



**UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO**

FACULTAD DE CIENCIAS

COMPRESION Y MANEJO DEL RAW DATA

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

ARTURO ROBERTO ISLAS ORTIZ

DIRECTOR DE TESIS: ACT. LEONARDO LOPEZ MONROY



2005

FACULTAD DE CIENCIAS
SERVICIO BIBLIOTECARIO

m. 348463



Universidad Nacional
Autónoma de México



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.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: Arturo Roberto Islas Ortiz
 FECHA: 27 septiembre 2005
 FIRMA: p.r. Gabriela Islas

ACT. MAURICIO AGUILAR GONZÁLEZ
 Jefe de la División de Estudios Profesionales de la
 Facultad de Ciencias
 Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

"Compresión y Manejo del Raw Data"

realizado por Arturo Roberto Islas Ortiz

con número de cuenta 09712172-9 , quien cubrió los créditos de la carrera de:
 Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
 Propietario

Act. Leonardo López Monroy

Propietario

M. en I. María de Luz Gasca Soto

Propietario

Mat. Laura Pastrana Ramírez

Laura Pastrana R

Suplente

Mat. Margarita Elvira Chávez Cano

Suplente

Dr. Luis Antonio Rincón Solís

Consejo Departamental de
 Matemáticas



Act. Jaime Vázquez Alamilla

FACULTAD DE CIENCIAS
 CONSEJO DEPARTAMENTAL
 DE
 MATEMÁTICAS

A Dios por permitirme llegar hasta aquí.

A mi Papá por su apoyo incondicional en mi vida, gracias Papá.

A mi Mamá por darme toda su paciencia, tiempo y amor, gracias

Mamá. A ambos, por educarme, confiar y creer siempre en mí.

A mis hermanos, Víctor, Gaby y Mónica que siempre me han apoyado y motivado a seguir adelante.

A todos mis familiares, amigos, compañeros y profesores que me han dado un espacio en su vida y me han enseñado a ser como soy.

A mi asesor Leonardo por su resistencia a todas esas tazas de café necesarias para terminar este trabajo.

Por último, a la Universidad Nacional Autónoma de México por darme todo lo que ninguna otra Institución me hubiera podido otorgar.

El uso de información es indispensable en cualquier ámbito; en la industria de la televisión el almacenamiento de los datos utilizados para calcular la Audiencia de los programas tiende a complicarse debido a que el volumen de estos datos crece rápidamente.

Mediante el uso de funciones simples y un buen ordenamiento podemos comprimir estos datos, con el propósito de guardar toda la información en forma más compacta y obtener una explotación más eficiente, haciendo posible un aprovechamiento más completo de éstos. El presente trabajo tiene por objeto mostrar la manera como se recolectan los datos para posteriormente desarrollar un reordenamiento de los mismos, con el fin de mejorar el almacenamiento así como el aprovechamiento.

En el primer capítulo mencionaremos una breve historia de la publicidad, comentando las formas de uso que el hombre le ha dado a través del tiempo dependiendo de sus distintas necesidades de información. En esta parte analizaremos los distintos medios que actualmente son utilizados para publicitar, destacando la efectividad de la televisión como medio para publicitar, así como la importancia de la acumulación y el manejo de los datos objeto de estudio.

En el segundo capítulo establecemos las definiciones que utilizaremos a lo largo del trabajo, en la primera sección profundizaremos en la descripción de la base de datos que contiene toda la información relativa al consumo televisivo, esta base se conoce con el nombre de Raw Data, en donde se pueden observar todos los datos crudos del consumo televisivo. Explicaremos lo mejor posible la composición y el arreglo de los datos que en él se encuentran, analizando sus defectos así como los atributos con los que cuenta. En la segunda sección desarrollamos las definiciones que sirven para el desarrollo del trabajo, de ahí la importancia del mismo.

