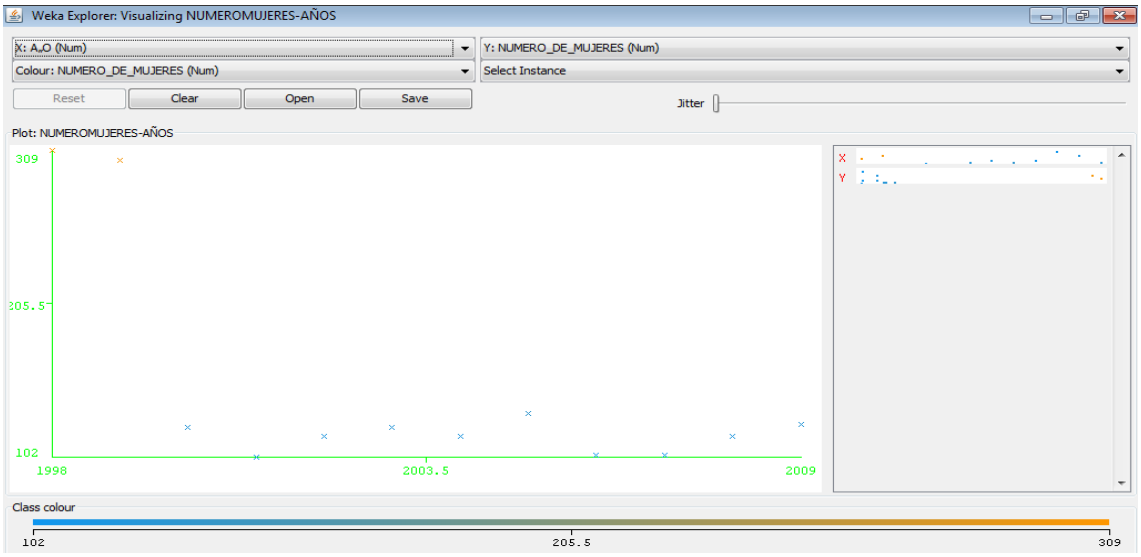


Figura 4.96 Visualizing WEKA

Una vez que se realizó el esquema general pasamos a un esquema mucho más particular (hablando de edades), primero que nada vamos a mencionar que esta vista se dividio en dos grupos el primero tomando en cuenta, mujeres con edades menores a 19 años y el segundo grupo de mujeres mayores a 38 años.

Analizaremos primero los resultados del primer grupo, empezando con el resultado de su visulización en WEKA, la cual nos muestra el siguiente resultado, observando la imagen en donde tenemos un resultado general de los 10 a los 18 años, y en el cual podemos ver que el mayor numero de incidencias se presentan de los 16 a los 18 años.

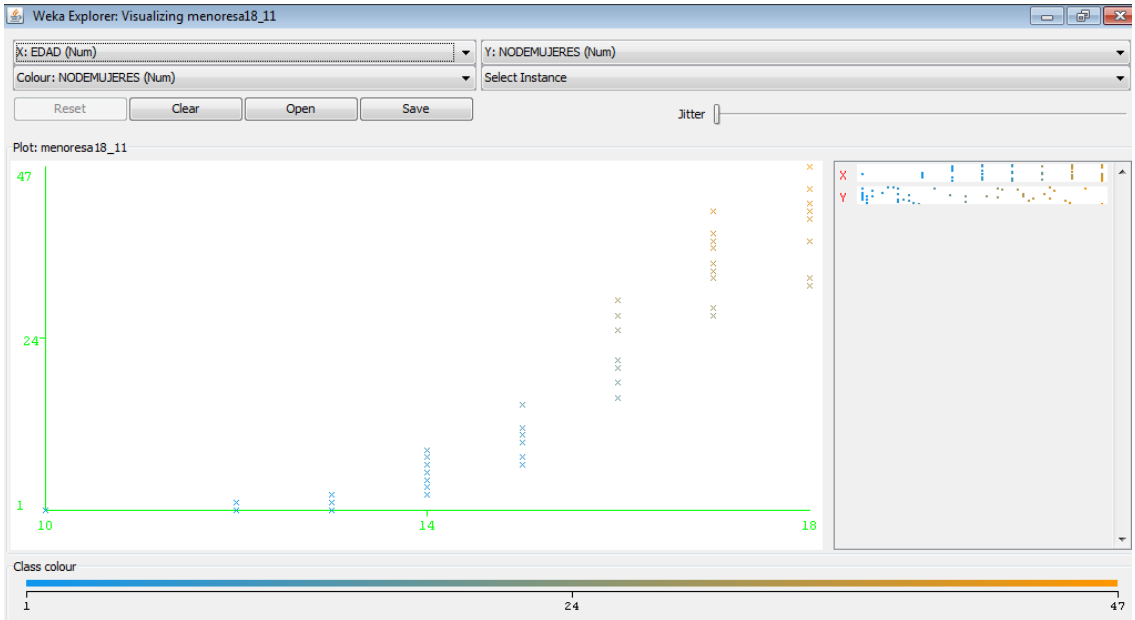


Figura 4.97 Visualizing WEKA

En una seguda imagen, pero ahora tomando solo las edades de lo 10 a 14 años, y observamos que existen más incidensias en las edades de los 13 a los 14 años, aunque no son numeros alarmantes, ya que hablamos de 9 mujeres con 14 años y de los 10 a los 12 años las incidencias son practicamnete nulas por ejemplo solo una mujer con 10 años.

El proceso que utilizamos para llevar a cabo la regresión lineal fue el siguiente, nos vamos a ***Classify _> Chosse _> Fuctions_> SimpleLinearRegresion.***

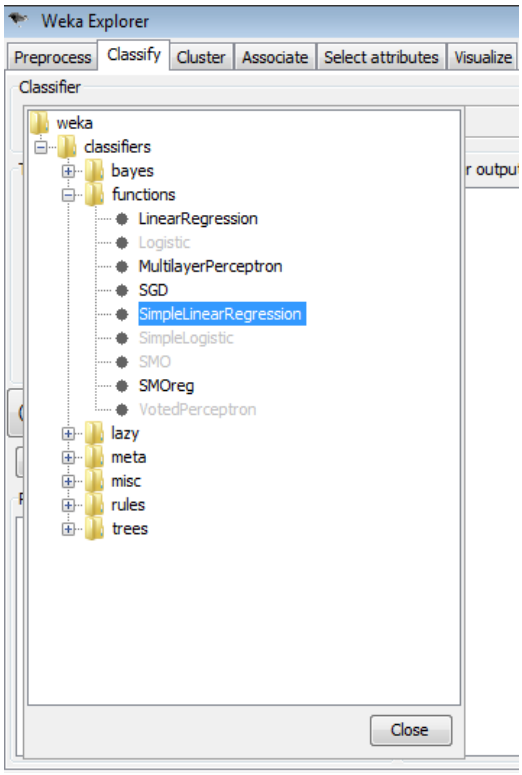


Figura 4.98 WEKA Explorer

La siguiente imagen nos muestra los resultados de lo obtenido después de aplicar la regresión lineal simple, se trabaja con 58 instancias, empezaremos por analizar el coeficiente de correlación, el cual nos permite determinar si, efectivamente, existe relación entre las dos variables (EDAD Y NO. DE MUJERES) en este caso, la correlación que existente entre estas dos variables es elevada, es de 0.9281 y como aparte es positivos indican una relación lineal directa o lo que es lo mismo la recta es creciente. Una vez que se concluye que sí existe relación, la regresión nos permite definir la recta que mejor se ajusta a esta nube de puntos. La ecuación que utilizamos es la siguiente:

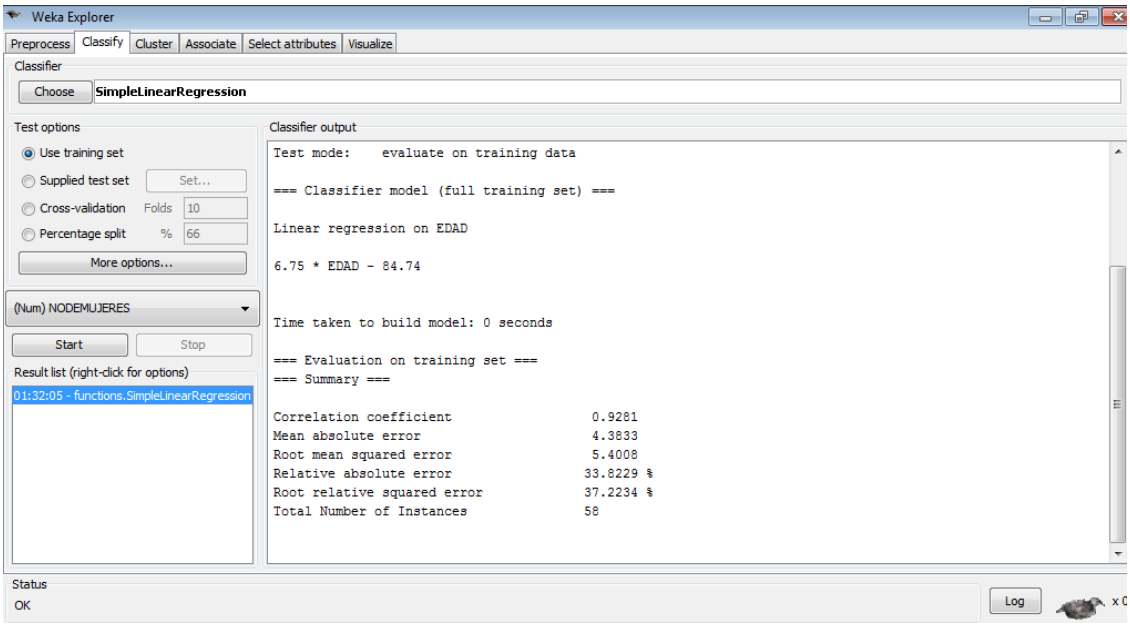


Figura 4.99 Classifier output

NO DE MUJERES= 6.75 * EDAD – 84.74

Donde la variable EDAD esta oscilando entre los 10 y 18 años.

Los valores obtenidos fueron:

- El punto donde la recta cruza el eje vertical (“a”) es (0,-82.74), es bueno aclara que esto viene siendo algo más numérico.
- La pendiente (“b”) de la recta es de 6.75 y como es positiva esto quiere decir que ambas variables crecen al mismo tiempo. Es decir entre menos edad tengan las mujeres embarazadas, existe menor número de incidencias.

Para comprobar que nuestro modelo (ecuación) es exacto lo que debemos hacer es validarlo con nuestro año de prueba (2009), lo que se hará es que con ayuda del SQL Server sacamos la cantidad de mujeres que fallecieron en el año 2009 con una edad determinada en nuestro ejemplo usaremos 2 edades (15 y 17 años).

- Con 15 años

Resultado obtenido con SQL Server = 14

Resultado con la ecuación aproximadamente de 16.51

Como podemos ver existe una margen de error, como lo habíamos mencionado antes los resultados que arroja la regresión siempre son aproximados, nunca exactos.

- Con 17 años

- Resultado obtenido con SQL Server = 33
- Resultado con la ecuación aproximadamente de 30.01

Para sustentarlo podemos aplicar el algoritmo M5P, seleccionado en WEKA como ***trees->m5->M5P,*** que lleva a cabo una regresión por tramos, con cada tramo determinado a partir de un árbol de regresión.

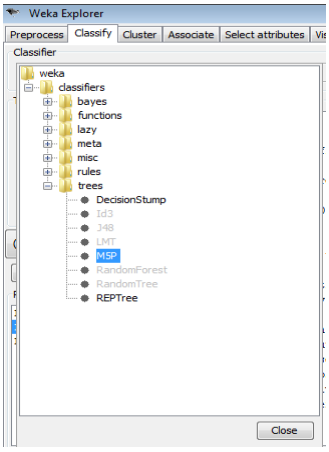


Figura 4.100 Classify M5P

Llegamos al siguiente resultado, que se muestra en la figura, observamos que nos muestra tres rectas cada una representada con su respectiva ecuación las ecuaciones son las siguientes:

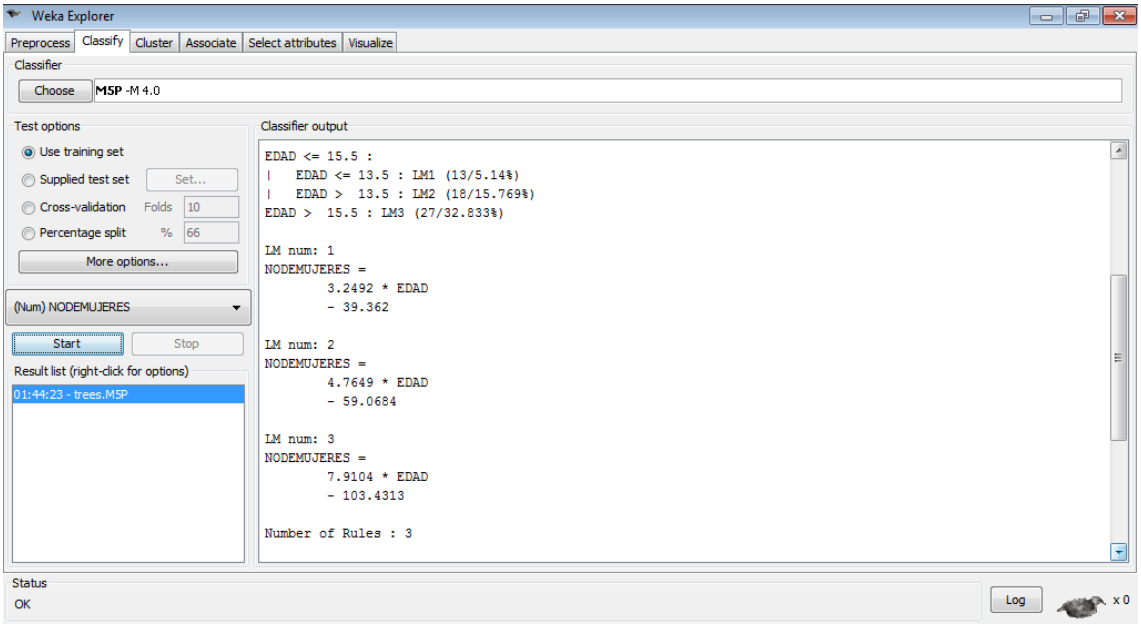


Figura 4.101 Classifier output

LM num: 1
NODEMUJERES = 3.2492 * EDAD - 39.362

Tomando un rango de edad de menores o iguales a 13.5 años

Punto donde la recta cruza el eje vertical (0, -39.362)

Pendiente es igual a 3.2492

LM num: 2
NODEMUJERES = 4.7649 * EDAD - 59.0684

Tomando un rango de edad de mayores de 13.5 años

Punto donde la recta cruza el eje (0, -59.0684)

Pendiente es igual a 4.7649

LM num: 3
NODEMUJERES = 7.9104 * EDAD – 103.4313

Tomando un rango de edad de mayores a 15.5 años

Punto donde la recta cruza el eje (0, -103.4313)

Pendiente es igual a 7.9104

Como nos podemos dar cuenta las tres pendientes son positivas esto quiere decir que las rectas son crecientes y a mayor edad la pendiente crece esto quiere decir que entre más edad mayor es el numero de incidencias, es decir el número de mujeres que fallecen por complicación en el embarazo es mayor.

Aquí nos dan la opción de poder visualizar un árbol de regresión, el cual se construye con cada tramo, es decir como se vio por cada regla se obtuvo una ecuación que nos mostraba una recta, esto es para que los resultados sean más exactos.

A continuación vemos el que resulta de usar el MP5, el cual nos dice que si son menores a 13.5 años existen 13 casos que en términos de porcentaje viene siendo un 5.14 %, si son mayores a 13.5 años hay 18 casos y su porcentaje es de 15.76 % y si son mayores a 15 .5 años hablamos de 27 casos con por lo cual tiene el porcentaje más alto y este es de un 32.83% esto nos muestra que si hay más incidencias en las mujeres que tienen mayor edad.

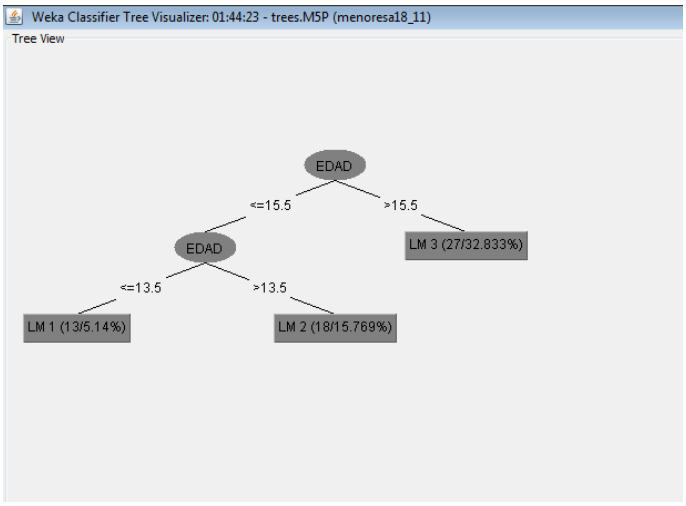


Figura 4.102 Árbol M5P

Con este método también realizamos los ejemplos de las edades anteriores 15 y 17 años los resultados fueron los siguientes:

- Con 15 años

Resultado obtenido con SQL Server = 14

Resultado con la ecuación aproximadamente de 12

Como podemos ver existe método un margen de error, como lo habíamos mencionado antes los resultados que arroja la regresión siempre son aproximados, nunca exactos.

- Con 17 años
- Resultado obtenido con SQL Server = 33
- Resultado con la ecuación aproximadamente de 31

Ambos métodos son buenos solo que se aproxima un poco más el método de M5P, esto era de suponerse ya que como va por rango es más lógico que se aproxime más.

Empezaremos a estudiar y analizar el segundo grupo de mujeres, como en el otro aquí también vamos a ver primero los resultados de la visualización que nos muestra WEKA. En la siguiente imagen observamos que el resultado es inverso al del primer grupo ya que entre menos edad tengan las mujeres embarazadas (hablando de un rango de 38 a 56 años), mayor es el numero de incidencias. El rango más crítico que nos muestra la imagen es de los 38 a los 45 años.

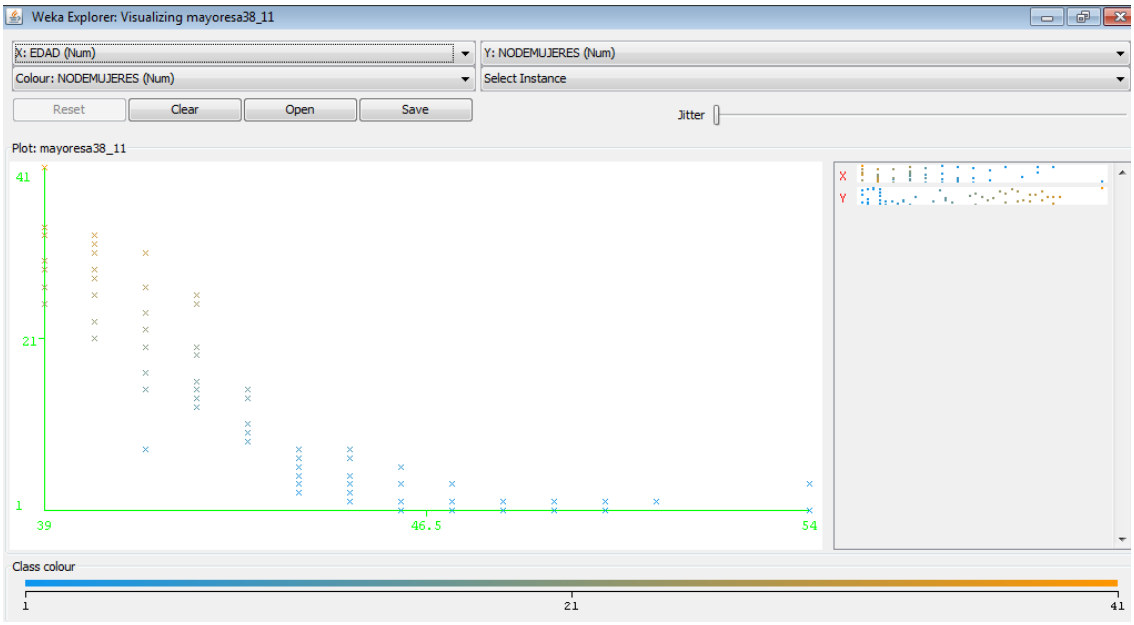


Figura 4.103 Visualizing WEKA

En este caso nuestra recta resultante con WEKA fue la siguiente:

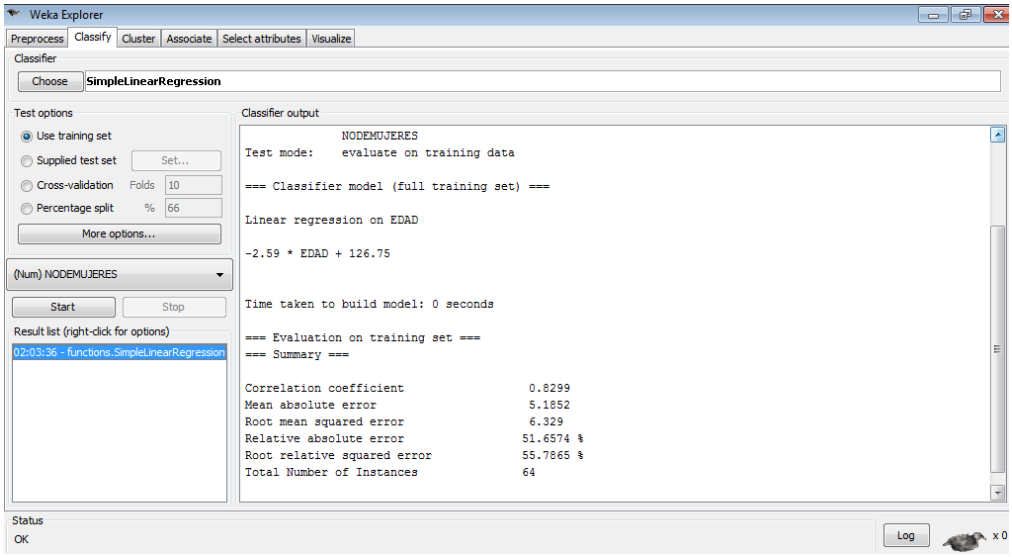


Figura 4.104 Simple Linear Regression WEKA

NO DE MUJERES= -2.59*EDAD + 126.75

Donde la variable EDAD esta oscilando entre los 38 a 54 años.

