



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA DE SISTEMAS – INVESTIGACIÓN DE OPERACIONES

Un modelo para la optimización de la eficiencia terminal de cursos en línea

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
ALEJANDRO FELIPE ZÁRATE PÉREZ

TUTOR PRINCIPAL
DRA. IDALIA FLORES DE LA MOTA, FACULTAD DE INGENIERÍA

CIUDAD UNIVERSITARIA, CDMX, FEBRERO 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

Presidente: Dra. Balderas Cañas Patricia E.

Secretario: Dra. Rodríguez Vázquez Katya

1 er. Vocal: Dra. Flores De La Mota Idalia

2 do. Vocal: M. I. Soler Anguiano Francisca Irene

3 er. Vocal: Dra. Segura Pérez Esther

Lugar donde se realizó la tesis: Ciudad de México

TUTOR DE TESIS:

Dra. Flores De La Mota Idalia



FIRMA

ÍNDICE

CAPITULO 1. PLANTEAMIENTO DE LA PROBLEMÁTICA	4
1.1 La educación a distancia	4
1.2 DGTIC Cursos en línea	9
1.3 Plataformas LMS	13
1.4 Analíticas del aprendizaje	17
1.5 Modelo de asignación	20
1.6 Objetivo general	22
CAPITULO 2. SITUACIÓN ACTUAL.....	24
2.1 Eficiencia terminal	24
2.2 Cuestionarios de opinión.....	27
2.3 Calendarización de cursos	32
CAPITULO 3. METODOLOGÍA PROPUESTA	40
3.1 Introducción	40
3.2 Metodología para el seguimiento de cursos	42
3.3 Análisis de los datos.....	44
CAPITULO 4. RESOLUCIÓN DEL CASO DE DGTIC	51
4.1 Eficiencia terminal de los asesores	51
4.2 Índice de apertura de los cursos	59
4.3 Modelo de asignación para los cursos.....	62
4.4 Análisis de resultados.....	71
CONCLUSIONES.....	73
APÉNDICE A	76
APÉNDICE B	88
REFERENCIAS	111

Capítulo 1. Planteamiento de la problemática

1.1 La educación a distancia

Desde ya varios años, la tecnología avanza cada vez más rápido y esto se ha reflejado también en la educación, un ejemplo de esto es la modalidad denominada “Educación a distancia”, la cual ha cobrado una mayor fuerza, al ser más accesible gracias a las nuevas tecnologías. Se pretende que con esta modalidad puedan acceder a la educación diversos sectores que no han podido ser atendidos, por situaciones como la geográfica, la laboral, la disponibilidad de tiempo, entre otras.

Sin embargo, es conveniente recordar que la educación a distancia no es tan nueva como se puede llegar a suponer, ya que ha existido desde hace algunos siglos.

En términos históricos, la enseñanza a distancia nace en el siglo XVIII, con el que parece ser el primer esfuerzo verificado en utilizar el correo postal como medio de comunicación entre profesor y alumno, fue en el año 1728, cuando en *The Boston Gazette*¹ el profesor C. Philipps, pone un anuncio en el que se ofrecía a enviar lecciones escritas por correo a los alumnos que estuvieran interesados, además de ofrecer tutorías por correspondencia.

Más adelante, en el siglo XIX se dan ejemplos parecidos en Europa Occidental y América del Norte, pero fue hasta el siglo XX, cuando a este tipo de enseñanza a distancia se le denominó, Enseñanza por correspondencia, la cual aparece como una forma de los países desarrollados para resolver el problema de educación de las minorías.

En este tipo de educación se hace uso de los medios de comunicación, es decir, fue producto de la modernización para resolver un problema social latente. De esta manera, las principales características de este proceso de enseñanza-aprendizaje están soportadas en los medios existentes con que cuenta cada sociedad. Por lo que, el primer medio productor de comunicación masiva fue la imprenta, siendo los libros el medio tradicional de soportar el conocimiento, para dar paso después, en los años sesenta y setenta, a los medios electrónicos como lo son la radio, la televisión y la computadora. Siendo la computadora “*el aparato mágico más espectacular e importante desde el punto de vista filosófico que la tecnología ha introducido en los últimos tiempos*”².

En México, el desarrollo de la Educación a distancia comienza a finales de los años sesenta, para cubrir el vacío de formación por parte de los docentes. Como primer dato histórico se parte del proyecto Radio Primaria que fue implementado para impartir la primaria fuera de la ciudad, en zonas rurales. También en los años sesenta se establece el Consejo Coordinador de Sistema Tecnológico Abierto y en los años setenta se establece el Consejo Coordinador de Sistemas Abiertos.

¹ *The Boston Gazette* (1719-1798) era un periódico publicado en Boston, Massachusetts, en las colonias británicas norteamericanas. Comenzó la publicación el 12 de diciembre de 1719.

² H.R. Fischer, 1997, p.138.

En 1973, la Universidad Nacional Autónoma de México establece el Sistema de Universidad Abierta, siendo el primer país de América Latina en el desarrollo de programas abiertos de educación superior.

En 1974, el Instituto Politécnico Nacional inicia el Sistema Tecnológico Abierto.

En 1976, con el fin de lograr que la educación básica tuviera acceso universal para los mexicanos se crea el Sistema de Telesecundaria bajo la supervisión de la Secretaría de Educación Pública.

Otros institutos como el Instituto Latinoamericano de la Comunicación Educativa (ILCE)³, fue creado en 1954 bajo la lógica internacional de la Conferencia General de la Organización de las Naciones Unidas (ONU) para la Educación, la Ciencia y la Cultura; de mejorar la educación en los países latinoamericanos a través de los medios y recursos audiovisuales. Y por otro lado el Instituto Nacional para la Educación de los Adultos (INEA), al igual que la Telesecundaria son proyectos de educación a distancia desarrollados en México que trabajan y extienden la educación por los medios disponibles.

Educación en línea

Las TIC (Tecnologías de la Información y la Comunicación), son las que se encargan del diseño, el desarrollo, fomento, mantenimiento y la administración de la información por medio de sistemas informáticos. Al decir, sistemas informáticos, no solamente se refiere a las computadoras, sino también las redes de telecomunicaciones, telemática, teléfonos celulares, televisión, radio, dispositivos portátiles, entre otros.

En la actualidad, con la incorporación de las TIC, se puede vislumbrar el alcance que estas representan para la Educación a distancia, desempeñando así un papel esencial, pues como resultado de la aplicación de estas nuevas tecnologías al ámbito educativo y de formación, surge lo que se denomina el “*e-learning*”.

El *e-learning* es una forma de utilizar las TIC como medio de distribución para materiales educativos y otros servicios, en la que además hay una interrelación entre profesores y alumnos. Siendo así, que en este nuevo entorno de enseñanza-aprendizaje se utiliza la *tecnología web* como la opción preferida, ya sea a través de una intranet/extranet como de Internet.

A menudo el término “*e-learning*”, es intercambiado por otros como: Enseñanza y Aprendizaje Digitales (EAD), Comunidades de Aprendizaje en Entornos Virtuales (CAEV), formación en línea, cursos en línea, formación virtual, formación a distancia, campus virtual, entre otros. En el sentido literal, del inglés, significa *aprendizaje electrónico*. Y como se puede observar, en el siguiente recuadro se muestran algunas definiciones:

“Término que cubre un amplio grupo de aplicaciones y procesos, tales como aprendizaje basado en web, aprendizaje basado en ordenadores, aulas virtuales y colaboración digital. Incluye entrega de contenidos vía Internet, intranet/extranet, audio y vídeo grabaciones, transmisiones satelitales, TV interactiva, CD-ROM y más”.

American Society of Training and Development

³ <http://www.ilce.edu.mx/v5/>

“Uso de las tecnologías basadas en internet para proporcionar un amplio despliegue de soluciones a fin de mejorar la adquisición de conocimientos y habilidades. El cual está basado en tres criterios fundamentales:

- 1) Que se produzca en red, lo que permite una actualización inmediata, almacenamiento y recuperación, distribución y capacidad de compartir los contenidos y la información,
- 2) Que llegue al usuario final a través de un ordenador, utilizando estándares tecnológicos de Internet,
- 3) Que esté centrado en la visión más amplia de soluciones para el aprendizaje que van más allá de los paradigmas tradicionales de la formación.”

*Definición por Marc J. Rosenberg (2001) en su libro **E-learning. Strategies for Delivering Knowledge in the Digital Age**⁴. McGraw-Hill.*

“Capacitación no presencial que, a través de plataformas tecnológicas, posibilita y flexibiliza el acceso y el tiempo en el proceso de enseñanza-aprendizaje, adecuándolos a las habilidades, necesidades y disponibilidades de cada discente, además de garantizar ambientes de aprendizaje colaborativos mediante el uso de herramientas de comunicación síncrona y asíncrona, potenciando en suma el proceso de gestión basado en competencias.”

*Definición del Dr. Francisco José García Peñalvo en su artículo **Estado actual de los sistemas e-learning**.*

Al hacer una revisión de estas definiciones, se puede comprender que el término *e-learning*, abarca muchas aplicaciones, por lo que se le conoce mejor por su vínculo con las TIC, así que para la finalidad de este trabajo se propone la siguiente explicación del término:

Es la modalidad de educación a distancia donde tanto el asesor como el alumno hacen uso de medios electrónicos para llevar a cabo el proceso de enseñanza-aprendizaje. Está basado en el uso de Internet e intranet/extranet como principales recursos y utiliza contenidos interactivos; además se utilizan herramientas como correo electrónico, blogs, foro, chat, páginas personales, videoconferencias, simuladores, entre otros; como apoyo para establecer la comunicación entre los participantes. Siendo uno de sus principales objetivos es mejorar el desarrollo académico de los estudiantes (habilidades, conocimientos, actitudes).

A pesar de que el término *e-learning* cubre un extenso conjunto de aplicaciones y procesos, está definido más estrechamente que la Educación a distancia, esta última puede incluir

⁴ Aprendizaje electrónico. Estrategias para distribuir el conocimiento en la era digital.

enseñanza en su forma más tradicional, como la basada en textos y cursos a través de correspondencia escrita.



Figura 1.1 Educación a distancia

Con respecto a los tipos de *e-learning* que existen, se encuentran los siguientes:

- CBT (*Computer Based Training*) o CAI (*Computer Assisted Instruction*), que es el aprendizaje a través de una computadora o instrucciones asistidas por computadora. Este aprendizaje fue implantando en varias instituciones educativas y organizaciones. Preferentemente se utilizaba software interactivo que se basaba en la lectura e incorporaba mecanismos de retroalimentación, convirtiendo al alumno en un ente más activo dentro de su propio proceso formativo.
- IBT (*Internet Based Training*), fue el siguiente gran paso evolutivo de los sistemas de aprendizaje por medio de una computadora (CBT), ya que, con la llegada de Internet, los contenidos podían llegar más fácilmente a sus destinatarios por medio de Internet o de la intranet.
- WBT (*Web Based Training*), consiste en la enseñanza-aprendizaje haciendo uso de la web, medio por el cual se reciben los contenidos. En este último tipo es donde se encuentra el campus virtual.

Además de los tipos, el *e-learning* tiene dos modalidades:

- *E-learning puro o virtual*, en el cual la formación se realiza completamente a distancia con soporte de las TIC.

Materiales disponibles en Internet.

Trabajo individual en línea.

Consultas a los tutores por correo o *chat*.

Intercambios grupales en foros, *chat* o videoconferencia por Internet.

Autoevaluaciones o evaluaciones en línea.⁵

- *B-learning (Blended learning)*. Sistemas mixtos Internet más otras tecnologías.

⁵ Kaplún, Gabriel, *Aprender y enseñar en tiempos de Internet. Formación profesional a distancia y nuevas tecnologías*, Montevideo: CINTERFOR/OIT, 2005, p. 147.

Materiales en Internet y/o en soportes físicos (CD, videos, impresos).

Trabajo individual y grupal a distancia y presencial (aula, taller, proyectos grupales).

Consultas a los tutores por correo, *chat*, teléfono o presencial.

Intercambios grupales en línea y presencial⁶.

Completa la formación presencial con la formación a través de las Nuevas Tecnologías de Información y Comunicación (NTIC).

En cuanto a las principales ventajas y desventajas del *e-learning*, se mencionan las siguientes:

VENTAJAS

- Puede llegar a cualquier lugar, se puede acceder a la información a cualquier hora y para cualquier persona.
- El participante avanza de acuerdo con su propio ritmo y capacidades, pero dentro de un margen de tiempo preestablecido.
- Se puede contar con un asesor que apoye al participante durante el curso.
- La comunicación entre el participante y el asesor es de tiempo remoto y en horarios no predeterminados.
- El participante puede consultar los materiales las veces que lo considere necesario, dándose además una autonomía de aprendizaje (aprendizaje por descubrimiento).
- Hay un aprendizaje colaborativo e interacción entre los participantes.
- Su precio es menor comparado con el traslado y pago por un curso presencial.
- En esta modalidad de educación a distancia, se pueden crear gran variedad de cursos de diferentes niveles de especialidad, además de poder ofrecer licenciaturas, diplomados, especialidades y maestrías.

DESVENTAJAS

- Que el participante no cuente con los conocimientos básicos del manejo de la computadora.
- Se requiere tener disponible un equipo de cómputo con el software y hardware adecuado, además de una estable conexión a Internet.
- Aunque el participante avanza de acuerdo con su propio ritmo y capacidades, se le debe dedicar varias horas para la revisión de materiales y realización de las actividades, para poder terminar dentro del margen de tiempo preestablecido.

⁶ Idem

- Si algún material creado para la modalidad presencial no funciona, tampoco servirá en esta modalidad.
- El temor al cambio de las formas tradicionales de educación.

Con esto, se puede visualizar que el *e-learning* da un giro completo con respecto a la educación tradicional, pues ya no se cuenta con el profesor que determine el qué, el cómo y en ocasiones hasta de que “humor”, se va a llevar a cabo la clase. Ya que ahora es el alumno el que lleva el centro del proceso y todas las NTIC deben estar orientadas al apoyo de su aprendizaje efectivo. Pero como no es lo mismo estar aprendiendo en un salón de clase, que hacerlo a través de algún medio como la televisión, la Internet, materiales impresos, entre otros; es aquí donde los especialistas en diseño instruccional cobran una gran importancia, tema que se abarcará más adelante.

Educación continua

Hay varias definiciones de educación continua, algunas de ellas varían de acuerdo con el país al que hagamos referencia. Sin embargo, para este trabajo tomaremos la definición de educación continua de la UNAM, la cual fue tomada desde el siguiente sitio: <https://educacioncontinua.unam.mx>. Y enuncia lo siguiente:

Es una modalidad educativa diseñada, organizada, sistematizada y programada que complementa la formación curricular y profundiza y amplía conocimientos en todos los campos del saber; capacita y actualiza profesionalmente y está dirigida a la comunidad universitaria y al público en general.

La educación escolarizada comprende los estudios de bachillerato, licenciatura, especialidad, maestría y doctorado. Por su parte, la educación continua es una modalidad que nutre y expande la formación curricular y sirve principalmente para la actualización u obtención de conocimientos complementarios. Algunas de las actividades de educación continua son: diplomados, cursos y talleres.

1.2 DGTIC Cursos en línea

La Dirección de Docencia en TIC (DDTIC), perteneciente a la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC) de la UNAM, imparte cursos en línea de educación continua y principalmente de cómputo, a través de la Coordinación de Capacitación Continua a Distancia, los cuales están dirigidos al público en general, la comunidad universitaria y a instituciones y empresas que los solicitan.

En el Reglamento General de Educación Continua de la UNAM, aprobado el 18 de marzo de 2016, se señala que la Educación Continua:

“Tiene como finalidad complementar la formación curricular, profundizar y ampliar conocimientos en todos los campos del saber, capacitar y actualizar profesionalmente para contribuir al bienestar y desarrollo individual y social, bajo los criterios de calidad y pertinencia distintivos de la Institución. Dado el compromiso social de la Universidad, la educación continua está dirigida a la comunidad universitaria y público en general.”

En el caso de la Dirección de Docencia en TIC (DDTIC), la oferta académica se enfoca principalmente en diversas áreas de conocimiento en cómputo e informática. En el año 2006 se creó un área para el desarrollo e impartición de cursos en línea, la cual comenzó con una oferta académica de 5 cursos, a lo largo de este tiempo, actualmente es la Coordinación de Capacitación Continua a Distancia (CCCD), cuya oferta supera los 100 cursos, impartiendo talleres, cursos, conferencias, seminarios, congresos y diplomados en formato semipresencial y virtual.

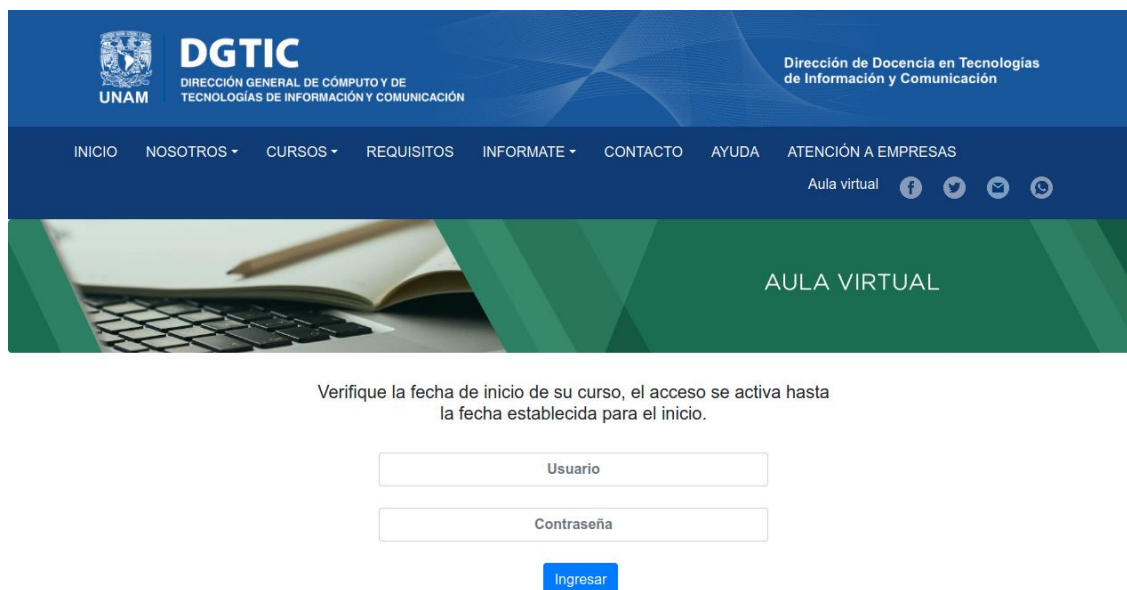


Figura 1.2 Sitio web <https://cursosenlinea.tic.unam.mx/aulavirtual.php>

Los cursos que se imparten se calendarizan de manera periódica, generalmente dos periodos mensuales, a los cuales se puede inscribir cualquier interesado en tomar un curso en línea.

Para cada uno de los grupos que se imparten por periodo se realiza la asignación de asesores para cada emisión, que son los encargados de la resolución de dudas, calificación de actividades, seguimiento de alumnos, entre otras actividades.

Modelo

La impartición de los cursos en línea se realiza a través de un sistema Moodle, todo el curso se realiza dentro de esta plataforma, por lo que todas las actividades quedan registradas dentro de la base de datos del sistema, entre estas actividades tenemos: revisión de materiales, entrega de tareas, participación en los foros, etc.

Para cada uno de los cursos se designa un asesor, el cual se encarga de responder dudas, incentivar la participación de los alumnos y realizar la evaluación de las tareas.

Todas las actividades dentro del curso deben ser realizadas dentro del periodo determinado para el curso, por lo cual, cada actividad está planeada para ser entregada de acuerdo con la planeación de cada curso. Con base en el diseño instruccional de cada curso, se planean de manera que se revise un tema por semana, lo que implica que las semanas de cada curso corresponden con los temas contenidos en cada uno de ellos.

Por lo anterior, cada curso impartido debe seguir un calendario de entrega y revisión de actividades por parte de los participantes del curso: alumnos y asesor. Este calendario debe cumplirse para que el curso finalice sin contratiempos, ya que un retraso por parte de alumnos o asesor puede terminar en un abandono del curso, lo cual genera que solo algunos alumnos finalicen su curso satisfactoriamente.

Para poder llevar a cabo un programa de formación basado en *e-learning*, se hace uso de herramientas tecnológicas como lo son las plataformas o sistemas de software que permiten la comunicación e interacción de alumnos, profesores y contenidos. Entre las más utilizadas están los *Sistemas de Administración de Aprendizaje (LMS, Learning Management Systems)* o también llamados *plataforma de gestión de aprendizaje*, los cuales se utilizan para impartir y dar seguimiento administrativo a los cursos en línea.

Deserción

Uno de los problemas de la educación, y en particular de la educación en línea, son los altos niveles de deserción que se presentan durante cada curso. Este es uno de los mayores problemas que presenta esta modalidad de cursos, ya que existen casos donde alcanza niveles alarmantes: más del 70% de deserción.

Entre las diversas causas de estos altos índices de deserción están las competencias tecnológicas de los estudiantes, de los asesores, la planeación de los cursos, diseños pedagógicos inadecuados, infraestructura tecnológica deficiente, etc.

De acuerdo con **Fuente especificada no válida.**, podemos atribuir causas a los estudiantes, otras a los docentes y algunas más a la institución.

Causas atribuibles a los estudiantes

- Falta de tiempo.
- Deficientes técnicas de estudio.
- Insuficiente motivación.
- Elección equivocada.
- Bajo rendimiento académico.
- Déficit en competencias digitales.
- Escasa capacidad para la autonomía y la autodisciplina.
- Temor al fracaso.
- Pobre integración académica y social

Causas atribuibles a los docentes

- Escaso seguimiento y supervisión docente.
- Los materiales de estudio carecen de un diseño adecuado.
- La rigidez curricular.
- Pruebas de evaluación poco adecuadas.

Causas atribuibles a la institución

- Deficiente información inicial sobre la carrera o curso.
- Ambigüedad en las directrices y orientaciones procedentes de la institución.
- Inexistencia de servicios de orientación. O servicios de orientación deficientes.
- Insuficiencia o inadecuación de los servicios tecnológicos.

En el caso de estudio de la DGTIC, el ingreso de los estudiantes no está regulado de ninguna manera, es decir, no se realiza ningún tipo de filtro, el interesado solo elige el curso y se inscribe al curso de su interés sin ninguna restricción. Por lo cual, ninguna de las causas mencionadas por **Fuente especificada no válida**, puede ser atendida de manera adecuada.

Sin embargo, las causas atribuibles a la institución y a los docentes si pudieran ser atendidas, ya que la DGTIC realiza la planeación de cursos y la asignación de docentes a cada uno de ellos según el periodo correspondiente.

Es por ello, que la CCCD realizó el desarrollo de un Sistema de Monitoreo Automática de Cursos (SMAC), a través de este sistema realiza un seguimiento diario a los cursos que se imparten, con el propósito de identificar de manera temprana los posibles problemas que se presentan y alcanzar una solución adecuada.

Además de los reportes que se generan para docentes y áreas encargadas de la gestión escolar de los cursos, el sistema guarda los registros diarios que fueron realizados en cada curso. Estos datos almacenados pueden convertirse en información para tomar medidas respecto de la deserción de cursos.

Para analizar esta deserción, se ha definido la eficiencia terminal de un curso, la cual se refiere al porcentaje de alumnos que termina satisfactoriamente un curso.

Lo cual nos lleva a analizar las causas atribuibles a los asesores, ya que según los datos que registra el SMAC, se ha observado que la eficiencia terminal de un curso presenta una mayor variación en la eficiencia terminal en el curso que en el asesor.

Esto podría implicar que la variación en la eficiencia terminal depende del asesor que haya sido asignado más que los materiales del curso, plataforma tecnológica, etc.

Monitoreo de cursos

El monitoreo de los cursos en línea se refiere al seguimiento del avance del curso durante el periodo en que se imparte, este monitoreo es sumamente importante porque permite a las áreas encargadas del seguimiento escolar detectar los posibles problemas que afectan el curso y realizar una pronta intervención para que los alumnos finalicen satisfactoriamente.

Este seguimiento debe realizarse para todos los cursos que se estén impartiendo y determinar aquellos en los cuales haya un retraso en el calendario programado y por lo tanto es necesaria una intervención. Asimismo, es conveniente determinar la causa que provoca el retraso, con el propósito de realizar la mejor intervención posible.

Sin embargo, este seguimiento puede ser una tarea laboriosa que puede demorar demasiado tiempo si se realiza de forma manual y si no se realiza de manera diaria se pierde la oportunidad de una intervención adecuada.

1.3 Plataformas LMS

Las plataformas de **Learning Management System (LMS)** son sistemas desarrollados específicamente para el *e-learning*, estos sistemas cuentan con todas las herramientas necesarias para realizar un proyecto de educación en línea.

Definición

Un **LMS** es un software basado en un servidor web que provee módulos con funcionalidades independientes, para los procesos administrativos y de seguimiento que se requieren en un proceso de enseñanza-aprendizaje virtual.

Los módulos van a permitir, por ejemplo, montar cursos, matricular alumnos, registrar profesores, asignar cursos a un alumno, entre otras funcionalidades que incluyen la logística necesaria para poder ofrecer cursos a través de Internet o de una intranet. También facilitan el aprendizaje con aulas virtuales sincrónicas (por ejemplo, el uso de videoconferencias, el chat, entre otros) o el acompañamiento asincrónico⁷ del alumno.

La manera en que el alumno interactúa con la plataforma es a través de una interfaz web, la cual le va a permitir seguir las unidades didácticas que conforman el curso, dentro de las cuales habrá materiales a revisar, actividades programadas a realizar, comunicarse con su profesor y con sus compañeros de grupo, así como un seguimiento a su progreso.

Principales plataformas

En general, todas las plataformas cuentan con funciones básicas que se han estado mencionando, sin embargo, la complejidad y capacidad varían de un sistema a otro. Algunas de estas, se describen a continuación:

Chamilo (<https://chamilo.org>)



⁷ En el aprendizaje asincrónico, se utilizan tecnologías de información y servicios de comunicación de internet como el correo electrónico, el foro, entre otras. El participante avanza de acuerdo con su propio ritmo y capacidades, dentro de un margen de tiempo preestablecido. El proceso de enseñanza-aprendizaje queda diferido en tiempo y espacio.