En el capítulo tercero exponemos el nuevo modelo de explotación para los datos del Consumo televisivo, en la primera sección delimitaremos los supuestos utilizados para nuestro modelo, así como el proceso seguido para el desarrollo del mismo, en esta sección asentamos la forma de abordar el problema presentado. La segunda sección desarrolla el caso en donde sólo tenemos un televidente, el cual cuenta únicamente con un canal para ver, esta sección sirve para que el lector empiece a utilizar las definiciones del capítulo dos, de ahí la importancia de éste ya que de ahora en adelante utilizamos las definiciones ahí estableci-

das. La tercera sección extiende el caso anterior, dando al televidente la posibilidad de ver dos canales, en él explicamos la mayoría de los casos posibles en el consumo televisivo. En esta sección justificamos el modelo desarrollado, dando las correspondientes pruebas matemáticas. En la cuarta sección generalizaremos el caso anterior, en esta parte desarrollaremos el modelo en donde el televidente puede ver un número fijo de canales, esta parte nos da la pauta para poder desarrollar el modelo general. En esta sección comprimimos toda la información del televidente en una matriz denominada “matriz de consumo”, no obstante que en los casos anteriores ya se había incluido esta matriz, es hasta este momento donde podemos realizar la construcción completa de ella. Finalmente, en la quinta sección desarrollamos el caso general, en donde abarcamos toda la información del Raw Data. Ya con toda la información describimos el proceso a seguir para calcular la Audiencia de cualquier programa o de cualquier periodo de tiempo que se desee conocer. En la implementación de nuestro modelo llegamos a la conclusión de que los números que utilizábamos excedían la capacidad de memoria de los sistemas de cómputo, motivo por el cual desarrollamos el cuarto capítulo, en donde analizamos el comportamiento de nues-

tras propuestas contra los sistemas reales, llevando a cabo los ajustes pertinentes para solucionar la problemática. Los ajustes resultaron ser simples, efectivos y no afectaron el desempeño del modelo. Como punto final realizamos las conclusiones del trabajo, exponiendo nuestros comentarios, resultados obtenidos y cuestionamientos.

Índice

1	Introducción	9
2	Definiciones	23
2.1	El Raw Data	23
2.2	Definiciones Generales	33
3	Un Nuevo Modelo de Explotación	41
3.1	Hipótesis y Metodología del Modelo	42
3.2	El caso inicial	45
3.3	Un televidente con dos canales	47
3.4	Un televidente con sistema de cable	53
3.5	El caso general	70
4	Consideraciones Numéricas	75

8

ÍNDICE

5 Conclusiones

79

Bibliografía

85

Capítulo 1

Introducción

La venta, intercambio y compra de bienes y servicios son las actividades económicas más importantes. Es curioso ver como estas actividades surgen como una necesidad casi natural del hombre desde una etapa muy primitiva, esto debido, a que un sólo hombre no puede llevar acabo todas las labores necesarias para satisfacer sus necesidades primarias e incluso sus lujos y caprichos personales, se ve forzado a empezar el trueque de mercancías, él cual beneficiaba a ambas partes participantes.

Más tarde se decide estandarizar el trueque mediante el establecimiento y la aceptación de moneda como forma de intercambio. Posteriormente, el trueque deja de ser una necesidad y se convierte en una actividad propia del hombre, naciendo así los primeros mercaderes.

Con el establecimiento de los primeros mercaderes se inicia, de manera paralela, una actividad conocida como Publicidad, la cual consiste en hacer llegar un mensaje de un cierto bien o servicio hacia los posibles compradores de éste, ya sea para lograr una mayor demanda, obtener la preferencia de los consumidores, entre otros.

Los inicios de la publicidad se observan en las civilizaciones de la antigüedad, en donde inicialmente se utilizaron los muros de las ciudades como medio para plasmar el mensaje dirigido a los posibles adquirientes; por ejemplo en los muros de la antigua Roma se han encontrado anuncios que indican la venta de una propiedad, de hecho desde hace muchos años los colores afuera de los establecimientos han sido utilizados por algunas personas como publicidad particular, basta observar que aún en nuestros tiempos las peluquerías conservan su publicidad de antaño; la cual consiste en poner franjas sesgadas de color blanco y azul que adornan sus paredes y que algunas veces también son colocados en cilindros giratorios afuera de las mismas. En este sentido es interesante reflexionar sobre la cantidad de publicidad que nos rodea diariamente y que forma parte de la vida cotidiana.