Los valores obtenidos fueron:

- El punto donde la recta cruza el eje vertical (“a”) es (0,126.75), es decir su punto de intersección.
- La pendiente (“b”) de la recta es de -2.59 y como es negativa quiere decir que cuando una variable crece la otra va a decrecer. Es decir entre menos edad tengan las mujeres embarazadas, existe mayor número de incidencias.

En este caso también usaremos un ejemplo para validar nuestra ecuación, tomaremos para esto las edades de 40 y 43 años, los resultados obtenidos son los siguientes:

- Con 40 años

Resultado obtenido con SQL Server = 14

Resultado con la ecuación aproximadamente de 23

Como podemos ver existe método un margen de error, como lo habíamos mencionado antes los resultados que arroja la regresión siempre son aproximados, nunca exactos, aunque también es bueno mencionar que en este caso con una edad de 40 años el margen si fue un poco más grande.

- Con 43 años

  Resultado obtenido con SQL Server = 11

  Resultado con la ecuación aproximadamente de 15

Usando el M5P obtenemos lo que se muestra en la siguiente imagen

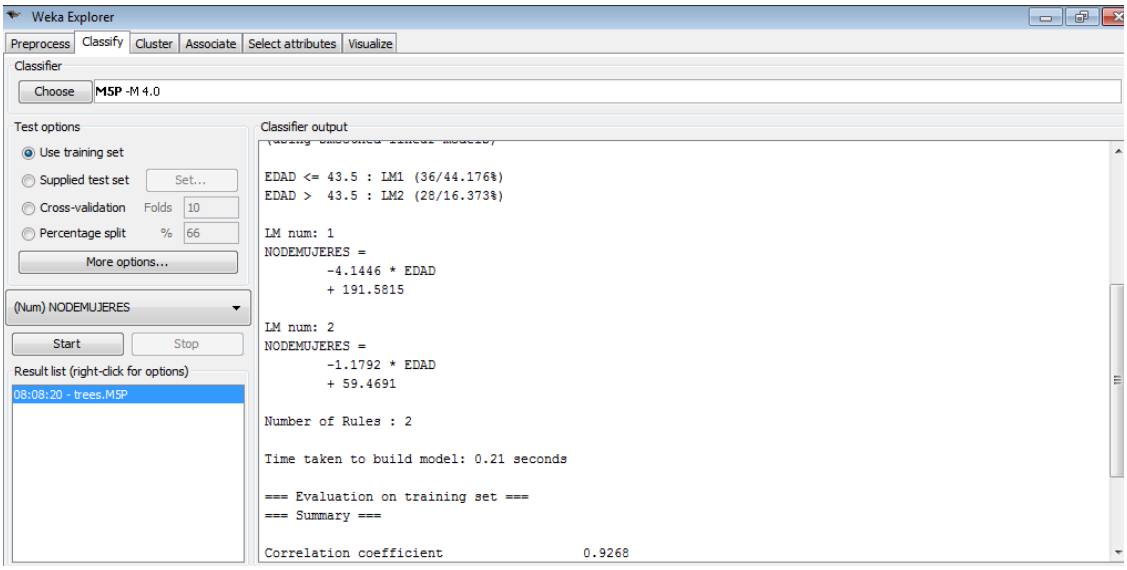


Figura 4.105 Resultado M5P

Su árbol quedo de la siguiente manera

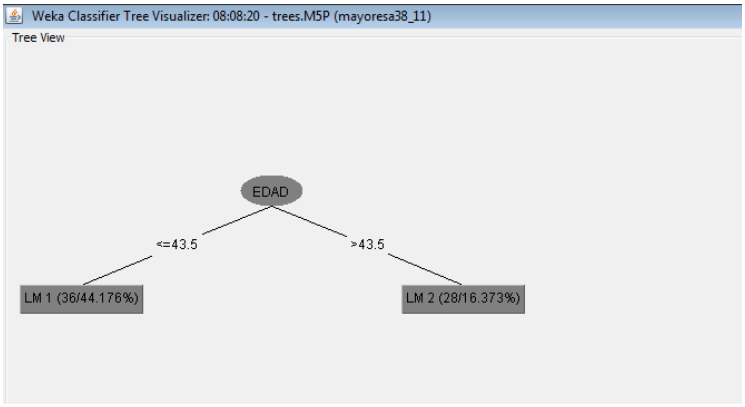


Figura 4.106 Árbol M5P

Para validar nuestra ecuación usaremos solo la primera ecuación y validaremos con las mismas edades, quedando de la siguiente manera:

- Con 40 años

Resultado obtenido con SQL Server = 14

Resultado con la ecuación aproximadamente de 25

Como podemos ver existe método un margen de error, como lo habíamos mencionado antes los resultados que arroja la regresión siempre son aproximados, nunca exactos, aunque también es bueno mencionar que en este caso con una edad de 40 años el margen si fue un poco más grande.

- Con 43 años

  Resultado obtenido con SQL Server = 11

  Resultado con la ecuación aproximadamente de 13

Conclusión "Mujeres Embarazadas"

Una vez que hemos pasado por todas las etapas llegamos a uno de los puntos más importantes ya que es con lo que se concluye la minería de datos, hablamos de el paso de difusión de conocimiento obtenido, este paso es de mucho cuidado ya que debemos de saber exactamente primero asea a que sector de la población se va a dirigir el conocimiento una vez que sabemos esto entonces debemos de buscar la manera de cómo difundirlo para qué así se pueda corregir o mejorar el problema existente.

En este caso vamos a estudiar la vista de mujeres embarazadas, para esto primero retomaremos cual es el problema de esta vista. Esta vista como sabemos habla del riegos para las mujeres embarazadas. Los subgrupos de población que se deben vigilar

por el riego que implica un embarazo son las adolecentes (menores de 18 años) y las añosas (mayores de 35 años). En el primer caso se supone que existe una competencia entre el feto y la madre (que aun está creciendo) por la obtención de nutrimentos. En el caso de las madres añosas es probable que exista un agotamiento de las reservas.

Como vemos dos son los grupos que corren más el riesgo o mujeres muy grandes o muy chicas de edad, con esto ya sabemos ase a que sector se tiene que dirigir dicho conocimiento obtenido, es decir hacia el sector salud, lo que prosigue es como difundiremos este conocimiento para así evitar en una totalidad ó que exista una disminución en el índice de defunciones de mujeres a causa de su edad (adolecentes o añosas), este conocimiento pasaría a ser parte del sector salud una vez que ellos lo hayan estudiado y analizado muy seguramente existirán muchas opciones para la mejora y corrección de este problema. Nosotros tomaremos un poco el papel de sector salud, mostrando algunas posibles soluciones:

- Como ya sabemos el concientizar a la gente es algo de lo que primero se debe de hacer, entonces una de nuestras soluciones es difundir por medio de las escuelas, centros de salud y campañas informativos, los riesgos que corre la madre ó el producto al concebirlo a una edad joven o avanzada.
- Se ha visto que la población no está hecha para seguir o escuchar los consejos y recomendaciones, por tal motivo vienen las soluciones alternativas es decir ya cuando la mujer esta embaraza, en este caso la solución sería poner en cada una de las clínicas, hospitales, etc. que existan en toda la republica mexicana un especialista que pueda atender embarazos de alto riesgo ó si es el caso especialistas en otras aéreas que sean requeridos y que sobre todo resguarde la vida de la madre. Y que cada especialista siga durante todo el proceso a las mujeres con embarazos de alto riego, es decir del proceso de concepción hasta el término de embarazo.

Para que nuestro modelo sea plenamente confiable tenemos que probarlo con el año 2009 que es nuestro año de prueba, para realizar esta prueba antes debemos de saber con ayuda del SQL Server cual fue la cantidad de mujeres fallecidas en 2009, para posteriormente comparar este resultado con lo que se obtendrá de la ecuación que no arrojo WEKA.

CONCLUSIONES

Después de haber realizado nuestra minería de datos podemos plasmar en este apartado las conclusiones a las que llegamos, por un lado podemos decir que el proceso KDD, es un ciclo muy completo que permite explorar los datos contenidos en repositorios de información, pero es necesario tener un contexto de la información, un análisis del negocio y sus necesidades, ya que con este análisis es como se decidirán las tareas de minería de datos que darán respuesta a los problemas.

La motivación de esta Tesis fue el aumento de volumen de bases de datos digitales que ha crecido en los últimos años, siendo este un almacén "memoria de las organizaciones", en nuestro caso la secretaria de salud y el INEGI conjuntamente, sabiendo que la mayoría de las decisiones de una institución se basan en información de experiencias pasadas, la minería de datos es una herramienta perfecta para la toma de decisiones.

Para la parte práctica de este trabajo, se analizaron las herramientas utilizadas WEKA y Rapid Miner, llegando a las ventajas o desventajas de cada herramienta:

WEKA tiene una parte muy amigable y útil que es el Visualizar los datos, antes de realizarles algún cambio o meterlos a alguna transformación, esta herramienta te permite gráficamente ver grupos o comportamientos en los datos. En cuento a una desventaja seria el problema de la memoria, siendo un poco limitada.
Rapid Miner es una herramienta con más posibilidades, como guardar los resultados en un archivo, o realizar un proyecto y guardarlo, sin necesidad de reconfigurarlo cada vez que se ejecute el contenido. Por lo que para las tareas descriptivas optaríamos por utilizar WEKA y para tareas predictivas utilizar Rapid Miner.

En cuanto al tema de nuestra minería es importante mencionar que nuestros sistemas de salud en México necesitan una renovación y atención significativa, en cuestión de atención a pacientes y mejores instalaciones.
Otro tema que nos sorprendió mucho fue el de las ocupaciones de las personas, teniendo a un sector de la población con más riesgo, las personas que se dedican al campo y los conductores de algún medio de transporte, se debe crear un seguro de vida para estas personas.

El tema de la predicción es una de las cosas más útiles en la minería de datos, el encontrar características de las personas para predecir si tienen riesgo de sufrir violencia familiar, esto se puede enfocar a trabajadoras sociales que pueden dar la atención necesaria a este tipo de casos para evitar que lleguen a extremos fatales.
Y por último a pesar de que desde hace ya varios años existe la educación en cuestión de planificación familiar y educación sexual en las escuelas sigue habiendo casos de embarazos a temprana edad y a edades muy avanzadas, por lo que se propone concientizar más a las personas sobre este tema.

Teniendo la minería de datos las siguientes ventajas: por un lado resulta ser un punto de encuentro entre las personas de sistemas y las personas que conocen más del negocio,

en este caso la Secretaria de salud, por otro si es bien aprovechada ahorraría a las empresas grandes cantidades de dinero a las organizaciones, con esto se demuestra el gran potencial que tiene la minería de datos. En cuanto a la toma de decisiones se toman decisiones más racionales basadas en resultados de la minería de datos teniendo un grado de confianza mayor.

Conclusiones Personales y Profesionales
Paulina Galván Castro

Primero detallare mis conclusiones personales, que fue lo que me dejo personalmente el haber realizado este trabajo, la mayor enseñanza fue, el valor de la perseverancia, para poder lograr mi objetivo que es la titulación, tuve que ser constante y no darme por vencida, ya que este fue un trabajo de meses, trabajar hasta que fuera concluido, y pienso que fue una buena prueba para empezar mi camino en el ámbito laboral, otra enseñanza que me dejo fue aprender a ser tolerante y aceptar las opiniones de los demás, en este caso de mi compañera Alejandra, ya que yo no siempre tenía la razón y debimos encontrar un punto medio entre las dos. Ahora que se ha concluido este trabajo, he cambiado en cuento a que soy más profesional, y ejerceré mi profesión con más dedicación y empeño, ya que este trabajo me enseño a esforzarme hasta poder lograr mis objetivos, a que uno como profesionista debe luchar, que nunca nadie nos regalara nada, hay que trabajar duro para lograrlo.

En cuanto al aprendizaje académico, me di cuenta que los trabajos, tareas y exámenes que realice durante toda mi carrera, son una base de lo que se puede presentar en el trabajo, pero en este caso la volumetría de las bases de datos con las que se trabajo son grandes y fue difícil el manejo de la información, por la cantidad de registros que se manejaron, ya que son datos de todo México, en un periodo de tiempo de 10 años, ya que uno como estudiante no tiene la dimensión de datos reales, hasta que manejas grandes cantidades de datos, como fue en este caso.

Conclusiones Personales y Profesionales
Alejandra Meza Mendoza

Es trabajo me ayudo a saber realmente que es un verdadero trabajo en equipo tanto en el ámbito personal como profesional ya que forma en mi un sentido de responsabilidad, esto porque el trabajo individual se verá reflejado en el trabajo en grupo, otro de las cosas que se refuerzan en un trabajo en equipo es la tolerancia y la convivencia ya que si estas dos se cumplen perfectamente el ambiente de trabajo será sano y esto dará mejores resultados al equipo.

Después de haber realizado este trabajo, me considero más profesional ya que cuando trabajas con datos reales debes tener presente que la información obtenida al final debe tener un marguen de error mínimo ó nulo ya que es información que no se quedara solo para nosotros, reforzó muchos cosas en mi como la del punto anterior que fue de trabajar en equipo, ser responsable saber paciente y tolerante, organizada creo que fue una experiencia buena en la que puedo decir que me formo un poco más como profesional.

La tesis es un trabajo que manejo grandes volúmenes de datos por lo cual me enseño la importancia de saber trabajar con grandes volúmenes de información y sobre todo de la gran necesidad de que una base de datos este limpia para así estar segura que la información que resulte de sea más fidedigna este trabajo de tesis me ayudo a ver la gran utilidad que tiene la minería de datos y de que si la sabes utilizar bien puede ser de gran ayuda en la empresas y en muchos otros lugares otra cosa que me dejo es el ser capaz de analizar y entender los resultados obtenidos para saber cómo presentarla a los expertos del tema.

REFERENCIAS

Bibliográficas

Hernández, O., Ramirez, M. y Ferri, C. (2000).introducción a la Minería de Datos. Madrid: Prentice Hall

Pérez, C. (2006). Data Mining Soluciones con Enterprise Miner". Madrid: Ra-Ma Primera Edición

Estrada Roció (comunicación personal).[Entrevista con Roció Estrada Doctora del Hospital General "Manuel Gea González" del servicio de patología]

Electrónicas

Waikato. (2011). WEKA. Extraído en Febrero del 2011 de

http://www.cs.waikato.ac.nz/ml/WEKA/

View. (2005).rapid-in. Extraído en Marzo del 2011 de

http://rapid-i.com/content/view/181/190/

Jorallo. (2009). WEKA. Extraído en Diciembre del 2011 de

http://users.dsic.upv.es/~jorallo/docent/doctorat/WEKA.pdf

INEGI. (1988). Metadatos. Extraído en Mayo del 2011 de

http://www.inegi.org.mx/est/contenidos/espanol/proyectos/metadatos/continuas/em_313.asp?s=est&c=11138 - top

El porvenir. (1986). Notas. Extraído en el 2011 de

http://elporvenir.com.mx/notas.asp?nota_id=267466v

IMG. (1986). Guía de Abordaje a Violencia Domestica en Servicios de Salud. Extraído en el 2011 de

http://www.mysu.org.uy/IMG/pdf/guia_de_abordaje_a_violencia_domestica_en_servicios_de_salud.pdf