Chamilo LMS es un campus virtual de código libre que se distribuye bajo licencia GNU/GPLv3, y que cualquier persona, institución o empresa puede usar libremente para la impartición de acciones formativas a través de internet.⁸

Detrás, encargada de velar por su correcta evolución, se encuentra la Asociación Chamilo, entidad sin ánimo de lucro fundada en 2010 cuyo principal objetivo es promover el desarrollo y uso del software garantizando su pervivencia como producto de código libre.

La Asociación Chamilo tiene como principal objetivo proveer la plataforma de *e-learning* más sencilla y usada del mundo bajo la fórmula del software libre haciendo de ella un motor para el desarrollo educativo de las personas de cualquier parte del mundo.

Entre los objetivos de la asociación se encuentran:

- Proveer un LMS libre de calidad que permita mejorar la educación a nivel mundial.
- Garantizar la continuidad del proyecto Chamilo LMS como un software libre accesible para todos de la manera más igualitaria posible.
- Contribuir a la reducción de la brecha digital y el acceso a la educación online de los países más pobres promoviendo el acceso a un LMS puntero.
- Desempeñar una dirección del proyecto con transparencia, ética y responsabilidad social.
- Respetar el trabajo de las empresas y usuarios de la comunidad que contribuyen al crecimiento y desarrollo del proyecto.

Blackboard (<http://www.blackboard.com/>)



Es un sistema comercial, fue fundado en 1997, con sede en Washington, DC, E.U. Desde entonces se ha fusionado con las compañías *CourseInfo LLC* (en 1998) y *WebCT* (2006).

Gran parte de las instituciones que usan el programa de *Blackboard*, lo utilizan para administrar aprendizaje en línea (*e-learning*), procesamiento de transacciones, comercio electrónico (*e-commerce*), y manejo de comunidades en línea.

El portafolio de productos y servicios de Blackboard abarca la enseñanza y el aprendizaje, la participación de la comunidad y los servicios, asegurando que la experiencia del alumno sea la mejor posible, tanto dentro como fuera del aula.⁹

⁸ <https://chamilo.org/es/chamilo/>

⁹ <https://www.blackboard.com/es-lac/about-us>



Figura.1.- Portafolio de productos y servicios de BlackBoard

CANVAS (<https://www.instructure.com/es>)



Canvas LMS es el sistema de gestión del aprendizaje abierto y extensible que potencia la genialidad de todos. Según su página web tiene las siguientes características:

- **Todo en un mismo lugar.** Contenido atractivo. Pruebas y calificaciones. Datos e información. Interacción entre estudiantes y profesores. Todo en Canvas LMS. Listo para apoyar el aprendizaje en persona, en línea y combinado.
- **Accesibilidad total.** Las expectativas han cambiado. Necesita un LMS que se adapte a lo que los estudiantes esperan hoy. Con Canvas LMS, estudiantes e instructores tienen acceso a todo lo que necesitan en un solo lugar.
- **Contenido al instante.** Cargue y comparta con rapidez la información del curso en una serie de formatos cautivadores. Y utilice pautas para impulsar el plan de estudios en toda su institución.
- **Aprendizaje móvil.** Canvas ofrece aplicaciones que combinan altas prestaciones y capacidad de respuesta para brindar accesibilidad incomparable y la seguridad de que su institución puede llegar a todos los estudiantes.

Sakai (<https://www.sakailms.org>)



El Proyecto Sakai es un software educativo de código abierto. El nombre de Sakai proviene del cocinero Hiroyuki Sakai. El Proyecto Sakai tiene su origen en la Universidad de Michigan y en la Universidad de Indiana, a las que se unieron el Instituto Tecnológico de Massachusetts y la Universidad Stanford, junto a la Iniciativa de Conocimiento Abierto (OKI) y el consorcio uPortal. El proyecto se consolidó con una donación de la Fundación Mellon.¹⁰

El objetivo del Proyecto Sakai es crear un entorno de colaboración y aprendizaje para la educación superior, que pueda competir con sus equivalentes comerciales Blackboard / WebCT y que mejore otras iniciativas de código abierto como Moodle.

De acuerdo con el proyecto Sakai, las características de las plataformas son:

- Diseñado para que el aprendizaje pueda ser colaborativo.
- Proporciona diferentes herramientas para la comunicación.
- Sakai es una arquitectura abierta, lo cual permite integrar herramientas de otras arquitecturas.
- Sakai adopta el estándar IMS (Learning Tools Interoperability).
- Alta flexibilidad a la hora de crear y diseñar el curso, y escoger las herramientas necesarias.

MOODLE (<http://moodle.org>)



Es un sistema de los más reconocidos, que se distribuye gratuitamente como Software libre (Open Source) (bajo la Licencia Pública GNU, cuya primera versión apareció el 20 de agosto de 2002).

Moodle es un paquete de software para la creación de cursos y sitios Web basados en Internet. Es un proyecto en desarrollo diseñado para dar soporte a un marco de educación social constructivista.

Su arquitectura y herramientas son apropiadas para clases en línea, así como también para complementar el aprendizaje presencial. Tiene una interfaz de navegador de tecnología sencilla, ligera, y compatible.

¹⁰ https://es.wikipedia.org/wiki/Proyecto_Sakai

Moodle tiene la confianza de instituciones y organizaciones grandes y pequeñas, incluyendo a Shell, La Escuela Londinense de Economía (London School of Economics), La Universidad Estatal de Nueva York, Microsoft y la Universidad Abierta del Reino Unido (Open University). El número de usuarios de Moodle a nivel mundial, de más de 200 millones de usuarios (en agosto del 2020), entre usuarios académicos y empresariales, lo convierten en la plataforma de aprendizaje más ampliamente utilizada del mundo y de México.¹¹

Según su página web algunas de las características que posee Moodle son:

- **Diseñado para soportar tanto la enseñanza como el aprendizaje.** Con más de 10 años de desarrollo guiado por la pedagogía de constructivismo social, Moodle proporciona un conjunto poderoso de herramientas centradas en el estudiante y ambientes de aprendizaje colaborativo, que le dan poder, tanto a la enseñanza como al aprendizaje.
- **Fácil de usar.** Una interfaz simple, características de arrastrar y soltar, y recursos bien documentados, junto con mejoras continuas en usabilidad, hacen a Moodle fácil de aprender y usar.
- **Gratuito, sin cargos por licenciamiento.** Moodle es proporcionado gratuitamente como programa de Código Abierto, bajo la Licencia Pública General GNU (GNU General Public License). Cualquier persona puede adaptar, extender o Modificar Moodle, tanto para proyectos comerciales como no-comerciales, sin pago de cuotas por licenciamiento, y beneficiarse del costo/beneficio, flexibilidad y otras ventajas de usar Moodle.
- **Siempre actualizado.** La implementación de Moodle en código abierto significa que Moodle es continuamente revisado y mejorado, para adecuarse a las necesidades actuales y cambiantes de sus usuarios.

En el caso de la DGTIC, desde 2008 la impartición de los cursos en línea se realiza a través de una plataforma Moodle, la cual está instalada en el Centro de datos de la dependencia y es administrada por el personal de la CCCD.

1.4 Analíticas del aprendizaje

A nivel técnico podemos afirmar que todos los Learning Management System (LMS) utilizan una base de datos para su correcto funcionamiento, en esta base se resguarda la información de las lecciones, actividades e incluso calificaciones de los participantes.

Sin embargo, no solo almacena la información didáctica, también recopilan la información de las interacciones de los participantes dentro de la plataforma, esta información puede ser útil ya que a través de ella podemos identificar como se desarrolla el proceso de enseñanza – aprendizaje.

¹¹ https://docs.moodle.org/all/es/Acerca_de_Moodle

Definición

Las analíticas del aprendizaje son un concepto relativamente nuevo, por lo que existen varias definiciones al respecto; algunas de ellas la mencionaremos a continuación.

La primera definición es de Long y Siemens de 2011:

“Uso de estos datos y de cualquier otra observación adicional de la que se pueda obtener información, para tener un impacto directo sobre los estudiantes, los profesores y el proceso de aprendizaje”

La segunda es de Buckingham y Ferguson, 2012 p. 4

“La medición, recopilación, análisis y reporte de datos sobre los alumnos y sus contextos, con el propósito de comprender y optimizar el aprendizaje y los entornos en los que se produce”

En estas dos definiciones el concepto se refiere a recopilar la información observada para comprender y realizar un cambio en los procesos de aprendizaje que se llevan a cabo.

Sin embargo, para este trabajo utilizaremos la definición dada por Durall et al en 2012 en “El Informe Horizon del The New Media Consortium y la Universitat Oberta de Catalunya”, la cual dice lo siguiente:

Las analíticas de aprendizaje consisten en la interpretación de un amplio rango de datos producidos y recogidos acerca de los estudiantes para orientar su progresión académica, predecir actuaciones futuras e identificar elementos problemáticos. El objetivo de la recolección, registro, análisis y presentación de estos datos es posibilitar que los profesores puedan adaptar de manera rápida y eficaz las estrategias educativas al nivel de necesidad y capacidad de cada alumno. Aun, en sus primeras etapas de desarrollo, las analíticas de aprendizaje responden a la necesidad de llevar a cabo el seguimiento y control de la actividad en el campus para la toma de decisiones estratégicas. Por otro lado, pretenden aprovechar la gran cantidad de datos producidos por los estudiantes en actividades académicas

Con esta definición realizaremos más adelante las consideraciones necesarias para la definición, recopilación, análisis y predicción de los datos del presente trabajo.

Ventajas y desventajas

Como cualquier herramienta que utilizamos, las analíticas del aprendizaje presentan ventajas y desventajas al utilizarlas, en esta parte revisaremos brevemente alguna de estas.

Las ventajas identificadas son:

- **Identificar el proceso de aprendizaje del alumno.** Cuando se diseña el curso se espera que los estudiantes sigan las actividades de acuerdo con el diseño planteado; sin embargo, algunas veces esto no sucede como estaba planeado, los datos observados nos permiten identificar de manera precisa como u estudiante lleva a cabo su proceso en particular.
- **Identificar alumnos en riesgo.** A través de los datos podemos identificar a los alumnos que por el comportamiento que están presentando pueden desertar del curso, con esta

identificación es posible establecer comunicación con el estudiante y prevenir su deserción.

- **Mejorar el diseño del curso.** Al identificar el proceso que sigue un alumno dentro del curso, es posible que este requiera cambios debido a los comportamientos observados, estos cambios podrían mejorar el funcionamiento de este en las siguientes emisiones del curso.
- **Recomendaciones a los asesores.** Con la información obtenida e interpretada es posible brindar recomendaciones, durante el desarrollo del curso, a los asesores encargados de guiar a los estudiantes; estas recomendaciones pueden ser desde actividades que tienen retrasados a los alumnos hasta posibles dudas que no han sido respondidas por el asesor.
- **Mejorar la toma de decisiones.** Una vez finalizado el curso, revisado e interpretado la información recolectada es posible tomar decisiones sobre el curso: quitar actividades, replantear preguntar, rediseñar materiales, etc.

Algunas desventajas:

- **Indicadores.** Actualmente no hay indicadores definidos que funcionen para cualquier proyecto de analíticas del aprendizaje, es por ello que para cada proyecto es necesario identificar los alcances del mismo y con base en ello definir los indicadores particulares.
- **Tiempo de implantación.** El tiempo para la realización de un proyecto es variable porque está determinado por los objetivos de este. Sin embargo, la recopilación, análisis e interpretación de los resultados puede tardar hasta más de 5 años en brindar los resultados esperados.
- **Recopilación de la información.** Aunque existen herramientas dentro de los LMS para generar casi cualquier tipo de reporte, algunas veces es necesario contar con los conocimientos técnicos especializados para la recolección de datos; ya sea porque la información necesaria no se encuentra en algún reporte predefinido o porque la cantidad de información a procesar es muy grande.
- **Análisis de la información.** Cuando se requiere un análisis más que descriptivo de la información es necesario utilizar herramientas o técnicas especializadas para obtener los resultados deseados; las cuales generalmente no están instaladas de manera predeterminada en los LMS utilizados.

Herramientas

Algunas de las herramientas utilizadas en las analíticas del aprendizaje son las siguientes:

- **Google Analytics.** Es una herramienta de analítica web de Google, ofrece información del tráfico que llega a los sitios web. Entre los análisis que proporciona están los grupos demográficos, geográficos, tiempo de la sesión, etc.
- **GISMO.** Es una herramienta de monitoreo gráfico interactivo que proporciona a los instructores una visualización útil de las actividades de los estudiantes en los cursos en línea. Con GISMO los instructores pueden examinar varios aspectos de los estudiantes a distancia, como la asistencia a los cursos, la lectura de materiales, la presentación de

tareas. Los usuarios del popular sistema de gestión del aprendizaje Moodle pueden beneficiarse de GISMO para sus actividades de enseñanza.¹²

- **KNIME.** Es una plataforma de análisis, una solución abierta para la innovación impulsada por datos, diseñada para descubrir el potencial oculto en los datos, extraer conocimientos nuevos o predecir nuevos futuros.¹³
- **Rapidminer.** Es una herramienta para el análisis y minería de datos. Permite el desarrollo de procesos de análisis de datos utilizando un entorno gráfico. Es utilizado en investigación, educación, prototipos y capacitación.

1.5 Modelo de asignación

El modelo de asignación es un muy estudiado dentro de la programación lineal, en el que los “asignados” son recursos que se destinan a la realización de tareas. Este modelo de asignación es un caso particular del problema de transporte, donde la oferta en cada origen y la demanda en cada destino son ambas de valor 1.

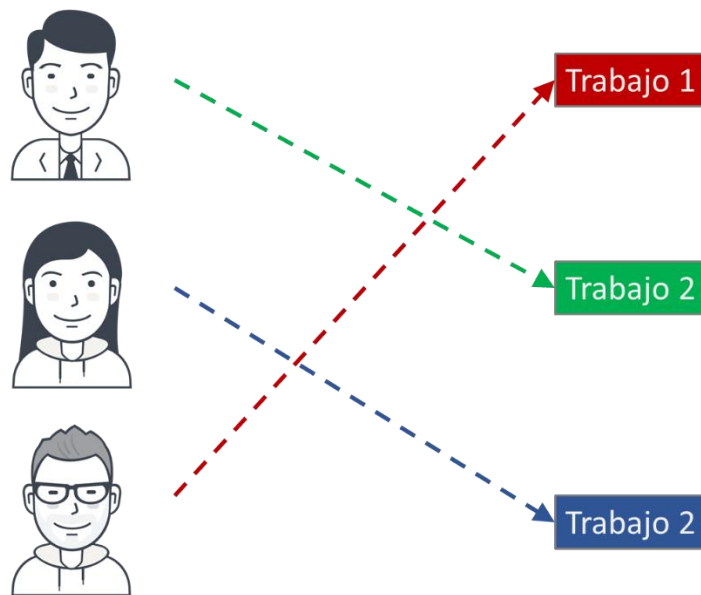


Figura 1.3 Una persona a cada trabajo

Aunque en el modelo de asignación se suele pensar que los que tenemos que asignar son personas a determinados trabajos, no necesariamente tiene que ser de esta forma, en lugar de personas pueden ser máquinas, vehículos, fabricas, etc.

El propósito de este problema es encontrar la manera óptima de asignar los recursos disponibles a la realización de tareas con el menor costo posible, tomando en cuenta que

¹² <http://gismo.sourceforge.net/>

¹³ <https://www.knime.com/about>

cada recurso se debe destinar a una sola tarea y que cada tarea es realizada por solo uno de los recursos.

Planteamiento del problema

Como mencionamos antes el problema de asignación es un caso particular del problema de transporte, en este caso se intenta asignar un número de orígenes (personas, tareas, vehículos, etc.) a un mismo número de destinos (tareas, fábricas, vehículos, etc.) con el objetivo de optimizar una función de costo.

La asignación debe realizarse con la condición de que cada origen tenga asignado un único destino y que cada destino este asignado a un único origen, donde asumimos que asignar un origen O_i a un Destino D_j tiene un costo c_{ij} .

Modelo matemático

Utilizando un modelo de programación lineal, en particular, un modelo binario de programación entera, tenemos lo siguiente.

$$x_{ij} \begin{cases} 1 & \text{Si el origen } O_i \text{ es asignado al destino } D_j \\ 0 & \text{En caso contrario} \end{cases}$$

Donde

$$i \in \{1, 2, \dots, n\} \text{ y } j \in \{1, 2, \dots, n\}$$

Entonces tenemos que la función objetivo es la siguiente:

$$\min z = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij}$$

Sujeto a

$$\sum_{j=1}^n x_{ij} = 1 \quad \forall i \in \{1, 2, \dots, n\}$$

$$\sum_{i=1}^n x_{ij} = 1 \quad \forall j \in \{1, 2, \dots, n\}$$

$$x_{ij} = 0, 1$$

Observe que las primeras n restricciones nos indican que cada origen debe estar asignado a un único destino, mientras que las siguientes n restricciones nos indican que cada destino tiene asignado un origen.

Cuando el número de orígenes no es igual al número de destinos, la forma de equilibrar el problema es agregar orígenes o destinos como necesitemos y como estos son ficticios el costo asociado a ellos será 0.

La tabla de costos se puede observar a continuación.

Tabla 1.1 Tabla de costos

	D ₁	D ₂	...	D _n
O ₁	C ₁₁	C ₁₂	...	C _{n1}
O ₂	C ₂₁	C ₂₂	...	C _{n2}
...
O _n	C _{n1}	C _{n2}	...	C _{nn}

Con este modelo de asignación finalizamos los conceptos y herramientas que necesitaremos a continuación.

1.6 Objetivo general

Hipótesis de investigación

Con base en todo lo anterior se parte del hecho de que no se cuenta con una metodología eficiente para dar seguimiento a los cursos en línea y por ello no existe un modelo para la mejora de la eficiencia terminal de los cursos.

Por esta razón se plantea el siguiente objetivo general

Analizar, evaluar y diagnosticar los cursos en línea de la DGTIC utilizando una metodología para el seguimiento de estos, obteniendo los datos necesarios para aplicar un modelo que maximice la **eficiencia terminal** de los cursos para un periodo en particular.

Para alcanzar este objetivo, a lo largo del trabajo, se propondrá una metodología para la recopilación de los datos, el procesamiento de estos, la construcción de los índices necesarios y un modelo de asignación que utilice estos datos para obtener el mejor resultado posible en la siguiente impartición.

De forma general la metodología propuesta se ve de la siguiente manera:



Figura 1.4 Metodología general

Donde para realizar cada una de estas partes es necesario:

- Analizar la situación actual respecto de la metodología utilizada actualmente.
- Definir una metodología para la extracción de datos desde el sistema donde se imparten los cursos en línea.
- Realizar el correspondiente análisis de los datos obtenidos.
- Definir y calcular los índices correspondientes con base en los datos obtenidos.
- Realizar un modelo de asignación que optimice la asignación de los asesores.

En el capítulo siguiente revisaremos como se evalúan actualmente los cursos y como se realiza la planeación de estos para un periodo en particular; desde la metodología utilizada para la evaluación hasta los criterios administrativos que se deben cumplir para la planeación de los cursos en los distintos periodos programados.

Capítulo 2. Situación actual

En este capítulo revisaremos el estatus actual del seguimiento que se realiza a los cursos en línea y a los asesores encargados de impartir dichos cursos; así como la calendarización actual de los cursos.

Para los análisis que realizaremos utilizaremos los datos recopilados durante el año 2020, en este año se impartieron 59 cursos, 196 grupos y 41 asesores. Asimismo, utilizaremos herramientas estadísticas para revisar la validez del seguimiento actual y propondremos un modelo para la asignación de cursos, ya que actualmente se realiza de forma manual.

Comenzaremos realizando la definición de **eficiencia terminal**, ya que a partir de esta trataremos de emplear los datos recopilados desde los **cuestionarios de opinión** utilizados actual y exclusivamente para evaluar el desempeño de un curso.

Asimismo, para estos cuestionarios de opinión describiremos la metodología actualmente utilizada y realizaremos un breve análisis de los datos obtenidos y como estos resultados afectan la eficiencia terminal de los cursos.

Finalmente, revisaremos como se realiza la **calendarización de los cursos**, los criterios que se toman en cuenta y utilizaremos un modelo de programación lineal entera para optimizar este proceso.

2.1 Eficiencia terminal

Las actividades de cada curso están predeterminadas desde el comienzo del curso, es decir, estas no se mueven una vez comenzado el mismo. Esto nos permite saber la cantidad de actividades que se realizarán en cada curso.



Figura 2.1 Ejemplo de actividades de un curso

Tomando esto en cuenta y para realizar el posterior análisis de información podemos decir que:

$$\text{Total de actividades curso} = \text{Número de alumnos} \times \text{Actividades planeadas del curso}$$

Por ejemplo, si para un grupo de 20 alumnos en un curso donde deben entregar 6 actividades (tareas) en todo el curso, tenemos que:

$$\text{Total de actividades curso} = 20 \times 6 = 120$$

Es decir, al finalizar el curso si todos los alumnos entregan todas las actividades tendremos 120 actividades calificadas, con lo que habremos alcanzado el 100% de las actividades esperadas para dicho curso.

Por lo anterior, podemos calcular el avance que esperamos para “cada semana”, en este caso decidimos “cada semana”, debido a que recordemos que así están calendarizadas las actividades en el diseño instruccional de los cursos analizados.

Para ello, primero calculemos el avance esperado semanal:

$$\text{Avance esperado semanal} = 100 / \text{Número total de semanas (N)}$$

Por ejemplo, siguiendo con el ejemplo anterior para un curso que dura 4 semanas, tenemos que:

$$\text{Avance esperado semanal} = 100 / 4 = 25$$

Entonces para cada semana esperamos se realicen aproximadamente el 25% de las actividades totales del curso.

Con este Avance esperado semanal podemos calcular que para un curso que dura N semanas el avance esperado hasta la semana i es el siguiente:

$$\text{Avance esperado hasta la semana } i = \text{Avance esperado semanal} \times i$$

Por ejemplo, en nuestro caso, para el grupo de 20 alumnos en un curso donde deben entregar 6 actividades (tareas) en todo el curso, podríamos decir que para la segunda semana tendríamos un avance esperado de:

$$\text{Avance esperado semana 2} = 25 \times 2 = 50$$

Es decir, a la mitad del curso esperamos que se hayan realizado el 50% de las actividades planeadas.

Cómo hemos mencionado antes, toda interacción entre los participantes se realiza dentro de la plataforma de cursos, por lo cual es posible calcular el **Avance real en el día correspondiente**. Este avance se refiere al conteo de las actividades finalizadas, es decir, aquellas actividades que los alumnos entregaron y que los asesores calificaron, si incumple alguna de estas condiciones no se considera finalizada.

Este grupo tiene **16 alumnos**, este curso está en la **última semana**. Para esta semana el avance esperado correspondiente es del **75%** del total del curso y en este momento tiene un avance real del **45%**.

Figura 2.2 Ejemplo de avance real de un curso

En este caso no existe una fórmula para realizar dicho cálculo ya que solo es un conteo que el sistema realiza de manera automática el sistema respecto de las actividades consideradas finalizadas.

Con este avance real es posible definir un **Índice de avance**, el cual calcularemos de la siguiente manera:

$$\text{Índice de avance} = \frac{\text{Avance real en el día correspondiente}}{\text{Avance esperado semana}}$$

Este nos permitirá observar la evolución de los avances a lo largo del curso, es decir, como van evolucionando las actividades realizadas por alumnos y asesores durante el periodo del curso.

Sin embargo, debemos tener en cuenta las siguientes consideraciones:

- Los primeros 7 días no se calcula el índice, esto se debe a que, de acuerdo con la planeación del curso, la primera semana no tiene un avance esperado, por lo cual división sería entre cero.
- El escenario ideal para este índice debe ser siempre 1, ya que esto nos indicaría que los alumnos y asesores van realizando las actividades de acuerdo con la planeación del curso.
- Un valor por debajo de 1 nos dice que los alumnos no están entregando sus actividades o el asesor tiene pendientes actividades por calificar, más adelante detallaremos este escenario.
- Un valor encima de 1 indica que se realizan más actividades de las esperadas, lo cual puede implicar que su avance es más rápido que el planeado originalmente.
- Al finalizar el curso, el avance real debería ser máximo 1, esto porque de antemano sabemos todas las actividades que se realizarán y en el mejor de los casos todos los alumnos entregarán sus actividades.

Tipos de retrasos

Cómo mencionamos antes, el índice de avance nos permite identificar los posibles retrasos que tiene un curso, por lo cual es necesario categorizar los posibles valores que se pueden obtener en este índice.

- **Sobresaliente.** Esta categoría se otorga cuando el valor del índice es mayor a 1, implica que las actividades se están realizando más rápido de lo planeado, por lo que el curso tiene un estado saludable.
- **Normal.** Esta categoría se otorga cuando el índice esta entre 0.75 y 1. Aunque el curso no ha alcanzado las actividades deseadas, el retraso no se considera como tal, ya que el día en que se obtiene el valor puede ser al principio de una nueva semana o al final cuando todos entregan y el asesor no ha tenido tiempo de revisarlas.
- **Atrasado.** Esta se otorga cuando el índice esta entre 0.5 y 0.75. Este valor ya se considera un retraso debido a que es muy probable que los participantes hayan dejado

de entregar o que, incluso, el asesor tenga demasiadas actividades pendientes sin calificar.

- **Retraso significativo.** Esta es la categoría más baja, cuando el valor del índice es menor a 0.5. Este caso es el más grave de todos ya que cuando un curso cae en esta categoría puede ser porque los alumnos han comenzado a abandonar el curso, han dejado de entregar actividades desde hace tiempo o el asesor tiene una gran cantidad de actividades sin evaluar.

Asimismo, a estos retrasos es posible atribuirles una causa, las cuales describimos a continuación:

- **Actividades sin calificar.** Esta causa implica que la responsabilidad en el retraso es atribuible al asesor, ya que tiene actividades pendientes que fueron entregadas por los alumnos y no han sido evaluadas por él.
- **Inasistencia de alumnos.** Esta causa de retraso implica que los alumnos han comenzado a dejar el curso, es decir, llevan algún tiempo sin ingresar y por lo tanto no han entregado las actividades programadas
- **No se entregan actividades.** Finalmente, esta causa implica que los alumnos continúan entrando al curso, el asesor no tiene actividades pendientes de evaluar, pero no se entregan las actividades correspondientes.

Finalmente, podemos definir la **eficiencia terminal** de un curso como el último *índice de avance* registrado al final del periodo de este. Por lo cual, el escenario ideal es que el curso alcance la mayor eficiencia terminal posible, es decir, el 100%.