Otra de las modalidades de la publicidad es la que consiste en promover los productos de viva voz, incluso actualmente hay muchos tipos de comercio que siguen utilizando la publicidad de esta manera, el claro ejemplo de esto es el comercio ilegal, por ejemplo la venta de discos en el metro o en la vía pública donde los comerciantes compiten por atraer la atención de la gente en su entorno; aunque también existen locales establecidos que utilizan específicamente esta publicidad.

La publicidad, como cualquier otra actividad ha tenido una evolución, el primer avance representativo de ésta se da con la llegada de la producción masiva ocasionada por la revolución industrial, se acrecenta la necesidad de transmitir de forma igualmente masiva la publicidad, la vanguardia de la imprenta acarrea un importante avance en la publicidad, donde era posible hacer un gran número de anuncios a un bajo costo, de esta forma se alcanzaba a mas posibles compradores mediante el volanteo y la distribución de carteles, sin embargo, esta evolución estaba limitada a aquellas personas que sabían leer, en esos tiempos los grandes conflictos mundiales provocaron que la educación pasara a segundo término para la gran mayoría de los países. En esta época en

donde las ciudades no eran tan grandes como lo son actualmente, el rol que desempeñan los pregoneros fue fundamental para lograr el desarrollo de la publicidad, ya que era posible elegir lugares estratégicos por donde pasara la mayoría de las personas, es ahí precisamente donde se conglomeraban los pregoneros a recitar los anuncios.

Posteriormente a la creación de la imprenta, la ciencia evolucionando dando lugar a la radio, medio de comunicación común en nuestros tiempos y por ende utilizado como vía de publicidad por su gran alcance. En sus inicios la radio no fue un medio utilizado para publicitar debido al fenómeno social que ocurre con todos los medios electrónicos en sus primeros años el cual consiste en una escasa demanda, lo cual obliga al medio a elevar sus costos de adquisición, todo ello lleva al medio a contar con un corto alcance; por lo anterior la radio no apoyó la evolución de la publicidad sino hasta que logró tener la penetración que la caracteriza actualmente. La radio solventó el problema de alfabetización que limitaba a los carteles, sin embargo, carecía del impacto visual logrado por los medios impresos.

No fue sino hasta la creación de la televisión (y específicamente cuando ésta logró un gran alcance) que se dio la más importante evolución de la publicidad, al principio este medio presentaba la misma limitante tecnológica que el radio, posteriormente, cuando superó el fenómeno descrito anteriormente no pasó mucho tiempo para que se notaran las virtudes de este medio para publicitar, poseía la virtud de lograr un impacto audio visual que no lograba ninguno de los medios anteriores, capturando así la atención del espectador, realizando un mayor impacto sobre éste, de forma tal que provocaba una intención de compra con mayor posibilidad de llevarse a cabo.

Actualmente la televisión sigue siendo el medio que captura en mayor medida la atención del espectador, razón por la cual se utiliza idóneamente para publicitar, no obstante que otros medios de comunicación como Internet están cobrando relativa importancia, el fácil acceso de la población a la televisión hace que conserve su preferencia, otra razón de la preferencia para publicitar en televisión es debido a que en ella se obtiene entretenimiento a un bajo costo [6]. Nótese que la televisión no siempre ofrece el mejor entretenimiento que la gente