Medicacenter. (1986). Factsheets. Extraído en el 2011 de

http://www.who.int/mediacentre/factsheets/fs239/es/index.html

ANEXO

SCRIPS DE VISTAS MINABLES

Consulta para la vista DERECHOHABIENTES

4.2

```
SELECT
( CASE
       WHEN    DERECH= 1      THEN 'NINGUNA'
       WHEN    DERECH= 2      THEN    'IMSS'
       WHEN    DERECH= 3      THEN    'ISSSTE'
       WHEN    DERECH= 4      THEN    'PEMEX'
       when    DERECH= 5      THEN    'FUERZAS ARMADAS'
               END) AS DERECHOHABIENCIA,
( CASE  WHEN    ocupa=  02     THEN    'No trabaja'
       WHEN    ocupa=  11     THEN    'Profesionistas'
       WHEN    ocupa=  12     THEN    'Técnicos'
       WHEN    ocupa=  13     THEN    'Trabajadores de la educación'
       WHEN    ocupa=  14     THEN    'Trabajadores del arte, espectáculos y deportes'
       WHEN    ocupa=  21     THEN    'Funcionarios y directivos'
       WHEN    ocupa=  41     THEN    'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
WHEN    ocupa=  51     THEN    'Personal de control producción industrial'
WHEN    ocupa=  52     THEN    'Trabajadores en la industria de la transformación'
       WHEN    ocupa=  53     THEN    'Operadores de maquinaria fija'
       WHEN    ocupa=  54     THEN    'Ayudantes producción industrial y artesanal'
WHEN    ocupa=  55     THEN    'Conductores de maquinaria móvil y medios de transporte'
WHEN    ocupa=  61     THEN    'Trabajadores administrativos de nivel intermedio'
WHEN    ocupa=  62     THEN    'Trabajadores administrativos de nivel   inferior'
WHEN    ocupa=  71     THEN    'Comerciantes, empleados de comercio, agentes de ventas'
       WHEN    ocupa=  72     THEN    'Vendedores ambulantes'
       WHEN    ocupa=  81     THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupa=  82     THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupa=  83     THEN    'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN    ocupa=  98     THEN    'No aplica'
       WHEN    ocupa=  99     THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupa=  00     THEN    'No especificado'     END) AS OCUPACION,
       ( CASE  WHEN    ASISMED=       1      THEN    'SI'
       WHEN    ASISMED=       2      THEN    'NO'
               END) AS ASISTENCIA,
( CASE  WHEN    EDOCIV= 1      THEN    'SOLTERO'
               WHEN    EDOCIV= 2      THEN    'CASADO'
               WHEN    EDOCIV= 3      THEN    'UNION LIBRE'
               WHEN    EDOCIV= 4      THEN    'SEPARADO'
               WHEN    EDOCIV= 5      THEN    'DIVORCIADO'
               WHEN    EDOCIV= 6      THEN    'VIUDO'
               END) AS ESTADOCIVIL,
(CASE WHEN ENTRES=01 THEN 'Aguascalientes'
WHEN ENTRES=02  THEN 'Baja California'
WHEN ENTRES=03 THEN 'Baja California Sur'
WHEN ENTRES=04 THEN 'Campeche'
WHEN ENTRES=05 THEN 'Coahuila de Zaragoza'
WHEN ENTRES=06 THEN 'Colima'
WHEN ENTRES=07 THEN 'Chiapas'
WHEN ENTRES=08 THEN 'Chihuahua'
WHEN ENTRES=09 THEN 'Distrito Federal'
WHEN ENTRES=10 THEN 'Durango'
WHEN ENTRES=11 THEN 'Guanajuato'
WHEN ENTRES=12 THEN 'Guerrero'
WHEN ENTRES=13 THEN 'Hidalgo'
WHEN ENTRES=14 THEN 'Jalisco'
WHEN ENTRES=15 THEN 'Edo Mex'
WHEN ENTRES=16 THEN 'Michoacan de Ocampo'
WHEN ENTRES=17 THEN 'Morelos'
WHEN ENTRES=18 THEN 'Nayarit'
```

```
WHEN ENTRES=19 THEN 'Nuevo Leon'
WHEN ENTRES=20 THEN 'Oaxaca'
WHEN ENTRES=21 THEN 'Puebla'
WHEN ENTRES=22 THEN 'Queretaro Arteaga'
WHEN ENTRES=23 THEN 'Quintana Roo'
WHEN ENTRES=24 THEN 'San Luis Potosi'
WHEN ENTRES=25 THEN 'Sinaloa'
WHEN ENTRES=26 THEN 'Sonora'
WHEN ENTRES=27 THEN 'Tabasco'
WHEN ENTRES=28 THEN 'Tamaulipas'
WHEN ENTRES=29 THEN 'Tlaxcala'
WHEN ENTRES=30 THEN 'Veracruz de Ignacio de la Llave'
WHEN ENTRES=31 THEN 'Yucatan'
WHEN ENTRES=32 THEN 'Zacatecas'
 END) AS ENTIDAD
FROM dbo.DEF2000
WHERE (ASISMED='1' OR ASISMED='2') AND
DERECH IN (1,2,3,4,5) and edad between 0 and 120
AND EDOCIV IN (1,2,3,4,5,6)

UNION all

SELECT
( CASE
       WHEN   DERECH= 1      THEN 'NINGUNA'
       WHEN   DERECH= 2      THEN   'IMSS'
       WHEN   DERECH= 3      THEN   'ISSSTE'
       WHEN   DERECH= 4      THEN   'PEMEX'
       when   DERECH= 5      THEN   'FUERZAS ARMADAS'
              END) AS DERECHOHABIENCIA,
( CASE WHEN   ocupa= 02     THEN   'No trabaja'
       WHEN   ocupa= 11     THEN   'Profesionistas'
       WHEN   ocupa= 12     THEN   'Técnicos'
       WHEN   ocupa= 13     THEN   'Trabajadores de la educación'
       WHEN   ocupa= 14     THEN   'Trabajadores del arte, espectáculos y deportes'
       WHEN   ocupa= 21     THEN   'Funcionarios y directivos'
       WHEN   ocupa= 41     THEN   'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
WHEN   ocupa= 51     THEN   'Personal de control producción industrial'
WHEN   ocupa= 52     THEN   'Trabajadores en la industria de la transformación'
       WHEN   ocupa= 53     THEN   'Operadores de maquinaria fija'
       WHEN   ocupa= 54     THEN   'Ayudantes producción industrial y artesanal'
WHEN   ocupa= 55     THEN   'Conductores de maquinaria móvil y medios de transporte'
WHEN   ocupa= 61     THEN   'Trabajadores administrativos de nivel intermedio'
WHEN   ocupa= 62     THEN   'Trabajadores administrativos de nivel   inferior'
WHEN   ocupa= 71     THEN   'Comerciantes, empleados de comercio, agentes de ventas'
       WHEN   ocupa= 72     THEN   'Vendedores ambulantes'
       WHEN   ocupa= 81     THEN   'Trabajadores servicios personales en
establecimientos'
       WHEN   ocupa= 82     THEN   'Trabajadores de servicio doméstico'
       WHEN   ocupa= 83     THEN   'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN   ocupa= 98     THEN   'No aplica'
       WHEN   ocupa= 99     THEN   'Ocupación no clasificada anteriormente'
       WHEN   ocupa= 00     THEN   'No especificado'     END) AS OCUPACION,
       ( CASE WHEN   ASISMED=       1      THEN   'SI'
       WHEN   ASISMED=       2      THEN   'NO'
              END) AS ASISTENCIA,
( CASE WHEN   EDOCIV= 1      THEN   'SOLTERO'
              WHEN   EDOCIV= 2      THEN   'CASADO'
              WHEN   EDOCIV= 3      THEN   'UNION LIBRE'
              WHEN   EDOCIV= 4      THEN   'SEPARADO'
              WHEN   EDOCIV= 5      THEN   'DIVORCIADO'
              WHEN   EDOCIV= 6      THEN   'VIUDO'
              END) AS ESTADOCIVIL,
(CASE WHEN ENTRES=01 THEN 'Aguascalientes'
WHEN ENTRES=02  THEN 'Baja California'
WHEN ENTRES=03 THEN 'Baja California Sur'
WHEN ENTRES=04 THEN 'Campeche'
WHEN ENTRES=05 THEN 'Coahuila de Zaragoza'
WHEN ENTRES=06 THEN 'Colima'
WHEN ENTRES=07 THEN 'Chiapas'
WHEN ENTRES=08 THEN 'Chihuahua'
WHEN ENTRES=09 THEN 'Distrito Federal'
WHEN ENTRES=10 THEN 'Durango'
```

```
WHEN ENTRES=11 THEN 'Guanajuato'
WHEN ENTRES=12 THEN 'Guerrero'
WHEN ENTRES=13 THEN 'Hidalgo'
WHEN ENTRES=14 THEN 'Jalisco'
WHEN ENTRES=15 THEN 'Edo Mex'
WHEN ENTRES=16 THEN 'Michoacan de Ocampo'
WHEN ENTRES=17 THEN 'Morelos'
WHEN ENTRES=18 THEN 'Nayarit'
WHEN ENTRES=19 THEN 'Nuevo Leon'
WHEN ENTRES=20 THEN 'Oaxaca'
WHEN ENTRES=21 THEN 'Puebla'
WHEN ENTRES=22 THEN 'Queretaro Arteaga'
WHEN ENTRES=23 THEN 'Quintana Roo'
WHEN ENTRES=24 THEN 'San Luis Potosi'
WHEN ENTRES=25 THEN 'Sinaloa'
WHEN ENTRES=26 THEN 'Sonora'
WHEN ENTRES=27 THEN 'Tabasco'
WHEN ENTRES=28 THEN 'Tamaulipas'
WHEN ENTRES=29 THEN 'Tlaxcala'
WHEN ENTRES=30 THEN 'Veracruz de Ignacio de la Llave'
WHEN ENTRES=31 THEN 'Yucatan'
WHEN ENTRES=32 THEN 'Zacatecas'
 END) AS ENTIDAD
FROM dbo.DEF2001
WHERE (ASISMED='1' OR ASISMED='2') AND
DERECH IN (1,2,3,4,5) and edad between 0 and 120
AND EDOCIV IN (1,2,3,4,5,6)
```

Consulta para conteo vista de DERECHOHABIENTES

```
select count(*) from dbo.DEF2000 where asismed='2'
and DERECH='1'
select count(*) from dbo.DEF2001 where asismed='2'
and DERECH='1'
select count(*) from dbo.DEF02 where asismed='2'
and DERECH='1'
select count(*) from dbo.DEF03 where asismed='2'
and DERECH='1'
select count(*) from dbo.DEF04 where asismed='2'
and DERECH='1'
select count(*) from dbo.DEF05 where asismed='2'
and DERECH='1'
select count(*) from dbo.DEF06 where asist='2'
and DERHAB='1'
select count(*) from dbo.DEF07 where asist='2'
and DERHAB='1'
select count(*) from dbo.DEF08 where asist='2'
and DERHAB='1'
```

Consulta para la vista OCUPACION PELIGROSA 4.3

```
/*2000*/

select CAUSA,EDAD,
( CASE  WHEN  ocupa= 02     THEN   'No trabaja'
       WHEN   ocupa= 11     THEN   'Profesionistas'
       WHEN   ocupa= 12     THEN   'Técnicos'
       WHEN   ocupa= 13     THEN   'Trabajadores de la educación'
       WHEN   ocupa= 14     THEN   'Trabajadores del arte, espectáculos y deportes'
       WHEN   ocupa= 21     THEN   'Funcionarios y directivos'
       WHEN   ocupa= 41     THEN   'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
       WHEN   ocupa= 51     THEN   'Personal de control producción industrial'
       WHEN   ocupa= 52     THEN   'Trabajadores en la industria de la transformación'
       WHEN   ocupa= 53     THEN   'Operadores de maquinaria fija'
       WHEN   ocupa= 54     THEN   'Ayudantes producción industrial y artesanal'
       WHEN   ocupa= 55     THEN   'Conductores de maquinaria móvil y medios de
transporte'
       WHEN   ocupa= 61     THEN   'Trabajadores administrativos de nivel intermedio'
       WHEN   ocupa= 62     THEN   'Trabajadores administrativos de nivel   inferior'
       WHEN   ocupa= 71     THEN   'Comerciantes, empleados de comercio, agentes de
ventas'
```

```
       WHEN   ocupa= 72     THEN   'Vendedores ambulantes'
       WHEN   ocupa= 81     THEN   'Trabajadores servicios personales en
establecimientos'
       WHEN   ocupa= 82     THEN   'Trabajadores de servicio doméstico'
       WHEN   ocupa= 83     THEN   'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN   ocupa= 98     THEN   'No aplica'
       WHEN   ocupa= 99     THEN   'Ocupación no clasificada anteriormente'
       WHEN   ocupa= 00     THEN   'No especificado'     END) AS OCUPACION,
       (CASE   WHEN  esco=  1      THEN   'Sin escolaridad'
                     WHEN  esco = 2      THEN   'Menos de tres años de primaria'
                     WHEN  esco = 3      THEN   'De tres a cinco años de primaria'
                     WHEN  esco = 4      THEN   'Primaria completa'
                     WHEN  esco = 5      THEN   'Secundaria o equivalente'
                     WHEN  esco = 6      THEN   'Preparatoria o equivalente'
                    WHEN   esco = 7      THEN   'Profesional'
              END) AS escolaridad
from def2000
where traba=1
and causa IN
('V496','V092','X598','X599','X958','V878','V899','X594','V892','V499','X954','V099')
and esco in(1,2,3,4,5,6,7)AND
EDAD BETWEEN '1' AND '130'

UNION ALL

/*2001*/

select CAUSA,EDAD,
( CASE  WHEN  ocupa= 02     THEN   'No trabaja'
       WHEN   ocupa= 11     THEN   'Profesionistas'
       WHEN   ocupa= 12     THEN   'Técnicos'
       WHEN   ocupa= 13     THEN   'Trabajadores de la educación'
       WHEN   ocupa= 14     THEN   'Trabajadores del arte, espectáculos y deportes'
       WHEN   ocupa= 21     THEN   'Funcionarios y directivos'
       WHEN   ocupa= 41     THEN   'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
       WHEN   ocupa= 51     THEN   'Personal de control producción industrial'
       WHEN   ocupa= 52     THEN   'Trabajadores en la industria de la transformación'
       WHEN   ocupa= 53     THEN   'Operadores de maquinaria fija'
       WHEN   ocupa= 54     THEN   'Ayudantes producción industrial y artesanal'
       WHEN   ocupa= 55     THEN   'Conductores de maquinaria móvil y medios de
transporte'
       WHEN   ocupa= 61     THEN   'Trabajadores administrativos de nivel intermedio'
       WHEN   ocupa= 62     THEN   'Trabajadores administrativos de nivel   inferior'
       WHEN   ocupa= 71     THEN   'Comerciantes, empleados de comercio, agentes de
ventas'
       WHEN   ocupa= 72     THEN   'Vendedores ambulantes'
       WHEN   ocupa= 81     THEN   'Trabajadores servicios personales en
establecimientos'
       WHEN   ocupa= 82     THEN   'Trabajadores de servicio doméstico'
       WHEN   ocupa= 83     THEN   'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN   ocupa= 98     THEN   'No aplica'
       WHEN   ocupa= 99     THEN   'Ocupación no clasificada anteriormente'
       WHEN   ocupa= 00     THEN   'No especificado'     END) AS OCUPACION,
       (CASE   WHEN  esco=  1      THEN   'Sin escolaridad'
                     WHEN  esco = 2      THEN   'Menos de tres años de primaria'
                     WHEN  esco = 3      THEN   'De tres a cinco años de primaria'
                     WHEN  esco = 4      THEN   'Primaria completa'
                     WHEN  esco = 5      THEN   'Secundaria o equivalente'
                     WHEN  esco = 6      THEN   'Preparatoria o equivalente'
                    WHEN   esco = 7      THEN   'Profesional'
              END) AS escolaridad
from def2001
where traba=1
and causa IN
('X954','V099','V499','V892','V899','X594','V878','X958','X599','X598','V093','X959','V0
92','W878','V496','Y094','W178','V489')
and esco in(1,2,3,4,5,6,7)AND
EDAD BETWEEN '1' AND '130'

/*2002*/

UNION ALL
```

```
select CAUSA,EDAD,
( CASE  WHEN   ocupa= 02      THEN    'No trabaja'
        WHEN    ocupa= 11     THEN    'Profesionistas'
        WHEN    ocupa= 12     THEN    'Técnicos'
        WHEN    ocupa= 13     THEN    'Trabajadores de la educación'
        WHEN    ocupa= 14     THEN    'Trabajadores del arte, espectáculos y deportes'
        WHEN    ocupa= 21     THEN    'Funcionarios y directivos'
        WHEN    ocupa= 41     THEN    'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
        WHEN    ocupa= 51     THEN    'Personal de control producción industrial'
        WHEN    ocupa= 52     THEN    'Trabajadores en la industria de la transformación'
        WHEN    ocupa= 53     THEN    'Operadores de maquinaria fija'
        WHEN    ocupa= 54     THEN    'Ayudantes producción industrial y artesanal'
        WHEN    ocupa= 55     THEN    'Conductores de maquinaria móvil y medios de
transporte'
        WHEN    ocupa= 61     THEN    'Trabajadores administrativos de nivel intermedio'
        WHEN    ocupa= 62     THEN    'Trabajadores administrativos de nivel   inferior'
        WHEN    ocupa= 71     THEN    'Comerciantes, empleados de comercio, agentes de
ventas'
        WHEN    ocupa= 72     THEN    'Vendedores ambulantes'
        WHEN    ocupa= 81     THEN    'Trabajadores servicios personales en
establecimientos'
        WHEN    ocupa= 82     THEN    'Trabajadores de servicio doméstico'
        WHEN    ocupa= 83     THEN    'Trabajadores de fuerzas armadas, protección y
vigilancia'
        WHEN    ocupa= 98     THEN    'No aplica'
        WHEN    ocupa= 99     THEN    'Ocupación no clasificada anteriormente'
        WHEN    ocupa= 00     THEN    'No especificado'     END) AS OCUPACION,
        (CASE    WHEN  esco=   1       THEN    'Sin escolaridad'
                        WHEN   esco =  2      THEN    'Menos de tres años de primaria'
                        WHEN   esco =  3      THEN    'De tres a cinco años de primaria'
                        WHEN   esco =  4      THEN    'Primaria completa'
                        WHEN   esco =  5      THEN    'Secundaria o equivalente'
                        WHEN   esco =  6      THEN    'Preparatoria o equivalente'
                       WHEN    esco =  7      THEN    'Profesional'
              END) AS escolaridad
from def02
where traba=1
and causa IN
('V099','X954','V892','V499','V899','V878','X594','X599','X958','V092','X959','X598')
and esco in(1,2,3,4,5,6,7)AND
EDAD BETWEEN '1' AND '130'

/*2003*/

UNION ALL

select CAUSA,EDAD,
( CASE  WHEN   ocupa= 02      THEN    'No trabaja'
        WHEN    ocupa= 11     THEN    'Profesionistas'
        WHEN    ocupa= 12     THEN    'Técnicos'
        WHEN    ocupa= 13     THEN    'Trabajadores de la educación'
        WHEN    ocupa= 14     THEN    'Trabajadores del arte, espectáculos y deportes'
        WHEN    ocupa= 21     THEN    'Funcionarios y directivos'
        WHEN    ocupa= 41     THEN    'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
        WHEN    ocupa= 51     THEN    'Personal de control producción industrial'
        WHEN    ocupa= 52     THEN    'Trabajadores en la industria de la transformación'
        WHEN    ocupa= 53     THEN    'Operadores de maquinaria fija'
        WHEN    ocupa= 54     THEN    'Ayudantes producción industrial y artesanal'
        WHEN    ocupa= 55     THEN    'Conductores de maquinaria móvil y medios de
transporte'
        WHEN    ocupa= 61     THEN    'Trabajadores administrativos de nivel intermedio'