2.2 Cuestionarios de opinión

Al finalizar un curso se presenta al alumno un cuestionario de opinión para que lo responda según su percepción de diversos aspectos del curso.

La última actividad del curso consiste en responder nuestro **cuestionario de opinión**, este está diseñado para el mejoramiento de los cursos en línea y no tiene ningún efecto en su calificación, por eso le solicitamos que elija la respuesta que mejor refleje su opinión con respecto al desarrollo del curso.



Cuestionario de opinión final

Figura 2.3 Presentación del cuestionario de opinión dentro del curso.

Los aspectos que se evalúan dentro de este cuestionario son: asesor, materiales, plataforma, evaluaciones, participante y curso; en cada una de estas categorías se presentan una serie de preguntas con el propósito de evaluar cada una de estas categorías.

Cuestionario de opinión final

Modo: Los nombres de los usuarios se mostrarán y registrarán con las respuestas

El asesor

- Cumplió con la forma de trabajo establecida al inicio del curso **!**
- (1) Nunca
 - (2) Casi nunca
 - (3) Algunas veces
 - (4) Casi siempre
 - (5) Siempre
- Respondió las dudas con claridad y en los tiempos acordados **!**
- (1) Nunca
 - (2) Casi nunca
 - (3) Algunas veces
 - (4) Casi siempre
 - (5) Siempre
- Retroalimentó de forma clara y con información de aciertos y errores **!**
- (1) Nunca
 - (2) Casi nunca
 - (3) Algunas veces
 - (4) Casi siempre
 - (5) Siempre

Figura 2.4 Preguntas de la categoría de asesor.

La plataforma

- Tuvo un rápido tiempo de respuesta **!**
- (1) Nunca
 - (2) Casi nunca
 - (3) Algunas veces
 - (4) Casi siempre
 - (5) Siempre
- Tuvo disponibles los contenidos en tiempo en forma **!**
- (1) Nunca
 - (2) Casi nunca
 - (3) Algunas veces
 - (4) Casi siempre
 - (5) Siempre
- Fue fácil de comprender y navegar **!**
- (1) Nunca
 - (2) Casi nunca
 - (3) Algunas veces
 - (4) Casi siempre
 - (5) Siempre

Figura 2.5 Preguntas de la categoría plataforma

Este cuestionario de opinión se aplica en todos los cursos, es decir, las mismas preguntas deben ser respondidas por cualquier alumno al finalizar su curso, no importa la temática de este ya que las preguntas fueron diseñadas con el propósito de evaluar cada aspecto sin importar el tipo o tema de curso.

La realización de este cuestionario no es obligatoria por parte del alumno, por lo cual depende en gran medida del recordatorio de parte del asesor a los alumnos para obtener una mayor cantidad de respuestas.

Descargar datos de tabla como Valores separados por comas (.CSV) Descargar

Imagen del usuario	Nombre / Apellido(s)	Grupos	Fecha	(1) Cumplió con la forma de ...	(2) Respondió las dudas con ...	(3) Retroalimentó de forma ...	(4) Fue amable y cordial en sus...	(5) Subió en los tiempos ...	(6) Evaluó de forma clara y ...	(7) Motivó el interés hacia el ...	(8) Ayudaron a cumplir los ...	(9) Facilitaron la comprensión ...
		2021-05-03_SQLBasico_CL_wsv	viernes, 4 de junio de 2021, 23:41	Siempre	Siempre	Siempre	Siempre	Siempre	Siempre	Siempre	Muy de acuerdo	Muy de acuerdo
		2021-05-03_SQLBasico_CL_wsv	sábado, 5 de junio de 2021, 11:08	Siempre	Siempre	Siempre	Siempre	Siempre	Siempre	Siempre	Muy de acuerdo	Muy de acuerdo
		2021-05-03_SQLBasico_CL_wsv	domingo, 30 de mayo de 2021, 19:56	Casi siempre	Casi siempre	Algunas veces	Casi siempre	Casi siempre	Casi siempre	Algunas veces	En desacuerdo	En desacuerdo
		2021-05-03_SQLBasico_CL_wsv	domingo, 30 de mayo de 2021, 23:35	Siempre	Casi siempre	Siempre	Siempre	Siempre	Siempre	Siempre	De acuerdo	De acuerdo

Figura 2.6 Respuestas de un cuestionario de opinión.

Procesamiento de las respuestas

En esta primera etapa se procesan las respuestas de cada alumno para convertirlas en un valor numérico que podrá ser analizado posteriormente.

Para ello se asigna un valor numérico a cada tipo de respuesta de la siguiente manera:

Respuesta	Valor
Siempre	4
Casi siempre	3
Algunas veces	2
Casi nunca	1
Nunca	0

Por lo cual el máximo valor posible es 4 para cada pregunta.

El siguiente paso será definir:

$$R_{ij} = \text{La respuesta } i \text{ en la categoría } j$$

Las categorías j son las definidas antes: asesor, materiales, plataforma, evaluaciones, participante y curso. Mientras que la cantidad de i 's dependerá de los n alumnos que hayan respondido el cuestionario.

Con la definición de la variable R_{ij} podemos definir los índices de cada categoría de la siguiente manera:

$$C_{ij} = \frac{\sum_{i=0}^n R_{ij}}{4 \cdot n}$$

Por la definición realizada el valor de C_{ij} será un valor entre 0 y 1, es decir, podemos definirlo como un porcentaje: **Índice de satisfacción**. Donde un valor de 100% indica que los alumnos del grupo estuvieron completamente satisfechos con la categoría j.

Asimismo, realizamos dos cálculos extra para este análisis del cuestionario:

$$\text{Respuestas} = \frac{\text{Cantidad de respuestas obtenidas}}{\text{Tamaño del grupo}}$$

General = Promedio de las categorías

Con estos cálculos, para cada curso podemos construir una tabla como la siguiente:

Num.	Nombre	Curso	Inicio	Fin	Respuestas	Asesor	Materiales	Plataforma	Evaluaciones	Participante	Curso	General
1	Miguel Ordaz	Conoce tu computadora	02 de septiembre del 2019	29 de septiembre del 2019	7.14%	43.88%	86.9%	83.33%	89.29%	86.61%	94.64%	80.78%

Figura 2.7 Ejemplo de resultados obtenidos

Este procesamiento de cuestionario se debe realizar para cada uno de los cursos impartidos, los datos obtenidos en cada uno de los grupos son depositados en una base de datos distinta que almacena los resultados históricos de cada grupo impartido.

Análisis estadístico

Para ejemplificar el análisis a realizar no utilizaremos los datos de todos los cursos, nos enfocaremos en un solo curso, sobre este realizaremos un análisis y obtendremos los resultados correspondientes. Sin embargo, el análisis de este curso se puede replicar hacia los demás cursos sin mayor problema.

De los resultados obtenidos al procesar los cuestionarios, utilizaremos tres columnas:

- Asesor. Se refiere a la percepción que del asesor.
- Materiales. Esta columna indica la calidad de los materiales por parte de los alumnos.
- Plataforma. Es la percepción de los alumnos respecto del sitio donde se imparte el curso.

A esta columna le agregaremos una adicional que es la eficiencia terminal de un curso, recordemos que este es el porcentaje de actividades que realizaron los alumnos al finalizar el curso.

Un porcentaje alto de este índice nos dice que todos los alumnos hicieron todas las actividades correspondientes, pero uno bajo nos indica que los alumnos desertaron, no pudieron realizar alguna actividad o encontraron los materiales necesarios para realizar las actividades.

Los datos elegidos de uno de los cursos se muestran en la siguiente tabla:

Eficiencia terminal	Asesor	Materiales	Plataforma
0.8	1	0.6667	0.9444
0.87	0.9241	0.7188	0.9688
0.7	0.1786	0.8333	1
0.56	0.625	0.7083	1
0.49	0.8643	0.65	0.9167
0.84	0.8661	0.6667	0.875
0.61	0.9583	0.8056	0.9722
0.65	0.5833	0.8333	1
0.39	0.8452	0.75	0.8333
0.74	0.7411	0.5	0.8333
0.71	1	0.8889	0.9722

Para realizar el análisis de los datos calcularemos las covarianzas de la Eficiencia terminal, Asesor, Materiales y Plataforma. Esto con el propósito de conocer si existe una relación entre los índices de los asesores, materiales y plataforma y el grado de finalización de un curso.

Las covarianzas se muestran a continuación:

Asesor	0.0037
Materiales	-0.0023
Plataforma	0.0015

De acuerdo con los resultados observados podemos ver que los resultados son muy bajos, por lo cual, la correlación entre el grado de finalización de este curso y las respuestas obtenidas por los cuestionarios de opinión es muy pequeña, esto porque el resultado de las covarianzas es cercano a cero

Análisis de los resultados

De acuerdo con procedimiento planteado, al procesamiento de las respuestas y al modelado de este; esperábamos que las variables recopiladas en los cuestionarios de opinión pudieran explicar el grado de finalización de un curso.

Sin embargo, esto no fue así ya que los resultados arrojados por las covarianzas no indicaron evidencia significativa de la relación entre estos. Aunque, en lo personal, esperaba que fueran a tener relación y que a través de estos datos podríamos encontrar una explicación a los grados de finalización de curso.

Un análisis más detallado puede explicar porque el resultado no fue el esperado, y la explicación no tiene que ver con los datos en sí mismos o en la forma en que fueron construidos sino en lo que tratamos de explicar con estos datos.

El problema con el diseño y la toma de datos es un problema desde la concepción de este, esto porque estamos asumiendo que todos los estudiantes respondieron este cuestionario, lo cual no es así. Como mencionamos antes este cuestionario se aplica al finalizar el curso y es opcional, por este motivo no podemos esperar que las respuestas de los cuestionarios nos expliquen el grado de finalización con una actividad que está dentro de lo que deseamos explicar.

En otras palabras, el diseño expuesto en este trabajo pretendía explicar las actividades realizadas con una actividad que deben realizar al finalizar el curso. Lo cual no tiene mucho sentido.

¿Qué hacer? ¿Cómo se puede explicar? Lo que se propondrá, más adelante, en este trabajo es recopilar información de variables que no implican la realización de una actividad por parte de los estudiantes, sino que recopilan la información de otras fuentes, por ejemplo, que tan puntual entregan sus actividades, tiempo dedicado a revisar materiales en el sitio, participación en las actividades grupales, cuanto tarda un profesor en calificar, constancia del profesor dentro del curso, entre otras.

Todas estas variables propuestas no implican que los estudiantes y asesores realicen alguna actividad extra sino es solo observar, medir y recopilar lo que está sucediendo, para después utilizar estas medidas y revisar estadísticamente si existe algún tipo de relación entre ellas.

De alguna manera lo que se pretende es modelar el comportamiento que deben tener los asesores y alumnos para conseguir que el grado de finalización sea satisfactorio y cuando el comportamiento se aparte de los parámetros deseados hacer saltar “las alarmas” para prevenir cualquier tipo de problema posterior.

2.3 Calendarización de cursos

La calendarización de cursos es la forma en que se muestran los posibles cursos a los cuales se puede inscribir cualquier interesado en tomar un curso en línea.



Figura 2.8 Calendario del sitio de cursos

Esta calendarización de cursos se muestra de manera mensual a los visitantes del sitio, pero es un proceso que se realiza semestralmente, es decir, todos los cursos del semestre ya se han programado para iniciar en un periodo específico.

Actualmente, esta calendarización de cursos se realiza de forma manual, es decir, a prueba y error se van asignando los cursos para acomodarlos de acuerdo con las características requeridas por la calendarización.

Sin embargo, en esta parte propondremos que esto se puede modelar utilizando un sencillo modelo de asignación, el cual puede cumplir con todas las características requeridas actualmente.

Para el desarrollo de este modelo asumiremos las siguientes características del calendario:

- A lo largo del semestre se tienen hasta 5 periodos, esto depende de si es el primer semestre del año, el segundo, los periodos vacacionales, etc.
- Los posibles cursos son los siguientes:

Categoría	Num (n_k)	Nombre del curso
1	1	Conoce tu computadora
1	2	Servicios de internet
2	1	Dibujando con PowerPoint
2	2	Elaboración de formatos para la oficina con WORD
2	3	Elaboración de gráficos con EXCEL
2	4	Manejo de máquinas virtuales
2	5	Tablas y gráficos dinámicos con EXCEL

Categoría	Num (n_k)	Nombre del curso
2	6	Taller de funciones con Excel
3	1	Creación de páginas web con HTML
3	2	Desarrollo de cursos con MOODLE
3	3	Hoja de cálculo EXCEL
3	4	Introducción a la programación
3	5	Macros con EXCEL
3	6	Manipulación de imágenes con PHOTOSHOP
3	7	SQL estándar básico
3	8	Técnicas avanzadas con Excel
3	9	Tecnologías para la capacitación a distancia

- De acuerdo con la categoría a la que pertenece cada curso se deben cumplir las siguientes condiciones:
 1. En cada periodo debe haber un sólo curso de la categoría uno.
 2. Deben programarse al menos 3 cursos de la categoría dos cada periodo.
 3. Deben programarse al menos 4 cursos de la categoría tres cada periodo.
 4. Los cursos programados en el periodo actual no pueden estar en la programación del siguiente periodo.

El objetivo de la calendarización es programar la mayor cantidad de cursos cumpliendo las condiciones dadas.

Modelado del problema

De acuerdo con los párrafos anteriores el problema de la calendarización de cursos es un problema de asignación, donde los cursos deben ser asignados a un periodo en particular para cumplir las restricciones propuestas: cantidad de cursos, orden, etc. Sin embargo, una característica del problema de calendarización con el que estamos tratando es que no tiene ninguna ponderación, es decir, basta con que se cumplan los criterios de las restricciones propuestas para alcanzar la solución deseada.

Más adelante en este trabajo utilizaremos como ponderación para esta calendarización la eficiencia terminal de los cursos descrita antes. Sin embargo, en esta parte nos limitaremos a modelar el problema como se realiza actualmente: sin ninguna ponderación.

Este problema es completamente lineal, por lo cual para resolverlo utilizaremos un modelo de **programación lineal**, en particular un modelo de **programación lineal entera binario**. Para lo cual definimos las siguientes variables:

$$\text{Sea } x_{kij} = \begin{cases} 1 & \text{Si se utiliza el curso } i \text{ que pertenece a la categoría } k \\ & \text{en el periodo } j. \\ 0 & \text{Si no se utiliza} \end{cases}$$

$$\forall k \in \{1, 2, 3\}$$

$$\forall i \in \{1, 2, \dots, n_k\} \text{ donde } n_k \text{ es la cantidad de cursos de la categoría } k$$

$$\forall j \in \{1, 2, \dots, 5\}$$

La **función objetivo** es la siguiente:

$$\text{Max } z = \sum_{k=1}^3 \sum_{i=1}^{n_k} \sum_{j=1}^5 x_{kij}$$

Esta función objetivo corresponde a la triple de suma de las variables binarias x_{kij} , la cual tiene el propósito de maximizar la cantidad de cursos que pueden programarse para todas las categorías y periodos que se tienen disponibles.

Sin embargo, esta función objetivo está sujeta a las siguientes restricciones:

Condición 1	Condición 2	Condición 3
$x_{111} + x_{121} = 1$	$\sum_{i=1}^6 x_{2i1} \geq 3$	$\sum_{i=1}^9 x_{3i1} \geq 4$
$x_{112} + x_{122} = 1$	$\sum_{i=1}^6 x_{2i2} \geq 3$	$\sum_{i=1}^9 x_{3i2} \geq 4$
$x_{113} + x_{123} = 1$	$\sum_{i=1}^6 x_{2i3} \geq 3$	$\sum_{i=1}^9 x_{3i3} \geq 4$
$x_{114} + x_{124} = 1$	$\sum_{i=1}^6 x_{2i4} \geq 3$	$\sum_{i=1}^9 x_{3i4} \geq 4$
$x_{115} + x_{125} = 1$	$\sum_{i=1}^6 x_{2i5} \geq 3$	$\sum_{i=1}^9 x_{3i5} \geq 4$

Condición 4			
$x_{111} + x_{112} = 1$	$x_{232} + x_{233} = 1$	$x_{313} + x_{314} = 1$	$x_{354} + x_{355} = 1$
$x_{112} + x_{113} = 1$	$x_{233} + x_{234} = 1$	$x_{314} + x_{315} = 1$	$x_{361} + x_{362} = 1$
$x_{113} + x_{114} = 1$	$x_{234} + x_{235} = 1$	$x_{321} + x_{322} = 1$	$x_{362} + x_{363} = 1$
$x_{114} + x_{115} = 1$	$x_{241} + x_{242} = 1$	$x_{322} + x_{323} = 1$	$x_{363} + x_{364} = 1$
$x_{121} + x_{122} = 1$	$x_{242} + x_{243} = 1$	$x_{323} + x_{324} = 1$	$x_{364} + x_{365} = 1$
$x_{122} + x_{123} = 1$	$x_{243} + x_{244} = 1$	$x_{324} + x_{325} = 1$	$x_{371} + x_{372} = 1$
$x_{123} + x_{124} = 1$	$x_{244} + x_{245} = 1$	$x_{331} + x_{332} = 1$	$x_{372} + x_{373} = 1$
$x_{124} + x_{125} = 1$	$x_{251} + x_{252} = 1$	$x_{332} + x_{333} = 1$	$x_{373} + x_{374} = 1$
$x_{211} + x_{212} = 1$	$x_{252} + x_{253} = 1$	$x_{333} + x_{334} = 1$	$x_{374} + x_{375} = 1$
$x_{212} + x_{213} = 1$	$x_{253} + x_{254} = 1$	$x_{334} + x_{335} = 1$	$x_{381} + x_{382} = 1$
$x_{213} + x_{214} = 1$	$x_{254} + x_{255} = 1$	$x_{341} + x_{342} = 1$	$x_{382} + x_{383} = 1$
$x_{214} + x_{215} = 1$	$x_{261} + x_{262} = 1$	$x_{342} + x_{343} = 1$	$x_{383} + x_{384} = 1$
$x_{221} + x_{222} = 1$	$x_{262} + x_{263} = 1$	$x_{343} + x_{344} = 1$	$x_{384} + x_{385} = 1$
$x_{222} + x_{223} = 1$	$x_{263} + x_{264} = 1$	$x_{344} + x_{345} = 1$	$x_{391} + x_{392} = 1$
$x_{223} + x_{224} = 1$	$x_{264} + x_{265} = 1$	$x_{351} + x_{352} = 1$	$x_{392} + x_{393} = 1$
$x_{224} + x_{225} = 1$	$x_{311} + x_{312} = 1$	$x_{352} + x_{353} = 1$	$x_{393} + x_{394} = 1$
$x_{231} + x_{232} = 1$	$x_{312} + x_{313} = 1$	$x_{353} + x_{354} = 1$	$x_{394} + x_{395} = 1$

Solución del problema

Para resolver este problema de programación lineal entera utilizamos el software Lindo, en su versión 6.1.

```

LINDO
File Edit Solve Reports Window Help
C:\USERS\ALEX\DESKTOP\MOPF.lbx
max x111 + x112 + x113 + x114 + x115 +
x121 + x122 + x123 + x124 + x125 +
x211 + x212 + x213 + x214 + x215 + x221 + x222 + x223 + x224 + x225 +
x231 + x232 + x233 + x234 + x235 + x241 + x242 + x243 + x244 + x245 +
x251 + x252 + x253 + x254 + x255 + x261 + x262 + x263 + x264 + x265 +
x311 + x312 + x313 + x314 + x315 + x321 + x322 + x323 + x324 + x325 +
x331 + x332 + x333 + x334 + x335 + x341 + x342 + x343 + x344 + x345 +
x351 + x352 + x353 + x354 + x355 + x361 + x362 + x363 + x364 + x365 +
x371 + x372 + x373 + x374 + x375 + x381 + x382 + x383 + x384 + x385 +
x391 + x392 + x393 + x394 + x395
st
x111 + x121 = 1
x112 + x122 = 1
x113 + x123 = 1
x114 + x124 = 1
x115 + x125 = 1
x211 + x221 + x231 + x241 + x251 + x261 >= 3
x212 + x222 + x232 + x242 + x252 + x262 >= 3
x213 + x223 + x233 + x243 + x253 + x263 >= 3
x214 + x224 + x234 + x244 + x254 + x264 >= 3
x215 + x225 + x235 + x245 + x255 + x265 >= 3

```

Figura 2.9 Planteamiento en Lingo

En el cual se introdujeron las 85 variables de la función objetivo, las 5 condiciones de la restricción 1, 2 y 3, respectivamente; y las 68 condiciones de la restricción 4.

En total, para cumplir las cuatro condiciones fue necesario utilizar 73 restricciones para realizar la calendarización de 17 cursos, divididos en 3 categorías a lo largo de 5 periodos.

El resultado fue el siguiente:

OBJECTIVE FUNCTION VALUE: 43.00000

VARIABLE	VALUE	VARIABLE	VALUE	VARIABLE	VALUE	VARIABLE	VALUE
X111	0	X234	0	X322	0	X365	0
X112	1	X235	1	X323	1	X371	1
X113	0	X241	1	X324	0	X372	0
X114	1	X242	0	X325	1	X373	1
X115	0	X243	1	X331	0	X374	0
X121	1	X244	0	X332	1	X375	1
X122	0	X245	1	X333	0	X381	0
X123	1	X251	0	X334	1	X382	1
X124	0	X252	1	X335	0	X383	0
X125	1	X253	0	X341	1	X384	1
X211	0	X254	1	X342	0	X385	0
X212	1	X255	0	X343	1	X391	1
X213	0	X261	0	X344	0	X392	0
X214	1	X262	1	X345	1	X393	1
X215	0	X263	0	X351	1	X394	0
X221	1	X264	1	X352	0	X395	1
X222	0	X265	0	X353	1		
X223	1	X311	0	X354	0		
X224	0	X312	1	X355	1		
X225	1	X313	0	X361	0		
X231	1	X314	1	X362	1		
X232	0	X315	0	X363	0		
X233	1	X321	1	X364	1		

Figura 2.10 Resumen de la solución arrojada por LINDO

Análisis de resultados

De acuerdo con los resultados arrojados por el software tenemos que la cantidad máxima de cursos que podemos calendarizar son 43.

Asimismo, los cursos deben ser programados de la siguiente forma:

Curso	Periodos				
	1	2	3	4	5
Conoce tu computadora		Sí		Sí	
Servicios de internet	Sí		Sí		Sí
Dibujando con PowerPoint		Sí		Sí	
Elaboración de formatos para la oficina con WORD	Sí		Sí		Sí
Elaboración de gráficos con EXCEL	Sí		Sí		Sí
Manejo de máquinas virtuales	Sí		Sí		Sí
Tablas y gráficos dinámicos con EXCEL		Sí		Sí	
Taller de funciones con Excel		Sí		Sí	
Creación de páginas web con HTML		Sí		Sí	
Desarrollo de cursos con MOODLE	Sí		Sí		Sí
Hoja de cálculo EXCEL		Sí		Sí	
Introducción a la programación	Sí		Sí		Sí
Macros con EXCEL	Sí		Sí		Sí
Manipulación de imágenes con PHOTOSHOP		Sí		Sí	
SQL estándar básico	Sí		Sí		Sí
Técnicas avanzadas con Excel		Sí		Sí	
Tecnologías para la capacitación a distancia	Sí		Sí		Sí

Realizando un resumen de la tabla anterior tenemos que la cantidad de cursos que se programaron en cada periodo por cada categoría son:

	Periodo 1	Periodo 2	Periodo 3	Periodo 4	Periodo 5
Categoría 1	1	1	1	1	1
Categoría 2	3	3	3	3	3
Categoría 3	5	4	5	4	5

Lo cual cumple con las condiciones solicitadas.

Actualmente, la calendarización se centra únicamente en conocer el acomodo de los cursos sin tomar en cuenta ninguna ponderación en cada uno de ellos. Sin embargo, podría incorporarse en la función objetivo del modelo algún tipo de ponderación para distinguir que curso podemos preferir sobre otro.

La ponderación que podría incorporarse puede ser el costo del curso, por lo cual preferiríamos colocar cursos cuyo costo sea mayor; otra opción puede ser utilizar los cursos que abren con mayor frecuencia, o los cursos que tienen más alumnos, etc.

Para el trabajo de tesis que se presenta se realizará de la siguiente manera:

Realizar una calendarización que maximice la eficiencia terminal de los cursos

Para realizar esta calendarización es necesario:

- Calcular la eficiencia terminal de los asesores. Esto nos permitirá asignar los asesores con la mayor eficiencia en los cursos que se abren regularmente.
- Calcular el índice de apertura de un curso. Este nos indicará la probabilidad de apertura de un curso según los datos observados.
- Crear un modelo de mayor complejidad que deberá tomar en cuenta las eficiencias terminales de los asesores, los índices de apertura de los cursos y las restricciones sobre las categorías de cursos mostradas en este capítulo.

Por esta razón en los siguientes capítulos se propone una metodología para recopilar la información de la eficiencia terminal y revisar que esta no es aleatoria, es decir, depende en gran medida de lo que sucede durante la impartición de un curso.

Asimismo, definiremos e índice de apertura de un curso y extenderemos el modelo de asignación para agregar más variables, lo cual incrementará la complejidad de este.

Capítulo 3. Metodología propuesta

En el capítulo anterior revisamos el concepto de eficiencia terminal, este concepto es fundamental para la óptima calendarización de los cursos, ya que sobre este centraremos el modelo de optimización de esta tesis.

Por lo cual durante este capítulo detallaremos la metodología general que utilizaremos para la recolección de datos, el procesamiento de estos y su uso para el modelo de asignación propuesto.

3.1 Introducción

La eficiencia terminal de un curso no va sola, a lo largo de la impartición de un curso existen muchos factores que pueden afectar esta eficiencia: la cantidad de alumnos, las faltas del asesor, los retrasos en las entregas, en las calificaciones, etc.

Estos factores no se reflejan en el cálculo descrito para dicha eficiencia, pero si los observamos, medimos y recopilamos podemos realizar un análisis posterior y determinar si son determinantes en el número obtenido al calcular la eficiencia terminal.

Por este motivo para la recopilación de los datos, para el cálculo de la eficiencia y de los factores que ocurren, es necesaria de una metodología que considere las reglas de impartición de los cursos y la definición realizada en el capítulo anterior.