busca, sin embargo su consumo se ha ido formando casi como una costumbre. Este consumo se ha ido clasificando cada vez más, al principio la televisión ofrecía un entretenimiento familiar (actualmente aún ofrece este tipo de entretenimiento) pero posteriormente, las necesidades de los anunciantes (y a consecuencia de las mismas televisoras) provocaron que los programas fueran dirigidos a un cierto grupo de personas con características similares, actualmente se pueden observar canales u horarios, o ambos, dirigidos a un grupo específico de personas, por ejemplo: niños, amas de casa, padres de familia, adolescentes, etcétera. Pese a que la radio también ofrece un fácil acceso, el consumo de este medio es distinto al televisivo, las personas utilizan normalmente la radio para acompañar otras actividades diarias, como por ejemplo: conducir, cocinar o hacer ejercicio y muchas veces no ponen tanta atención en ella como lo hacen en la televisión, motivo por el cual la publicidad en este medio no impacta del mismo modo. Por su parte los medios impresos como revistas y periódicos son utilizados como vías de información, el consumo en este medio es más específico, las personas sólo leen las secciones o artículos de su interés y muchas veces pasan por alto los mensajes que no logran llamar su atención. [7]

Considerando la distribución que tiene la televisión (de alrededor del 98 por ciento de la población en todo el país), y del tipo de impacto que causa en los consumidores, podemos afirmar que la publicidad en este medio es de gran importancia para los anunciantes y para las propias televisoras ya que para ambos representa un gran negocio. [1]

Pese a que todos estos medios llegaban a una gran cantidad de personas se desconocía la cantidad de gente que entraba en contacto con ellos, por esto surge la necesidad de crear una medida que estime dicha cantidad; la medida a utilizar es conocida como Audiencia o Rating (por su nombre en Inglés). La Audiencia se calcula de forma distinta para cada medio y muchas veces este cálculo no es tan preciso como se desearía. Las fallas en el cálculo surgen por problemas básicos como conocer que es lo que ve, oye o lee cada persona, menos aún saber que percibe o que es lo que realmente atrapó su atención. Por lo anterior se recurre a la toma de una muestra que otorga una idea global de como es dicho comportamiento. De forma general, la Audiencia ha sido adoptada por cada medio como un estándar para medir el número de personas que son alcanzadas por un mensaje publicitario.

En el presente trabajo nos enfocaremos a la Audiencia del medio televisivo, por ello desde ahora, cuando hablemos del medio, se sobreentenderá que estamos hablando del medio televisivo y cuando mencionemos a los participantes del medio, nos referiremos a las cadenas televisivas, agencias de publicidad, empresas y demás personas que utilizan a la televisión como vía para publicitarse. A continuación describiremos brevemente la forma utilizada por el medio para calcular la Audiencia.

La Audiencia es calculada de forma distinta en cada país [2], en México se utilizaron distintos medidores de consumo televisivo, algunos de ellos fueron los “Set Meter’s”, los cuales eran antenas que capturaban todas las señales en un área o conjunto de personas determinado, generando una *muestra* regional que pretende describir el comportamiento global de la población, los más recientes medidores son los “People Meter Móviles”, los cuales son aparatos portátiles que captan las señales a las que está expuesta una persona a lo largo del día, con estos medidores se pretende capturar también el consumo de la radio. Los medidores que se utilizan actualmente son los “People Meter”,

los cuales describiremos posteriormente. En otros países la Audiencia se calcula mediante el clásico levantamiento de encuestas, en las que se pregunta a la gente cuales son los programas vistos recientemente, capturando así los modos y formas de consumo televisivo, dichas encuestas son levantadas en diferentes lapsos de tiempo, ya sea de forma diaria, semanal, quincenal, etcétera. En ambos casos se presentan ventajas y desventajas propias del proceso de cálculo, generalmente estos problemas son de difícil corrección.