        WHEN    ocupa= 62     THEN    'Trabajadores administrativos de nivel   inferior'
        WHEN    ocupa= 71     THEN    'Comerciantes, empleados de comercio, agentes de
ventas'
        WHEN    ocupa= 72     THEN    'Vendedores ambulantes'
        WHEN    ocupa= 81     THEN    'Trabajadores servicios personales en
establecimientos'
        WHEN    ocupa= 82     THEN    'Trabajadores de servicio doméstico'
        WHEN    ocupa= 83     THEN    'Trabajadores de fuerzas armadas, protección y
vigilancia'
        WHEN    ocupa= 98     THEN    'No aplica'
        WHEN    ocupa= 99     THEN    'Ocupación no clasificada anteriormente'
```

```
       WHEN   ocupa= 00     THEN   'No especificado'     END) AS OCUPACION,
       (CASE   WHEN  esco=  1      THEN   'Sin escolaridad'
                      WHEN  esco = 2      THEN   'Menos de tres años de primaria'
                      WHEN  esco = 3      THEN   'De tres a cinco años de primaria'
                      WHEN  esco = 4      THEN   'Primaria completa'
                      WHEN  esco = 5      THEN   'Secundaria o equivalente'
                      WHEN  esco = 6      THEN   'Preparatoria o equivalente'
                     WHEN   esco = 7      THEN   'Profesional'
              END) AS escolaridad
from def03
where traba=1
and causa IN
('V099','X954','V892','V499','V899','V878','X958','X599','X958','X594','V092')
and esco in(1,2,3,4,5,6,7)AND
EDAD BETWEEN '1' AND '130'

/*2004*/
UNION ALL

select causab,EDAD,
( CASE  WHEN  ocupa= 02     THEN   'No trabaja'
       WHEN   ocupa= 11     THEN   'Profesionistas'
       WHEN   ocupa= 12     THEN   'Técnicos'
       WHEN   ocupa= 13     THEN   'Trabajadores de la educación'
       WHEN   ocupa= 14     THEN   'Trabajadores del arte, espectáculos y deportes'
       WHEN   ocupa= 21     THEN   'Funcionarios y directivos'
       WHEN   ocupa= 41     THEN   'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
       WHEN   ocupa= 51     THEN   'Personal de control producción industrial'
       WHEN   ocupa= 52     THEN   'Trabajadores en la industria de la transformación'
       WHEN   ocupa= 53     THEN   'Operadores de maquinaria fija'
       WHEN   ocupa= 54     THEN   'Ayudantes producción industrial y artesanal'
       WHEN   ocupa= 55     THEN   'Conductores de maquinaria móvil y medios de
transporte'
       WHEN   ocupa= 61     THEN   'Trabajadores administrativos de nivel intermedio'
       WHEN   ocupa= 62     THEN   'Trabajadores administrativos de nivel   inferior'
       WHEN   ocupa= 71     THEN   'Comerciantes, empleados de comercio, agentes de
ventas'
       WHEN   ocupa= 72     THEN   'Vendedores ambulantes'
       WHEN   ocupa= 81     THEN   'Trabajadores servicios personales en
establecimientos'
       WHEN   ocupa= 82     THEN   'Trabajadores de servicio doméstico'
       WHEN   ocupa= 83     THEN   'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN   ocupa= 98     THEN   'No aplica menore sde 12 años'
       WHEN   ocupa= 99     THEN   'Ocupación no clasificada anteriormente'
       WHEN   ocupa= 00     THEN   'No especificado'     END) AS OCUPACION,
       (CASE   WHEN  esco=  1      THEN   'Sin escolaridad'
                      WHEN  esco = 2      THEN   'Menos de tres años de primaria'
                      WHEN  esco = 3      THEN   'De tres a cinco años de primaria'
                      WHEN  esco = 4      THEN   'Primaria completa'
                      WHEN  esco = 5      THEN   'Secundaria o equivalente'
                      WHEN  esco = 6      THEN   'Preparatoria o equivalente'
                     WHEN   esco = 7      THEN   'Profesional'
                     WHEN   esco = 8      THEN   'Menores de 6 años'
                     WHEN   esco = 0      THEN   'No Especificado'
              END) AS escolaridad
from def04
where trabajo=1
and causab IN ('X95','V89','V09','X59','V49','V87','W87','W17','X99','W20','Y09','V48')
and esco in(1,2,3,4,5,6,7,0,8)AND
EDAD BETWEEN '1' AND '130'

/*2005*/

UNION ALL

select causab,EDAD,
( CASE  WHEN  ocupa= 02     THEN   'No trabaja'
       WHEN   ocupa= 11     THEN   'Profesionistas'
       WHEN   ocupa= 12     THEN   'Técnicos'
       WHEN   ocupa= 13     THEN   'Trabajadores de la educación'
       WHEN   ocupa= 14     THEN   'Trabajadores del arte, espectáculos y deportes'
       WHEN   ocupa= 21     THEN   'Funcionarios y directivos'
```

```
       WHEN    ocupa= 41     THEN   'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
       WHEN    ocupa= 51     THEN   'Personal de control producción industrial'
       WHEN    ocupa= 52     THEN   'Trabajadores en la industria de la transformación'
       WHEN    ocupa= 53     THEN   'Operadores de maquinaria fija'
       WHEN    ocupa= 54     THEN   'Ayudantes producción industrial y artesanal'
       WHEN    ocupa= 55     THEN   'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupa= 61     THEN   'Trabajadores administrativos de nivel intermedio'
       WHEN    ocupa= 62     THEN   'Trabajadores administrativos de nivel   inferior'
       WHEN    ocupa= 71     THEN   'Comerciantes, empleados de comercio, agentes de
ventas'
       WHEN    ocupa= 72     THEN   'Vendedores ambulantes'
       WHEN    ocupa= 81     THEN   'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupa= 82     THEN   'Trabajadores de servicio doméstico'
       WHEN    ocupa= 83     THEN   'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN    ocupa= 98     THEN   'No aplica menore sde 12 años'
       WHEN    ocupa= 99     THEN   'Ocupación no clasificada anteriormente'
       WHEN    ocupa= 00     THEN   'No especificado'     END) AS OCUPACION,
       (CASE    WHEN  esco=   1      THEN   'Sin escolaridad'
                      WHEN   esco =  2      THEN   'Menos de tres años de primaria'
                      WHEN   esco =  3      THEN   'De tres a cinco años de primaria'
                      WHEN   esco =  4      THEN   'Primaria completa'
                      WHEN   esco =  5      THEN   'Secundaria o equivalente'
                      WHEN   esco =  6      THEN   'Preparatoria o equivalente'
                     WHEN    esco =  7      THEN   'Profesional'
                     WHEN    esco =  8      THEN   'Menores de 6 años'
                     WHEN    esco =  0      THEN   'No Especificado'
             END) AS escolaridad
from def05
where trabajo=1
and causab IN
('X95','V89','V09','X59','V49','V87','W87','W17','X99','Y09','W20','V48','W19','W13')
and esco in(1,2,3,4,5,6,7,0,8)AND
EDAD BETWEEN '1' AND '130'

/*2006*/

UNION ALL

select causadef,EDADVALOR,
( CASE  WHEN   ocupacion=    02     THEN   'No trabaja'
       WHEN    ocupacion=    11     THEN   'Profesionistas'
       WHEN    ocupacion=    12     THEN   'Técnicos'
       WHEN    ocupacion=    13     THEN   'Trabajadores de la educación'
       WHEN    ocupacion=    14     THEN   'Trabajadores del arte, espectáculos y
deportes'
       WHEN    ocupacion=    21     THEN   'Funcionarios y directivos'
       WHEN    ocupacion=    41     THEN   'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
       WHEN    ocupacion=    51     THEN   'Personal de control producción industrial'
       WHEN    ocupacion=    52     THEN   'Trabajadores en la industria de la
transformación'
       WHEN    ocupacion=    53     THEN   'Operadores de maquinaria fija'
       WHEN    ocupacion=    54     THEN   'Ayudantes producción industrial y
artesanal'
       WHEN    ocupacion=    55     THEN   'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupacion=    61     THEN   'Trabajadores administrativos de nivel
intermedio'
       WHEN    ocupacion=    62     THEN   'Trabajadores administrativos de nivel
inferior'
       WHEN    ocupacion=    71     THEN   'Comerciantes, empleados de comercio,
agentes de ventas'
       WHEN    ocupacion=    72     THEN   'Vendedores ambulantes'
       WHEN    ocupacion=    81     THEN   'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupacion=    82     THEN   'Trabajadores de servicio doméstico'
       WHEN    ocupacion=    83     THEN   'Trabajadores de fuerzas armadas, protección
y vigilancia'
       WHEN    ocupacion=    98     THEN   'No aplica menore sde 12 años'
       WHEN    ocupacion=    99     THEN   'Ocupación no clasificada anteriormente'
       WHEN    ocupacion=    00     THEN   'No especificado'     END) AS OCUPACION,
```

```
       (CASE    WHEN  escol= 1      THEN   'Sin escolaridad'
                       WHEN  escol = 2      THEN   'Menos de tres años de primaria'
                       WHEN  escol = 3      THEN   'De tres a cinco años de primaria'
                       WHEN  escol = 4      THEN   'Primaria completa'
                       WHEN  escol = 5      THEN   'Secundaria o equivalente'
                       WHEN  escol = 6      THEN   'Preparatoria o equivalente'
                      WHEN   escol = 7      THEN   'Profesional'
                      WHEN   escol = 8      THEN   'Menores de 6 años'
                      WHEN   escol = 0      THEN   'No Especificado'
              END) AS escolaridad
from def06
where trabajo=1
and causadef IN ('X59','X95','V89','V09','V49','V87','W87','W17','X99','Y09','X09')
and escol in(1,2,3,4,5,6,7,0,8) AND
EDADVALOR BETWEEN 1 AND 130


/*2007*/

UNION ALL

select causadef,EDADVALOR,
( CASE  WHEN  ocupacion=     02     THEN   'No trabaja'
       WHEN   ocupacion=     11     THEN   'Profesionistas'
       WHEN   ocupacion=     12     THEN   'Técnicos'
       WHEN   ocupacion=     13     THEN   'Trabajadores de la educación'
       WHEN   ocupacion=     14     THEN   'Trabajadores del arte, espectáculos y
deportes'
       WHEN   ocupacion=     21     THEN   'Funcionarios y directivos'
       WHEN   ocupacion=     41     THEN   'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
       WHEN   ocupacion=     51     THEN   'Personal de control producción industrial'
       WHEN   ocupacion=     52     THEN   'Trabajadores en la industria de la
transformación'
       WHEN   ocupacion=     53     THEN   'Operadores de maquinaria fija'
       WHEN   ocupacion=     54     THEN   'Ayudantes producción industrial y
artesanal'
       WHEN   ocupacion=     55     THEN   'Conductores de maquinaria móvil y medios de
transporte'
       WHEN   ocupacion=     61     THEN   'Trabajadores administrativos de nivel
intermedio'
       WHEN   ocupacion=     62     THEN   'Trabajadores administrativos de nivel
inferior'
       WHEN   ocupacion=     71     THEN   'Comerciantes, empleados de comercio,
agentes de ventas'
       WHEN   ocupacion=     72     THEN   'Vendedores ambulantes'
       WHEN   ocupacion=     81     THEN   'Trabajadores servicios personales en
establecimientos'
       WHEN   ocupacion=     82     THEN   'Trabajadores de servicio doméstico'
       WHEN   ocupacion=     83     THEN   'Trabajadores de fuerzas armadas, protección
y vigilancia'
       WHEN   ocupacion=     98     THEN   'No aplica menore sde 12 años'
       WHEN   ocupacion=     99     THEN   'Ocupación no clasificada anteriormente'
       WHEN   ocupacion=     00     THEN   'No especificado'     END) AS OCUPACION,
       (CASE    WHEN  escol= 1      THEN   'Sin escolaridad'
                       WHEN  escol = 2      THEN   'Menos de tres años de primaria'
                       WHEN  escol = 3      THEN   'De tres a cinco años de primaria'
                       WHEN  escol = 4      THEN   'Primaria completa'
                       WHEN  escol = 5      THEN   'Secundaria o equivalente'
                       WHEN  escol = 6      THEN   'Preparatoria o equivalente'
                      WHEN   escol = 7      THEN   'Profesional'
                      WHEN   escol = 8      THEN   'Menores de 6 años'
                      WHEN   escol = 0      THEN   'No Especificado'
              END) AS escolaridad
from def07
where trabajo=1
and causadef IN ('X59','X95','V89','V09','V49','V87','W87','W17','X99','Y09','X09')
and escol in(1,2,3,4,5,6,7,0,8) AND
EDADVALOR BETWEEN 1 AND 130

/*2008*/

UNION ALL

select causadef,EDADVALOR,
```

```
( CASE  WHEN   ocupacion=     02      THEN    'No trabaja'
       WHEN    ocupacion=     11      THEN    'Profesionistas'
       WHEN    ocupacion=     12      THEN    'Técnicos'
       WHEN    ocupacion=     13      THEN    'Trabajadores de la educación'
       WHEN    ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
       WHEN    ocupacion=     21      THEN    'Funcionarios y directivos'
       WHEN    ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
       WHEN    ocupacion=     51      THEN    'Personal de control producción industrial'
       WHEN    ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
       WHEN    ocupacion=     53      THEN    'Operadores de maquinaria fija'
       WHEN    ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
       WHEN    ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
       WHEN    ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
       WHEN    ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
       WHEN    ocupacion=     72      THEN    'Vendedores ambulantes'
       WHEN    ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
       WHEN    ocupacion=     98      THEN    'No aplica menore sde 12 años'
       WHEN    ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
       (CASE    WHEN   escol= 1       THEN    'Sin escolaridad'
                       WHEN   escol = 2       THEN    'Menos de tres años de primaria'
                       WHEN   escol = 3       THEN    'De tres a cinco años de primaria'
                       WHEN   escol = 4       THEN    'Primaria completa'
                       WHEN   escol = 5       THEN    'Secundaria o equivalente'
                       WHEN   escol = 6       THEN    'Preparatoria o equivalente'
                      WHEN    escol = 7       THEN    'Profesional'
                      WHEN    escol = 8       THEN    'Menores de 6 años'
                      WHEN    escol = 0       THEN    'No Especificado'
              END) AS escolaridad
from def08
where trabajo=1
and causadef IN ('X95','V89','X59','V09','V87','V49','W87','W17','X99','W19','W20')
and escol in(1,2,3,4,5,6,7,0,8)AND
EDADVALOR BETWEEN 1 AND 130

/*2009*/

UNION ALL

select causadef,EDADVALOR,
( CASE  WHEN   ocupacion=     02      THEN    'No trabaja'
       WHEN    ocupacion=     11      THEN    'Profesionistas'
       WHEN    ocupacion=     12      THEN    'Técnicos'
       WHEN    ocupacion=     13      THEN    'Trabajadores de la educación'
       WHEN    ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
       WHEN    ocupacion=     21      THEN    'Funcionarios y directivos'
       WHEN    ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
       WHEN    ocupacion=     51      THEN    'Personal de control producción industrial'
       WHEN    ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
       WHEN    ocupacion=     53      THEN    'Operadores de maquinaria fija'
       WHEN    ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
       WHEN    ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
       WHEN    ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
```

```
       WHEN    ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
       WHEN    ocupacion=     72      THEN    'Vendedores ambulantes'
       WHEN    ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
       WHEN    ocupacion=     98      THEN    'No aplica menore sde 12 años'
       WHEN    ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
       (CASE    WHEN   escol=  1       THEN    'Sin escolaridad'
                       WHEN   escol = 2       THEN    'Menos de tres años de primaria'
                       WHEN   escol = 3       THEN    'De tres a cinco años de primaria'
                       WHEN   escol = 4       THEN    'Primaria completa'
                       WHEN   escol = 5       THEN    'Secundaria o equivalente'
                       WHEN   escol = 6       THEN    'Preparatoria o equivalente'
                      WHEN    escol = 7       THEN    'Profesional'
                      WHEN    escol = 8       THEN    'Menores de 6 años'
                      WHEN    escol = 0       THEN    'No Especificado'
              END) AS escolaridad
from def_09
where trabajo=1
and causadef IN ('X95','X59','V89','V09','V49','V87','W87','W17','X99','Y09','W20')
and escol in(1,2,3,4,5,6,7,0,8)AND
EDADVALOR BETWEEN 1 AND 130
```