Asimismo, en un principio este proceso de recolección de datos podrá parecer simple; pero, como se describirá más adelante, este proceso debe realizarse para todos los cursos que se impartan en cada uno de los días que dicho curso este vigente. Lo cual ocasionará que la simplicidad planteada al principio se convierta en un proceso complejo debido a la cantidad de datos que se generan y que se deben procesar y calcular.

De forma general la metodología propuesta se ve de la siguiente manera.

1. Apertura de cursos. No todos los cursos que son programados se abren, por lo cual en este punto conocemos cuales son los cursos que se impartirán. En el siguiente capítulo de este trabajo utilizaremos los datos de los cursos abiertos para crear un índice que utilizaremos en el planteamiento del modelo de optimización.
2. Seguimiento de cursos. Durante la impartición de los cursos se debe realizar un seguimiento de lo que sucede durante este. En este capítulo desarrollaremos la metodología para realizar este seguimiento.
3. Procesamiento de datos. Después de la recolección de datos durante el seguimiento del curso es necesario procesarlos para calcular la Eficiencia terminal del curso, la cual, aunque es solo observar las actividades resultantes al final de un curso, es necesario determinar que lo sucedido durante la impartición del curso tiene un impacto en la Eficiencia terminal final del curso, para lo cual utilizaremos un modelo de regresión línea múltiple.
4. Modelo de optimización. Una vez que los datos han sido recolectados y procesados el siguiente paso es utilizarlos para realizar un modelo de optimización que maximice la eficiencia terminal de un conjunto de periodos. Para ello utilizaremos

un modelo de asignación de cursos de forma que este modelo nos indique los asesores que deben estar asignados en cada curso desde la calendarización de cursos para un conjunto de periodos dado.

5. Calendarización de cursos. Esta programación de cursos para un conjunto de periodos nos indica los tiempos en que cada curso y asesor deben ser programados para maximizar la eficiencia terminal cumpliendo con las condiciones requeridas para realizar dicha programación: cantidad de cursos, periodos, categorías de cursos, etc. Esta última parte es el objetivo general del trabajo porque una vez que hemos realizado el seguimiento y procesamiento de los datos los utilizamos para obtener una programación óptima de los cursos.

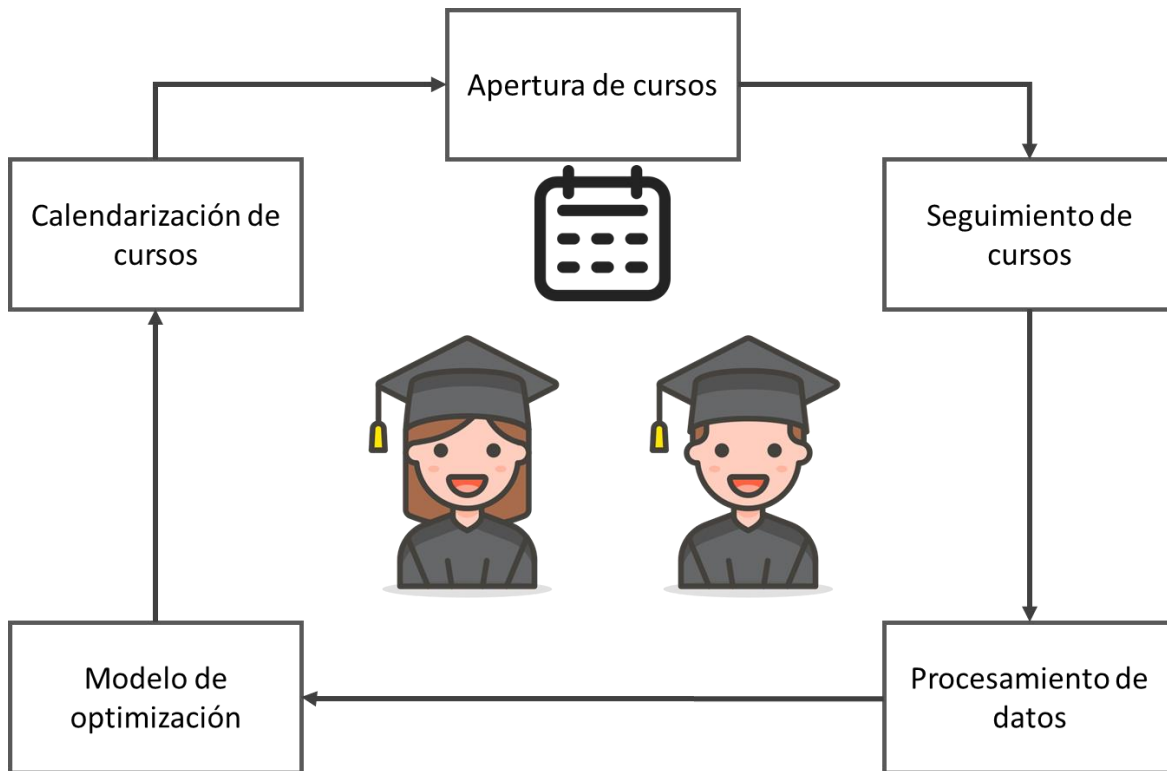


Figura 3.1 Diagrama de la metodología propuesta.

En este capítulo nos centraremos en revisar la metodología para el seguimiento de cursos: recopilación de datos, procesamiento de los datos obtenidos y realizaremos un análisis de descriptivo de los mismos.

Antes de comenzar con la metodología se recomienda revisar el Apéndice A, en este se describen los conceptos y herramientas utilizados por las diversas etapas de la metodología propuesta: sistemas de información, procesos para la recopilación de información, modelo de regresión, modelo de optimización.

Comenzaremos con el seguimiento de cursos porque a partir de los datos obtenidos durante la impartición de un curso tendremos los elementos necesarios para alimentar el modelo de optimización general.

3.2 Metodología para el seguimiento de cursos

Como se mencionó antes son muchos los factores que influyen en el desempeño de un curso, por lo cual para el presente trabajo se han seleccionado los siguientes factores, los cuales pueden ser recolectados directamente del sistema donde se imparten los cursos en línea.

1. **Eficiencia terminal (Ef)**. Último *índice de avance* registrado al final del periodo del curso.
2. **Alumnos (A)**. Cantidad de alumnos que tiene el curso.
3. **Retraso 1 (R1)**. Cantidad de retrasos del tipo “No se entregan actividades” observados.
4. **Retraso 2 (R2)**. Cantidad de retrasos del tipo “Actividades sin calificar” observados.
5. **Retraso 3 (R3)**. Cantidad de retrasos del tipo “Inasistencia de alumnos” observados.

Para realizar la recolección de estos factores en cada curso fue necesario realizar la extracción de la información almacenada en la base de datos y procesarla a través de un lenguaje de programación.

En este caso la información de la base de datos estaba almacenada en un manejador MaríaDB y se utilizó el lenguaje de programación PHP para desplegar los resultados en una página web.

El sistema donde se almacena la información es un LMS Moodle, cuya base de datos tiene 455 tablas, de las cuales solo se utilizaron las siguientes para obtener la información necesaria:

- mdl_grade_items. Esta tabla contiene las actividades que serán calificadas.
- mdl_user. Esta tabla almacena los datos de los usuarios de la plataforma.
- mdl_grade_grades. Esta tabla almacena las calificaciones asignadas a los usuarios en las actividades que requieren calificación.
- mdl_course. Esta tabla contiene la información acerca de los cursos almacenados en la plataforma.
- mdl_groups_members. Esta tabla almacena los miembros de un grupo dentro de un curso.
- mdl_groups. Esta tabla contiene la definición de los grupos dentro de un curso.
- mdl_role_assignments. Esta tabla almacena las asignaciones de rol para los usuarios dentro de un curso en particular.
- mdl_role. Esta tabla contiene la definición de los roles dentro del sitio.

Proceso ETL del seguimiento

El primer paso es la recolección de los datos, el proceso para realizar la extracción de cada curso es el siguiente:

1. Obtener los miembros del grupo.
2. Obtener las actividades que deben entregarse.
3. Obtener las actividades entregadas por los alumnos.
4. Realizar el conteo de las entregas.
5. Calcular el índice de avance del curso.
6. Clasificar el resultado del índice.

El proceso anterior se repite para todos los cursos que tenga un grupo impartándose al momento de solicitar el estatus de los cursos.

Existe un reporte que puede ser generado en cualquier momento, basta con ingresar al enlace correspondiente para realizar lo cálculos antes descritos y transformar los datos almacenados en información sobre el estado de los cursos.

Un ejemplo del resultado de este proceso se muestra a continuación:



Figura 3.2 Reporte de estado de los cursos

Aunque el reporte se genera en cualquier momento, para realizar el análisis de los datos que se presenta en este trabajo se ha establecido una tarea programada para que esta recolección se genere todos los días a las 7 de la mañana para cualquier curso vigente.

El resultado de esta ejecución se guarda en una base de datos distinta, la cual funciona como un *Data Warehouse* con un esquema de estrella, donde las tablas de dimensiones son Curso, Asesor, Actividades y Tiempo. Mientras que la tabla de hechos registra las siguientes métricas:

- Idcurso

- NombreCurso
- Iniciocurso
- idasesor
- Día del curso
- Alumnos
- Accesos
- Faltas
- Indice de avance
- Tiporetraso
- Causa
- Fecha en que se originó el evento.

idCurso	NombreCurso	inicioCurso	idAsesor	día	alumnos	Accesos	Faltas	indiceAvance	tipoRetraso	causa	fecha
25	Creación de páginas web con HTML	2020-01-20	40	8	2	1	0	0.52	Atrasado	Actividades sin calificar	27/01/2020
43	Creación de imágenes con Illustrator	2020-01-20	44	8	3	1	0	0.80	Normal	No se entregan actividades	27/01/2020
7	Hoja de cálculo Excel 2016	2020-01-20	57	8	3	3	0	0.88	Normal	No se entregan actividades	27/01/2020
11	Taller de fórmulas con Excel 2016	2020-01-20	214	8	6	0	1	0.12	Significativo	No se entregan actividades	27/01/2020
26	Taller de ortografía	2020-01-20	55	8	4	1	0	0.32	Significativo	No se entregan actividades	27/01/2020

Figura 3.3 Ejemplo de hechos.

Esta tarea programada genera observaciones diarias para cada curso, las cuales deben ser consolidadas al finalizar el curso con la intención de tener integrados los valores de cada factor para cada curso que haya sido impartido.

3.3 Análisis de los datos

Utilizando el método antes descrito realizaremos el análisis de los datos cargados en el Data Warehouse, exclusivamente sobre los datos correspondientes a los retrasos ya que a partir de las observaciones podremos observar la importancia de los asesores dentro de la impartición de un curso.

En los datos recolectados tenemos un conjunto de registros correspondientes a todo el año 2020, durante este año se registraron 4776 hechos, es decir, se observó algún tipo de incidencia: faltas, retrasos, no se entregan actividades, etc. Los cuales corresponden a 61 cursos impartidos en 27 periodos distribuidos a lo largo del año.

Fecha	Cursos	Alumno	Significativo	Atrasado	Normal	Sobresaliente	Actividades sin calificar	Inasistencia	No se entregan actividades
27/01/2020	5	18	2	1	2	0	1	0	4
28/01/2020	7	34	1	3	3	0	2	3	2
29/01/2020	9	42	2	1	5	1	1	3	4
30/01/2020	9	42	2	1	5	1	1	3	4
31/01/2020	10	46	3	1	3	3	3	2	2
01/02/2020	10	46	3	1	3	3	3	2	2
02/02/2020	10	46	3	1	3	3	4	3	0
03/02/2020	9	35	6	1	2	0	4	1	4
04/02/2020	9	35	4	3	1	1	2	5	1
05/02/2020	9	35	3	3	2	1	2	4	2
06/02/2020	9	35	2	2	4	1	1	3	4

Figura 3.4 Datos analizados

A partir de los datos recolectados y del procesamiento para analizar los retrasos observados realizamos un análisis descriptivo y visual de estos.

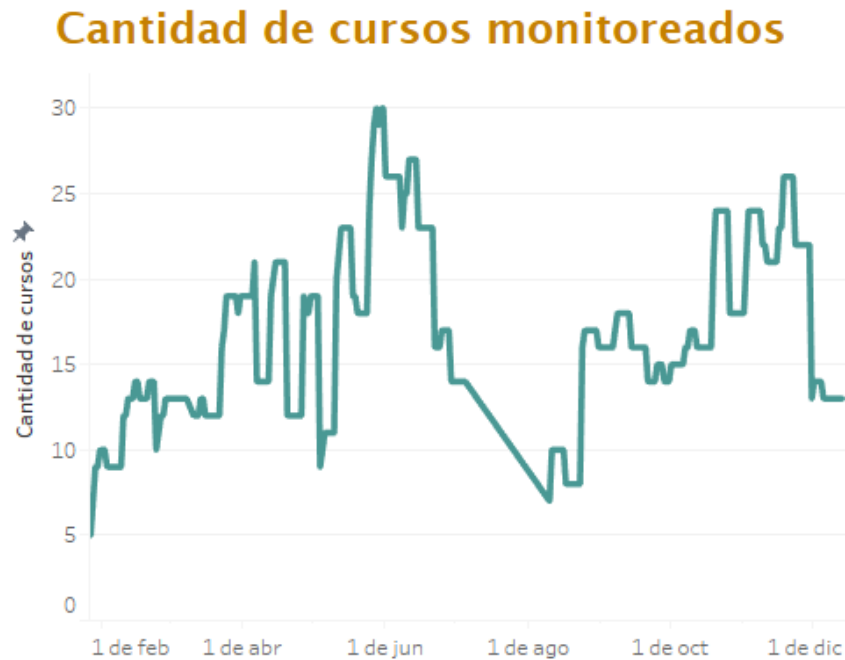


Figura 3.5 Cursos durante 2020.

La figura 3.5 muestra la cantidad de cursos que requirieron un monitoreo constante durante el año 2020, como se puede observar hay claramente dos periodos definidos dentro del año, esto se debe a que durante las vacaciones administrativas de la Universidad se suspende todo tipo de actividad, incluso en línea.

Asimismo, vemos que durante algunos periodos es necesario realizar el monitoreo de más de 25 cursos, lo cual podría requerir de personas dedicadas por completo a ello.

En la siguiente figura se muestran los índices observados durante el año 2020, los cuales fueron contados después haber sido catalogados según los criterios antes mencionados.

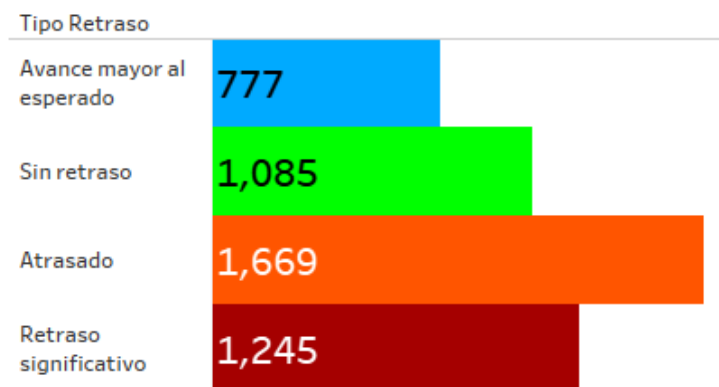


Figura 3.6 Índice de avance observados durante en el 2020

Lo que podemos observar es que el 61% de todas las observaciones registradas presentó algún tipo de retraso, incluso, el 26% de las observaciones corresponden a un retraso significativo, lo que probablemente implica que más de la cuarta parte de los cursos impartidos no alcanzaron un índice de avance satisfactorio al final del curso.

Lo anterior nos lleva a realizar una revisión del mismo tipo, pero por cada uno de los cursos impartidos durante el año.

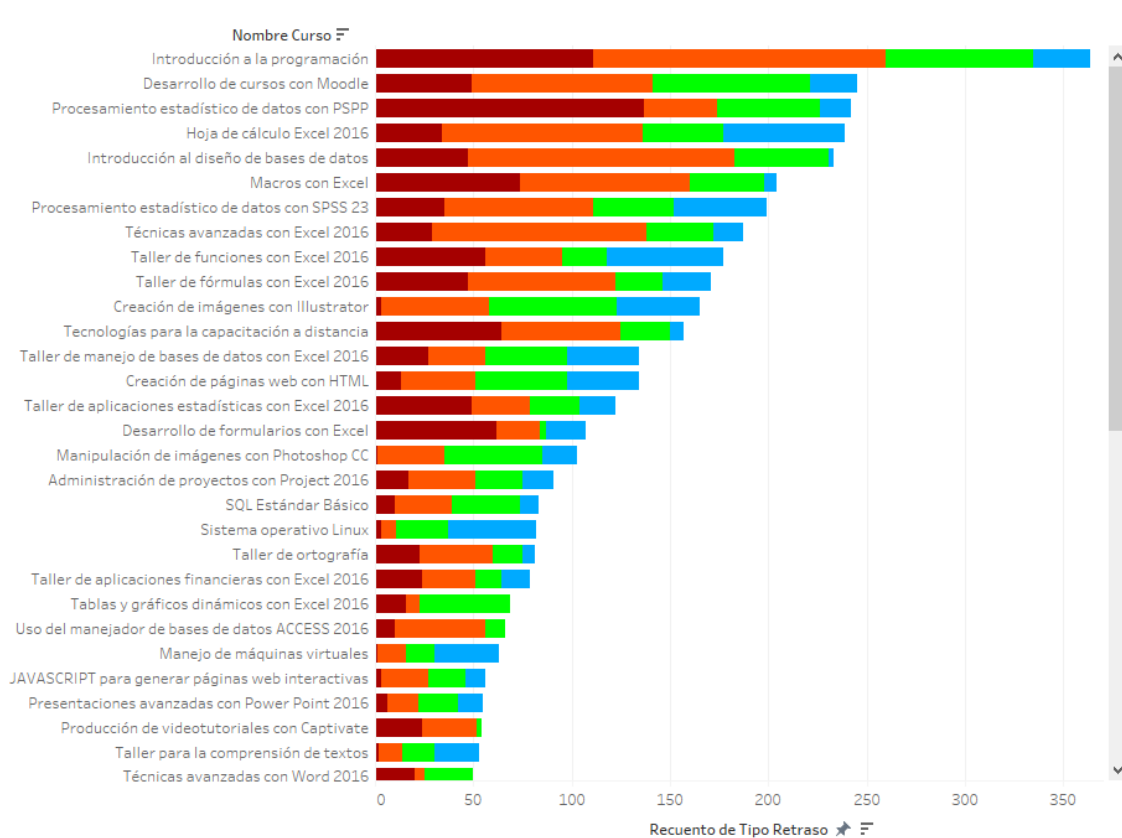


Figura 3.7 Retrasos observados por curso.

La figura 18 nos muestra los cursos ordenados por la cantidad de observaciones realizadas, en este caso un mayor número observaciones corresponden a un mayor número de veces que fue impartido.

La codificación de colores por curso corresponde a la figura 4, por lo que podemos observar de acuerdo con estos colores, aquellos cursos que presentan una mayor cantidad de retrasos, e incluso, cursos que siempre presentan algún tipo de retraso.

De forma análoga realizamos el mismo análisis, pero esta vez por asesor, la siguiente figura muestra los resultados.

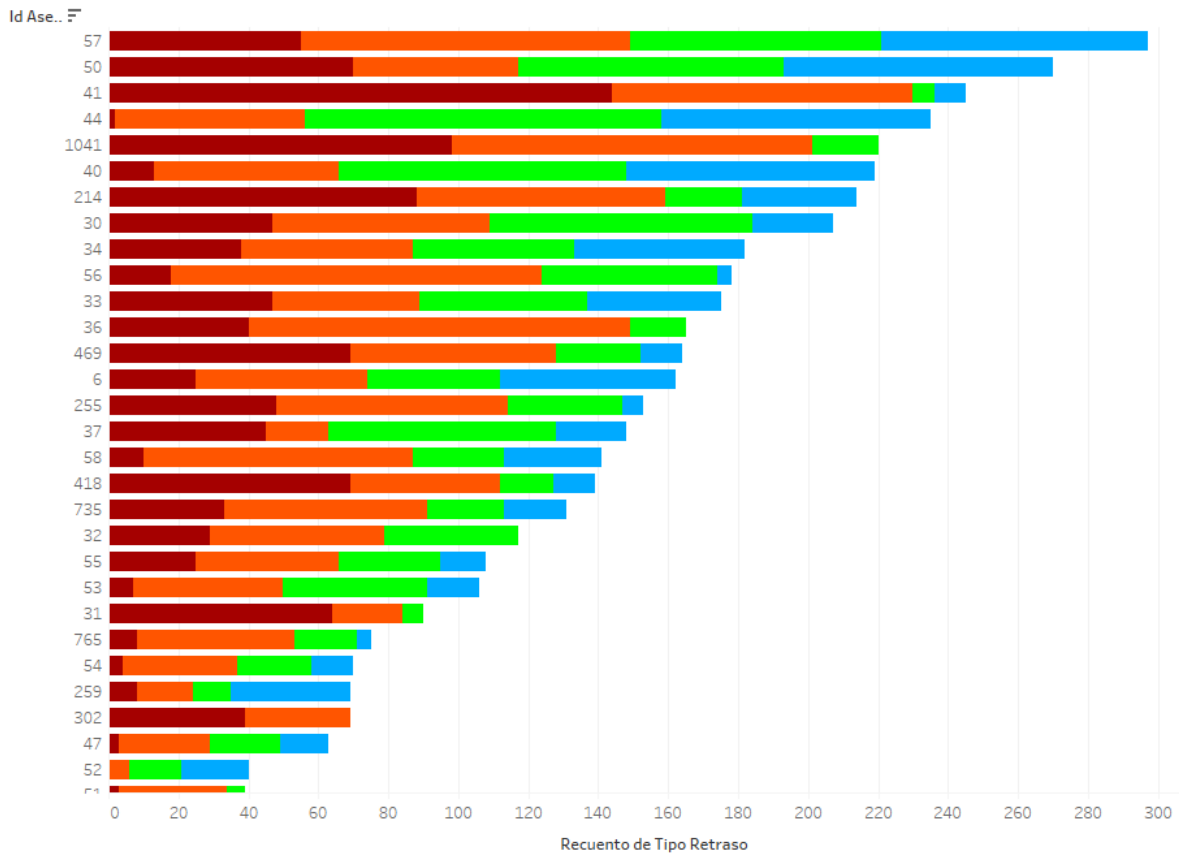


Figura 3.8 Retrasos observados por asesor

Esta figura 19 nos permite identificar los valores observados en los retrasos de los asesores, al igual que pasaba con los cursos, el tamaño de barra implica una mayor cantidad de cursos asesorados; y de acuerdo con la codificación de colores, esperamos que un asesor tenga un mayor tamaño de colores verde y azul, ya que estos implican que no tuvo retrasos observados.

Sin embargo, observamos que tenemos muchos casos donde la mayor parte de observaciones corresponden a colores rojo y naranja, es decir, algún tipo de retraso.

El siguiente análisis que realizaremos será de las causas que originaron los retrasos.



Figura 3.9 Causas observadas durante 2020

De manera general, en la figura 20 podemos observar que la inasistencia de alumnos es la mayor causa observada, más del 45%; es decir, la deserción de alumnos es muy alta.

La figura 21 muestra las causas que originaron los retrasos en las figuras anteriores, de manera general observamos dichas causas por mes.

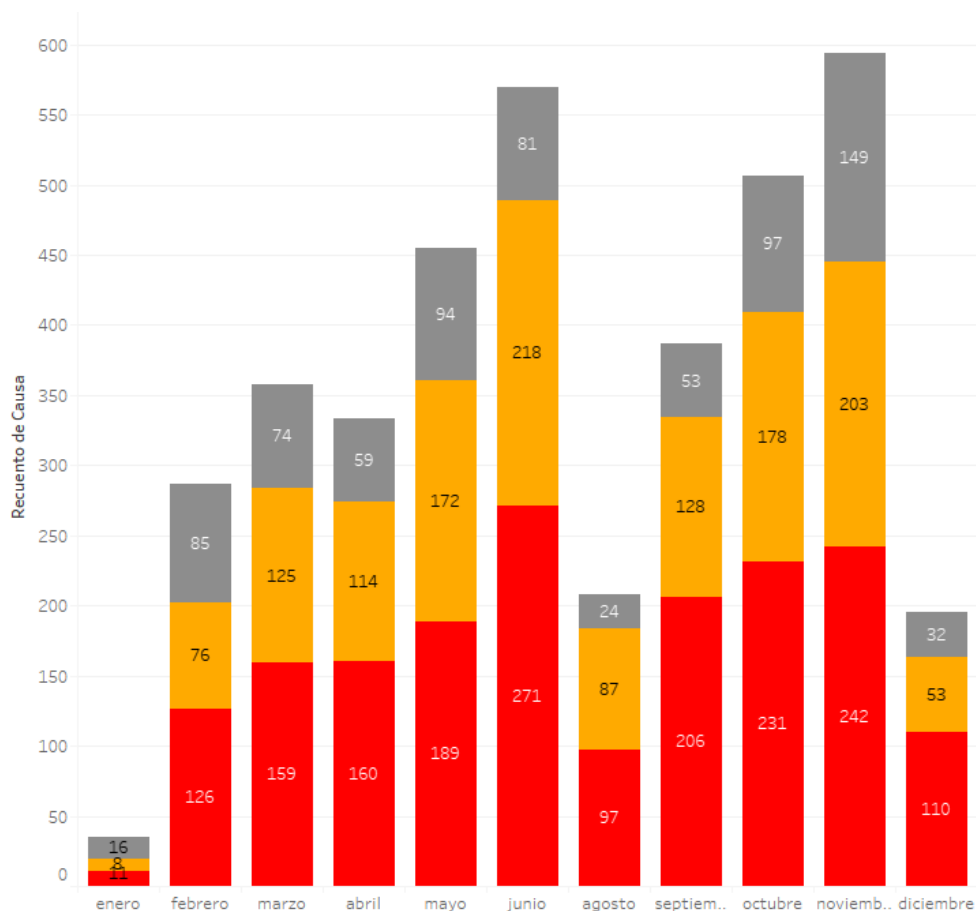


Figura 3.10 Causas de los retrasos por mes

En esta vista mensual observamos que la inasistencia de alumnos se mantuvo de manera constante como el principal motivo de los retrasos durante todo el año. Sin embargo, la causa de retrasos en la calificación también es un motivo importante en los retrasos, y este motivo es atribuible a la tardanza en la evaluación de las actividades por parte de los asesores.

Para revisar lo ocurrido con los retrasos por curso tenemos la siguiente figura.