Consulta para la vista VIOLENCIA FAMILIAR 4.4

(Mujeres)

```
/*2004*/
SELECT CVEEDAD,EDAD,
( CASE  WHEN   ocupa=  02      THEN    'No trabaja'
       WHEN    ocupa=  11      THEN    'Profesionistas'
       WHEN    ocupa=  12      THEN    'Técnicos'
       WHEN    ocupa=  13      THEN    'Trabajadores de la educación'
       WHEN    ocupa=  14      THEN    'Trabajadores del arte, espectáculos y deportes'
       WHEN    ocupa=  21      THEN    'Funcionarios y directivos'
       WHEN    ocupa=  41      THEN    'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
       WHEN    ocupa=  51      THEN    'Personal de control producción industrial'
       WHEN    ocupa=  52      THEN    'Trabajadores en la industria de la transformación'
       WHEN    ocupa=  53      THEN    'Operadores de maquinaria fija'
       WHEN    ocupa=  54      THEN    'Ayudantes producción industrial y artesanal'
       WHEN    ocupa=  55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupa=  61      THEN    'Trabajadores administrativos de nivel intermedio'
       WHEN    ocupa=  62      THEN    'Trabajadores administrativos de nivel   inferior'
       WHEN    ocupa=  71      THEN    'Comerciantes, empleados de comercio, agentes de
ventas'
       WHEN    ocupa=  72      THEN    'Vendedores ambulantes'
       WHEN    ocupa=  81      THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupa=  82      THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupa=  83      THEN    'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN    ocupa=  98      THEN    'No aplica menore sde 12 años'
       WHEN    ocupa=  99      THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupa=  00      THEN    'No especificado'     END) AS OCUPACION,
( CASE  WHEN    edocony=       0       THEN    'SE IGNORA'
               WHEN    edocony=       1       THEN    'SOLTERO'
               WHEN    edocony=       2       THEN    'VIUDO'
               WHEN    edocony=       3       THEN    'DIVORSIADO'
               WHEN    edocony=       4       THEN    'UNION LIBRE'
               WHEN    edocony=       5       THEN    'CASADO'
               WHEN    edocony=       8       THEN    'MENOR DE 12 AÑOS'
               END) AS ESTADOCIVIL,
(CASE    WHEN   esco=   1       THEN    'Sin escolaridad'
                       WHEN   esco =  2       THEN    'Menos de tres años de primaria'
                       WHEN   esco =  3       THEN    'De tres a cinco años de primaria'
                       WHEN   esco =  4       THEN    'Primaria completa'
```

```
                        WHEN   esco =  5       THEN    'Secundaria o equivalente'
                        WHEN   esco =  6       THEN    'Preparatoria o equivalente'
                       WHEN    esco =  7       THEN    'Profesional'
                       WHEN    esco =  8       THEN    'Menores de 6 años'
                       WHEN    esco =  0       THEN    'No Especificado'
                END) AS escolaridad,
(CASE WHEN  SEXO=2 THEN 'MUJER' END) AS SEXO,
RIEGO=CASE when edocony IN (5,4) OR
                                                       OCUPA=2 AND
                                                       ESCO=1 then 'MUY ALTO'
                                                when edocony=1 OR
                                                       OCUPA=2 AND
                                                       ESCO IN (4,5)then 'ALTO'
                                                when edocony=3 OR
                                                       ESCO=3 then 'BAJO'
                                                when edocony=2 OR
                                                       ESCO in(2,7) then 'MUY BAJO' END
FROM DEF04
where violfam=1
AND SEXO=2 AND
EDAD BETWEEN 15 AND 120

UNION

/*2005*/

SELECT CVEEDAD,EDAD,

( CASE  WHEN   ocupa=  02      THEN    'No trabaja'
       WHEN    ocupa=  11      THEN    'Profesionistas'
       WHEN    ocupa=  12      THEN    'Técnicos'
       WHEN    ocupa=  13      THEN    'Trabajadores de la educación'
       WHEN    ocupa=  14      THEN    'Trabajadores del arte, espectáculos y deportes'
       WHEN    ocupa=  21      THEN    'Funcionarios y directivos'
       WHEN    ocupa=  41      THEN    'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
       WHEN    ocupa=  51      THEN    'Personal de control producción industrial'
       WHEN    ocupa=  52      THEN    'Trabajadores en la industria de la transformación'
       WHEN    ocupa=  53      THEN    'Operadores de maquinaria fija'
       WHEN    ocupa=  54      THEN    'Ayudantes producción industrial y artesanal'
       WHEN    ocupa=  55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupa=  61      THEN    'Trabajadores administrativos de nivel intermedio'
       WHEN    ocupa=  62      THEN    'Trabajadores administrativos de nivel   inferior'
       WHEN    ocupa=  71      THEN    'Comerciantes, empleados de comercio, agentes de
ventas'
       WHEN    ocupa=  72      THEN    'Vendedores ambulantes'
       WHEN    ocupa=  81      THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupa=  82      THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupa=  83      THEN    'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN    ocupa=  98      THEN    'No aplica menore sde 12 años'
       WHEN    ocupa=  99      THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupa=  00      THEN    'No especificado'     END) AS OCUPACION,
( CASE  WHEN    edocony=       0       THEN    'SE IGNORA'
               WHEN    edocony=       1       THEN    'SOLTERO'
               WHEN    edocony=       2       THEN    'VIUDO'
               WHEN    edocony=       3       THEN    'DIVORSIADO'
               WHEN    edocony=       4       THEN    'UNION LIBRE'
               WHEN    edocony=       5       THEN    'CASADO'
               WHEN    edocony=       8       THEN    'MENOR DE 12 AÑOS'
               END) AS ESTADOCIVIL,
(CASE    WHEN   esco=   1       THEN    'Sin escolaridad'
                        WHEN   esco =  2       THEN    'Menos de tres años de primaria'
                        WHEN   esco =  3       THEN    'De tres a cinco años de primaria'
                        WHEN   esco =  4       THEN    'Primaria completa'
                        WHEN   esco =  5       THEN    'Secundaria o equivalente'
                        WHEN   esco =  6       THEN    'Preparatoria o equivalente'
                       WHEN    esco =  7       THEN    'Profesional'
                       WHEN    esco =  8       THEN    'Menores de 6 años'
                       WHEN    esco =  0       THEN    'No Especificado'
               END) AS escolaridad,
(CASE WHEN  SEXO=2 THEN 'MUJER' END) AS SEXO,
RIEGO=CASE when edocony IN (5,4) OR
```

```
                                                        OCUPA=2 AND
                                                        ESCO=1 then 'MUY ALTO'
                                                when edocony=1 OR
                                                        OCUPA=2 AND
                                                        ESCO IN (4,5)then 'ALTO'
                                                when edocony=3 OR
                                                        ESCO=3 then 'BAJO'
                                                when edocony=2 OR
                                                        ESCO in(2,7) then 'MUY BAJO' END
FROM DEF05
where violfam=1
AND SEXO=2 AND
EDAD BETWEEN 15 AND 120

UNION

/*2006*/
SELECT edaduni ,edadvalor,
( CASE  WHEN   ocupacion=     02      THEN    'No trabaja'
        WHEN    ocupacion=     11      THEN    'Profesionistas'
        WHEN    ocupacion=     12      THEN    'Técnicos'
        WHEN    ocupacion=     13      THEN    'Trabajadores de la educación'
        WHEN    ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
        WHEN    ocupacion=     21      THEN    'Funcionarios y directivos'
        WHEN    ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
        WHEN    ocupacion=     51      THEN    'Personal de control producción industrial'
        WHEN    ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
        WHEN    ocupacion=     53      THEN    'Operadores de maquinaria fija'
        WHEN    ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
        WHEN    ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
        WHEN    ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
        WHEN    ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
        WHEN    ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
        WHEN    ocupacion=     72      THEN    'Vendedores ambulantes'
        WHEN    ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
        WHEN    ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
        WHEN    ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
        WHEN    ocupacion=     98      THEN    'No aplica menore sde 12 años'
        WHEN    ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
        WHEN    ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
(CASE   WHEN   escol= 1       THEN    'Sin escolaridad'
                        WHEN   escol = 2       THEN    'Menos de tres años de primaria'
                        WHEN   escol = 3       THEN    'De tres a cinco años de primaria'
                        WHEN   escol = 4       THEN    'Primaria completa'
                        WHEN   escol = 5       THEN    'Secundaria o equivalente'
                        WHEN   escol = 6       THEN    'Preparatoria o equivalente'
                        WHEN   escol = 7       THEN    'Profesional'
                        WHEN   escol = 8       THEN    'NO APLICA A MENOR DE 6 AÑOS'
                        WHEN   escol = 0       THEN    'NO ESPECIFICADO'
                        END) AS ESCOLARIDAD,
( CASE  WHEN    edocivil=      0       THEN    'SE IGNORA'
               WHEN    edocivil=      1       THEN    'SOLTERO'
               WHEN    edocivil=      2       THEN    'VIUDO'
               WHEN    edocivil=      3       THEN    'DIVORSIADO'
               WHEN    edocivil=      4       THEN    'UNION LIBRE'
               WHEN    edocivil=      5       THEN    'CASADO'
               WHEN    edocivil=      8       THEN    'MENOR DE 12 AÑOS'
               END) AS ESTADOCIVIL,
(CASE WHEN  SEXO=2 THEN 'MUJER' END) AS SEXO, RIESGO=case
                                                when edocivil IN (5,4) OR
                                                        OCUPACION=2 AND
                                                        ESCOL=1 then 'MUY ALTO'
                                                when edocivil=1 OR
                                                        OCUPACION=2 AND
                                                        ESCOL IN (3,5)then 'ALTO'
```

```
                                        when edocivil=3 OR
                                                ESCOL=4 then 'BAJO'
                                        when edocivil=2 OR
                                                ESCOL in(6,7) then 'MUY BAJO' END
FROM DEF06
where violfam=1
AND SEXO=2 AND
EDADVALOR BETWEEN 15 AND 120

UNION
/*2007*/

SELECT edaduni ,edadvalor,
( CASE  WHEN   ocupacion=     02      THEN    'No trabaja'
        WHEN    ocupacion=     11      THEN    'Profesionistas'
        WHEN    ocupacion=     12      THEN    'Técnicos'
        WHEN    ocupacion=     13      THEN    'Trabajadores de la educación'
        WHEN    ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
        WHEN    ocupacion=     21      THEN    'Funcionarios y directivos'
        WHEN    ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
        WHEN    ocupacion=     51      THEN    'Personal de control producción industrial'
        WHEN    ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
        WHEN    ocupacion=     53      THEN    'Operadores de maquinaria fija'
        WHEN    ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
        WHEN    ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
        WHEN    ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
        WHEN    ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
        WHEN    ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
        WHEN    ocupacion=     72      THEN    'Vendedores ambulantes'
        WHEN    ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
        WHEN    ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
        WHEN    ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
        WHEN    ocupacion=     98      THEN    'No aplica menore sde 12 años'
        WHEN    ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
        WHEN    ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
(CASE   WHEN   escol= 1       THEN    'Sin escolaridad'
                      WHEN   escol = 2       THEN    'Menos de tres años de primaria'
                      WHEN   escol = 3       THEN    'De tres a cinco años de primaria'
                      WHEN   escol = 4       THEN    'Primaria completa'
                      WHEN   escol = 5       THEN    'Secundaria o equivalente'
                      WHEN   escol = 6       THEN    'Preparatoria o equivalente'
                      WHEN   escol = 7       THEN    'Profesional'
                      WHEN   escol = 8       THEN    'NO APLICA A MENOR DE 6 AÑOS'
                      WHEN   escol = 0       THEN    'NO ESPECIFICADO'
                      END) AS ESCOLARIDAD,
( CASE  WHEN    edocivil=      0       THEN    'SE IGNORA'
               WHEN    edocivil=      1       THEN    'SOLTERO'
               WHEN    edocivil=      2       THEN    'VIUDO'
               WHEN    edocivil=      3       THEN    'DIVORSIADO'
               WHEN    edocivil=      4       THEN    'UNION LIBRE'
               WHEN    edocivil=      5       THEN    'CASADO'
               WHEN    edocivil=      8       THEN    'MENOR DE 12 AÑOS'
               END) AS ESTADOCIVIL,
(CASE WHEN  SEXO=2 THEN 'MUJER' END) AS SEXO, RIESGO=case
                                        when edocivil IN (5,4) OR
                                                OCUPACION=2 AND
                                                ESCOL=1 then 'MUY ALTO'
                                        when edocivil=1 OR
                                                OCUPACION=2 AND
                                                ESCOL IN (3,5)then 'ALTO'
                                        when edocivil=3 OR
                                                ESCOL=4 then 'BAJO'
                                        when edocivil=2 OR
                                                ESCOL in(6,7) then 'MUY BAJO' END
FROM DEF07
```

```
where violfam=1
AND SEXO=2 AND
EDADVALOR BETWEEN 15 AND 120

UNION
/*2008*/

SELECT edaduni ,edadvalor,
( CASE  WHEN   ocupacion=     02      THEN    'No trabaja'
        WHEN    ocupacion=     11      THEN    'Profesionistas'
        WHEN    ocupacion=     12      THEN    'Técnicos'
        WHEN    ocupacion=     13      THEN    'Trabajadores de la educación'
        WHEN    ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
        WHEN    ocupacion=     21      THEN    'Funcionarios y directivos'
        WHEN    ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
        WHEN    ocupacion=     51      THEN    'Personal de control producción industrial'
        WHEN    ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
        WHEN    ocupacion=     53      THEN    'Operadores de maquinaria fija'
        WHEN    ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
        WHEN    ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
        WHEN    ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
        WHEN    ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
        WHEN    ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
        WHEN    ocupacion=     72      THEN    'Vendedores ambulantes'
        WHEN    ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
        WHEN    ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
        WHEN    ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
        WHEN    ocupacion=     98      THEN    'No aplica menore sde 12 años'
        WHEN    ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
        WHEN    ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
(CASE   WHEN   escol= 1       THEN    'Sin escolaridad'
                      WHEN   escol = 2       THEN    'Menos de tres años de primaria'
                      WHEN   escol = 3       THEN    'De tres a cinco años de primaria'
                      WHEN   escol = 4       THEN    'Primaria completa'
                      WHEN   escol = 5       THEN    'Secundaria o equivalente'
                      WHEN   escol = 6       THEN    'Preparatoria o equivalente'
                      WHEN   escol = 7       THEN    'Profesional'
                      WHEN   escol = 8       THEN    'NO APLICA A MENOR DE 6 AÑOS'
                      WHEN   escol = 0       THEN    'NO ESPECIFICADO'
                      END) AS ESCOLARIDAD,
( CASE WHEN    edocivil=      0       THEN    'SE IGNORA'
               WHEN    edocivil=      1       THEN    'SOLTERO'
               WHEN    edocivil=      2       THEN    'VIUDO'
               WHEN    edocivil=      3       THEN    'DIVORSIADO'
               WHEN    edocivil=      4       THEN    'UNION LIBRE'
               WHEN    edocivil=      5       THEN    'CASADO'
               WHEN    edocivil=      8       THEN    'MENOR DE 12 AÑOS'
               END) AS ESTADOCIVIL,
(CASE WHEN  SEXO=2 THEN 'MUJER' END) AS SEXO, RIESGO=case
                                        when edocivil IN (5,4) OR
                                                OCUPACION=2 AND
                                                ESCOL=1 then 'MUY ALTO'
                                        when edocivil=1 OR
                                                OCUPACION=2 AND
                                                ESCOL IN (3,5)then 'ALTO'
                                        when edocivil=3 OR
                                                ESCOL=4 then 'BAJO'
                                        when edocivil=2 OR
                                                ESCOL in(6,7) then 'MUY BAJO' END
FROM DEF08
where violfam=1
AND SEXO=2 AND
EDADVALOR BETWEEN 15 AND 120

/*2009*/
```

```
SELECT edaduni ,edadvalor,
( CASE  WHEN  ocupacion=     02      THEN   'No trabaja'
       WHEN   ocupacion=     11      THEN   'Profesionistas'
       WHEN   ocupacion=     12      THEN   'Técnicos'
       WHEN   ocupacion=     13      THEN   'Trabajadores de la educación'
       WHEN   ocupacion=     14      THEN   'Trabajadores del arte, espectáculos y
deportes'
       WHEN   ocupacion=     21      THEN   'Funcionarios y directivos'
       WHEN   ocupacion=     41      THEN   'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
       WHEN   ocupacion=     51      THEN   'Personal de control producción industrial'
       WHEN   ocupacion=     52      THEN   'Trabajadores en la industria de la
transformación'
       WHEN   ocupacion=     53      THEN   'Operadores de maquinaria fija'
       WHEN   ocupacion=     54      THEN   'Ayudantes producción industrial y
artesanal'
       WHEN   ocupacion=     55      THEN   'Conductores de maquinaria móvil y medios de
transporte'
       WHEN   ocupacion=     61      THEN   'Trabajadores administrativos de nivel
intermedio'
       WHEN   ocupacion=     62      THEN   'Trabajadores administrativos de nivel
inferior'
       WHEN   ocupacion=     71      THEN   'Comerciantes, empleados de comercio,
agentes de ventas'
       WHEN   ocupacion=     72      THEN   'Vendedores ambulantes'
       WHEN   ocupacion=     81      THEN   'Trabajadores servicios personales en
establecimientos'
       WHEN   ocupacion=     82      THEN   'Trabajadores de servicio doméstico'
       WHEN   ocupacion=     83      THEN   'Trabajadores de fuerzas armadas, protección
y vigilancia'
       WHEN   ocupacion=     98      THEN   'No aplica menore sde 12 años'
       WHEN   ocupacion=     99      THEN   'Ocupación no clasificada anteriormente'
       WHEN   ocupacion=     00      THEN   'No especificado'     END) AS OCUPACION,
(CASE   WHEN   escol= 1       THEN   'Sin escolaridad'
                      WHEN   escol = 2       THEN   'Menos de tres años de primaria'
                      WHEN   escol = 3       THEN   'De tres a cinco años de primaria'
                      WHEN   escol = 4       THEN   'Primaria completa'
                      WHEN   escol = 5       THEN   'Secundaria o equivalente'
                      WHEN   escol = 6       THEN   'Preparatoria o equivalente'
                      WHEN   escol = 7       THEN   'Profesional'
                      WHEN   escol = 8       THEN   'NO APLICA A MENOR DE 6 AÑOS'
                      WHEN   escol = 0       THEN   'NO ESPECIFICADO'
                      END) AS ESCOLARIDAD,
( CASE WHEN   edocivil=      0       THEN   'SE IGNORA'
              WHEN   edocivil=      1       THEN   'SOLTERO'
              WHEN   edocivil=      2       THEN   'VIUDO'
              WHEN   edocivil=      3       THEN   'DIVORSIADO'
              WHEN   edocivil=      4       THEN   'UNION LIBRE'
              WHEN   edocivil=      5       THEN   'CASADO'
              WHEN   edocivil=      8       THEN   'MENOR DE 12 AÑOS'
              END) AS ESTADOCIVIL,
(CASE WHEN  SEXO=2 THEN 'MUJER' END) AS SEXO, RIESGO=case
                                          when edocivil IN (5,4) OR
                                                  OCUPACION=2 AND
                                                  ESCOL=1 then 'MUY ALTO'
                                          when edocivil=1 OR
                                                  OCUPACION=2 AND
                                                  ESCOL IN (3,5)then 'ALTO'
                                          when edocivil=3 OR
                                                  ESCOL=4 then 'BAJO'
                                          when edocivil=2 OR
                                                  ESCOL in(6,7) then 'MUY BAJO' END
FROM def_09
where violfam=1
AND SEXO=2 AND
EDADVALOR BETWEEN 15 AND 120
```