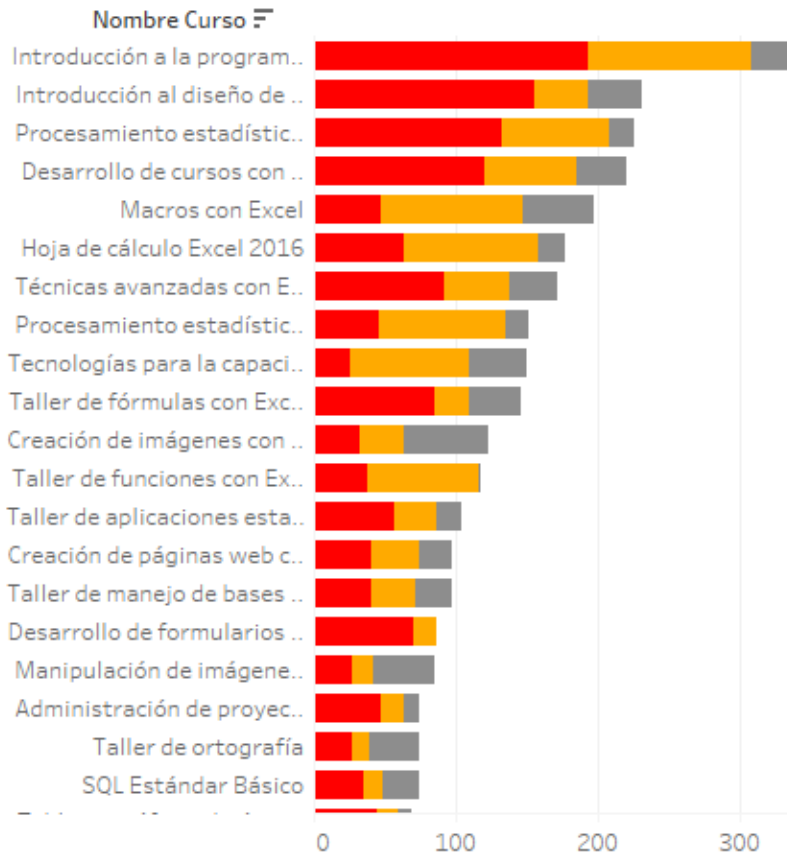


Figura 3.11 Causas observadas por curso

En la figura 11 observamos que la inasistencia de alumnos no es siempre la causa principal para los retrasos, vemos cursos donde la causa principal es el retraso en las calificaciones de las actividades por parte de los asesores.

Por lo cual la siguiente figura muestra las causas observadas por asesor.

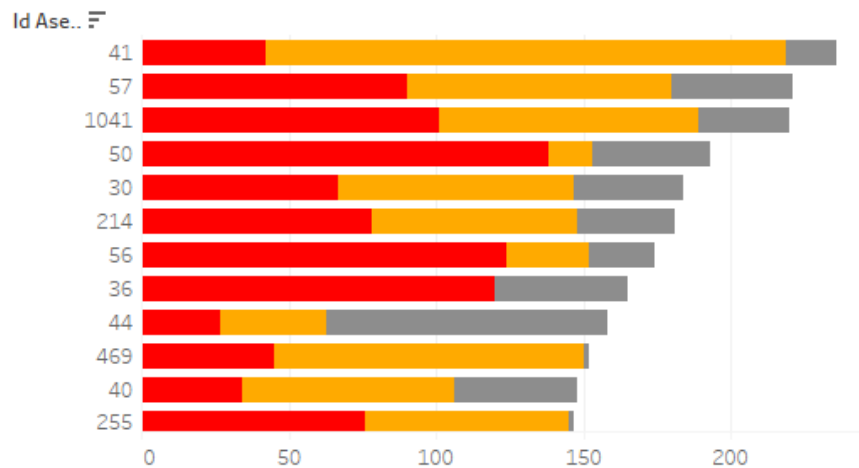


Figura 3.12 Causas observadas por asesor

La figura 23 nos muestra que algunos asesores tienen como principal causa el retraso en sus calificaciones. Esta causa debería siempre ser la menor de ellas, ya que depende por completo del asesor, si este realiza las evaluaciones en tiempo y forma no debería, existir retrasos observables en los cursos por causa de este motivo.

Como se ha observado hasta ahora los asesores son de vital importancia para un curso, ya que el asesor podría reducir de manera significativa los retrasos, los cuales podrían impactar en las otras causas: deserción e inactividad de los alumnos.

Por lo cual, en el siguiente capítulo, a partir de los datos observados en los retrasos añadiremos algunos factores para realizar un análisis de regresión y estimar las relaciones entre los factores elegidos y la eficiencia terminal de cada asesor.

Asimismo, con los datos procesados en el siguiente capítulo realizaremos la propuesta del modelo de optimización general para la asignación de cursos en periodos completos.

Capítulo 4. Resolución del caso de DGTIC

En este capítulo integraremos los datos obtenidos antes, realizaremos un análisis de regresión para verificar que los datos recopilados pueden explicar la eficiencia terminal de los cursos. Finalmente, se propone un modelo para realizar una asignación que optimice la eficiencia terminal desde la programación de los cursos con los datos obtenidos durante el seguimiento y un índice que integraremos en este capítulo.

4.1 Eficiencia terminal de los asesores

Con los datos obtenidos durante el modelo de seguimiento propuesto en el capítulo anterior obtuvimos 59 cursos impartidos en 194 grupos y 41 asesores, un fragmento de estos datos los podemos ver en la siguiente imagen.

cursoPeriodo	curso	periodo	idAsesor	indiceAvance	alumnos	Significativo	Atrasado	Normal	Sobresaliente	Actividades sin calificar	Inasistencia	No se entregan actividades
49_2020-03-30_34	49	2020-03-30	34	100%	2	0	0	2	19	2	0	0
51_2020-03-16_52	51	2020-03-16	52	100%	2	0	2	5	19	0	1	6
25_2020-03-16_40	25	2020-03-16	40	83%	5	5	8	4	8	15	1	1
39_2020-03-30_6	39	2020-03-30	6	97%	8	0	2	4	15	5	0	1
9_2020-03-30_50	9	2020-03-30	50	100%	2	1	0	4	30	3	0	2
43_2020-03-16_44	43	2020-03-16	44	100%	3	1	5	7	14	2	1	10
10_2020-03-30_469	40	2020-03-30	469	76%	14	1	0	9	12	9	1	0
47_2020-05-18_35	47	2020-05-18	35	100%	2	0	0	0	2	0	0	0
24_2020-06-15_35	24	2020-06-15	35	100%	4	0	0	0	4	0	0	0
34_2020-10-12_259	54	2020-10-12	259	100%	2	0	0	0	7	0	0	0
48_2020-05-04_38	48	2020-05-04	38	100%	3	0	0	0	17	0	0	0
6_2020-08-03_33	6	2020-08-03	33	100%	3	0	0	14	14	2	0	12
17_2020-10-26_46	17	2020-10-26	46	100%	4	0	0	1	27	0	0	1
18_2020-05-04_259	18	2020-05-04	259	100%	2	0	0	0	27	0	0	0
25_2020-01-20_40	25	2020-01-20	40	100%	2	1	7	15	5	4	5	14
25_2020-02-17_50	25	2020-02-17	50	100%	4	0	1	6	19	0	0	7
26_2020-01-20_55	26	2020-01-20	55	100%	4	4	3	15	6	3	3	16
27_2020-10-26_40	27	2020-10-26	40	100%	2	2	9	13	4	18	0	6
36_2020-05-18_40	36	2020-05-18	40	100%	2	1	3	5	19	1	1	7
36_2020-09-28_40	36	2020-09-28	40	100%	2	0	3	7	18	2	0	8
40_2020-08-17_6	40	2020-08-17	6	100%	2	0	1	0	27	1	0	0
13_2020-05-18_735	43	2020-05-18	735	100%	2	0	6	11	11	6	1	10
43_2020-10-26_44	43	2020-10-26	44	100%	2	0	4	13	11	3	0	14
44_2020-01-20_41	44	2020-01-20	41	100%	2	15	5	1	5	17	0	4
44_2020-08-03_57	44	2020-08-03	57	100%	4	1	6	11	10	14	1	3

Figura 4.1 Datos recopilados por el seguimiento

Los datos completos los puede encontrar en el Apéndice B, ahí encontrará los datos organizados como se muestran en la imagen anterior.

Con estos datos realizamos un análisis de regresión, el propósito de realizar este análisis es estimar la eficiencia terminal de los asesores de acuerdo con los factores observados durante la impartición de un curso y determinar cuáles son significativos.

Esto es importante porque si logramos determinar cuáles son los factores significativos podremos concluir que el seguimiento de un curso es importante, ya que lo que sucede durante el curso determina la eficiencia terminal de este al finalizar el periodo de impartición.

Por lo cual si asignamos a los asesores con las mejores eficiencias terminales podríamos mejorar de forma global la eficiencia terminal de los cursos en todo un periodo.

Coeficientes del modelo de regresión

De acuerdo con los datos obtenidos tenemos los siguientes posibles coeficientes:

- Alumnos. Cantidad de alumnos en el grupo impartido.

- Significativo. Cantidad de retrasos de nivel significativo observados.
- Atrasado. Cantidad de retrasos de nivel atrasado observados.
- Normal. Cantidad de retrasos de nivel normal observados.
- Sobresaliente. Cantidad de retrasos de nivel sobresaliente observados.
- Actividades sin calificar. De los retrasos observados, cuantos de estos fueron actividades que no fueron calificadas.
- Inasistencia. De los retrasos observados, cuantos de estos fueron inasistencias de los alumnos.
- No se entregan actividades. De los retrasos observados, cuantos de estos fueron actividades que no entregaron los alumnos.

Los primeros que podemos descartar son los coeficientes Significativo, Atrasado, Normal y Sobresaliente, los cuales dejaremos fuera sin necesidad de realizar algún análisis. Esto lo podemos realizar porque los coeficientes de “Actividades sin calificar”, Inasistencia y “No se entregan actividades” se calculan a partir de los primeros, esto se debe a que los primeros solo miden la cantidad de los retrasos, pero los segundos miden la causa de que haya sido contado como retraso.

Por esta razón vamos a preferir quedarnos con los segundos como coeficientes para nuestro modelo de regresión, es decir, preferimos los coeficientes que nos indican las causas sobre aquellos que solo nos dicen la cantidad de retrasos.

Finalmente, el modelo queda de la siguiente manera:

$$\text{EficienciaTerminal} = \beta_0 + \beta_1 \text{Alumnos} + \beta_2 \text{ActividadesSinCalifica} + \beta_3 \text{Inasistencia} + \beta_4 \text{NoSeEntreganActividades} + \epsilon$$

Análisis de regresión

Para realizar el análisis para cada uno de los asesores tomaremos en cuenta solo a los asesores que haya impartido al menos 5 cursos, lo cual nos deja 22 asesores que cumplen con esta restricción, como se muestra en la tabla 1.

Tabla 4.1 Cantidad de grupos por asesor.

Asesor	Grupos
57	10
50	9
469	9
33	9
255	9

Asesor	Grupos
214	9
34	8
40	8
44	8
37	8
30	8
34	8
1041	8
56	8
41	7
6	6
36	6
35	5
259	5
735	5
58	5
418	5

Para ejemplificar el análisis realizado para cada uno de los asesores utilizaremos al asesor 57, el procedimiento que se mostrara a continuación se replicó con cada uno de los demás asesores.

Asimismo, en todos los casos utilizo un nivel de confianza del 95%.

Los datos para este asesor 57 se presentan en la tabla 2:

Tabla 4.2 Resultados obtenidos para el asesor 57.

Eficiencia Terminal	Alumnos	Actividades sin calificar	Inasistencia	No se entregan actividades
1.00	4	14	1	3

Eficiencia Terminal	Alumnos	Actividades sin calificar	Inasistencia	No se entregan actividades
1.00	2	0	4	14
1.00	3	28	0	6
0.89	3	0	7	1
0.88	20	0	7	0
0.75	4	22	5	0
0.70	15	0	14	0
0.63	8	11	11	2
0.61	3	1	11	7
0.56	2	11	13	4

El resultado de la regresión se muestra en la tabla 3:

Tabla 4.3 Estadísticas de la regresión

Coeficiente de correlación múltiple	0.960183462
Coeficiente de determinación R ²	0.921952281
R ² ajustado	0.859514105
Error típico	0.06483045
Observaciones	10

De este resultado podemos observar que el valor de R² ajustado es bastante bueno, superior al 85%, lo cual implica que los datos se ajustan muy bien a la línea de regresión, es decir, las causas en los retrasos explican el valor final de la eficiencia terminal.

Los datos resultados del Análisis de varianza se muestran en la tabla 4.

Tabla 4.4 Análisis de varianza

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	4	0.248242591	0.062060648	14.76584275	0.005624143

Residuos	5	0.021014936	0.004202987		
Total	9	0.269257527			

Del análisis de varianza tenemos un estadístico F de 14.76 y un p-value de 0.0056, el cual es menor al 5% que estamos utilizando. Por lo tanto, los coeficientes elegidos son de manera conjunta distintos de cero, es decir, de manera conjunta los coeficientes son significativos. Esto es importante porque podemos asegurar que al menos uno de los coeficientes es distinto de cero y por lo tanto, al menos alguno de ellos afecta a la eficiencia terminal.

Sin embargo, es necesario revisar que sucede manera individual con estos coeficientes, ya que a pesar de que de manera conjunta son significativos, el escenario ideal sería que también fueran significativos individualmente.

Tabla 4.5 Resultados individuales de los coeficientes

	Coeficientes	Error típico	Estadístico t	Probabilidad
Intercepción	1.109933826	0.095651371	11.60395106	8.34303E-05
Alumnos	0.004468533	0.004821107	0.926868803	0.396531824
Actividades sin calificar	-0.006179674	0.002966683	-2.083024849	0.091708772
Inasistencia	-0.039392847	0.005684845	-6.929449417	0.000960459
No se entregan actividades	0.001337233	0.006790079	0.196939265	0.851630198

De estos resultados podemos observar que, de forma individual, únicamente, el coeficiente de Inasistencia es significativo, ya que los otros superan el 5% establecido. Al superar el 5% no podemos asegurar que los valores sean distintos de 0 y por lo tanto podrían no aportar nada a la eficiencia terminal.

Repetiremos el análisis, pero esta vez eliminando el coeficiente **“No se entregan actividades”**, los resultados son los siguientes:

Tabla 4.6 Análisis de varianza.

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	3	0.248079578	0.082693193	23.42810224	0.001032474
Residuos	6	0.021177949	0.003529658		
Total	9	0.269257527			

Como podemos observar, para este caso el p-value es menor a 5% e incluso menor que el valor con cuatro coeficientes; por lo cual, sus coeficientes también son significativos de forma global.

Tabla 4.7 Resultados individuales de los coeficientes.

	Coeficientes	Error típico	Estadístico t	Probabilidad
Intercepción	1.124333159	0.056515451	19.89426145	1.0466E-06
Alumnos	0.003882833	0.003477305	1.116621152	0.306866644
Actividades sin calificar	-0.006474331	0.002347614	-2.75783516	0.032948252
Inasistencia	-0.039822925	0.004809897	-8.279371095	0.000168097

Para el caso de los coeficientes analizados de manera individual observamos que el coeficiente de **Alumnos** es mayor al 5%, por lo cual no nos asegura un valor distinto de cero.

Repetiremos el proceso, pero esta vez quitando “**Alumnos**” y dejando “**No se entregan actividades**”, los resultados son los siguientes

Tabla 4.8 Análisis de varianza

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	3	0.244631865	0.081543955	19.86804339	0.001614948
Residuos	6	0.024625662	0.004104277		
Total	9	0.269257527			

Como podemos observar, para este caso el p-value es menor a 5% e incluso menor que el valor con cuatro coeficientes; por lo cual, sus coeficientes también son significativos de forma global.

Tabla 4.9 Resultados individuales de los coeficientes.

	Coeficientes	Error típico	Estadístico t	Probabilidad
Intercepción	1.17000823	0.069513094	16.83147975	2.80977E-06
Actividades sin calificar	-0.007534753	0.002550935	-2.953721631	0.025489599
Inasistencia	-0.040121899	0.005563659	-7.21142351	0.000360186

	Coeficientes	Error típico	Estadístico t	Probabilidad
No se entregan actividades	-0.002545072	0.005281082	-0.481922518	0.646935592

Para el caso de los coeficientes analizados de manera individual observamos que el coeficiente de **No se entregan actividades** es mayor al 5%, por lo cual no nos asegura un valor distinto de cero.

Finalmente, repetiremos el proceso utilizando únicamente los coeficientes de “**Actividades sin calificar**” e **Inasistencia**. Los resultados se muestran a continuación:

Tabla 4.10 Análisis de varianza

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	0.243678649	0.121839325	33.34295093	0.000264239
Residuos	7	0.025578878	0.003654125		
Total	9	0.269257527			

Nuevamente observamos un valor menor al 5%, incluso este valor es mucho menor que los anteriores, por lo que al igual que en los casos anteriores podemos afirmar que de forma global los coeficientes son significativos.

Tabla 4.11 Resultados individuales de los coeficientes.

	Coeficientes	Error típico	Estadístico t	Probabilidad
Intercepción β_0	1.149960168	0.052548336	21.88385499	1.05027E-07
Actividades sin calificar β_1	-0.007175148	0.002301703	-3.117321995	0.016906962
Inasistencia β_2	-0.039094127	0.004848703	-8.062801612	8.6707E-05

En este caso observamos que los coeficientes si son menores al 5%, por lo cual podemos asegurar que son distintos de CERO, es decir, los coeficientes son significativos de manera individual, es decir, los dos aportan a la eficiencia terminal, en este caso al ser menores que cero afectan negativamente a la eficiencia terminal.

Por lo tanto, de acuerdo con estos resultados la ecuación de regresión para este caso (asesor 57) se vería de la siguiente forma:

$$\text{EficienciaTerminal} = 1.1499 - 0.0071\text{ActividadesSinCalificar} - 0.039\text{Inasistencia} + 0.0604$$

Este análisis nos ha permitido identificar que los coeficientes de Alumnos y de “No se entregan actividades” no son significativos, lo cual implica que no inciden en el resultado de la Eficiencia Terminal del curso.

En la práctica esto quiere decir que no es significativa la cantidad de alumnos que tenga un curso, esto dentro del rango que se maneja para los cursos: máximo 20 alumnos. Asimismo, el coeficiente de “No se entregan actividades” tampoco es significativo, esto podría ser algo esperado porque de acuerdo con lo revisado para esta causa, sucede cuando en un curso ya no sucede nada, es decir, siguen entrando, pero no se realizan las actividades programadas; lo cual, a reserva de un análisis más profundo, puede ser una consecuencia de que las actividades no se calificaron a tiempo.

Los resultados de los análisis de regresión de cada uno de los asesores se encuentran en el Apéndice C, ahí encontrará una tabla con el resumen de los principales valores resultantes en cada regresión para cada asesor.

Finalmente, después de verificar que los coeficientes de “Actividades sin calificar” e Inasistencia son significativos global e individualmente podemos afirmar que el seguimiento realizado es determinante para la Eficiencia Terminal de los cursos, ya que si logramos disminuir la cantidad de observaciones de estos coeficientes podríamos aumentar esta Eficiencia Terminal.

Asimismo, con los valores de esta Eficiencia Terminal para cada uno de los asesores modelados con el análisis de regresión podemos utilizar los valores observados para realizar una asignación de cursos a cada asesor de manera que optimicemos la Eficiencia terminal no solo por curso sino por periodos completos.

Los valores que utilizaremos, más adelante, para cada asesor en el modelo de asignación son los siguientes.

Tabla 4.12 Eficiencia terminal promedio por asesor

Asesor	Eficiencia Terminal Promedio
57	80.21%
50	78.44%
469	65.78%
33	63.78%
255	60.81%
214	57.22%
34	65.38%

Asesor	Eficiencia Terminal Promedio
40	91.86%
44	88.86%
37	73.17%
30	71.78%
34	65.38%
1041	58.13%
56	66.88%
41	66.86%
6	76.20%
36	55.33%
35	54.33%
259	59.89%
735	64.53%
58	76.60%
418	49.80%

4.2 Índice de apertura de los cursos

Este **Índice De Apertura De Cursos (IPC)** lo definiremos como la probabilidad de que un curso X se haya abierto una vez que haya sido calendarizado, más adelante este índice nos permitirá preferir los cursos de acuerdo con los valores observados en periodos anteriores.

Para el cálculo de este índice utilizaremos las observaciones realizadas a los cursos que se han calendarizado durante los años 2019 y 2020 y realizaremos el siguiente calculo:

$$IPC = \frac{\text{Cantidad de veces que que abrió el curso}}{\text{Cantidad de veces que se calendarizó el curso}}$$

Por lo cual el valor de este se encuentra entre cero y uno. Donde uno nos indica que el curso se abre cada vez que se programa y cero que nunca se ha abierto a pesar de haber sido programado alguna vez.

De acuerdo con el periodo observado tenemos 61 cursos cuyos índices se encuentran entre 0 y 95%. De la misma forma que hicimos con las observaciones de los asesores en las eficiencias terminales, para los cursos utilizaremos solo aquellos que han sido calendarizados al menos en cinco ocasiones.

Por lo cual, el rango observado para el índice será entre 27% y 95% en 43 cursos que al menos han sido programados cinco veces.

Los datos obtenidos se muestran a continuación:

Tabla 4.13 Índices de apertura por curso

Núm.	Nombre	Cantidad de veces calendarizado	Aperturas	Índice de apertura
1	Hoja de cálculo EXCEL	19	18	94.74%
2	Desarrollo de cursos con MOODLE	18	17	94.44%
3	Introducción a la programación	14	13	92.86%
4	Tablas y gráficos dinámicos con EXCEL	22	20	90.91%
5	Administración de proyectos con PROJECT	9	8	88.89%
6	Taller de manejo de bases de datos con Excel	14	12	85.71%
7	Taller de diseño instruccional en educación en línea	6	5	83.33%
8	Técnicas avanzadas con Excel	17	14	82.35%
9	Macros con EXCEL	16	13	81.25%
10	Taller de fórmulas con Excel	20	16	80.00%
11	Taller de introducción al diseño instruccional	5	4	80.00%
12	Creación de imágenes con ILLUSTRATOR	17	13	76.47%
13	Taller de funciones con Excel	16	12	75.00%
14	Introducción al diseño de bases de datos	16	12	75.00%

Núm.	Nombre	Cantidad de veces calendarizado	Aperturas	Índice de apertura
15	Taller de aplicaciones estadísticas con EXCEL	14	10	71.43%
16	Taller de ortografía	16	11	68.75%
17	Taller para la comprensión de textos	11	7	63.64%
18	Tecnologías para la capacitación a distancia	13	8	61.54%
19	Uso del manejador de bases de datos ACCESS	12	7	58.33%
20	Creación de formularios con Google Forms	18	10	55.56%
21	Procesamiento estadístico de datos con SPSS	15	8	53.33%
22	Procesamiento estadístico de datos con PSPP	15	8	53.33%
23	Creación de páginas web con HTML	17	9	52.94%
24	Elaboración de gráficos con EXCEL	16	8	50.00%
25	Creación de formularios de captura con Word	14	7	50.00%
26	Elaboración de plantillas y temas con POWERPOINT	14	7	50.00%
27	SQL estándar básico	14	7	50.00%
28	Presentaciones avanzadas con POWERPOINT	13	6	46.15%
29	JAVASCRIPT para generar páginas web interactivas	11	5	45.45%
30	Estadística inferencial con SPSS	11	5	45.45%
31	Edición de publicaciones editoriales con INDESIGN	11	5	45.45%
32	Sistema operativo Linux	18	8	44.44%

Núm.	Nombre	Cantidad de veces calendarizado	Aperturas	Índice de apertura
33	Manipulación de imágenes con PHOTOSHOP	18	8	44.44%
34	Introducción al desarrollo de contenidos dinámicos con HTML5	11	4	36.36%
35	Taller de aplicaciones contables y financieras con Excel	12	4	33.33%
36	Procesador de texto WORD	16	5	31.25%
37	Técnicas avanzadas con WORD	16	5	31.25%
38	Combinación de correspondencia con WORD	13	4	30.77%
39	Manejo de máquinas virtuales	13	4	30.77%
40	Uso del manejador de bases de datos MYSQL	13	4	30.77%
41	Manejo de estilos con WORD	14	4	28.57%
42	Almacenamiento virtual	18	5	27.78%

4.3 Modelo de asignación para los cursos

Finalmente, con los datos de la eficiencia terminal de los asesores y el índice de apertura de cursos retomaremos la planeación de cursos revisada en el capítulo 2 e incorporaremos la eficiencia terminal y el índice de apertura dentro de la función objetivo.

Lo anterior tiene como propósito no solo realizar una planeación adecuada de los cursos según las condiciones planteadas sino, además, asignar directamente a los mejores asesores (en términos de eficiencia terminal) en un curso y periodo para alcanzar la mejor eficiencia terminal en los periodos propuestos.

Planteamiento del problema

En el capítulo describimos de manera particular el problema de la asignación de cursos, mencionamos que se realiza de manera semestral, los cursos no se pueden repetir en el siguiente periodo, debe haber una cantidad mínima de cursos por cada periodo, etc.

Ahora generalizaremos el modelo de la siguiente manera para cada conjunto de periodos a planificar:

- Existen j periodos a planificar.
- Tenemos i cursos posibles para asignar.
- Los cursos se agrupan dentro de k categorías.
- Un asesor solo puede ser asignado a un curso i en un periodo j .
- Se deben asignar al menos (o máximo) X cursos de la categoría k en cada periodo j .
- Un curso i asignado al periodo j no debe repetirse durante n periodos.
- Para toda la planeación un asesor no puede tener más de n cursos asignados.

Con estas condiciones planteadas podemos realizar la modelación de este problema.

Modelado del problema

Para resolver este problema utilizaremos un modelo de programación lineal, en particular un modelo de programación lineal entera binario. Para lo cual definimos las siguientes variables:

$$\text{Sea } x_{kij} = \begin{cases} 1 & \text{Si se utiliza el curso } i \text{ en el periodo } j \text{ con el asesor } k \\ 0 & \text{Si no se utiliza} \end{cases}$$

$$\forall i \in \{1, 2, \dots, n_i\}$$

$$\forall j \in \{1, 2, \dots, n_j\}$$

$$\forall k \in \{1, 2, \dots, n_k\}$$

Asimismo, definimos las siguientes variables:

$$a_k = \text{Eficiencia terminal del asesor } k$$

$$p_i = \text{Índice de apertura para el curso } i$$

Por lo tanto, la **función objetivo** de nuestro nuevo problema es la siguiente:

$$\text{Max } z = \sum_{n=1}^{n_j} \sum_{i=1}^{n_i} \sum_{k=1}^{n_k} p_i a_k x_{ijk}$$

La cual está sujeta a las siguientes restricciones:

Un solo asesor por curso y periodo

Cursos	Periodos			
	$\sum_{k=1}^{n_k} x_{11k} \leq 1$	$\sum_{k=1}^{n_k} x_{12k} \leq 1$	$\sum_{k=1}^{n_k} x_{1n_j k} \leq 1$

	$\sum_{k=1}^{n_k} x_{21k} \leq 1$	$\sum_{k=1}^{n_k} x_{22k} \leq 1$	$\sum_{k=1}^{n_k} x_{2n_jk} \leq 1$

	$\sum_{k=1}^{n_k} x_{n_i1k} \leq 1$	$\sum_{k=1}^{n_k} x_{n_i2k} \leq 1$	$\sum_{k=1}^{n_k} x_{n_in_jk} \leq 1$

Las siguientes restricciones tienen que ver con categoría de cursos, para definirlo de manera general suponemos que tenemos n_j cursos, los cuales están agrupados en m categorías, es decir, tenemos las categorías C_1, C_2, \dots, C_m .

Por lo tanto, el tamaño de cada categoría será:

$$n_{c_1}, n_{c_2}, \dots, n_{c_m}$$

Con estas definiciones procedemos a definir las restricciones de los cursos por categoría.

Hasta y_1 cursos de la categoría 1 por periodo

$$\sum_{k=1}^{n_k} \sum_{i=1}^{n_{c_1}} x_{i1k} \leq y_1$$

$$\sum_{k=1}^{n_k} \sum_{i=1}^{n_{c_1}} x_{i2k} \leq y_1$$

...

$$\sum_{k=1}^{n_k} \sum_{i=1}^{n_{c_1}} x_{in_jk} \leq y_1$$

Al menos y_2 cursos de la categoría 2 por periodo

$$\sum_{k=1}^{n_k} \sum_{i=n_{c_1}+1}^{n_{c_2}} x_{i1k} \geq y_2$$

$$\sum_{k=1}^{n_k} \sum_{i=n_{c_1}+1}^{n_{c_2}} x_{i2k} \geq y_2$$

...

$$\sum_{k=1}^{n_k} \sum_{i=n_{c_1}+1}^{n_{c_2}} x_{in_jk} \geq y_2$$

Por lo tanto, de manera general podemos tener tantas restricciones para categorías como sea necesario, el término “Hasta” y “Al menos” definirá el sentido de la igualdad.

Entonces de manera general podemos definir el conjunto de restricciones con el término “Hasta” para cualquier categoría de la siguiente forma:

$$\sum_{k=1}^{n_k} \sum_{i=n_{c_{m-1}}+1}^{n_{c_m}} x_{i1k} \leq y_m$$

$$\sum_{k=1}^{n_k} \sum_{i=n_{c_{m-1}}+1}^{n_{c_m}} x_{i2k} \leq y_m$$

...

$$\sum_{k=1}^{n_k} \sum_{i=n_{c_{m-1}}+1}^{n_{c_m}} x_{in_jk} \leq y_m$$

O con la desigualdad invertida si el término es “Al menos”.

A continuación, están las restricciones correspondientes a que un curso no puede repetirse durante h periodos.

Para realizar la definición general de las restricciones definiremos d como el resultado de $n_j \% h$, por lo cual h siempre debe ser menor que n_j .

Un curso no se repite por h periodos

Cursos	Periodos			
	$\sum_{j=1}^h \sum_{k=1}^{n_k} x_{1jk} = 1$	$\sum_{j=h}^{2h} \sum_{k=1}^{n_k} x_{1jk} = 1$	$\sum_{j=n_j-d-1}^{n_j} \sum_{k=1}^{n_k} x_{1jk} = 1$
	$\sum_{j=1}^h \sum_{k=1}^{n_k} x_{2jk} = 1$	$\sum_{j=h}^{2h} \sum_{k=1}^{n_k} x_{2jk} = 1$	$\sum_{j=n_j-d-1}^{n_j} \sum_{k=1}^{n_k} x_{2jk} = 1$

$\sum_{j=1}^h \sum_{k=1}^{n_k} x_{n_jk} = 1$	$\sum_{j=h}^{2h} \sum_{k=1}^{n_k} x_{n_jk} = 1$	$\sum_{j=n_j-d-1}^{n_j} \sum_{k=1}^{n_k} x_{n_jk} = 1$	

Finalmente, tenemos las restricciones respecto del asesor, estas restricciones limitan el número de cursos (n_c) que puede tener un asesor para el ciclo de periodo planeado.

Máximo número de cursos asignados a un asesor

$$\sum_{j=1}^{n_j} \sum_{i=1}^{n_i} x_{ij1} \leq n_c$$

$$\sum_{j=1}^{n_j} \sum_{i=1}^{n_i} x_{ij2} \leq n_c$$

...

$$\sum_{j=1}^{n_j} \sum_{i=1}^{n_i} x_{ijn_k} \leq n_c$$

En caso de que no se requiera un máximo número de cursos sino un mínimo solo es necesario invertir la desigualdad

Como se puede observar la cantidad de restricciones y variables depende por completo del número de cursos, periodos y asesores que tengamos, en general podemos decir que:

$$\text{Cantidad de variables} = n_i * n_j * n_k$$

Ejemplo

Para ejemplificar el funcionamiento del modelo utilizaremos un problema con las siguientes características:

- Tenemos disponibles cinco cursos y cinco asesores.
- Los cursos están agrupados en dos categorías de la siguiente manera:

Categoría 1 = {Curso 1, Curso 2}

Categoría 2 = {Curso 3, Curso 4, Curso 5}

- Únicamente tenemos cuatro periodos de cursos.
- Para la categoría 1 solo podemos tener un curso por periodo.
- Para la categoría 2 debemos tener al menos un curso por periodo.
- Los cursos no pueden programarse dos veces seguidas.

De acuerdo con estos datos tenemos 100 variables.

Para los casos de la eficiencia terminal e índice de apertura tenemos los siguientes datos:

Eficiencia terminal de los asesores = {0.8, 0.9, 0.88, 0.97, 0.91}

Índice de apertura de los cursos = {0.9, 1, 0.79, 0.95, 0.85}

Por lo tanto, la tabla de costos de Asesores (Eficiencia terminal) y Cursos (Índice de apertura) se muestra en la siguiente tabla:

Tabla 4.14 Costos para el ejemplo.

		Cursos				
		1	2	3	4	5
Asesores		0.9	1	0.79	0.95	0.85
	1 0.8	0.72	0.8	0.632	0.76	0.68
	2 0.9	0.81	0.9	0.711	0.855	0.765
	3 0.88	0.792	0.88	0.6952	0.836	0.748
	4 0.97	0.873	0.97	0.7663	0.9215	0.8245
	5 0.91	0.819	0.91	0.7189	0.8645	0.7735

La función objetivo es la siguiente:

$$\begin{aligned}
 & \max 0.72x_{111} + 0.81x_{112} + 0.792x_{113} + 0.873x_{114} + 0.819x_{115} + 0.8x_{211} + 0.9x_{212} + \\
 & 0.88x_{213} + 0.97x_{214} + 0.91x_{215} + 0.632x_{311} + 0.711x_{312} + 0.6952x_{313} + 0.7663x_{314} + \\
 & 0.7189x_{315} + 0.76x_{411} + 0.855x_{412} + 0.836x_{413} + 0.9215x_{414} + 0.8645x_{415} + 0.68x_{511} \\
 & + 0.765x_{512} + 0.748x_{513} + 0.8245x_{514} + 0.7735x_{515} + 0.72x_{121} + 0.81x_{122} + \\
 & 0.792x_{123} + 0.873x_{124} + 0.819x_{125} + 0.8x_{221} + 0.9x_{222} + 0.88x_{223} + 0.97x_{224} + \\
 & 0.91x_{225} + 0.632x_{321} + 0.711x_{322} + 0.6952x_{323} + 0.7663x_{324} + 0.7189x_{325} + 0.76x_{421} \\
 & + 0.855x_{422} + 0.836x_{423} + 0.9215x_{424} + 0.8645x_{425} + 0.68x_{521} + 0.765x_{522} + \\
 & 0.748x_{523} + 0.8245x_{524} + 0.7735x_{525} + 0.72x_{131} + 0.81x_{132} + 0.792x_{133} + 0.873x_{134} \\
 & + 0.819x_{135} + 0.8x_{231} + 0.9x_{232} + 0.88x_{233} + 0.97x_{234} + 0.91x_{235} + 0.632x_{331} + \\
 & 0.711x_{332} + 0.6952x_{333} + 0.7663x_{334} + 0.7189x_{335} + 0.76x_{431} + 0.855x_{432} + \\
 & 0.836x_{433} + 0.9215x_{434} + 0.8645x_{435} + 0.68x_{531} + 0.765x_{532} + 0.748x_{533} + \\
 & 0.8245x_{534} + 0.7735x_{535} + 0.72x_{141} + 0.81x_{142} + 0.792x_{143} + 0.873x_{144} + 0.819x_{145} \\
 & + 0.8x_{241} + 0.9x_{242} + 0.88x_{243} + 0.97x_{244} + 0.91x_{245} + 0.632x_{341} + 0.711x_{342} + \\
 & 0.6952x_{343} + 0.7663x_{344} + 0.7189x_{345} + 0.76x_{441} + 0.855x_{442} + 0.836x_{443} + \\
 & 0.9215x_{444} + 0.8645x_{445} + 0.68x_{541} + 0.765x_{542} + 0.748x_{543} + 0.8245x_{544} + \\
 & 0.7735x_{545}
 \end{aligned}$$

Sujeto a

! Solo un asesor por curso y periodo:

$$x_{111} + x_{112} + x_{113} + x_{114} + x_{115} \leq 1$$

$$x_{211}+x_{212}+x_{213}+x_{214}+x_{215} \leq 1$$

$$x_{311}+x_{312}+x_{313}+x_{314}+x_{315} \leq 1$$

$$x_{411}+x_{412}+x_{413}+x_{414}+x_{415} \leq 1$$

$$x_{511}+x_{512}+x_{513}+x_{514}+x_{515} \leq 1$$

$$x_{121}+x_{122}+x_{123}+x_{124}+x_{125} \leq 1$$

$$x_{221}+x_{222}+x_{223}+x_{224}+x_{225} \leq 1$$

$$x_{321}+x_{322}+x_{323}+x_{324}+x_{325} \leq 1$$

$$x_{421}+x_{422}+x_{423}+x_{424}+x_{425} \leq 1$$

$$x_{521}+x_{522}+x_{523}+x_{524}+x_{525} \leq 1$$

$$x_{131}+x_{132}+x_{133}+x_{134}+x_{135} \leq 1$$

$$x_{231}+x_{232}+x_{233}+x_{234}+x_{235} \leq 1$$

$$x_{331}+x_{332}+x_{333}+x_{334}+x_{335} \leq 1$$

$$x_{431}+x_{432}+x_{433}+x_{434}+x_{435} \leq 1$$

$$x_{531}+x_{532}+x_{533}+x_{534}+x_{535} \leq 1$$

$$x_{141}+x_{142}+x_{143}+x_{144}+x_{145} \leq 1$$

$$x_{241}+x_{242}+x_{243}+x_{244}+x_{245} \leq 1$$

$$x_{341}+x_{342}+x_{343}+x_{344}+x_{345} \leq 1$$

$$x_{441}+x_{442}+x_{443}+x_{444}+x_{445} \leq 1$$

$$x_{541}+x_{542}+x_{543}+x_{544}+x_{545} \leq 1$$

! Exactamente un curso de categoría 1 por periodo:

$$x_{111}+x_{112}+x_{113}+x_{114}+x_{115} + x_{211}+x_{212}+x_{213}+x_{214}+x_{215} = 1$$

$$x_{121}+x_{122}+x_{123}+x_{124}+x_{125} + x_{221}+x_{222}+x_{223}+x_{224}+x_{225} = 1$$

$$x_{131}+x_{132}+x_{133}+x_{134}+x_{135} + x_{231}+x_{232}+x_{233}+x_{234}+x_{235} = 1$$

$$x_{141}+x_{142}+x_{143}+x_{144}+x_{145} + x_{241}+x_{242}+x_{243}+x_{244}+x_{245} = 1$$

! Al menos un curso de categoría 2 por periodo:

$$x_{311}+x_{312}+x_{313}+x_{314}+x_{315} + x_{411}+x_{412}+x_{413}+x_{414}+x_{415} + \\ x_{511}+x_{512}+x_{513}+x_{514}+x_{515} \geq 1$$

$$x_{321}+x_{322}+x_{323}+x_{324}+x_{325} + x_{421}+x_{422}+x_{423}+x_{424}+x_{425} + \\ x_{521}+x_{522}+x_{523}+x_{524}+x_{525} \geq 1$$

$$x331+x332+x333+x334+x335 + x431+x432+x433+x434+x435 + \\ x531+x532+x533+x534+x535 \geq 1$$

$$x341+x342+x343+x344+x345 + x441+x442+x443+x444+x445 + \\ x541+x542+x543+x544+x545 \geq 1$$

! Los cursos no se pueden programar dos veces seguidas:

$$x111+x112+x113+x114+x115 + x121+x122+x123+x124+x125 = 1$$

$$x121+x122+x123+x124+x125 + x131+x132+x133+x134+x135 = 1$$

$$x131+x132+x133+x134+x135 + x141+x142+x143+x144+x145 = 1$$

$$x211+x212+x213+x214+x215 + x221+x222+x223+x224+x225 = 1$$

$$x221+x222+x223+x224+x225 + x231+x232+x233+x234+x235 = 1$$

$$x231+x232+x233+x234+x235 + x241+x242+x243+x244+x245 = 1$$

$$x311+x312+x313+x314+x315 + x321+x322+x323+x324+x325 = 1$$

$$x321+x322+x323+x324+x325 + x331+x332+x333+x334+x335 = 1$$

$$x331+x332+x333+x334+x335 + x341+x342+x343+x344+x345 = 1$$

$$x411+x412+x413+x414+x415 + x421+x422+x423+x424+x425 = 1$$

$$x421+x422+x423+x424+x425 + x431+x432+x433+x434+x435 = 1$$

$$x431+x432+x433+x434+x435 + x441+x442+x443+x444+x445 = 1$$

$$x511+x512+x513+x514+x515 + x521+x522+x523+x524+x525 = 1$$

$$x521+x522+x523+x524+x525 + x531+x532+x533+x534+x535 = 1$$

$$x531+x532+x533+x534+x535 + x541+x542+x543+x544+x545 = 1$$

! Un asesor no puede impartir más de 2 cursos en todo el ciclo

$$x111+x211+x311+x411+x511 + x121+x221+x321+x421+x521 + \\ x131+x231+x331+x431+x531 + x141+x241+x341+x441+x541 \leq 2$$

$$x112+x212+x312+x412+x512 + x122+x222+x322+x422+x522 + \\ x132+x232+x332+x432+x532 + x142+x242+x342+x442+x542 \leq 2$$

$$x_{113}+x_{213}+x_{313}+x_{413}+x_{513} + x_{123}+x_{223}+x_{323}+x_{423}+x_{523} + x_{133}+x_{233}+x_{333}+x_{433}+x_{533} + x_{143}+x_{243}+x_{343}+x_{443}+x_{543} \leq 2$$

$$x_{114}+x_{214}+x_{314}+x_{414}+x_{514} + x_{124}+x_{224}+x_{324}+x_{424}+x_{524} + x_{134}+x_{234}+x_{334}+x_{434}+x_{534} + x_{144}+x_{244}+x_{344}+x_{444}+x_{544} \leq 2$$

$$x_{115}+x_{215}+x_{315}+x_{415}+x_{515} + x_{125}+x_{225}+x_{325}+x_{425}+x_{525} + x_{135}+x_{235}+x_{335}+x_{435}+x_{535} + x_{145}+x_{245}+x_{345}+x_{445}+x_{545} \leq 2$$

Para la resolución de este problema utilizamos LINDO, obteniendo el siguiente resultado:

```

LP OPTIMUM FOUND AT STEP      58
OBJECTIVE VALUE =      8.04899979

NEW INTEGER SOLUTION OF      8.04899979      AT BRANCH      0 PIVOT      58
BOUND ON OPTIMUM:      8.049000
ENUMERATION COMPLETE. BRANCHES=      0 PIVOTS=      58

LAST INTEGER SOLUTION IS THE BEST FOUND
RE-INSTALLING BEST SOLUTION...

OBJECTIVE FUNCTION VALUE
1)      8.049000

```

Figura 4.2 Solución encontrada por LINDO

Los valores encontrados para las variables se muestran en la siguiente imagen.

VARIABLE	VALUE	REDUCED COST
X111	0.000000	-0.720000
X112	1.000000	-0.810000
X113	0.000000	-0.792000
X114	0.000000	-0.873000
X115	0.000000	-0.819000
X211	0.000000	-0.800000
X212	0.000000	-0.900000
X213	0.000000	-0.880000
X214	0.000000	-0.970000
X215	0.000000	-0.910000
X311	1.000000	-0.632000
X312	0.000000	-0.711000
X313	0.000000	-0.695200
X314	0.000000	-0.766300
X315	0.000000	-0.718900
X411	0.000000	-0.760000
X412	0.000000	-0.855000
X413	0.000000	-0.836000
X414	0.000000	-0.921500

Figura 4.3 Algunas de los valores encontrados

Sin embargo, así como están presentados, los valores en sí mismos no nos dicen nada, hay que traducirlos en una tabla con curso, periodo y el asesor correspondiente

Tabla 4.15 Resultados encontrados para la asignación.

	Periodo 1	Periodo 2	Periodo 3	Periodo 4
Curso 1	Asesor 2		Asesor 2	
Curso 2		Asesor 4		Asesor 4
Curso 3	Asesor 1		Asesor 1	
Curso 4	Asesor 5		Asesor 5	
Curso 5		Asesor 3		Asesor 3

Por lo tanto, de acuerdo con la tabla resultante tenemos la asignación que cumple con todos los criterios establecidos de antemano: cursos saltados, asesores en distintos periodos, etc.

Por lo cual podemos observar que el modelo de asignación planteado realizó de forma adecuada la asignación maximizando las eficiencias terminales de acuerdo con la probabilidad de apertura de cada curso.

4.4 Análisis de resultados

Como se pudo observar con el ejemplo anterior, el modelo de asignación que se planteo realizó correctamente la programación de cursos, asesores y periodos de acuerdo con los criterios establecidos

Sin embargo, el mayor problema que encontramos es la cantidad de variables que se deben utilizar, en este caso con solo cinco asesores, cinco cursos y 4 periodos necesitamos de 100 variables para plantear el problema. Lo cual resulta en un obstáculo porque no todos los programas que resuelven problemas de este tipo lo realizan con las licencias estándar, en muchos casos será necesario utilizar licencias especializadas o realizar una solución a la medida.

Esto es necesario ya que los cursos, por lo general, se programan para 8 periodos dentro de los cuales se utilizan 50 asesores para la misma cantidad de cursos. Con estos números tendríamos 20,000 variables para realizar la programación de un semestre completo.

Alguna de las soluciones que pueden utilizarse es dividir esta programación en periodos más pequeños: mensual, bimestral, trimestral, etc. Lo cual implicaría que la cantidad de cursos también puede reducirse y por lo tanto las variables utilizadas se reducirían considerablemente.

La opción de utilizar una licencia especializada para el programa con el que se ejemplifico el uso del modelo no necesariamente es la mejor, al menos no de la manera en que se utilizó en el ejemplo. Esto se debe a que no solo las variables utilizadas son demasiadas, sino que también la cantidad de ecuaciones necesarias para el planteamiento crecen de manera significativa, lo cual nos lleva a que el solo planteamiento del modelo pueda

convertirse rápidamente en algo intratable, donde el más pequeño error durante la captura de los datos del modelo significaría una solución errónea o incluso no alcanzar una solución.

Por lo que, en el caso de una licencia especializada, la mejor opción podría ser utilizar las API's¹⁴ disponibles para construir una interfaz que nos permita la captura óptima de los datos del modelo y luego con esta captura utilizar las funciones disponibles del software licenciado.

Finalmente, la opción de utilizar una solución a la medida implicaría el desarrollo de un programa o conjunto de programas hechos explícitamente para el modelo de asignación planteado. Exploraremos con más detalle esta última opción en las conclusiones del trabajo, sobre todo en el trabajo futuro que será necesario realizar para implementar por completo el modelo de asignación planteado en las operaciones cotidianas de la programación de cursos.

¹⁴ “Una API es una **interfaz de programación de aplicaciones** (del inglés API: Application Programming Interface). Es un conjunto de rutinas que provee acceso a funciones de un determinado software.” https://es.wikipedia.org/wiki/Web_API

Conclusiones

La educación en línea, en particular la educación continua en línea tiene graves problemas con la deserción de los alumnos que se inscriben en los cursos. Las causas de esta deserción pueden ser diversas: falta de conocimientos tecnológicos, deficiencias en los contenidos didácticos, plataformas tecnológicas inadecuadas, asesores/profesores deficientes, etc.

En este trabajo abordamos el problema desde la perspectiva de los asesores, enfocándonos principalmente en el trabajo que realizan durante la impartición de un curso, el cual se basa en los siguientes aspectos:

- Resolución de dudas a los alumnos.
- Calificación y retroalimentación de las actividades.
- Seguimiento de los alumnos.

Es por esto, que con base en las actividades de los asesores sobre un curso definimos la eficiencia terminal de los cursos, la cual indica el porcentaje de finalización de las actividades planeadas de un grupo para un curso en línea. Esta eficiencia terminal se la asociamos a los asesores para el posterior análisis de datos.

Tradicionalmente, la evaluación de un curso se realiza a través de un cuestionario de opinión que los alumnos deben responder al finalizarlo de forma opcional. Es por ello que el primer análisis que realizamos en este trabajo fue utilizar los datos de las respuestas de dichos cuestionarios y compararlas con los datos de las eficiencias terminales.

Sin embargo, como se observó en los resultados, no hay ninguna relación entre ellos porque el cuestionario que debe explicar los resultados es parte de las actividades a realizar, lo cual es contradictorio.

Por esta razón, y aprovechando que los cursos se imparten por completo en línea cambiamos el enfoque, en este caso realizamos un seguimiento puntual de lo que sucede diariamente para monitorear las actividades de los asesores y revisar si las actividades del asesor influyen en la eficiencia terminal de los cursos.

Para realizar este monitoreo definimos un modelo para el seguimiento diario de los cursos, a través del cual recopilamos la información de cada día referente a las actividades del asesor.

Con toda la información recabada establecimos tres causas por las cuales existen retrasos durante un curso:

- Actividades sin calificar.
- Inasistencia.
- No se entregan actividades.

Para cada una de estas causas realizamos un conteo de las veces que ocurren durante la impartición de un curso. Con estos datos realizamos un análisis de regresión de la eficiencia terminal explicada por estas causas.

Luego del análisis de los modelos de regresión concluimos que las causas de Actividades sin calificar e Inasistencia influyen estadísticamente en dicha eficiencia, por lo cual asumimos que el asesor influye de forma significativa en el grado de finalización de un curso.

Dado que el asesor influye en la eficiencia terminal de un curso entonces el asesor que sea designado para impartir un curso puede obtener una mayor o menor eficiencia terminal para el curso asignado.

Por lo tanto, si necesitamos que los cursos tengan la mayor eficiencia terminal posible entonces debemos designar a los asesores con las más altas eficiencias terminales para los cursos que se impartan.

Sin embargo, la planeación realizada no implica que todos los cursos dentro de la planeación se impartan, esto se debe a que los cursos se imparten al público en general y por lo tanto están sujetos a la demanda de estos.

Por este motivo se definió un índice de apertura de cursos con el propósito de que la planeación prefiera los cursos con el índice más alto de apertura para asignarles el asesor con la mejor eficiencia terminal.

Con el índice de apertura y la eficiencia terminal de los asesores se procedió a generar un modelo de programación entera, en particular un modelo de asignación para la calendarización de cursos, el cual tuvo como propósito maximizar la eficiencia terminal de los cursos con mayor probabilidad de apertura dentro de los periodos considerados por la calendarización. Todo esto sujeto a las condiciones particulares descritas en el capítulo 4.

Con el modelo generado ejemplificamos su uso, resultando en una asignación adecuada que cumple con todas las condiciones descritas para el caso planteado.

Sin embargo, como se mencionó en el capítulo 4 el mayor problema encontrado no es la modelación de este sino la resolución de este, lo cual se debe a que la cantidad de variables y ecuaciones de las restricciones crecen muy rápidamente cuando el número de asesores, cursos y periodos se incrementa. Incluso, para la asignación semestral que se realiza, la cantidad de variables necesarias llega a ser de al menos veinte mil, lo cual excede por mucho la capacidad de las versiones gratuitas de algunos programas especializados en la resolución de problemas de programación lineal.

Al final del capítulo 4 se exploraron superficialmente algunas soluciones a este problema, pero considero que para implementar el modelo de asignación planteado en este trabajo en las operaciones cotidianas de DGTIC es necesario continuar con el desarrollo de uno o varios métodos de solución para resolver la calendarización a través del modelo realizado.

Para ello considero que las principales vías para la resolución podrían ser las siguientes: dividir el problema en periodos más pequeños y/o crear una solución a la medida. Estas dos opciones no son excluyentes sino todo lo contrario, ya que resolver problemas más pequeños con una solución creada explícitamente reduciría la complejidad de modelo y la solución se encontraría más rápidamente.

Una de las ventajas de dividir el problema no solo es reducir la complejidad del mismo, sino que podríamos tener los resultados de las eficiencias terminales de periodos más cortos,

es decir, no tendríamos que esperar un semestre completo para revisar "como les fue" a los asesores sino realizar las mediciones mensualmente o bimestralmente, para que con base en estos nuevos datos realicemos una nueva asignación óptima.

Respecto de una solución a la medida, esta podría ser un sistema completo que obtenga los datos directamente del Data Warehouse para realizar los cálculos descritos en el capítulo 3 y prepararlos su uso en el modelo.

Con los datos preparados, el mismo sistema solo esperaría los cursos y asesores requeridos para un siguiente periodo y procedería a encontrar la solución de este.