(General)

```
/*2004*/
SELECT CVEEDAD,EDAD,

( CASE  WHEN  ocupa= 02     THEN   'No trabaja'
       WHEN   ocupa= 11     THEN   'Profesionistas'
```

```
       WHEN    ocupa= 12      THEN    'Técnicos'
       WHEN    ocupa= 13      THEN    'Trabajadores de la educación'
       WHEN    ocupa= 14      THEN    'Trabajadores del arte, espectáculos y deportes'
       WHEN    ocupa= 21      THEN    'Funcionarios y directivos'
       WHEN    ocupa= 41      THEN    'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
       WHEN    ocupa= 51      THEN    'Personal de control producción industrial'
       WHEN    ocupa= 52      THEN    'Trabajadores en la industria de la transformación'
       WHEN    ocupa= 53      THEN    'Operadores de maquinaria fija'
       WHEN    ocupa= 54      THEN    'Ayudantes producción industrial y artesanal'
       WHEN    ocupa= 55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupa= 61      THEN    'Trabajadores administrativos de nivel intermedio'
       WHEN    ocupa= 62      THEN    'Trabajadores administrativos de nivel   inferior'
       WHEN    ocupa= 71      THEN    'Comerciantes, empleados de comercio, agentes de
ventas'
       WHEN    ocupa= 72      THEN    'Vendedores ambulantes'
       WHEN    ocupa= 81      THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupa= 82      THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupa= 83      THEN    'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN    ocupa= 98      THEN    'No aplica menore sde 12 años'
       WHEN    ocupa= 99      THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupa= 00      THEN    'No especificado'     END) AS OCUPACION,
( CASE  WHEN    edocony=       0       THEN    'SE IGNORA'
               WHEN    edocony=       1       THEN    'SOLTERO'
               WHEN    edocony=       2       THEN    'VIUDO'
               WHEN    edocony=       3       THEN    'DIVORSIADO'
               WHEN    edocony=       4       THEN    'UNION LIBRE'
               WHEN    edocony=       5       THEN    'CASADO'
               WHEN    edocony=       8       THEN    'MENOR DE 12 AÑOS'
               END) AS ESTADOCIVIL,
(CASE    WHEN   esco=  1       THEN    'Sin escolaridad'
                       WHEN   esco =  2       THEN    'Menos de tres años de primaria'
                       WHEN   esco =  3       THEN    'De tres a cinco años de primaria'
                       WHEN   esco =  4       THEN    'Primaria completa'
                       WHEN   esco =  5       THEN    'Secundaria o equivalente'
                       WHEN   esco =  6       THEN    'Preparatoria o equivalente'
                      WHEN    esco =  7       THEN    'Profesional'
                      WHEN    esco =  8       THEN    'Menores de 6 años'
                      WHEN    esco =  0       THEN    'No Especificado'
              END) AS escolaridad,
(case when edad BETWEEN  0 and 9 then 'MUY ALTO'
                                        when edad BETWEEN  10 and 44 then 'ALTO'
                                        when edad BETWEEN  45 and 120 then 'BAJO'
END) AS RIESGO
FROM DEF04
where violfam=1

UNION

/*2005*/

SELECT CVEEDAD,EDAD,

( CASE  WHEN   ocupa= 02      THEN    'No trabaja'
       WHEN    ocupa= 11      THEN    'Profesionistas'
       WHEN    ocupa= 12      THEN    'Técnicos'
       WHEN    ocupa= 13      THEN    'Trabajadores de la educación'
       WHEN    ocupa= 14      THEN    'Trabajadores del arte, espectáculos y deportes'
       WHEN    ocupa= 21      THEN    'Funcionarios y directivos'
       WHEN    ocupa= 41      THEN    'Trabajadores en actividades, agrícolas, ganaderas,
caza y pesca'
       WHEN    ocupa= 51      THEN    'Personal de control producción industrial'
       WHEN    ocupa= 52      THEN    'Trabajadores en la industria de la transformación'
       WHEN    ocupa= 53      THEN    'Operadores de maquinaria fija'
       WHEN    ocupa= 54      THEN    'Ayudantes producción industrial y artesanal'
       WHEN    ocupa= 55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupa= 61      THEN    'Trabajadores administrativos de nivel intermedio'
       WHEN    ocupa= 62      THEN    'Trabajadores administrativos de nivel   inferior'
       WHEN    ocupa= 71      THEN    'Comerciantes, empleados de comercio, agentes de
ventas'
       WHEN    ocupa= 72      THEN    'Vendedores ambulantes'
```

```
       WHEN    ocupa= 81      THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupa= 82      THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupa= 83      THEN    'Trabajadores de fuerzas armadas, protección y
vigilancia'
       WHEN    ocupa= 98      THEN    'No aplica menore sde 12 años'
       WHEN    ocupa= 99      THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupa= 00      THEN    'No especificado'     END) AS OCUPACION,
( CASE WHEN    edocony=       0       THEN    'SE IGNORA'
               WHEN    edocony=       1       THEN    'SOLTERO'
               WHEN    edocony=       2       THEN    'VIUDO'
               WHEN    edocony=       3       THEN    'DIVORSIADO'
               WHEN    edocony=       4       THEN    'UNION LIBRE'
               WHEN    edocony=       5       THEN    'CASADO'
               WHEN    edocony=       8       THEN    'MENOR DE 12 AÑOS'
               END) AS ESTADOCIVIL,
(CASE   WHEN   esco=  1       THEN    'Sin escolaridad'
                      WHEN   esco = 2       THEN    'Menos de tres años de primaria'
                      WHEN   esco = 3       THEN    'De tres a cinco años de primaria'
                      WHEN   esco = 4       THEN    'Primaria completa'
                      WHEN   esco = 5       THEN    'Secundaria o equivalente'
                      WHEN   esco = 6       THEN    'Preparatoria o equivalente'
                     WHEN    esco = 7       THEN    'Profesional'
                     WHEN    esco = 8       THEN    'Menores de 6 años'
                     WHEN    esco = 0       THEN    'No Especificado'
              END) AS escolaridad,
(case when edad BETWEEN  0 and 9 then 'MUY ALTO'
                                            when edad BETWEEN  10 and 44 then 'ALTO'
                                            when edad BETWEEN  45 and 120 then 'BAJO'
END) AS RIESGO
FROM DEF05
where violfam=1

UNION

/*2006*/
SELECT edaduni ,edadvalor,
( CASE  WHEN   ocupacion=     02      THEN    'No trabaja'
       WHEN    ocupacion=     11      THEN    'Profesionistas'
       WHEN    ocupacion=     12      THEN    'Técnicos'
       WHEN    ocupacion=     13      THEN    'Trabajadores de la educación'
       WHEN    ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
       WHEN    ocupacion=     21      THEN    'Funcionarios y directivos'
       WHEN    ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
       WHEN    ocupacion=     51      THEN    'Personal de control producción industrial'
       WHEN    ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
       WHEN    ocupacion=     53      THEN    'Operadores de maquinaria fija'
       WHEN    ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
       WHEN    ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
       WHEN    ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
       WHEN    ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
       WHEN    ocupacion=     72      THEN    'Vendedores ambulantes'
       WHEN    ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
       WHEN    ocupacion=     98      THEN    'No aplica menore sde 12 años'
       WHEN    ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
(CASE   WHEN   escol= 1       THEN    'Sin escolaridad'
                      WHEN   escol = 2      THEN    'Menos de tres años de primaria'
                      WHEN   escol = 3      THEN    'De tres a cinco años de primaria'
                      WHEN   escol = 4      THEN    'Primaria completa'
                      WHEN   escol = 5      THEN    'Secundaria o equivalente'
                      WHEN   escol = 6      THEN    'Preparatoria o equivalente'
```

```
                        WHEN   escol = 7       THEN    'Profesional'
                        WHEN   escol = 8       THEN    'NO APLICA A MENOR DE 6 AÑOS'
                        WHEN   escol = 0       THEN    'NO ESPECIFICADO'
                        END) AS ESCOLARIDAD,
( CASE  WHEN    edocivil=      0       THEN    'SE IGNORA'
                WHEN    edocivil=      1       THEN    'SOLTERO'
                WHEN    edocivil=      2       THEN    'VIUDO'
                WHEN    edocivil=      3       THEN    'DIVORSIADO'
                WHEN    edocivil=      4       THEN    'UNION LIBRE'
                WHEN    edocivil=      5       THEN    'CASADO'
                WHEN    edocivil=      8       THEN    'MENOR DE 12 AÑOS'
                END) AS ESTADOCIVIL,
(case when edadvalor BETWEEN  0 and 9 then 'MUY ALTO'
                                               when edadvalor BETWEEN  10 and 44 then
'ALTO'
                                               when edadvalor BETWEEN  45 and 120 then
'BAJO' END) AS RIESGO
FROM DEF06
where violfam=1

UNION
/*2007*/

SELECT edaduni ,edadvalor,
( CASE   WHEN   ocupacion=     02      THEN    'No trabaja'
        WHEN    ocupacion=     11      THEN    'Profesionistas'
        WHEN    ocupacion=     12      THEN    'Técnicos'
        WHEN    ocupacion=     13      THEN    'Trabajadores de la educación'
        WHEN    ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
        WHEN    ocupacion=     21      THEN    'Funcionarios y directivos'
        WHEN    ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
        WHEN    ocupacion=     51      THEN    'Personal de control producción industrial'
        WHEN    ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
        WHEN    ocupacion=     53      THEN    'Operadores de maquinaria fija'
        WHEN    ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
        WHEN    ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
        WHEN    ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
        WHEN    ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
        WHEN    ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
        WHEN    ocupacion=     72      THEN    'Vendedores ambulantes'
        WHEN    ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
        WHEN    ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
        WHEN    ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
        WHEN    ocupacion=     98      THEN    'No aplica menore sde 12 años'
        WHEN    ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
        WHEN    ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
(CASE    WHEN   escol= 1       THEN    'Sin escolaridad'
                        WHEN   escol = 2       THEN    'Menos de tres años de primaria'
                        WHEN   escol = 3       THEN    'De tres a cinco años de primaria'
                        WHEN   escol = 4       THEN    'Primaria completa'
                        WHEN   escol = 5       THEN    'Secundaria o equivalente'
                        WHEN   escol = 6       THEN    'Preparatoria o equivalente'
                        WHEN   escol = 7       THEN    'Profesional'
                        WHEN   escol = 8       THEN    'NO APLICA A MENOR DE 6 AÑOS'
                        WHEN   escol = 0       THEN    'NO ESPECIFICADO'
                        END) AS ESCOLARIDAD,
( CASE  WHEN    edocivil=      0       THEN    'SE IGNORA'
                WHEN    edocivil=      1       THEN    'SOLTERO'
                WHEN    edocivil=      2       THEN    'VIUDO'
                WHEN    edocivil=      3       THEN    'DIVORSIADO'
                WHEN    edocivil=      4       THEN    'UNION LIBRE'
                WHEN    edocivil=      5       THEN    'CASADO'
                WHEN    edocivil=      8       THEN    'MENOR DE 12 AÑOS'
                END) AS ESTADOCIVIL,
(case when edadvalor BETWEEN  0 and 9 then 'MUY ALTO'
```

```
                                          when edadvalor BETWEEN  10 and 44 then
'ALTO'
                                          when edadvalor BETWEEN  45 and 120 then
'BAJO' END) AS RIESGO
FROM DEF07
where violfam=1

UNION

/*2008*/

SELECT edaduni ,edadvalor,
( CASE  WHEN   ocupacion=     02      THEN    'No trabaja'
        WHEN   ocupacion=     11      THEN    'Profesionistas'
        WHEN   ocupacion=     12      THEN    'Técnicos'
        WHEN   ocupacion=     13      THEN    'Trabajadores de la educación'
        WHEN   ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
        WHEN   ocupacion=     21      THEN    'Funcionarios y directivos'
        WHEN   ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
        WHEN   ocupacion=     51      THEN    'Personal de control producción industrial'
        WHEN   ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
        WHEN   ocupacion=     53      THEN    'Operadores de maquinaria fija'
        WHEN   ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
        WHEN   ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
        WHEN   ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
        WHEN   ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
        WHEN   ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
        WHEN   ocupacion=     72      THEN    'Vendedores ambulantes'
        WHEN   ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
        WHEN   ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
        WHEN   ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
        WHEN   ocupacion=     98      THEN    'No aplica menore sde 12 años'
        WHEN   ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
        WHEN   ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
(CASE   WHEN   escol= 1       THEN    'Sin escolaridad'
                      WHEN   escol = 2       THEN    'Menos de tres años de primaria'
                      WHEN   escol = 3       THEN    'De tres a cinco años de primaria'
                      WHEN   escol = 4       THEN    'Primaria completa'
                      WHEN   escol = 5       THEN    'Secundaria o equivalente'
                      WHEN   escol = 6       THEN    'Preparatoria o equivalente'
                      WHEN   escol = 7       THEN    'Profesional'
                      WHEN   escol = 8       THEN    'NO APLICA A MENOR DE 6 AÑOS'
                      WHEN   escol = 0       THEN    'NO ESPECIFICADO'
                      END) AS ESCOLARIDAD,
( CASE WHEN    edocivil=      0       THEN    'SE IGNORA'
              WHEN    edocivil=      1       THEN    'SOLTERO'
              WHEN    edocivil=      2       THEN    'VIUDO'
              WHEN    edocivil=      3       THEN    'DIVORSIADO'
              WHEN    edocivil=      4       THEN    'UNION LIBRE'
              WHEN    edocivil=      5       THEN    'CASADO'
              WHEN    edocivil=      8       THEN    'MENOR DE 12 AÑOS'
              END) AS ESTADOCIVIL,
(case when edadvalor BETWEEN  0 and 9 then 'MUY ALTO'
                                          when edadvalor BETWEEN  10 and 44 then
'ALTO'
                                          when edadvalor BETWEEN  45 and 120 then
'BAJO' END) AS RIESGO
FROM DEF08
where violfam=1

union all

/*2009*/

SELECT edaduni ,edadvalor,
```

```
( CASE  WHEN   ocupacion=     02      THEN    'No trabaja'
       WHEN    ocupacion=     11      THEN    'Profesionistas'
       WHEN    ocupacion=     12      THEN    'Técnicos'
       WHEN    ocupacion=     13      THEN    'Trabajadores de la educación'
       WHEN    ocupacion=     14      THEN    'Trabajadores del arte, espectáculos y
deportes'
       WHEN    ocupacion=     21      THEN    'Funcionarios y directivos'
       WHEN    ocupacion=     41      THEN    'Trabajadores en actividades, agrícolas,
ganaderas, caza y pesca'
       WHEN    ocupacion=     51      THEN    'Personal de control producción industrial'
       WHEN    ocupacion=     52      THEN    'Trabajadores en la industria de la
transformación'
       WHEN    ocupacion=     53      THEN    'Operadores de maquinaria fija'
       WHEN    ocupacion=     54      THEN    'Ayudantes producción industrial y
artesanal'
       WHEN    ocupacion=     55      THEN    'Conductores de maquinaria móvil y medios de
transporte'
       WHEN    ocupacion=     61      THEN    'Trabajadores administrativos de nivel
intermedio'
       WHEN    ocupacion=     62      THEN    'Trabajadores administrativos de nivel
inferior'
       WHEN    ocupacion=     71      THEN    'Comerciantes, empleados de comercio,
agentes de ventas'
       WHEN    ocupacion=     72      THEN    'Vendedores ambulantes'
       WHEN    ocupacion=     81      THEN    'Trabajadores servicios personales en
establecimientos'
       WHEN    ocupacion=     82      THEN    'Trabajadores de servicio doméstico'
       WHEN    ocupacion=     83      THEN    'Trabajadores de fuerzas armadas, protección
y vigilancia'
       WHEN    ocupacion=     98      THEN    'No aplica menore sde 12 años'
       WHEN    ocupacion=     99      THEN    'Ocupación no clasificada anteriormente'
       WHEN    ocupacion=     00      THEN    'No especificado'     END) AS OCUPACION,
(CASE   WHEN   escol= 1       THEN    'Sin escolaridad'
                      WHEN   escol = 2       THEN    'Menos de tres años de primaria'
                      WHEN   escol = 3       THEN    'De tres a cinco años de primaria'
                      WHEN   escol = 4       THEN    'Primaria completa'
                      WHEN   escol = 5       THEN    'Secundaria o equivalente'
                      WHEN   escol = 6       THEN    'Preparatoria o equivalente'
                      WHEN   escol = 7       THEN    'Profesional'
                      WHEN   escol = 8       THEN    'NO APLICA A MENOR DE 6 AÑOS'
                      WHEN   escol = 0       THEN    'NO ESPECIFICADO'
                      END) AS ESCOLARIDAD,
( CASE  WHEN    edocivil=      0       THEN    'SE IGNORA'
               WHEN    edocivil=      1       THEN    'SOLTERO'
               WHEN    edocivil=      2       THEN    'VIUDO'
               WHEN    edocivil=      3       THEN    'DIVORSIADO'
               WHEN    edocivil=      4       THEN    'UNION LIBRE'
               WHEN    edocivil=      5       THEN    'CASADO'
               WHEN    edocivil=      8       THEN    'MENOR DE 12 AÑOS'
               END) AS ESTADOCIVIL,
(case when edadvalor BETWEEN  0 and 9 then 'MUY ALTO'
                                          when edadvalor BETWEEN  10 and 44 then
'ALTO'
                                          when edadvalor BETWEEN  45 and 120 then
'BAJO' END) AS RIESGO
FROM dbo.def_09
where violfam=1
```

MUJERES EMBARAZADAS 4.5

```
select edad as EDAD, count(edad)as NODEMUJERES from dbo.def2000
WHERE CAUSA LIKE 'O%' AND EDAD <> 998
AND edad<19
group by edad
UNION ALL
select edad as EDAD, count(edad)as NODEMUJERES from dbo.DEF2001
WHERE CAUSA LIKE 'O%' AND EDAD <> 998
AND edad<19
group by edad
--ORDER BY count(edad) DESC
UNION ALL
select edad as EDAD, count(edad)as NODEMUJERES from dbo.def02
WHERE CAUSA LIKE 'O%' AND EDAD <> 998
AND edad<19
```

```
group by edad
UNION ALL
select edad as EDAD, count(edad)as NODEMUJERES from dbo.DEF03
WHERE CAUSA LIKE 'O%' AND EDAD <> 998
AND edad<19
group by edad
--ORDER BY count(edad) DESC
UNION ALL
select edad as EDAD, count(edad)as NODEMUJERES from dbo.def04
WHERE CAUSAB LIKE 'O%' AND EDAD <> 998
AND edad<19
group by edad
UNION ALL
select edad as EDAD, count(edad)as NODEMUJERES from dbo.DEF05
WHERE CAUSAB LIKE 'O%' AND EDAD <> 998
AND edad<19
group by edad
--ORDER BY count(edad) DESC
UNION ALL
select edadVALOR as EDADVALOR, count(edadVALOR)as NODEMUJERES from dbo.DEF06
WHERE CAUSADEF LIKE 'O%' AND EDADVALOR <> 998
AND edadVALOR<19
group by edadVALOR
--ORDER BY count(edadVALOR) DESC
UNION ALL
select edadVALOR as EDADVALOR, count(edadVALOR)as NODEMUJERES from dbo.def07
WHERE CAUSADEF LIKE 'O%' AND EDADVALOR <> 998
AND edadVALOR<19
group by edadVALOR
---ORDER BY count(edadVALOR),count(edadVALOR) DESC
UNION ALL
select edadVALOR as EDADVALOR, count(edadVALOR)as NODEMUJERES from dbo.def08
WHERE CAUSADEF LIKE 'O%' AND EDADVALOR <> 998
AND edadVALOR<19
group by edadVALOR
UNION ALL
select edadVALOR as EDADVALOR, count(edadVALOR)as NODEMUJERES from dbo.def_09
WHERE CAUSADEF LIKE 'O%' AND EDADVALOR <> 998
AND edadVALOR<19
group by edadVALOR
```

/*Mayores a 35*/

```
select edad as EDAD, count(edad)as CONTEOPOREDAD from dbo.def2000
WHERE CAUSA LIKE 'O%' AND EDAD <> 998
AND edad>35
group by edad
UNION ALL
select edad as EDAD, count(edad)as CONTEOPOREDAD from dbo.DEF2001
WHERE CAUSA LIKE 'O%' AND EDAD <> 998
AND edad>35
group by edad
--ORDER BY count(edad) DESC
UNION ALL
select edad as EDAD, count(edad)as CONTEOPOREDAD from dbo.def02
WHERE CAUSA LIKE 'O%' AND EDAD <> 998
AND edad>35
group by edad
UNION ALL
select edad as EDAD, count(edad)as CONTEOPOREDAD from dbo.DEF03
WHERE CAUSA LIKE 'O%' AND EDAD <> 998
AND edad>35
group by edad
--ORDER BY count(edad) DESC
UNION ALL
select edad as EDAD, count(edad)as CONTEOPOREDAD from dbo.def04
WHERE CAUSAB LIKE 'O%' AND EDAD <> 998
AND edad>35
group by edad
UNION ALL
select edad as EDAD, count(edad)as CONTEOPOREDAD from dbo.DEF05
WHERE CAUSAB LIKE 'O%' AND EDAD <> 998
AND edad>35
group by edad
```

```
--ORDER BY count(edad) DESC
UNION ALL
select edadVALOR as EDADVALOR, count(edadVALOR)as CONTEOPOREDAD from dbo.DEF06
WHERE CAUSADEF LIKE 'O%' AND EDADVALOR <> 998
AND edadVALOR>35
group by edadVALOR
--ORDER BY count(edadVALOR) DESC
UNION ALL
select edadVALOR as EDADVALOR, count(edadVALOR)as CONTEOPOREDAD from dbo.def07
WHERE CAUSADEF LIKE 'O%' AND EDADVALOR <> 998
AND edadVALOR>35
group by edadVALOR
---ORDER BY count(edadVALOR),count(edadVALOR) DESC
UNION ALL
select edadVALOR as EDADVALOR, count(edadVALOR)as CONTEOPOREDAD from dbo.def08
WHERE CAUSADEF LIKE 'O%' AND EDADVALOR <> 998
AND edadVALOR>35
group by edadVALOR
UNION ALL
select edadVALOR as EDADVALOR, count(edadVALOR)as CONTEOPOREDAD from dbo.def_09
WHERE CAUSADEF LIKE 'O%' AND EDADVALOR <> 998
AND edadVALOR>35
group by edadVALOR
--ORDER BY count(edadVALOR)
```

DICCIONARIO DE DATOS

No	Etiqueta	Definición
1	Control	Es un identificador único, es del tipo de dato String, y tiene una longitud de 11.
2	Entidad de registro	Entidad federativa de registro donde se inscribe el hecho vital y el tipo de dato que se maneja es String, con una longitud de 2.
3	Municipio de registro	Municipio o delegación donde se inscribe el hecho vital y el tipo de dato que se maneja es String, con una longitud de 3.
4	Localidad de registro	Todo lugar ocupado con una o más viviendas, las cuales pueden estar o no habitadas, y que es conocido con un nombre dado por la ley o la costumbre, y en donde se inscribe el hecho vital, es del tipo de dato String, y tiene una longitud de 4.
5	Entidad de residencia	Es la entidad federativa de residencia donde la persona tiene su domicilio particular, principal o permanente y es del tipo de dato String y su longitud es de 2.
6	Municipio de residencia	Es el municipio o delegación de residencia donde la persona tiene su domicilio particular, principal o permanente y es del tipo de dato String y su longitud es de 3.
7	Localidad de residencia	Todo lugar ocupado con una o más viviendas, las cuales pueden estar o no habitadas, y que es conocido con un nombre dado por la ley o la costumbre, y en donde la persona tiene su domicilio particular, principal o permanente, es del tipo de dato String, y tiene una longitud de 4.
8	Tamaño de localidad residencia	Clasificación de las localidades de acuerdo al número al número de las personas que las habitan y donde la persona tiene su domicilio particular. Principal o permanente y es del tipo de dato Numeric y su longitud es de 2.
9	Entidad de defunción	Es la entidad federativa donde ocurrió el hecho vital ya sea nacimiento, parto, lesión, ceremonia o trámite, es del tipo de dato String y su longitud es de 2.
10	Municipio de defunción	Es el municipio o delegación donde ocurrió el hecho vital ya sea nacimiento, parto, lesión, ceremonia o trámite, es del tipo de dato String y su longitud es de 3.
11	Localidad de defunción	Todo lugar ocupado con una o más viviendas, las cuales pueden estar o no habitadas, y que es conocido con un nombre dado por la ley o la costumbre, y en donde ocurrió el hecho vital ya sea nacimiento, parto, lesión, ceremonia o tramite, es del tipo de dato String, y tiene una longitud de 4.

No	Etiqueta	Definición
12	Tamaño de localidad de defunción	Clasificación de las localidades de acuerdo al número de personas que habitan y donde ocurrió el hecho vital o el acontecimiento, ya sea nacimiento, parto, lesión, ceremonia o trámite, es del tipo de dato Numeric y su longitud es de 2.
13	Causa	Especifica la causa de la defunción de acuerdo a la Clasificación Internacional de Enfermedades (CIE-10 aplica para los años 1998 en adelante), es del tipo de dato String y su longitud es de 4.
14	Sexo	Condición biológica que distingue a las personas en hombres y mujeres, es del tipo de dato Numeric y su longitud es de 1.
15	Clave de edad	Es un identificador, donde se usan la letras para asignarle un rango de edades a cada una de edades, es del tipo de dato String, y tiene una longitud de 1.
16	Edad	Tiempo transcurrido entre la fecha de nacimiento de la persona y la del momento en que ocurre o se registra el hecho vital, y el tipo de dato que maneja es un Numeric, con una longitud de 3.
17	Año de defunción	Año en que ocurrió el hecho vital, es del tipo de dato String y tiene una longitud de 4.
18	Mes de defunción	Mes en el que ocurrió el hecho vital, es del tipo de dato String ycon una longitud de 2. .
19	Día de defunción	Día en el que ocurrió el hecho vital, es del tipo de dato String y con una longitud de 2.
20	Año de registro	Año en el que se inscribe el hecho vital en la institución correspondiente, según su competencia, es del tipo de datos String y con una longitud de 4.
21	Mes de registro	Mes en que se inscribe el hecho vital en la institución correspondiente, según su competencia, es del tipo de dato String y con una longitud de 2.
22	Día de registro	Día en que se inscribe el hecho vital en la institución correspondiente, según su competencia, es del tipo de dato String y con longitud de 2.
23	Año de nacimiento	Es el año en el que nació la persona, es del tipo de dato String y tiene una longitud de 4.
24	Mes de nacimiento	Es el mes en el que nació la persona, es del tipo de dato String, con una longitud de 2.
25	Día de nacimiento	Es el día en el que nació la persona, es del tipo de dato String, con una longitud de 2.
26	Ocupación	Realización de una actividad económica, ya sea de manera independiente o subordinada, en el catalogo de 1990 (utilizado en 1992 a la fecha) se dividen en dos clasificaciones: “Trabajadores administrativos de nivel intermedio” y “Trabajadores administrativos de nivel inferior”. Por lo anterior se clasificaron en los “Trabajadores administrativos de nivel intermedio”, es del tipo de dato String y con una longitud de 2
27	Escolaridad	Último grado aprobado en el ciclo de instrucción alcanzando que declare haber cursado la persona en el Sistema Educativo Nacional o su equivalente en el caso de estudios en el extranjero al momento de registrar el hecho vital, es del tipo de dato String y con una longitud de 1.
28	Estado civil	Situación de las personas en relación con los derechos y obligaciones legales y de costumbre del país, respecto de la unión o matrimonio, incluye por lo tanto, las condiciones de hechos y derechos, es del tipo de dato String y con una longitud de 1 .
29	Presunto	Especifica la presunción de accidente, homicidio, suicidio u operaciones legales y de guerra. Hasta 2003 las operaciones legales y de guerra se incluían en homicidios. Apartir del 2004 se incluyen en homicidios las secuelas de agresiones y en suicidios las secuelas se lesiones autoinflingidas, es del tipo de dato que se maneja es el String y tiene una longitud

No	Etiqueta	Definición
		de 1.
30	Ocurrió desempeño trabajo	Especifica si la persona falleció durante el desarrollo de alguna actividad con el desempeño de su trabajo, es del tipo de dato String y con una longitud de 1.
31	Lugar donde ocurrió la lesión	Espacio físico donde tuvo lugar el hecho vital, ya sea nacimiento, parto, lesión, ceremonia o trámite, es del tipo de dato String y con una longitud de 1.
32	Necropsia	Especifica si se realizó el examen del cadáver que incluye el de órganos y estructuras internas, después de la disección para precisar la causa de la muerte o el carácter de cambios patológicos. Nota: Hasta 2003 la variable “Necropsia” se captaba exclusivamente para las muertes accidentales y violentas, a partir de 2004 se capta para defunciones generales, es del tipo de dato String y con una longitud de 1.
33	Atención médica	Situación que distingue a la persona, según haya recibido o no asistencia médica antes de la muerte, es del tipo de dato String y tiene una longitud de 1.
34	Sitio ocurrencia	Espacio físico donde se realizo la sistencia medica y si es el caso en donde ocurrió la defunción, es del tipo de dato String y con una longitud de 1.
35	Certificada por	Individuo autorizado por la ley que expide el certificado con los hechos relacionados con el suceso y las características del difunto, es del tipo de dato String y con una longitud de 1.
36	Certificante no médico	Es el tipo de certificado médico o no médico que se expidio con los hechos relacionados con el suceso, es del tipo de dato String, con una longitud de 1.
37	Nacionalidad	Condición legal particular que adquieren las personas por nacimiento o naturalización en una nación determinada, que permite clasificar a los habitantes de un país, en ciudades nacionales o extranjeros, es del tipo de dato String y tiene una longitud de 1.
38	Derechohabiencia	Organismo o establecimiento médico, público o privado al cual encuentra afiliada la persona, es del tipo de dato, y tiene una longitud de 2 .
39	Condición de embarazo	Muerte de una mujer mientras está embarazada o dentro de los 42 días siguientes a la terminación del embarazo, independientemente de la duración y el sitio del embarazo, debida a cualquier causa relacionada con o agravada por el embarazo mismo o su atención, pero no por causas accidentales o incidentales, es del tipo de dato Numeric, y tiene una longitud de 1.
40	Relación con el embarazo	Cualquier causa relacionada con o agravada por el embarazo mismo o su atención, pero no por causas accidentales o incidentales, es decir que no tuvieron relación las causas, es del tipo de dato Numeric, y tiene una longitud de 1.
41	Hora de la defunción	Especifica el tiempo (en hora) en que ocurrió la defunción, es del tipo de dato String, y tiene una longitud de 2.
42	Minuto de la defunción	Especifica el tiempo (en minutos) en que ocurrió la defunción, es del tipo de dato String, y tiene una longitud de 2..
43	Violencia	Especifica si existió violencia solo en caso de homicidios, es del tipo de dato Numeric, y tiene una longitud de 1.
44	Categoría de Lista mexica Mexicana	La lista general de enfermedades que ocasionaron la defunción (aplica para los años de 1998 en adelante), es del tipo de dato Numeric, y tiene una longitud de 4.