Y en este punto, para encontrar la solución sería necesario realizar la programación de ésta utilizando métodos exactos, heurísticos, algoritmos evolutivos o simulación. Incluso considero que el escenario ideal sería una mezcla de los tres.

La mezcla es posible porque los datos de las eficiencias terminales y los índices de apertura son estocásticos, lo cual nos da la oportunidad de simularlos cuando tengamos las suficientes observaciones para asociarlos a una función de distribución.

Y utilizar estos datos simulados en el modelo de asignación propuesto, para el cual podemos comenzar la resolución de este utilizando métodos heurísticos (búsqueda tabú, genéticos, colonia de hormigas, etc.) y luego de un tiempo razonable pasar la solución encontrada por el método heurístico a alguno de los algoritmos exactos para agilizar búsqueda del óptimo correspondiente.

Finalmente, como hemos podido observar el trabajo propone un método para maximizar la eficiencia terminal los cursos, lo cual implica que si esta eficiencia se incrementa también se reduce la deserción de alumnos.

Sin embargo, aunque el trabajo describe como realizar el seguimiento de los cursos y realiza un modelo de asignación para optimizar esta eficiencia terminal nos encontramos con el problema de que hallar la solución óptima no es una tarea sencilla, esto se debe a que la complejidad del modelo crece muy rápidamente con la cantidad de periodos, cursos y asesores.

Apéndice A

En esta sección definir algunos conceptos y herramientas utilizadas por las diversas etapas de la metodología propuesta en el capítulo 3: sistemas de información, procesos para la recopilación de información, modelo de regresión, modelo de optimización, etc.

Bases de datos

Antes de definir que es una base de datos, es necesario definir una serie de conceptos inherentes a las bases de datos.

Dato

Un dato es una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos, sucesos y entidades.

Información

Conjunto de datos interrelacionados entre sí. La información son los datos que han sido organizados y presentados de modo que los patrones que muestran sean claros. Además de contar con un sentido y ambiente particular.

Por ejemplo: el carro amarillo, la pelota naranja, el árbol verde, etc.



Figura 4.4 Ejemplos de información

Cualidades de la información

- **Precisión.** Es el grado de aproximación entre la información almacenada y el valor real.
- **Oportunidad.** Tiempo transcurrido entre el momento en que se produjo el hecho que originó el dato y el momento en que está disponible la información.
- **Compleción.** Implica que ha de ser completa para poder cumplir sus fines. Debe contener todos los elementos necesarios para tomar una decisión.
- **Significativa.** Debe de ser comprensible e interesante, además debe de ser proporcionada en cantidades asimilables.

- **Coherencia.** La información debe de ser consistente con las reglas semánticas propias del mundo real al que ha de representar con la mayor fidelidad.

Tipos de datos

- **Datos estructurados.** Tienen una estructura definida, generalmente los encontramos en las bases de datos relacionales. Son archivos de texto que se almacenan con formato de tabla con títulos en cada categoría que permiten identificarlos.

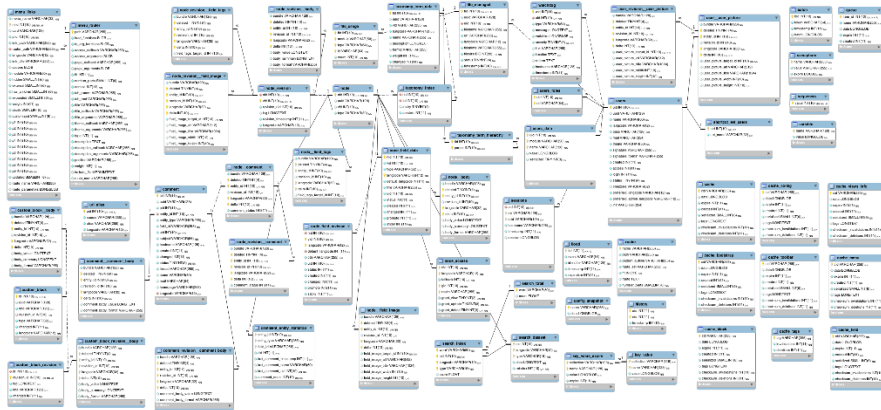


Figura 4.5 Datos estructurados

- **Datos no estructurados.** Son datos que no tienen una estructura fácilmente identificable. Es un conjunto masivo y desorganizado de varios objetos que no tienen valor hasta que se identifican y almacenan de manera organizada.



Figura 4.6 Aplicaciones que generan datos no estructurados

Con los conceptos anteriores definidos podemos realizar la definición de una base de datos:

Una base de datos es un conjunto de datos (estructurados) interrelacionados entre sí para satisfacer los requisitos de información de una organización.

Cada base de datos es única, esto porque cada base de datos se diseña según las necesidades particulares de cada organización.

Actualmente, el modelo más común para el diseño de bases de datos es el modelo relacional, este modelo fue creado por el Dr. Edgar Frank Codd en 1970 y hasta el día de hoy se sigue utilizando en las organizaciones para manipular su información.

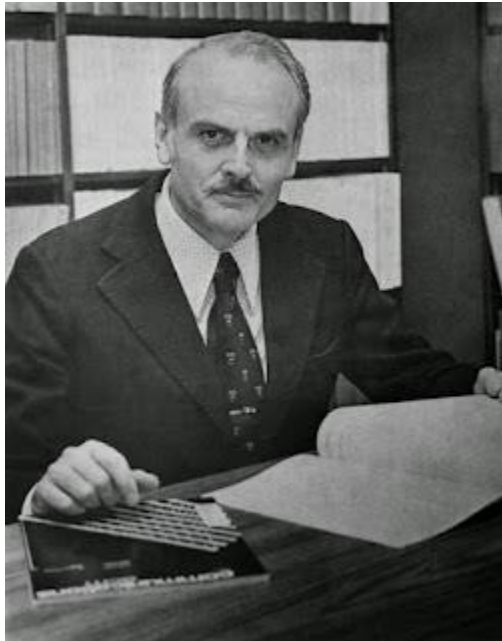


Figura 4.7 Dr. Edgar Frank Codd

Para la manipulación de la información dentro de una base de datos se utiliza SQL (Structured Query Language) este es el lenguaje de las bases de datos y es el único que puede administrar y manipular los datos, esto según las reglas definidas por Codd para el modelo relacional.

Sistemas manejadores de bases de datos

El modelo relacional es un modelo matemático que no especifica ningún tipo de herramienta informática para su implementación, por lo que para la implementación de las bases de datos se utilizan herramientas informáticas que fueron creadas exclusivamente para implementar y administrar bases de datos, a estos programas se les conoce como Sistemas manejadores de bases de datos, comúnmente llamados solamente **DBMS**, por sus siglas en ingles.

Algunos de los principales manejadores de bases de datos los revisaremos a continuación.

PostgreSQL

Este manejador multiplataforma es un proyecto de código abierto, es de los manejadores de software libre más grandes que existe actualmente, sus primeras versiones se remontan a 1986.

De acuerdo con su página web (<https://www.postgresql.org>), algunas de sus principales características son:

- Alta concurrencia.
- Protege la integridad de los datos.
- Permite crear entornos tolerantes a fallos.
- Altamente extensible.
- Intenta apegarse al estándar SQL.

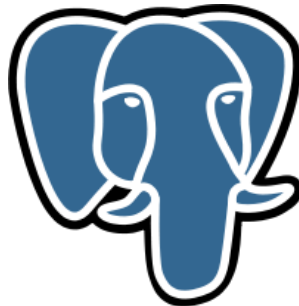


Figura 4.8 Logotipo de PostgreSQL

SQL Server

Este es un manejador desarrollado por la empresa Microsoft, originalmente estaba desarrollado para funcionar únicamente en entornos con sistemas Windows, pero a partir de 2016 funciona en entornos Linux. La primera versión de este manejador data de 1989.

Una de las principales ventajas de este manejador es que proporciona una interfaz gráfica robusta para la administración del manejador. Asimismo, se integra por completo con todas las herramientas desarrolladas por Microsoft, actualmente se puede utilizar como una máquina virtual dentro de la nube de Azure.



Figura 4.9 Logotipo de SQL Server.

MySQL

Este manejador está desarrollado bajo licencia dual: Licencia pública general/Licencia comercial por Oracle Corporation y antes de ser parte de Oracle era la base de datos de código abierto más popular del mundo.

MySQL es muy utilizado en aplicaciones web, como Joomla, Wordpress, Drupal o phpBB, en plataformas (Linux/Windows-Apache-MySQL-PHP/Perl/Python), su popularidad siempre ha estado muy relacionada con el lenguaje de programación PHP, ya que las aplicaciones web suelen desarrollarse con estos entornos.



Figura 4.10 Logotipo de MySQL

MariaDB

Este es un manejador derivado de MySQL con licencia GPL (General Public License). El desarrollo de este manejador es realizado por la fundación MariaDB y la comunidad de desarrollo de software libre.

Este manejador surge luego de la compra de Sun Microsystems (propietaria de MySQL) por parte de Oracle, por lo cual es una bifurcación directa de MySQL. Esto se realizó para tener una versión de MySQL con licencia GPL y a partir de entonces se ha continuado el desarrollo del manejador como software libre.



Figura 4.11 Logotipo de MariaDB

Procesos ETL

ETL es un acrónimo de las palabras en inglés: Extract, Transform y Load. Por lo cual proceso ETL es un proceso de tres pasos (extraer, transformar y cargar) que se utiliza para procesar la información de una o varias fuentes de datos.

Este proceso se realiza para facilitar el análisis posterior de los datos, ya que a través de este proceso podemos guardar los datos procesados en otro lugar, generalmente un almacén de datos.

Guardar los datos procesados en otro lugar es una gran ventaja ya que la disponibilidad y el análisis de estos será muy rápido debido a que no se tiene que realizar ninguna operación en la fuente de datos original.

A continuación, describiremos de manera breve cada uno de estos pasos.

Extraer

Esta primera parte del proceso consiste en la extracción de los datos desde el o los sistemas de origen. Durante este paso se definen los lugares, generalmente tablas y campos, de donde se obtendrá la información; en este mismo paso se convierten los tipos de datos necesarios para el procesamiento posterior de los mismos.

Asimismo, en esta fase se revisa si los datos cumplen con las reglas necesarias para ser tomados en cuenta: tipo, valor, estructura, etc. De no cumplir las reglas los datos no son tomados en cuenta y se rechazan.

Transformar

Durante esta fase se aplican los cálculos sobre los datos extraídos para convertirlos en datos que pueden ser cargados. Algunas veces los datos extraídos no requieren de muchos cálculos o la transformación es casi directa; sin embargo, en otras ocasiones este proceso de transformación implica un proceso en sí mismo porque es necesario realizar diversos cálculos o mezclar diversas fuentes para obtener uno solo de los datos que serán cargados.

Por lo anterior en esta fase puede ser necesario realizar una o varias de las siguientes acciones:

- Seleccionar los datos a transformar de los datos extraídos que están disponibles.
- Realizar la codificación de los valores extraídos.
- Unir o dividir los valores extraídos para su procesamiento.
- Realizar los cálculos para obtener nuevos valores: promedios, totales, máximos, mínimos, etc.
- Realizar la validación de los datos transformados.
- Ejecutar políticas para el tratamiento de valores erróneos.

Cargar

La última parte de este proceso es cuando los datos transformados se guardan en el almacén designado para este fin. El destino de estos datos depende de las políticas de la organización, pueden ir a una base de datos donde se sobrescribe la información o a un datawarehouse donde se almacenan para mantener un historial de estos.

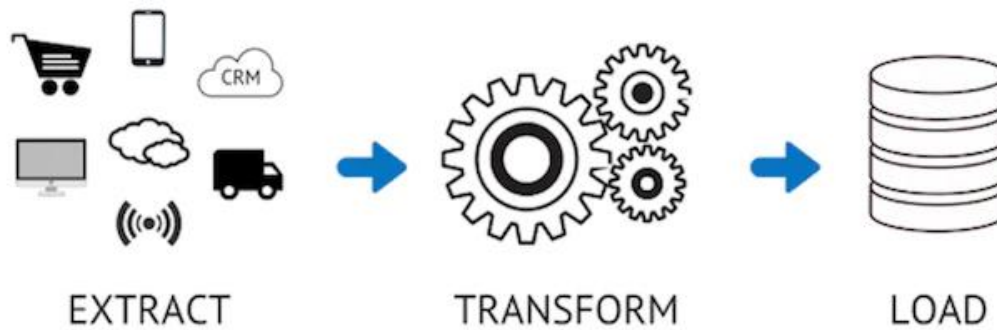


Figura 4.12 Diagrama de un proceso ETL.

Data Warehouse

La traducción de Data Warehouse es Almacén de datos, por lo cual un sistema de Data Warehouse se utiliza para almacenar los datos, generalmente, obtenidos desde un proceso ETL.

Un data warehouse nos permite facilitar el análisis de los datos históricos de una organización, ya que al tener únicamente los datos que ya fueron procesados, el análisis de estos se limita a utilizar los datos almacenados sin utilizar fuentes que requieran de un procesamiento previo.

Ventajas

Entre las ventajas de un data warehouse tenemos:

- Rápido acceso a la información.
- Disponibilidad de la información.
- Información relevante para la organización.
- Información confiable para aplicar técnicas de análisis de datos.
- Facilita la toma de decisiones.
- Permite la comunicación entre varias áreas de la organización.

Desventajas

Algunas desventajas de un data warehouse son:

- No proporciona información en tiempo real de la situación de la organización.
- Requiere de procesos ETL, los cuales pueden ser muy complejos.
- Poca flexibilidad para el cambio, una vez diseñado un data warehouse es muy difícil modificar su diseño, ya sea para añadir nuevos tipos de datos o nuevas fuentes de información.
- Necesita mantenimiento continuo.
- Por lo general tiene costos muy altos.

Modelo de estrella

Aunque existen diversos modelos para realiza un data warehouse, para este trabajo nos centraremos únicamente en describir el modelo de estrella. Este modelo de datos divide los datos en hechos y dimensiones, la tabla de hechos tiene los datos para el análisis y está rodeada de las tablas de dimensiones, las cuales son atributos que describen los hechos; por lo tanto, la tabla de hecho es una gran tabla rodeada de tablas más pequeñas, esto asemeja a una estrella.

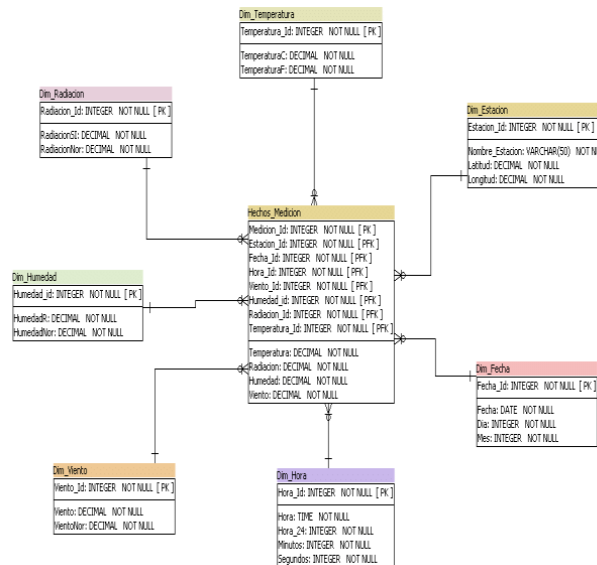


Figura 4.13 Esquema de un modelo de estrella

Tabla de Hechos

La tabla de hechos contiene métricas de un evento específico, por ejemplo, en el caso los cursos en línea un evento puede ser el día de un curso donde contamos cuantas veces ingresó el asesor, cuantos alumnos ingresaron ese día, cuantas tareas se entregaron, cuantas calificaciones faltan y cuál es el avance real del curso ese día.

Esta tabla generalmente almacena valores numéricos asociados con el evento y las claves foráneas de las tablas de dimensiones que describen la información, por lo que esta tabla registra eventos a nivel atómico. Esto ocasiona que la tabla contenga un número muy grande de registros a lo largo del tiempo, de ahí resulta que la tabla es muy grande.

Tablas de dimensiones

A diferencia de la tabla de hechos, en las tablas de dimensiones hablamos en plural, esto se debe a que dentro del esquema pueden existir varias dimensiones que deben ser tomadas en cuenta. Por la naturaleza de estas tablas, el número de registros que almacena es mucho menor ya que solo se concentran en describir las características de la dimensión respectiva.

Por ejemplo, en el caso de los cursos las dimensiones podrían ser los cursos en sí mismos, el periodo en que se impartió en curso, el asesor asignado en el curso, etc. En este caso, la dimensión del curso puede describir el nombre del curso, la duración en semanas de este, el tipo de cursos (en línea o a distancia), para que tipo de público se dirigió, etc.

La principal razón por la que se prefiere este esquema de estrella para este trabajo es su simplicidad, ya que puede ser utilizado para el análisis con cubos OLAP, Datamarts, etc.

Minería de datos

La minería de datos (Data Mining) es un conjunto de técnicas que permiten explorar fuentes de información, de manera automática o semiautomática, con el propósito de encontrar patrones y correlaciones en los conjuntos de datos para explicar e identificar las causas de estos y utilizar estos resultados para predecir lo que sucederá según lo observado.

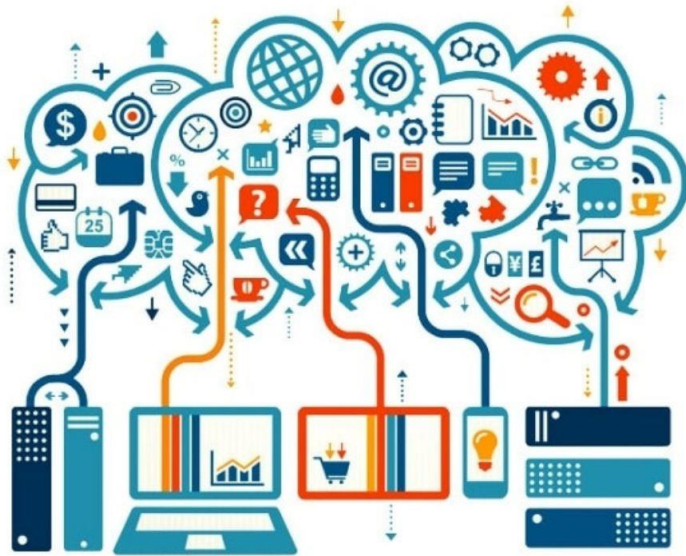


Figura 4.14 Datos generados desde diversas fuentes

Aunque el concepto de minería de datos es muy utilizado actualmente, este no es nuevo, este término apareció en los años sesenta, tomando mayor relevancia en los años ochenta y siendo muy popular en la última década con el auge de las redes sociales y la gran cantidad de información que se genera diariamente.



Figura 4.15 Redes sociales

Técnicas de minería de datos

Existen una gran cantidad de técnicas que pueden utilizarse para realizar la minería de datos, pero algunas de las más representativas son las siguientes:

- **Modelos de regresión.** Son técnicas estadísticas utilizadas para modelar la relación entre variables. El propósito es establecer como una o varias variables dependientes se comportan respecto a una o más variables independientes.
- **Redes neuronales.** Es un modelo que está formado por un conjunto de nodos que están conectados y transmiten información entre sí, esta información se transmite desde la entrada hasta generar una salida. El objetivo del modelo es aprender modificándose automáticamente a si mismo de manera que pueda llegar a realizar tareas complejas que no pueden ser realizadas mediante la programación tradicional.
- **Arboles de decisión.** Es un modelo de predicción donde dado un conjunto de datos predicen o clasifican observaciones futuras basándose en un conjunto de reglas de decisión.
- **Agrupamientos o clustering.** Este modelo consiste en agrupar elementos (ítems) en grupos con características similares y se utiliza para determinar patrones, el objetivo es formar grupos cerrados y homogéneos a partir de un conjunto de elementos que tienen diferentes características o propiedades, pero que comparten ciertas similitudes.

Modelo de regresión lineal múltiple

Un modelo de regresión es un modelo estadístico que nos permite evaluar la relación entre variables, generalmente entre una variable X y otras variables Y's. En este modelo se intenta determinar el valor de X a partir de las variables predictoras Y's.

El modelo de regresión múltiple suele utilizarse para predecir el valor de la variable dependiente (X) o para evaluar la influencia que tienen los predictores (Y's) sobre ella (X).

La forma general de estos modelos es:

$$X_i = \beta_0 + \beta_1 Y_{1i} + \beta_{12} Y_{2i} + \dots + \beta_n Y_{ni} + \varepsilon_i$$

Donde

β_0 : Es la ordenada de origen, el valor de X cuando los predictores sean cero.

β_i : Es la magnitud del efecto que los predictores (Y's) tienen sobre la variable dependiente X.

ε_i : Es el residuo o error, es decir, la diferencia entre el valor observado y el estimado por el modelo.

Para la estimación de los parámetros β_i se pueden utilizar diversos métodos, el más común para los modelos de regresión es utilizar el método de mínimos cuadrados. Sin embargo, la estimación de los parámetros no es un problema ya que se utilizan métodos analíticos disponibles en casi cualquier software para el análisis de datos.

Supuestos del modelo

Para que este modelo sea confiable es necesario establecer los supuestos o condiciones que deben cumplirse para que el modelo sea válido.

Los supuestos son los siguientes:

- **Linealidad.** Implica que la relación entre las variables sea lineal.
- **Independencia.** Los residuos son independientes entre sí.
- **Homocedasticidad.** Que la varianza de los errores sea constante.
- **Normalidad.** Que las variables sigan una distribución normal.
- **No colinealidad.** Que no exista correlación entre las variables.

Con base en estos supuestos se utilizan las pruebas de hipótesis que revisaremos a continuación. Si estos supuestos no se cumplen será necesario utilizar pruebas de hipótesis distintas, por ejemplo, pruebas no paramétricas.

Pruebas de hipótesis

Uno de los problemas que nos podemos encontrar con estos parámetros es que estos sean cero, lo cual implica que el predictor no funciona para el modelo planteado. Por lo cual, durante el análisis de regresión se realizan siempre dos pruebas estadísticas para revisar esto: una general y otra particular.

La primera prueba estadística que realizamos es verificar que existe al menos una β_i distinta de cero, por lo tanto, las hipótesis son:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_i = 0$$

$$H_1: \text{Existe } \beta_i \neq 0$$

Esta prueba de hipótesis realiza un análisis de varianza mediante el cálculo del estadístico F, a partir del cual se obtiene el valor crítico (p-value) a partir de una distribución F. Cuanto más alejado esté el estadístico F de 1, mayor será la evidencia de que al menos uno de los predictores en el modelo es útil, rechazando así la hipótesis nula. La regla de decisión es la siguiente:

Si el Valor crítico F < Nivel de confianza entonces rechazamos H₀ y aceptamos H ₁ .
--

Si pasa esta prueba podemos continuar, ya que sabemos que al menos una de β 's es distinta de cero.

Por lo tanto, la siguiente prueba de hipótesis se aplica sobre cada uno de los predictores para revisar si el valor es cero o distinto de cero. Las hipótesis son:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

A diferencia de la prueba anterior en esta utilizamos una distribución t, donde vamos a preferir que el valor resultante sea el más pequeño, de manera general la regla de decisión será:

Si el Valor crítico t < Nivel de confianza entonces rechazamos H₀ y aceptamos H ₁ .
--

El escenario ideal es que todos los predictores elegidos sean distintos de cero ya que eso implica que si influyen de en el modelo planteado; de lo contrario, si alguno de los predictores no supera esta prueba será necesario replantear su pertenencia al modelo.

Coefficiente de determinación de ajuste R^2

Este coeficiente mide la bondad de ajuste del modelo, corresponde al cuadrado de la correlación entre la variable respuesta y el modelo lineal ajustado.

El valor de este coeficiente esta entre 0 y 1, por lo cual entre más cerca este el valor de 1 mayor será el ajuste del modelo propuesto. El coeficiente se calcula de la siguiente manera:

$$R^2 = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

Sin embargo, uno de los problemas con este coeficiente de determinación es que no penaliza la inclusión de predictores no significativos; es decir, si al modelo le añadimos predictores que aportan muy poco al valor de X de todas formas el coeficiente R^2 aumentará, haciéndonos creer que el modelo se ajusta mejor.

Por lo que se define el coeficiente de determinación ajustado (\bar{R}^2 ajustado), el cual si penaliza la inclusión de predictores. Este se calcula de la siguiente manera:

$$\bar{R}^2 = 1 - \frac{N - 1}{N - k - 1} [1 - R^2]$$

Donde N es el tamaño de la muestra y k es el número de predictores. Por lo cual, si tenemos muchos predictores y pocos datos el valor se acercará a 0. Y esto es una observación muy importante, la muestra siempre debe ser mayor que el número de predictores utilizados.