No	Etiqueta	Definición
45	Capítulo de Lista mexica Mexicana	Especifica las causas de defunción de acuerdo a la Lista Mexicana y Lista Básica de Enfermedades, su información identifica la enfermedad o lesión que inició la cadena de acontecimientos patológicos que condujeron directamente a la muerte o las circunstancias del accidente o violencia que produjo la lesión fatal, es del tipo de dato Numeric, y con una longitud de 4.
46	Lista GBD	Es la lista que mide la carga de las enfermedades, sobre dicha población, The Global Burden diseeases and Injury (GBD), es del tipo de datoNumeric, y tiene una longitud de 4.
47	Peso	Es el peso con que murió la persona que sufrió el hecho vital
48	Violencia Familiar	Especifica si existió violencia familiar solo en caso de homicidios, es del tipo de dato Numeric, y tiene una longitud de 1.
49	Lugar donde ocurrió la defunción	Espacio físico donde se realizo la sistencia medica y si es el caso en donde ocurrió la defunción, es del tipo de dato String y con una longitud de 1.
50	Año de Certificación	Es el año en elque se expide el certificado con los hechos relacionados, con el suceso y las características del difunto, es del tipo de dato String y tiene una longitud de 4.
51	Mes de certificación	Es el mes en elque se expide el certificado con los hechos relacionados, con el suceso y las características del difunto, es del tipo de dato String y tiene una longitud de 4.
52	Día de certificación	Es el día en elque se expide el certificado con los hechos relacionados, con el suceso y las características del difunto, es del tipo de dato String y tiene una longitud de 4.
53	Condición de embarazo (Condicio)	Muerte de una mujer mientras está embarazada o dentro de los 42 días siguientes a la terminación del embarazo, independientemente de la duración y el sitio del embarazo, debida a cualquier causa relacionada con o agravada por el embarazo mismo o su atención, pero no por causas accidentales o incidentales, es del tipo de dato Numeric, y tiene una longitud de 1..
54	Fueron complicaciones (Corrobora)	Cualquier causa relacionada con o agravada por el embarazo mismo o su atención, pero no por causas accidentales o incidentales, es decir que no tuvieron relación las causas, es del tipo de dato Numeric, y tiene una longitud de 1.
55	complicaron el embarazo (Corrobora)	Causa relacionada con o agravada por el embarazo mismo o su atención, pero no por causas accidentales o incidentales, que complicaron el embarazo en el caso de que existieran, es del tipo de dato Numeric, y tiene una longitud de 1.
56	Desdoblamiento de la causa básica	FALTA
57	La causa básica	Causa primaria por la cual sucedió el hecho vital

Etiqueta	2000 2001	2002 2003	2004 2005	2006, 2007, 2008,2009
Control	control	control		
Entidad de registro		ent_reg	entreg	entregf
Municipio de registro		mun_reg	munreg	munregf
Localidad de registro		loc_reg		
Entidad de residencia	entres	entres	entres	entrh
Municipio de residencia	mpores	mpores	mpores	entrh
Localidad de residencia		loc_res	locres	locrh
Tamaño de localidad residencia	tloc	tloc	tamloc_res	tamlocrh
Entidad de defunción	entdef	entdef	entdef	entocu

Etiqueta	2000 2001	2002 2003	2004 2005	2006, 2007, 2008,2009
Municipio de defunción	mpodef	mpodef	mpodef	munocu
Localidad de defunción		loc_def	locdef	lococu
Tamaño de localidad de defunción		t_loc_def	tamloc_def	tamlococ
Causa	causa	causa	causab	causadef
Sexo	sexo	sexo	sexo	sexo
Clave de edad	cveedad	cveedad	cveedad	edaduni
Edad	edad	edad	edad	edadvalor
Año de defunción	anodef	anodef	anodef	aniodef
Mes de defunción	mesdef	mesdef	mesdef	mesdef
Día de defunción	diadef	diadef	diadef	diadef
Año de registro	anoreg	anoreg	anoreg	anioreg
Mes de registro	mesreg	mesreg	mesreg	mesreg
Día de registro	diareg	diareg	diareg	diareg
Año de nacimiento	anonac	anonac	anonac	anionac
Mes de nacimiento	mesnac	mesnac	mesnac	mesnac
Día de nacimiento	dianac	dianac	dianac	dianac
Ocupación	ocupa	ocupa	ocupa	ocupacion
Escolaridad	esco	esco	esco	escol
Estado civil	edociv	edociv	edocony	edocivil
Presunto	presunto	presunto	presunto	presunto
Ocurrió desempeño trabajo	traba	traba	trabajo	trabajo
Lugar donde ocurrió la lesión	lugles	lugles	luglesion	lugles
Necropsia	necrop	necrop	necrop	necropcia
Atención médica	asismed	asismed	asismed	asist
Sitio ocurrencia	sitioo	sitioo		
Certificada por	certif	certif	certifican	certif
Certificante no médico	tipocert	tipocert		
Nacionalidad	nacion	nacion	nacion	nacion
Derechohabiencia	derech	derech	derech	derechab
Condición de embarazo	conemb	conemb		
Relación con el embarazo	relemb	relemb		
Hora de la defunción	horadef	horadef	horadef	horadef
Minuto de la defunción	mindef	mindef	mindef	mindef
Violencia	violenc	violenc		
Categoría de Lista mexica Mexicana		lm2002		
Capítulo de Lista mexica Mexicana		cap_lm2002		
Lista GBD	gbd	gbd		
Peso			peso	peso
Violencia Familiar			violfam	violfam

Etiqueta	2000 2001	2002 2003	2004 2005	2006, 2007, 2008,2009
Lugar donde ocurrió la defunción			sitioo	sitio_les
Año de Certificación			anocert	aniocert
Mes de certificación			mescert	mescert
Día de certificación			diacert	diacert
Condición de embarazo (Condicio)			condemb	condemba
Fueron complicaciones (Corrobora)			compemb	rel_emba
complicaron el embarazo (Corrobora)			com_emb	complicaro
Desdoblamiento de la causa basica			causad	desdobla
La causa basica				

CAUSAS

CODIGO	DESCRIPCIÓN
O00	Embarazo ectópico
O000	Embarazo abdominal
O001	Embarazo tubárico
O002	Embarazo ovárico
O008	Otros embarazos ectópicos
O009	Embarazo ectópico, no especificado
O01	Mola hidatiforme
O010	Mola hidatiforme clásica
O011	Mola hidatiforme, incompleta o parcial
O019	Mola hidatiforme, no especificada
O02	Otros productos anormales de la concepción
O020	Detención del desarrollo del huevo y mola no hidatiforme
O021	Aborto retenido
O028	Otros productos anormales especificados de la concepción
O029	Producto anormal de la concepción, no especificado
O03	Aborto espontáneo
O030	Aborto espontáneo incompleto, complicado con infección genital y pelviana
O031	Aborto espontáneo incompleto, complicado por hemorragia excesiva o tardía
O032	Aborto espontáneo incompleto, complicado por embolia
O033	Aborto espontáneo incompleto, con otras complicaciones especificadas y las no especificadas
O034	Aborto espontáneo incompleto, sin complicación
O035	Aborto espontáneo completo o no especificado, complicado con infección genital y pelviana
O036	Aborto espontáneo completo o no especificado, complicado por hemorragia excesiva o tardía
O037	Aborto espontáneo completo o no especificado, complicado por embolia
O038	Aborto espontáneo completo o no especificado, con otras complicaciones especificadas y las no especificadas
O039	Aborto espontáneo completo o no especificado, sin complicación
O04	Aborto médico

CODIGO	DESCRIPCIÓN
O040	Aborto médico incompleto, complicado con infección genital y pelviana
O041	Aborto médico incompleto, complicado por hemorragia excesiva o tardía
O042	Aborto médico incompleto, complicado por embolia
O043	Aborto médico incompleto, con otras complicaciones especificadas y las no especificadas
O044	Aborto médico incompleto, sin complicación
O045	Aborto médico completo o no especificado, complicado con infección genital y pelviana
O046	Aborto médico completo o no especificado, complicado por hemorragia excesiva o tardía
O047	Aborto médico completo o no especificado, complicado por embolia
O048	Aborto médico completo o no especificado, con otras complicaciones especificadas y las no especificadas
O049	Aborto médico completo o no especificado, sin complicación
O05	Otro aborto
O050	Otro aborto incompleto, complicado con infección genital y pelviana
O051	Otro aborto incompleto, complicado por hemorragia excesiva o tardía
O052	Otro aborto incompleto, complicado por embolia
O053	Otro aborto incompleto, con otras complicaciones especificadas y las no especificadas
O054	Otro aborto incompleto, sin complicación
O055	Otro aborto completo o no especificado, complicado con infección genital y pelviana
O056	Otro aborto completo o no especificado, complicado por hemorragia excesiva o tardía
O057	Otro aborto completo o no especificado, complicado por embolia
O058	Otro aborto completo o no especificado, con otras complicaciones especificadas y las no especificadas
O059	Otro aborto completo o no especificado, sin complicación
O06	Aborto no especificado
O060	Aborto no especificado incompleto, complicado con infección genital y pelviana
O061	Aborto no especificado incompleto, complicado por hemorragia excesiva o tardía
O062	Aborto no especificado incompleto, complicado por embolia
O063	Aborto no especificado incompleto, con otras complicaciones especificadas y las no especificadas
O064	Aborto no especificado incompleto, sin complicación
O065	Aborto no especificado completo o no especificado, complicado con infección genital y pelviana
O066	Aborto no especificado completo o no especificado, complicado por hemorragia excesiva o tardía
O067	Aborto no especificado completo o no especificado, complicado por embolia
O068	Aborto no especificado completo o no especificado, con otras complicaciones especificadas y las no especificadas
O069	Aborto no especificado completo o no especificado, sin complicación
O07	Intento fallido de aborto
O070	Falla de la inducción médica del aborto, complicado por infección genital y pelviana
O071	Falla de la inducción médica del aborto, complicado por hemorragia excesiva o tardía
O072	Falla de la inducción médica del aborto, complicado por embolia
O073	Falla de la inducción médica del aborto, con otras complicaciones y las no especificadas
O074	Falla de la inducción médica del aborto, sin complicación
O075	Otros intentos fallidos de aborto y los no especificados, complicados por infección genital y pelviana
O076	Otros intentos fallidos de aborto y los no especificados, complicados por hemorragia excesiva o tardía
O077	Otros intentos fallidos de aborto y los no especificados, complicados por embolia
O078	Otros intentos fallidos de aborto y los no especificados, con otras complicaciones y las no especificadas
O079	Otros intentos fallidos de aborto y los no especificados, sin complicación
O08	Complicaciones consecutivas al aborto, al embarazo ectópico y al embarazo molar
O080	Infección genital y pelviana consecutiva al aborto, al embarazo ectópico y al embarazo molar
O081	Hemorragia excesiva o tardía consecutiva al aborto, al embarazo ectópico y al embarazo molar
O082	Embolia consecutiva al aborto, al embarazo ectópico y al embarazo molar
O083	Choque consecutivo al aborto, al embarazo ectópico y al embarazo molar

CODIGO	DESCRIPCIÓN
O084	Insuficiencia renal consecutiva al aborto, al embarazo ectópico y al embarazo molar
O085	Trastorno metabólico consecutivo al aborto, al embarazo ectópico y al embarazo molar
O086	Lesión de órganos o tejidos de la pelvis consecutivo al aborto, al embarazo ectópico y al embarazo molar
O087	Otras complicaciones venosas consecutivas al aborto, al embarazo ectópico y al embarazo molar
O088	Otras complicaciones consecutivas al aborto, al embarazo ectópico y al embarazo molar
O089	Complicación no especificada consecutiva al aborto, al embarazo ectópico y al embarazo molar
O10	Hipertensión preexistente que complica el embarazo, el parto y el puerperio
O100	Hipertensión esencial preexistente que complica el embarazo, el parto y el puerperio
O101	Enfermedad cardíaca hipertensiva preexistente que complica el embarazo, el parto y el puerperio
O102	Enfermedad renal hipertensiva preexistente que complica el embarazo, el parto y el puerperio
O103	Enfermedad cardio-renal hipertensiva preexistente que complica el embarazo, el parto y el puerperio
O104	Hipertensión secundaria preexistente que complica el embarazo, el parto y el puerperio
O109	Hipertensión preexistente no especificada, que complica el embarazo, el parto y el puerperio
O11X	Trastornos hipertensivos preexistentes, con proteinuria agregada
O12	Edema y proteinuria gestacionales [inducidos por el embarazo] sin hipertensión
O120	Edema gestacional
O121	Proteinuria gestacional
O122	Edema gestacional con proteinuria
O13X	Hipertensión gestacional [inducida por el embarazo] sin proteinuria significativa
O14	Hipertensión gestacional [inducida por el embarazo] con proteinuria significativa
O140	Preeclampsia moderada
O141	Preeclampsia severa
O149	Preeclampsia, no especificada
O15	Eclampsia
O150	Eclampsia en el embarazo
O151	Eclampsia durante el trabajo de parto
O152	Eclampsia en el puerperio
O159	Eclampsia, en período no especificado
O16X	Hipertensión materna, no especificada
O20	Hemorragia precoz del embarazo
O200	Amenaza de aborto
O208	Otras hemorragias precoces del embarazo
O209	Hemorragia precoz del embarazo, sin otra especificación
O21	Vómitos excesivos del embarazo
O210	Hiperemesis gravídica leve
O211	Hiperemesis gravídica con trastornos metabólicos
O212	Hiperemesis gravídica tardía
O218	Otros vómitos que complican el embarazo
O219	Vómitos del embarazo, no especificados
O22	Complicaciones venosas del embarazo
O220	Venas varicosas de los miembros inferiores en el embarazo
O221	Várices genitales en el embarazo
O222	Tromboflebitis superficial en el embarazo
O223	Flebotrombosis profunda en el embarazo
O224	Hemorroides en el embarazo
O225	Trombosis venosa cerebral en el embarazo
O228	Otras complicaciones venosas en el embarazo
O229	Complicación venosa no especificada en el embarazo
O23	Infección de las vías genitourinarias en el embarazo
O230	Infección del riñón en el embarazo

CODIGO	DESCRIPCIÓN
O231	Infección de la vejiga urinaria en el embarazo
O232	Infección de la uretra en el embarazo
O233	Infección de otras partes de las vías urinarias en el embarazo
O234	Infección no especificada de las vías urinarias en el embarazo
O235	Infección genital en el embarazo
O239	Otras Infecciones y las no especificadas de las vías genitourinarias en el embarazo
O24	Diabetes mellitus en el embarazo
O240	Diabetes mellitus preexistente insulinodependiente, en el embarazo
O241	Diabetes mellitus preexistente no insulinodependiente, en el embarazo
O242	Diabetes mellitus preexistente relacionada con desnutrición, en el embarazo
O243	Diabetes mellitus preexistente, sin otra especificación, en el embarazo
O244	Diabetes mellitus que se origina con el embarazo
O249	Diabetes mellitus no especificada, en el embarazo
O25X	Desnutrición en el embarazo
O26	Atención a la madre por otras complicaciones principalmente relacionadas con el embarazo
O260	Aumento excesivo de peso en el embarazo
O261	Aumento pequeño de peso en el embarazo
O262	Atención del embarazo en una abortadora habitual
O263	Retención de dispositivo anticonceptivo intrauterino en el embarazo
O264	Herpes gestacional
O265	Síndrome de hipotensión materna
O266	Trastornos del hígado en el embarazo, el parto y el puerperio
O267	Subluxación de la sínfisis (del pubis) en el embarazo, el parto y el puerperio
O268	Otras complicaciones especificadas relacionadas con el embarazo
O269	Complicación relacionada con el embarazo, no especificada
O28	Hallazgos anormales en el examen prenatal de la madre
O280	Hallazgo hematológico anormal en el examen prenatal de la madre
O281	Hallazgo bioquímico anormal en el examen prenatal de la madre
O282	Hallazgo citológico anormal en el examen prenatal de la madre
O283	Hallazgo ultrasónico anormal en el examen prenatal de la madre
O284	Hallazgo radiológico anormal en el examen prenatal de la madre
O285	Hallazgo cromosómico o genético anormal en el examen prenatal de la madre
O288	Otros hallazgos anormales en el examen prenatal de la madre
O289	Hallazgo anormal no especificado en el examen prenatal de la madre
O29	Complicaciones de la anestesia administrada durante el embarazo
O290	Complicaciones pulmonares de la anestesia administrada durante el embarazo
O291	Complicaciones cardíacas de la anestesia administrada durante el embarazo
O292	Complicaciones del sistema nervioso central debidas a la anestesia administrada durante el embarazo
O293	Reacción tóxica a la anestesia local administrada durante el embarazo
O294	Cefalalgia inducida por la anestesia espinal o epidural administradas durante el embarazo
O295	Otras complicaciones de la anestesia espinal o epidural administradas durante el embarazo
O296	Falla o dificultad en la intubación durante el embarazo
O298	Otras complicaciones de la anestesia administrada durante el embarazo
O299	Complicación no especificada de la anestesia administrada durante el embarazo
O30	Embarazo múltiple
O300	Embarazo doble
O301	Embarazo triple
O302	Embarazo cuádruple
O308	Otros embarazos múltiples
O309	Embarazo múltiple, no especificado

CODIGO	DESCRIPCIÓN
O31	Complicaciones específicas del embarazo múltiple
O310	Feto papiráceo
O311	Embarazo que continúa después del aborto de un feto o más
O312	Embarazo que continúa después de la muerte intrauterina de un feto o más
O318	Otras complicaciones específicas del embarazo múltiple
O32	Atención materna por presentación anormal del feto, conocida o presunta
O320	Atención materna por posición fetal inestable
O321	Atención materna por presentación de nalgas
O322	Atención materna por posición fetal oblicua o transversa
O323	Atención materna por presentación de cara, de frente o de mentón
O324	Atención materna por cabeza alta en gestación a término
O325	Atención materna por embarazo múltiple con presentación anormal de un feto o más
O326	Atención materna por presentación compuesta
O328	Atención materna por otras presentaciones anormales del feto
O329	Atención materna por presentación anormal no especificada del feto
O33	Atención materna por desproporción conocida o presunta
O330	Atención materna por desproporción debida a deformidad de la pelvis ósea en la madre
O331	Atención materna por desproporción debida a estrechez general de la pelvis
O332	Atención materna por desproporción debida a disminución del estrecho superior de la pelvis
O333	Atención materna por desproporción debida a disminución del estrecho inferior de la pelvis
O334	Atención materna por desproporción fetopelviana de origen mixto, materno y fetal
O335	Atención materna por desproporción debida a feto demasiado grande
O336	Atención materna por desproporción debida a feto hidrocefálico
O337	Atención materna por desproporción debida a otra deformidad fetal
O338	Atención materna por desproporción de otro origen
O339	Atención materna por desproporción de origen no especificado
O34	Atención materna por anormalidades conocidas o presuntas de los órganos pelvianos de la madre
O340	Atención materna por anomalía congénita del útero
O341	Atención materna por tumor del cuerpo del útero
O342	Atención materna por cicatriz uterina debida a cirugía previa
O343	Atención materna por incompetencia del cuello uterino
O344	Atención materna por otra anormalidad del cuello uterino
O345	Atención materna por otras anormalidades del útero grávido
O346	Atención materna por anormalidad de la vagina
O347	Atención materna por anormalidad de la vulva y del perineo
O348	Atención materna por otras anormalidades de los órganos pelvianos
O349	Atención materna por anormalidad no especificada de órgano pelviano
O35	Atención materna por anormalidad o lesión fetal, conocida o presunta
O350	Atención materna por (presunta) malformación del sistema nervioso central en el feto
O351	Atención materna por (presunta) anormalidad cromosómica en el feto
O352	Atención materna por (presunta) enfermedad hereditaria en el feto
O353	Atención materna por (presunta) lesión fetal debida a enfermedad vírica en la madre
O354	Atención materna por (presunta) lesión al feto debida al alcohol
O355	Atención materna por (presunta) lesión fetal debida a drogas
O356	Atención materna por (presunta) lesión al feto debida a radiación
O357	Atención materna por (presunta) lesión fetal debida a otros procedimientos médicos
O358	Atención materna por otras (presuntas) anormalidades y lesiones fetales
O359	Atención materna por (presunta) anormalidad y lesión fetal no especificada
O36	Atención materna por otros problemas fetales conocidos o presuntos
O360	Atención materna por isoinmunización rhesus

CODIGO	DESCRIPCIÓN
O361	Atención materna por otra isoinmunización
O362	Atención materna por hidropesía fetal
O363	Atención materna por signos de hipoxia fetal
O364	Atención materna por muerte intrauterina
O365	Atención materna por déficit del crecimiento fetal
O366	Atención materna por crecimiento fetal excesivo
O367	Atención materna por feto viable en embarazo abdominal
O368	Atención materna por otros problemas fetales especificados
O369	Atención materna por problemas fetales no especificados
O40X	Polihidramnios
O41	Otros trastornos del líquido amniótico y de las membranas
O410	Oligohidramnios
O411	Infección de la bolsa amniótica o de las membranas
O418	Otros trastornos especificados del líquido amniótico y de las membranas
O419	Trastorno del líquido amniótico y de las membranas, no especificado
O42	Ruptura prematura de las membranas
O420	Ruptura prematura de las membranas, e inicio del trabajo de parto dentro de las 24 horas
O421	Ruptura prematura de las membranas, e inicio del trabajo de parto después de las 24 horas
O422	Ruptura prematura de las membranas, trabajo de parto retrasado por la terapéutica
O429	Ruptura prematura de las membranas, sin otra especificación
O43	Trastornos placentarios
O430	Síndrome de transfusión placentaria
O431	Malformación de la placenta
O438	Otros trastornos placentarios
O439	Trastorno de la placenta, no especificado
O44	Placenta previa
O440	Placenta previa con especificación de que no hubo hemorragia
O441	Placenta previa con hemorragia
O45	Desprendimiento prematuro de la placenta [Abruptio placentae]
O450	Desprendimiento prematuro de la placenta con defecto de la coagulación
O458	Otros desprendimientos prematuros de la placenta
O459	Desprendimiento prematuro de la placenta, sin otra especificación
O46	Hemorragia anteparto, no clasificada en otra parte
O460	Hemorragia anteparto con defecto de la coagulación
O468	Otras hemorragias anteparto
O469	Hemorragia anteparto, no especificada
O47	Falso trabajo de parto
O470	Falso trabajo de parto antes de las 37 semanas completas de gestación
O471	Falso trabajo de parto a las 37 y más semanas completas de gestación
O479	Falso trabajo de parto, sin otra especificación
O48X	Embarazo prolongado
O60X	Parto prematuro
O61	Fracaso de la inducción del trabajo de parto
O610	Fracaso de la inducción médica del trabajo de parto
O611	Fracaso de la inducción instrumental del trabajo de parto
O618	Otros fracasos de la inducción del trabajo de parto
O619	Fracaso no especificado de la inducción del trabajo de parto
O62	Anormalidades de la dinámica del trabajo de parto
O620	Contracciones primarias inadecuadas
O621	Inercia uterina secundaria

CODIGO	DESCRIPCIÓN
O622	Otras inercias uterinas
O623	Trabajo de parto precipitado
O624	Contracciones uterinas hipertónicas, incoordinadas y prolongadas
O628	Otras anomalías dinámicas del trabajo de parto
O629	Anomalía dinámica del trabajo de parto, no especificada
O63	Trabajo de parto prolongado
O630	Prolongación del primer período (del trabajo de parto)
O631	Prolongación del segundo período (del trabajo de parto)
O632	Retraso de la expulsión del segundo gemelo, del tercero, etc.
O639	Trabajo de parto prolongado, no especificado
O64	Trabajo de parto obstruido debido a mala posición y presentación anormal del feto
O640	Trabajo de parto obstruido debido a rotación incompleta de la cabeza fetal
O641	Trabajo de parto obstruido debido a presentación de nalgas
O642	Trabajo de parto obstruido debido a presentación de cara
O643	Trabajo de parto obstruido debido a presentación de frente
O644	Trabajo de parto obstruido debido a presentación de hombro
O645	Trabajo de parto obstruido debido a presentación compuesta
O648	Trabajo de parto obstruido debido a otras presentaciones anormales del feto
O649	Trabajo de parto obstruido debido a presentación anormal del feto no especificada
O65	Trabajo de parto obstruido debido a anormalidad de la pelvis materna
O650	Trabajo de parto obstruido debido a deformidad de la pelvis
O651	Trabajo de parto obstruido debido a estrechez general de la pelvis
O652	Trabajo de parto obstruido debido a disminución del estrecho superior de la pelvis
O653	Trabajo de parto obstruido debido a disminución del estrecho inferior de la pelvis
O654	Trabajo de parto obstruido debido a desproporción fetopelviana, sin otra especificación
O655	Trabajo de parto obstruido debido a anomalías de los órganos pelvianos maternos
O658	Trabajo de parto obstruido debido a otras anomalías pelvianas maternas
O659	Trabajo de parto obstruido debido a anomalía pelviana no especificada
O66	Otras obstrucciones del trabajo de parto
O660	Trabajo de parto obstruido debido a distocia de hombros
O661	Trabajo de parto obstruido debido a distocia gemelar
O662	Trabajo de parto obstruido debido a distocia por feto inusualmente grande
O663	Trabajo de parto obstruido debido a otras anormalidades del feto
O664	Fracaso de la prueba del trabajo de parto, no especificada
O665	Fracaso no especificado de la aplicación de fórceps o de ventosa extractora
O668	Otras obstrucciones especificadas del trabajo de parto
O669	Trabajo de parto obstruido, sin otra especificación
O67	Trabajo de parto y parto complicados por hemorragia intraparto, no clasificados en otra parte
O670	Hemorragia intraparto con defectos de la coagulación
O678	Otras hemorragias intraparto
O679	Hemorragia intraparto, no especificada
O68	Trabajo de parto y parto complicados por sufrimiento fetal
O680	Trabajo de parto y parto complicados por anomalía de la frecuencia cardíaca fetal
O681	Trabajo de parto y parto complicados por la presencia de meconio en el líquido amniótico
O682	Trabajo de parto y parto complicados por anomalía de la frecuencia cardíaca fetal asociada con presencia de meconio en el líquido amniótico
O683	Trabajo de parto y parto complicados por evidencia bioquímica de sufrimiento fetal
O688	Trabajo de parto y parto complicados por otras evidencias de sufrimiento fetal
O689	Trabajo de parto y parto complicados por sufrimiento fetal, sin otra especificación
O69	Trabajo de parto y parto complicados por problemas del cordón umbilical

CODIGO	DESCRIPCIÓN
O690	Trabajo de parto y parto complicados por prolapso del cordón umbilical
O691	Trabajo de parto y parto complicados por circular pericervical del cordón, con compresión
O692	Trabajo de parto y parto complicados por otros enredos del cordón
O693	Trabajo de parto y parto complicados por cordón umbilical corto
O694	Trabajo de parto y parto complicados por vasa previa
O695	Trabajo de parto y parto complicados por lesión vascular del cordón
O698	Trabajo de parto y parto complicados por otros problemas del cordón umbilical
O699	Trabajo de parto y parto complicados por problemas no especificados del cordón umbilical
O70	Desgarro perineal durante el parto
O700	Desgarro perineal de primer grado durante el parto
O701	Desgarro perineal de segundo grado durante el parto
O702	Desgarro perineal de tercer grado durante el parto
O703	Desgarro perineal de cuarto grado durante el parto
O709	Desgarro perineal durante el parto, de grado no epecificado
O71	Otro trauma obstétrico
O710	Ruptura del útero antes del inicio del trabajo de parto
O711	Ruptura del útero durante el trabajo de parto
O712	Inversión del útero, postparto
O713	Desgarro obstétrico del cuello uterino
O714	Desgarro vaginal obstétrico alto, sólo
O715	Otros traumatismos obstétricos de los órganos pelvianos
O716	Traumatismo obstétrico de los ligamentos y articulaciones de la pelvis
O717	Hematoma obstétrico de la pelvis
O718	Otros traumas obstétricos especificados
O719	Trauma obstétrico, no especificado
O72	Hemorragia postparto
O720	Hemorragia del tercer período del parto
O721	Otras hemorragias postparto inmediatas
O722	Hemorragia postparto secundaria o tardía
O723	Defecto de la coagulación postparto
O73	Retención de la placenta o de las membranas, sin hemorragia
O730	Retención de la placenta sin hemorragia
O731	Retención de fragmentos de la placenta o de las membranas, sin hemorragia
O74	Complicaciones de la anestesia administrada durante el trabajo de parto y el parto
O740	Neumonitis por aspiración debida a la anestesia administrada durante el trabajo de parto y el parto
O741	Otras complicaciones pulmonares debidas a la anestesia administrada durante el trabajo de parto y el parto
O742	Complicaciones cardíacas de la anestesia administrada durante el trabajo de parto y el parto
O743	Complicaciones del sistema nervioso central por la anestesia administrada durante el trabajo de parto y el parto
O744	Reacción tóxica a la anestesia local administrada durante el trabajo de parto y el parto
O745	Cefalalgia inducida por la anestesia espinal o epidural administradas durante el trabajo de parto y el parto
O746	Otras complicaciones de la anestesia espinal o epidural administradas durante el trabajo de parto y el parto
O747	Falla o dificultad en la intubación durante el trabajo de parto y el parto
O748	Otras complicaciones de la anestesia administrada durante el trabajo de parto y el parto
O749	Complicación no especificada de la anestesia administrada durante el trabajo de parto y el parto
O75	Otras complicaciones del trabajo de parto y del parto, no clasificadas en otra parte
O750	Sufrimiento materno durante el trabajo de parto y el parto
O751	Choque durante o después del trabajo de parto y el parto
O752	Pirexia durante el trabajo de parto, no clasificada en otra parte

CODIGO	DESCRIPCIÓN
O753	Otras infecciones durante el trabajo de parto
O754	Otras complicaciones de la cirugía y otros procedimientos obstétricos
O755	Retraso del parto después de la ruptura artificial de las membranas
O756	Retraso del parto después de la ruptura espontánea o no especificada de las membranas
O757	Parto vaginal posterior a una cesárea previa
O758	Otras complicaciones especificadas del trabajo de parto y del parto
O759	Complicación no especificada del trabajo de parto y del parto
O80	Parto único espontáneo
O800	Parto único espontáneo, presentación cefálica de vértice
O801	Parto único espontáneo, presentación de nalgas o podálica
O808	Parto único espontáneo, otras presentaciones
O809	Parto único espontáneo, sin otra especificación
O81	Parto único con fórceps y ventosa extractora
O810	Parto con fórceps bajo
O811	Parto con fórceps medio
O812	Parto con fórceps medio con rotación
O813	Parto con fórceps de otros tipos y los no especificados
O814	Parto con ventosa extractora
O815	Parto con combinación de fórceps y ventosa extractora
O82	Parto único por cesárea
O820	Parto por cesárea electiva
O821	Parto por cesárea de emergencia
O822	Parto por cesárea con histerectomía
O828	Otros partos únicos por cesárea
O829	Parto por cesárea, sin otra especificación
O83	Otros partos únicos asistidos
O830	Extracción de nalgas
O831	Otros partos únicos asistidos, de nalgas
O832	Otros partos únicos con ayuda de manipulación obstétrica
O833	Parto de feto viable en embarazo abdominal
O834	Operación destructiva para facilitar el parto
O838	Otros partos únicos asistidos especificados
O839	Parto único asistido, sin otra especificación
O84	Parto múltiple
O840	Parto múltiple, todos espontáneos
O841	Parto múltiple, todos por fórceps y ventosa extractora
O842	Parto múltiple, todos por cesárea
O848	Otros partos múltiples
O849	Parto múltiple, no especificado
O85X	Sepsis puerperal
O86	Otras infecciones puerperales
O860	Infección de herida quirúrgica obstétrica
O861	Otras infecciones genitales consecutivas al parto
O862	Infección de las vías urinarias consecutiva al parto
O863	Otras infecciones de las vías genitourinarias consecutivas al parto
O864	Pirexia de origen desconocido consecutiva al parto
O868	Otras infecciones puerperales especificadas
O87	Complicaciones venosas en el puerperio
O870	Tromboflebitis superficial en el puerperio
O871	Flebotrombosis profunda en el puerperio

CODIGO	DESCRIPCIÓN
O872	Hemorroides en el puerperio
O873	Trombosis venosa cerebral en el puerperio
O878	Otras complicaciones venosas en el puerperio
O879	Complicación venosa en el puerperio, no especificada
O88	Embolia obstétrica
O880	Embolia gaseosa, obstétrica
O881	Embolia de líquido amniótico
O882	Embolia de coágulo sanguíneo, obstétrica
O883	Embolia séptica y piémica, obstétrica
O888	Otras embolias obstétricas
O89	Complicaciones de la anestesia administrada durante el puerperio
O890	Complicaciones pulmonares de la anestesia administrada durante el puerperio
O891	Complicaciones cardíacas de la anestesia administrada durante el puerperio
O892	Complicaciones del sistema nervioso central debidas a la anestesia administrada durante el puerperio
O893	Reacción tóxica a la anestesia local administrada durante el puerperio
O894	Cefalalgia inducida por la anestesia espinal o epidural administradas durante el puerperio
O895	Otras complicaciones de la anestesia espinal o epidural administradas durante el puerperio
O896	Falla o dificultad de intubación durante el puerperio
O898	Otras complicaciones de la anestesia administrada durante el puerperio
O899	Complicación no especificada de la anestesia administrada durante el puerperio
O90	Complicaciones del puerperio, no clasificadas en otra parte
O900	Dehiscencia de sutura de cesárea
O901	Dehiscencia de sutura obstétrica perineal
O902	Hematoma de herida quirúrgica obstétrica
O903	Cardiomiopatía en el puerperio
O904	Insuficiencia renal aguda postparto
O905	Tiroiditis postparto
O908	Otras complicaciones puerperales, no clasificadas en otra parte
O909	Complicación puerperal, no especificada
O91	Infecciones de la mama asociadas con el parto
O910	Infecciones del pezón asociados con el parto
O911	Absceso de la mama asociado con el parto
O912	Mastitis no purulenta asociada con el parto
O92	Otros trastornos de la mama y de la lactancia asociados con el parto
O920	Retracción del pezón asociada con el parto
O921	Fisuras del pezón asociadas con el parto
O922	Otros trastornos de la mama y los no especificados asociados con el parto
O923	Agalactia
O924	Hipogalactia
O925	Supresión de la lactancia
O926	Galactorrea
O927	Otros trastornos y los no especificados de la lactancia
O95X	Muerte obstétrica de causa no especificada
O96X	Muerte materna debida a cualquier causa obstétrica que ocurre después de 42 días pero antes de un año del parto
O97X	Muerte por secuelas de causas obstétricas directas
O98	Enfermedades maternas infecciosas y parasitarias clasificables en otra parte, pero que complican el embarazo, el parto y el puerperio
O980	Tuberculosis que complica el embarazo, el parto y el puerperio
O981	Sífilis que complica el embarazo, el parto y el puerperio

CODIGO	DESCRIPCIÓN
O982	Gonorrea que complica el embarazo, el parto y el puerperio
O983	Otras infecciones con un modo de transmisión predominantemente sexual que complican el embarazo, parto y puerperio
O984	Hepatitis viral que complica el embarazo, el parto y el puerperio
O985	Otras enfermedades virales que complican el embarazo, el parto y el puerperio
O986	Enfermedades causadas por protozoarios que complican el embarazo, el parto y el puerperio
O988	Otras enfermedades infecciosas y parasitarias maternas que complican el embarazo, el parto y el puerperio
O989	Enfermedad infecciosa y parasitaria materna no especificada que complica el embarazo, el parto y el puerperio
O99	Otras enfermedades maternas clasificables en otra parte, pero que complican el embarazo, el parto y el puerperio
O990	Anemia que complica el embarazo, el parto y el puerperio
O991	Otras enfermedades de la sangre y órganos hematopoyéticos y ciertos trastornos que afectan el sistema inmunitario cuando complican el embarazo, el parto y puerperio
O992	Enfermedades endocrinas, de la nutrición y del metabolismo que complican el embarazo, el parto y el puerperio
O993	Trastornos mentales y enfermedades del sistema nervioso que complican el embarazo, el parto y el puerperio
O994	Enfermedades del sistema circulatorio que complican el embarazo, el parto y el purperio
O995	Enfermedades del sistema respiratorio que complican el embarazo, el parto y el puerperio
O996	Enfermedades del sistema digestivo que complican el embarazo, el parto y el puerperio
O997	Enfermedades de la piel y del tejido subcutáneo que complican el embarazo, el parto y el puerperio
O998	Otras enfermedades especificadas y afecciones que complican el embarazo, el parto y el puerperio

TABLA DE CARACTERÍSTICAS DE LOS ATRIBUTOS

ATRIBUTO	TABLA	TIPO	# TOTAL	# NULOS	# DISTS	MEDIA	DESVIACIÓN ESTANDAR	MINIMO	MAXIMO	MODA
entdef	DEF2000	NUMERICO	437667	0	35	16.4077461	8.083719983	1	35	9

mpodef	DEF2000	NUMERICO	437667	0	566	50.6046789	66.22569344	0	570	39
causa	DEF2000	NOMINAL	437667	0	3703					I219
sexo	DEF2000	NUMERICO	437667	0	3	1.44129669	0.497057633	0	2	1
cveedad	DEF2000	NOMINAL	437667	0	4					A
edad	DEF2000	NUMERICO	437667	0	121	62.684228	71.94305378	1	998	1
ocupa	DEF2000	NUMERICO	437667	0	22	31.0276237	34.31439637	0	99	2
esco	DEF2000	NUMERICO	437667	0	9	3.34010332	2.345401721	0	8	1
edociv	DEF2000	NUMERICO	437667	0	8	3.55359211	2.420716768	0	8	2
traba	DEF2000	NUMERICO	437667	0	4	7.17991532	2.253417467	0	8	8
lugles	DEF2000	NUMERICO	437667	0	8	7.29812163	2.012593095	0	8	8
asismed	DEF2000	NUMERICO	437667	0	3	1.08351555	0.429731304	0	2	1
sitioo	DEF2000	NUMERICO	437667	0	5	2.128006	1.107143676	0	4	3
derech	DEF2000	NUMERICO	437667	0	14	1.55158374	1.124170181	0	36	1
conemb	DEF2000	NUMERICO	437667	0	4	0.0038134	0.064638633	0	3	0
relemb	DEF2000	NUMERICO	437667	0	3	0.00422467	0.073512549	0	2	0
violenc	DEF2000	NUMERICO	437667	0	3	7.80419588	1.234874234	0	8	8