Apéndice B

curso Período	curso	período	idAsesor	índiceAvance	alumnos	Significativo	Atrasado	Normal	Sobresaliente	Actividades sin calificar	Inasistencia	No se entregan actividades
49_2 020-03-30_34	49	2020-03-30	34	100%	2	0	0	2	19	2	0	0
51_2 020-03-16_52	51	2020-03-16	52	100%	2	0	2	5	19	0	1	6
25_2 020-03-16_40	25	2020-03-16	40	83%	5	5	8	4	8	15	1	1
39_2 020-03-30_6	39	2020-03-30	6	97%	8	0	2	4	15	5	0	1
9_20 20-03-30_50	9	2020-03-30	50	100%	2	1	0	4	30	3	0	2
43_2 020-03-16_44	43	2020-03-16	44	100%	3	1	5	7	14	2	1	10
40_2 020-03-30_469	40	2020-03-30	469	76%	14	1	0	9	12	9	1	0
47_2 020-05-18_35	47	2020-05-18	35	100%	2	0	0	0	2	0	0	0

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dades sin calific ar	Inasi stenc ia	No se entre gan activi dades
24_2 020- 06- 15_3 5	24	2020- 06-15	35	100%	4	0	0	0	4	0	0	0
54_2 020- 10- 12_2 59	54	2020- 10-12	259	100%	2	0	0	0	7	0	0	0
48_2 020- 05- 04_3 8	48	2020- 05-04	38	100%	3	0	0	0	17	0	0	0
6_20 20- 08- 03_3 3	6	2020- 08-03	33	100%	3	0	0	14	14	2	0	12
17_2 020- 10- 26_4 6	17	2020- 10-26	46	100%	4	0	0	1	27	0	0	1
18_2 020- 05- 04_2 59	18	2020- 05-04	259	100%	2	0	0	0	27	0	0	0
25_2 020- 01- 20_4 0	25	2020- 01-20	40	100%	2	1	7	15	5	4	5	14
25_2 020- 02- 17_5 0	25	2020- 02-17	50	100%	4	0	1	6	19	0	0	7
26_2 020- 01-	26	2020- 01-20	55	100%	4	4	3	15	6	3	3	16

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dades sin calific ar	Inasi stenc ia	No se entre gan activi dades
20_5 5												
27_2 020- 10- 26_4 0	27	2020- 10-26	40	100%	2	2	9	13	4	18	0	6
36_2 020- 05- 18_4 0	36	2020- 05-18	40	100%	2	1	3	5	19	1	1	7
36_2 020- 09- 28_4 0	36	2020- 09-28	40	100%	2	0	3	7	18	2	0	8
40_2 020- 08- 17_6	40	2020- 08-17	6	100%	2	0	1	0	27	1	0	0
43_2 020- 05- 18_7 35	43	2020- 05-18	735	100%	2	0	6	11	11	6	1	10
43_2 020- 10- 26_4 4	43	2020- 10-26	44	100%	2	0	4	13	11	3	0	14
44_2 020- 01- 20_4 1	44	2020- 01-20	41	100%	2	15	5	1	5	17	0	4
44_2 020- 08- 03_5 7	44	2020- 08-03	57	100%	4	1	6	11	10	14	1	3
45_2 020-	45	2020- 05-04	57	100%	2	0	3	15	8	0	4	14

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dades sin calific ar	Inasi stenc ia	No se entre gan activi dades
05- 04_5 7												
50_2 020- 05- 04_5 8	50	2020- 05-04	58	100%	3	0	8	2	16	2	0	8
51_2 020- 10- 26_5 0	51	2020- 10-26	50	100%	3	7	12	3	5	0	11	11
73_2 020- 09- 28_5 0	73	2020- 09-28	50	100%	2	0	4	8	16	0	3	9
9_20 20- 10- 12_5 7	9	2020- 10-12	57	100%	3	0	21	13	8	28	0	6
30_2 020- 08- 17_4 4	30	2020- 08-17	44	100%	6	0	0	7	35	0	4	3
36_2 020- 02- 17_4 7	36	2020- 02-17	47	99%	6	2	2	14	8	0	5	13
51_2 020- 04- 20_3 2	51	2020- 04-20	32	98%	4	3	7	17	0	16	2	9
43_2 020- 01- 20_4 4	43	2020- 01-20	44	97%	3	0	7	18	3	0	6	19

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stencia	No se entre gan activi dade s
28_2 020- 03- 30_3 7	28	2020- 03-30	37	96%	2	0	15	13	7	0	6	22
24_2 020- 04- 20_3 0	24	2020- 04-20	30	96%	6	3	0	4	0	3	3	1
50_2 020- 03- 16_5 5	50	2020- 03-16	55	96%	4	2	4	14	7	2	14	4
21_2 020- 11- 09_3 7	21	2020- 11-09	37	94%	4	5	0	10	13	7	4	4
9_20 20- 05- 04_3 0	9	2020- 05-04	30	94%	2	1	15	16	9	21	2	9
55_2 020- 10- 12_5 8	55	2020- 10-12	58	93%	5	1	19	15	0	4	5	26
55_2 020- 02- 03_3 4	55	2020- 02-03	34	90%	2	0	7	20	6	0	7	20
28_2 020- 05- 18_1 041	28	2020- 05-18	1041	90%	2	10	21	8	0	27	0	12
59_2 020- 03-	59	2020- 03-30	53	89%	2	0	4	9	9	0	1	12

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stenc ia	No se entre gan activi dade s
30_5 3												
4_20 20- 08- 31_3 7	4	2020- 08-31	37	89%	9	0	0	7	0	0	3	4
23_2 020- 06- 15_5 2	23	2020- 06-15	52	89%	9	0	1	6	0	0	4	3
40_2 020- 06- 01_5 7	40	2020- 06-01	57	89%	3	0	0	8	20	0	7	1
54_2 020- 08- 31_2 59	54	2020- 08-31	259	88%	2	0	0	1	0	0	0	1
4_20 20- 04- 20_5 7	4	2020- 04-20	57	88%	20	1	0	6	0	0	7	0
4_20 20- 06- 15_3 7	4	2020- 06-15	37	88%	13	1	0	6	0	1	5	1
4_20 20- 10- 26_4 69	4	2020- 10-26	469	88%	7	5	0	2	0	5	2	0
24_2 020- 08- 31_3 5	24	2020- 08-31	35	88%	2	0	0	7	0	0	0	7

curso Período	curso	período	idAse sor	indic eAvance	alumnos	Significativo	Atrasado	Normal	Sobresaliente	Actividades sin calificar	Inasistencia	No se entregan actividades
25_2020-08-03_40	25	2020-08-03	40	88%	4	3	8	13	4	15	8	1
29_2020-06-01_44	29	2020-06-01	44	88%	4	0	6	20	2	0	5	21
73_2020-01-20_56	73	2020-01-20	56	87%	6	3	12	11	0	4	20	2
10_2020-11-09_53	10	2020-11-09	53	85%	3	1	3	20	4	0	8	16
33_2020-02-17_30	33	2020-02-17	30	85%	2	1	2	22	14	7	8	10
6_2020-03-30_49	6	2020-03-30	49	84%	4	0	1	13	8	0	14	0
6_2020-05-18_41	6	2020-05-18	41	84%	4	11	10	3	4	18	3	3
4_2020-01-20_255	4	2020-01-20	255	83%	11	0	1	5	0	1	5	0
4_2020-03-16	4	2020-03-16	37	83%	9	0	1	6	0	0	5	2

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stenc ia	No se entre gan activi dade s
16_3 7												
4_20 20- 09- 28_4 69	4	2020- 09-28	469	83%	7	1	0	6	0	1	5	1
11_2 020- 02- 17_2 14	11	2020- 02-17	214	83%	2	3	1	6	8	0	8	2
11_2 020- 04- 20_3 4	11	2020- 04-20	34	83%	3	0	0	8	12	0	2	6
43_2 020- 04- 20_4 4	43	2020- 04-20	44	83%	4	1	11	14	1	20	0	6
87_2 020- 08- 31_6	87	2020- 08-31	6	83%	9	1	4	24	13	5	21	3
4_20 20- 05- 18_4 69	4	2020- 05-18	469	82%	10	3	0	4	0	3	3	1
39_2 020- 11- 09_3 3	39	2020- 11-09	33	82%	13	5	15	15	0	19	14	2
4_20 20- 11- 23_3 0	4	2020- 11-23	30	80%	15	3	0	4	0	3	4	0
41_2 020-	41	2020- 10-12	37	80%	5	3	2	23	0	0	9	19

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dades sin calific ar	Inasi stenc ia	No se entre gan activi dades
10- 12_3 7												
28_2 020- 09- 14_3 1	28	2020- 09-14	31	80%	4	26	12	4	0	37	5	0
23_2 020- 08- 17_5 2	23	2020- 08-17	52	79%	7	0	3	4	0	0	6	1
29_2 020- 11- 09_4 4	29	2020- 11-09	44	79%	3	0	11	11	6	3	10	9
46_2 020- 10- 26_4 0	46	2020- 10-26	40	79%	3	0	7	12	8	3	11	5
45_2 020- 03- 16_3 4	45	2020- 03-16	34	79%	3	0	7	10	10	9	8	0
6_20 20- 08- 31_3 3	6	2020- 08-31	33	78%	4	0	6	12	10	0	11	7
39_2 020- 08- 17_5 6	39	2020- 08-17	56	78%	5	2	14	11	0	1	18	8
51_2 020- 01- 20_6	51	2020- 01-20	6	78%	5	8	8	11	0	1	22	4

curso Período	curso	período	idAsesor	índiceAvance	alumnos	Significativo	Atrasado	Normal	Sobresaliente	Actividades sin calificar	Inasistencia	No se entregan actividades
51_2 020-08-03_3 2	51	2020-08-03	32	78%	16	2	14	12	0	19	9	0
7_20 20-03-16_2 55	7	2020-03-16	255	77%	2	0	7	2	17	3	3	3
10_2 020-06-01_4 0	10	2020-06-01	40	76%	5	1	8	13	5	14	8	0
70_2 020-08-17_2 14	70	2020-08-17	214	76%	17	24	1	3	0	5	23	0
53_2 020-10-12_4 9	53	2020-10-12	49	75%	3	0	0	7	0	0	7	0
54_2 020-05-18_2 43	54	2020-05-18	243	75%	2	0	3	3	0	0	6	0
29_2 020-05-04_4 4	29	2020-05-04	44	75%	3	0	10	12	5	8	1	13
39_2 020-05-04_1 041	39	2020-05-04	1041	75%	4	7	15	4	0	11	2	13
51_2 020-09-	51	2020-09-28	50	75%	4	19	4	5	0	0	27	1

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stenc ia	No se entre gan activi dade s
28_5 0												
81_2 020- 08- 31_5 7	81	2020- 08-31	57	75%	4	19	1	7	0	22	5	0
38_2 020- 11- 09_5 6	38	2020- 11-09	56	75%	7	3	10	22	0	9	25	1
51_2 020- 08- 31_3 2	51	2020- 08-31	32	74%	16	3	16	9	0	12	15	1
87_2 020- 05- 18_5 4	87	2020- 05-18	54	74%	6	1	13	16	12	1	25	4
51_2 020- 08- 03_2 55	51	2020- 08-03	255	73%	15	7	16	4	0	14	13	0
38_2 020- 06- 01_5 3	38	2020- 06-01	53	73%	4	1	14	11	2	0	18	8
23_2 020- 11- 09_5 6	23	2020- 11-09	56	71%	7	0	5	0	0	0	5	0
95_2 020- 09- 14_7 35	95	2020- 09-14	735	71%	3	3	23	2	0	11	9	8

curso Período	curso	período	idAsesor	indic eAvance	alumnos	Significativo	Atrasado	Normal	Sobresaliente	Actividades sin calificar	Inasistencia	No se entregan actividades
110_2020-11-09_255	110	2020-11-09	255	71%	17	2	15	9	2	1	25	0
11_2020-06-15_57	11	2020-06-15	57	70%	15	1	12	1	0	0	14	0
49_2020-11-09_34	49	2020-11-09	34	70%	3	1	8	5	0	0	9	5
26_2020-11-09_55	26	2020-11-09	55	70%	2	6	20	0	0	3	4	19
51_2020-05-18_50	51	2020-05-18	50	70%	5	0	13	14	0	2	21	4
86_2020-02-03_255	86	2020-02-03	255	70%	21	10	10	8	0	27	1	0
87_2020-02-17_6	87	2020-02-17	6	70%	5	14	25	0	0	1	29	9
43_2020-08-03_735	43	2020-08-03	735	68%	4	1	22	2	2	0	25	0
28_2020-10-10-	28	2020-10-12	302	68%	4	28	13	0	0	19	15	7

curso Período	curso	período	idAse sor	indic eAvance	alumnos	Significativo	Atrasado	Normal	Sobresaliente	Actividades sin calificar	Inasistencia	No se entregan actividades
12_302												
22_2020-10-26_34	22	2020-10-26	34	67%	3	4	8	0	2	0	12	0
11_2020-05-18_0	11	2020-05-18	0	67%	3	9	11	0	0	3	12	5
7_2020-08-03_58	7	2020-08-03	58	67%	12	0	15	0	12	6	8	1
8_2020-06-01_302	8	2020-06-01	302	67%	22	11	17	0	0	17	11	0
49_2020-06-01_33	49	2020-06-01	33	67%	3	0	7	7	14	0	14	0
21_2020-03-16_259	21	2020-03-16	259	67%	2	1	16	10	0	0	24	3
29_2020-03-30_735	29	2020-03-30	735	67%	3	1	7	7	5	4	11	0
74_2020-04-20_214	74	2020-04-20	214	66%	7	0	7	0	20	0	7	0
33_2020-08-03	33	2020-08-03	50	66%	4	1	11	28	2	4	34	2

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Normal	Sobr esali ente	Activi dades sin calificar	Inasi stencia	No se entre gan activi dades
08-03_50												
39_2020-06-01_1041	39	2020-06-01	1041	65%	6	6	19	0	0	4	21	0
89_2020-03-09_32	89	2020-03-09	32	65%	21	21	13	0	0	33	1	0
38_2020-05-04_36	38	2020-05-04	36	65%	6	4	22	1	0	0	27	0
38_2020-02-03_36	38	2020-02-03	36	64%	9	2	18	8	0	0	23	5
73_2020-04-13_729	73	2020-04-13	729	64%	16	0	6	5	0	0	11	0
12_2020-08-03_418	12	2020-08-03	418	63%	4	2	5	0	0	2	4	1
111_2020-11-30_56	111	2020-11-30	56	63%	16	3	11	0	0	4	10	0
7_2020-09-28_57	7	2020-09-28	57	63%	8	0	18	6	4	11	11	2

curso Período	curso	período	idAsesor	índiceAvance	alumnos	Significativo	Atrasado	Normal	Sobresaliente	Actividades sin calificar	Inasistencia	No se entregan actividades
25_2 020-08-31_5 8	25	2020-08-31	58	63%	4	4	14	9	0	1	26	0
27_2 020-08-31_4 7	27	2020-08-31	47	63%	4	1	15	6	6	0	19	3
95_2 020-11-09_3 1	95	2020-11-09	31	63%	2	21	5	0	0	8	3	15
7_20 20-04-20_2 55	7	2020-04-20	255	62%	9	6	9	5	4	14	6	0
26_2 020-06-01_5 5	26	2020-06-01	55	62%	5	13	14	0	0	6	20	1
55_2 020-05-04_4 1	55	2020-05-04	41	62%	5	29	5	0	0	16	13	5
12_2 020-06-01_3 3	12	2020-06-01	33	61%	3	0	7	0	0	0	7	0
12_2 020-10-26_4 69	12	2020-10-26	469	61%	3	5	2	0	0	5	2	0
7_20 20-01-	7	2020-01-20	57	61%	3	0	14	5	9	1	11	7

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stenc ia	No se entre gan activi dade s
20_5 7												
39_2 020- 02- 03_1 041	39	2020- 02-03	1041	61%	6	5	21	0	0	4	18	4
39_2 020- 10- 12_5 6	39	2020- 10-12	56	61%	3	4	23	0	0	2	19	6
55_2 020- 09- 14_4 69	55	2020- 09-14	469	61%	4	18	17	0	0	27	8	0
87_2 020- 05- 04_7 65	87	2020- 05-04	765	61%	11	0	25	16	0	0	40	1
87_2 020- 03- 16_3 0	87	2020- 03-16	30	60%	5	4	20	17	0	11	26	4
4_20 20- 08- 03_3 0	4	2020- 08-03	30	60%	12	2	5	0	0	2	5	0
85_2 020- 02- 10_4	85	2020- 02-10	4	60%	18	1	7	2	0	8	2	0
38_2 020- 09- 14_2 14	38	2020- 09-14	214	60%	5	7	21	0	0	12	4	12
92_2 020-	92	2020- 04-20	58	60%	22	5	21	0	0	15	11	0

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dades sin calific ar	Inasi stenc ia	No se entre gan activi dades
04- 20_5 8												
34_2 020- 10- 26_3 6	34	2020- 10-26	36	59%	3	4	24	0	0	0	7	21
87_2 020- 04- 20_4 1	87	2020- 04-20	41	59%	7	31	10	0	0	34	4	3
11_2 020- 09- 28_2 14	11	2020- 09-28	214	58%	2	5	14	0	0	3	2	14
55_2 020- 03- 02_4 1	55	2020- 03-02	41	58%	4	6	24	2	0	28	4	0
33_2 020- 05- 18_4 18	33	2020- 05-18	418	58%	3	19	19	2	0	36	4	0
38_2 020- 08- 17_5 3	38	2020- 08-17	53	57%	7	5	22	1	0	1	24	3
9_20 20- 03- 02_5 1	9	2020- 03-02	51	57%	2	3	31	5	0	7	32	0
6_20 20- 10- 26_5 7	6	2020- 10-26	57	56%	2	16	12	0	0	11	13	4

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dades sin calific ar	Inasi stenc ia	No se entre gan activi dades
7_20 20- 08- 31_4 18	7	2020- 08-31	418	56%	6	3	11	7	7	15	4	2
73_2 020- 05- 04_7 29	73	2020- 05-04	729	56%	15	14	12	0	0	12	14	0
87_2 020- 10- 26_4 1	87	2020- 10-26	41	56%	10	30	12	0	0	34	8	0
16_2 020- 10- 12_3 0	16	2020- 10-12	30	55%	4	3	11	9	0	3	7	13
10_2 020- 09- 28_3 6	10	2020- 09-28	36	55%	3	8	18	2	0	0	19	9
14_2 020- 10- 12_0	14	2020- 10-12	0	54%	7	10	19	5	0	11	23	0
97_2 020- 09- 21_2 55	97	2020- 09-21	255	53%	16	14	11	0	0	13	12	0
14_2 020- 03- 02_3 6	14	2020- 03-02	36	53%	2	0	27	5	0	0	29	3
16_2 020- 03-	16	2020- 03-30	50	52%	4	1	2	8	5	6	1	4

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stenc ia	No se entre gan activi dade s
30_5 0												
55_2 020- 08- 17_4 69	55	2020- 08-17	469	52%	6	20	14	1	0	25	10	0
45_2 020- 09- 28_1 041	45	2020- 09-28	1041	51%	5	16	8	0	0	7	17	0
23_2 020- 03- 16_3 3	23	2020- 03-16	33	50%	2	0	1	0	0	0	0	1
23_2 020- 05- 04_3 3	23	2020- 05-04	33	50%	3	0	1	0	0	0	0	1
48_2 020- 10- 12_3 5	48	2020- 10-12	35	50%	3	0	7	0	0	0	7	0
22_2 020- 11- 23_6	22	2020- 11-23	6	50%	2	2	11	0	0	0	13	0
11_2 020- 01- 20_2 14	11	2020- 01-20	214	50%	6	1	6	9	5	0	13	3
11_2 020- 03- 16_2 14	11	2020- 03-16	214	50%	4	9	9	0	0	11	5	2
11_2 020-	11	2020- 10-26	56	50%	3	2	18	0	0	0	16	4

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dades sin calific ar	Inasi stenc ia	No se entre gan activi dades
10- 26_5 6												
11_2 020- 11- 23_2 55	11	2020- 11-23	255	50%	5	17	4	0	0	8	13	0
7_20 20- 02- 17_5 6	7	2020- 02-17	56	50%	7	1	13	6	4	8	11	1
40_2 020- 03- 02_4 69	40	2020- 03-02	469	50%	4	2	21	2	0	11	14	0
44_2 020- 06- 01_3 4	44	2020- 06-01	34	50%	2	8	16	1	0	0	19	6
7_20 20- 10- 26_1 041	7	2020- 10-26	1041	49%	7	17	7	4	0	22	5	1
87_2 020- 09- 28_4 1	87	2020- 09-28	41	49%	6	22	20	0	0	30	10	2
7_20 20- 05- 18_4 18	7	2020- 05-18	418	48%	4	7	8	6	5	15	4	2
41_2 020- 03- 02_3 1	41	2020- 03-02	31	48%	4	17	3	2	0	18	0	4

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stencia	No se entre gan activi dade s
40_2 020- 05- 04_2 14	40	2020- 05-04	214	47%	13	11	12	4	0	11	16	0
12_2 020- 03- 02_3 3	12	2020- 03-02	33	46%	4	5	0	0	0	0	5	0
45_2 020- 08- 17_1 041	45	2020- 08-17	1041	44%	4	16	12	0	0	13	14	1
9_20 20- 08- 17_3 0	9	2020- 08-17	30	44%	3	30	9	3	0	30	12	0
33_2 020- 10- 26_5 0	33	2020- 10-26	50	43%	2	41	0	0	0	0	41	0
87_2 020- 04- 01_7 65	87	2020- 04-01	765	42%	6	8	20	2	4	0	30	0
33_2 020- 09- 28_3 3	33	2020- 09-28	33	40%	2	37	5	0	0	4	32	6
40_2 020- 11- 09_4 69	40	2020- 11-09	469	39%	3	14	5	0	0	19	0	0
38_2 020- 03-	38	2020- 03-30	36	36%	2	22	0	0	0	0	15	7

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stenc ia	No se entre gan activi dade s
30_3 6												
52_2 020- 05- 18_3 7	52	2020- 05-18	37	34%	29	21	0	0	0	0	21	0
90_2 020- 03- 09_8	90	2020- 03-09	8	32%	17	18	2	0	0	20	0	0
74_2 020- 05- 18_1 63	74	2020- 05-18	163	31%	31	16	12	0	0	17	11	0
74_2 020- 09- 14_1 041	74	2020- 09-14	1041	30%	2	21	0	3	0	0	24	0
53_2 020- 05- 18_3 5	53	2020- 05-18	35	25%	2	5	0	0	0	0	5	0
54_2 020- 03- 30_2 59	54	2020- 03-30	259	25%	2	7	0	0	0	4	0	3
40_2 020- 10- 12_2 14	40	2020- 10-12	214	25%	2	28	0	0	0	28	0	0
33_2 020- 01- 20_4 18	33	2020- 01-20	418	24%	4	38	0	0	0	24	14	0
52_2 020-	52	2020- 10-12	37	21%	3	15	0	0	0	2	9	4

curso Perio do	curso	perio do	idAse sor	indic eAva nce	alum nos	Signif icativ o	Atras ado	Norm al	Sobr esali ente	Activi dade s sin calific ar	Inasi stenc ia	No se entre gan activi dade s
10- 12_3 7												
74_2 020- 05- 18_3 4	74	2020- 05-18	34	19%	30	25	3	0	0	0	28	0
93_2 020- 05- 18_7 35	93	2020- 05-18	735	17%	22	28	0	0	0	27	1	0
58_2 020- 06- 29_2 55	58	2020- 06-29	255	8%	16	1	0	0	0	0	1	0

Referencias

- (01 de 07 de 2022). Obtenido de ¿Qué es una base de datos?: <https://www.oracle.com/mx/database/what-is-database/>
- (01 de 07 de 2022). Obtenido de ¿Qué es una base de datos relacional?: <https://aws.amazon.com/es/relational-database/>
- (01 de 07 de 2022). Obtenido de ¿Qué es una base de datos relacional?: <https://cloud.google.com/learn/what-is-a-relational-database>
- Aretio, L. G. (2019). El problema del abandono en estudios a distancia. Respuestas desde el Diálogo Didáctico Mediado. *Revista Iberoamericana de Educación a Distancia*, 245-270.
- Dalaarg, P. (2008). *Introductory Statistics with R* (Segunda ed.). (D. H. J. Chambers, Ed.) Dinamarca: Springer.
- Daysi García-Tinizaray, K. O.-B.-D. (2014). Learning analytics para predecir la deserción de estudiantes a distancia. *Campus Virtuales*, 1(3), 120-126.
- Ed Fincham, A. W.-W.-P. (2019). Counting Clicks is Not Enough: Validating a Theorized Model of Engagement in Learning Analytics. *The 9th International Learning Analytics & Knowledge Conference*, 501-510.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenge. *International Journal of Technology Enhanced Learning*, 304-317.
- Gaspars-Wieloch. (2021). The Assignment Problem in Human Resource Project Management under Uncertainty. *Risks*.
- Janice K. Winch, J. Y. (2014). Case Article—Class Scheduling with Linear Programming. *INFORMS Transactions on Education*, 143-147.
- John J. Dinkel, J. M. (1989). OR Practice—An Efficient Decision Support System for Academic Course Scheduling. *Operations Research* 37, 853-864.
- K., E. H. (s.f.). Modelos de análisis de la deserción estudiantil en la educación estudiantil . *Retención y movilidad estudiantil*, 91-108.
- Llanos Navarrete, S., & Zárate Pérez, A. F. (2011). *Sistematización para la creación de unidades didácticas para cursos en línea*. Universidad Nacional Autónoma de México.
- M. König, C. K. (2012). Intelligent BIM-based construction scheduling using discrete event simulation. *Proceedings of the 2012 Winter Simulation Conference*, 1-12.
- Madrid, J. L. (2016). Factores que promueven la deserción del aula virtual. *Orbis, Revista Científica Ciencias Humanas*, 12(35), 18-40.
- Marqués, M., & Universitat Jaume I. (2011). *Bases de Datos*. Castelló de la Plana.
- Montgomery, D. C. (2004). *Diseño y análisis de experimentos* (Segunda ed.). New York: Limusa Wiley.

- Mulvey, J. M. (1982). A classroom/time assignment model. *European Journal of Operational Research* Volume 9, Issue 1, 64-70.
- Mulvey, J. M. (1982). A classroom/time assignment model. *European Journal of Operational Research*, Volume 9, Issue 1, 64-70.
- Rao, S. S. (2009). *Engineering Optimization: Theory and Practice* . New Jersey : JOHN WILEY & SONS.
- Ravindra K. Ahuja, T. L. (1993). *Network Flows: Theory, algorithms*. New Jersey, Upper Saddle River: Prentice Hall.
- Rockafellar, R. T. (1998). *Network Flows and Monotropic Optimization*. Belmont: Athena Scientific .
- Sabulsky, G. (2019). Analíticas de aprendizaje para mejorar la enseñanza y el seguimiento a través de entornos virtuales. *Revista Iberoamericana de Educación*, 80(1), 13-30.
- Sahni, E. H. (1978). *Fundamentals of computer algorithms*. Maryland: Computer Science Press.
- Syakinah Faudzi, S. A.-R. (2018). An Assignment Problem and Its Application in Education Domain: A Review and Potential Path. *Advances in Operations Research*. Obtenido de <https://doi.org/10.1155/2018/8958393>
- Thensen, A. (1978). *Computer methods in operations research*. (J. W. Schmidt, Ed.) New York: Academic Press.
- Tillett, P. (1989). An operations research approach to the assignment of teachers to courses. *Socio-Economic Planning Sciences*, Volume 9, Issues 3–4.
- Zarate Perez, A. F., & Flores de la Mota, I. (2021). Un modelo para el seguimiento de cursos de capacitación. *Undécima Conferencia de Directores de Tecnología de Información y Comunicación en Instituciones de Educación Superior, TICAL2021 y 5° Encuentro Latinoamericano de e-Ciencia “Repensando La Universidad Impulsada por las Tecnologías Digitales”* (págs. 208-221). RedClara.