Los "People Meter" son aparatos conectados a los televisores, los cuales capturan los canales vistos en cada minuto del día por el televidente, esta información es almacenada en varias bases de datos, cada una de ellas destinada a un uso específico, de esta forma se crean distintas bases de acuerdo a sus necesidades de explotación, por cuestiones de volumen o espacio, en el proceso de construcción de estas bases se limitan los datos en ellas, desechando información que podría ser útil para investigaciones detalladas. Cada una de estas bases por construcción, es accesada por un sistema específico, debido a que los participantes requieren usar la información de los mismos se ven obli-

gados a adquirirlos todos. Adicionalmente a los sistemas con sus bases particulares, se ofrece a los participantes del medio la información del Raw Data, sin embargo, debido a su tamaño y al desconocimiento de la posible explotación de ésta es despreciada ya que es considerada complementaria, sin darse cuenta de que en realidad es la base general en la que recaen todos sus sistemas y de hecho, con esta base de datos se pueden realizar más y mejores análisis de los que se pueden obtener de los sistemas mencionados anteriormente. [2]

El Raw Data tiene una estructura peculiar, la cual se describirá posteriormente, por ahora basta decir que a causa de su estructura se dificulta el aprovechamiento de los datos que contiene, esto ocasiona un problema ya que se están almacenando datos que presentan dificultad de utilización, por tal motivo es que se propone la construcción de una nueva base más compacta que contenga implícitamente todos los datos del Raw Data, que proporcione mejoras tanto en el almacenamiento como en la extracción y el manejo de datos, es decir, una base que sea consistente con su explotación y de esta forma tener el mejor manejo posible.

Este punto es bastante importante si consideramos que cualquier empresa del medio publicitario requiere del cálculo de la Audiencia y demás variables del medio, a consecuencia sufre este problema ya que, gracias a la necesidad imperante del uso de esta información están obligadas a almacenar la misma en bases que muchas veces no son tan compactas como se quisiera. Es por esto que para ellas es indispensable contar con una forma de acumulación de los datos de forma inteligente, en el sentido de poder obtener ventajas utilizables sobre la nueva base.

Observe que un anunciante elige el medio masivo más adecuado para cubrir sus objetivos publicitarios, por otra parte el medio televisivo utiliza sus niveles de Audiencia para atraer a los anunciantes, utilizando la Audiencia de forma directa o indirecta para establecer sus precios; es claro que los participantes que fijan sus precios de venta de forma directa con la Audiencia son los más interesados en lograr una mejor puntuación. Tal es el caso de la televisora TV-Azteca, la cual basa sus ventas publicitarias sobre un concepto conocido como CPR [4], el cual denotaremos como CPA; por su parte la otra televisora, aunque sí toma en cuenta la Audiencia, no lo hace de forma tan rigurosa como

su contrincante, esta televisora asigna una tarifa fija a cada uno de sus programas, de acuerdo a la Audiencia registrada estadísticamente en periodos pasados, obligando al anunciante a comprar esos espacios sin la garantía de que su mensaje será visto por un determinado número de personas, de hecho, puede suceder que su mensaje no sea visto por la cantidad de gente que él hubiese planeado; pero gracias a la amplia cobertura que Televisa mantiene en la población, se puede dar el lujo de establecer estas políticas. [5]

El costo por punto de Audiencia o CPA, como su nombre lo indica es el precio que se pacta pagar por unidad de tiempo por cada punto porcentual de Audiencia que se obtenga, de esta manera, el precio de un anuncio publicitario depende de la cantidad de CPA que alcance, es decir, el precio es una función de la Audiencia.

Las empresas publicitan en algunos programas de acuerdo a su nivel de Audiencia y al tipo de gente al que se pretende llegar, esto es, las empresas buscan el programa más adecuado para su producto, de esta forma, una marca de cosméticos preferirá un programa dirigido a mujeres, por ejemplo telenovelas, mientras que una cervecería elegirá un

programa enfocado a hombres, como los eventos deportivos. En consecuencia, la Audiencia de los programas es de gran importancia para las cadenas ya que de ésta depende el interés de los anunciantes y el grado de sus ingresos. Además, gracias a la cobertura tan amplia con la que actualmente cuenta la televisión, los anunciantes deciden utilizarla para publicitarse ya que en ella obtienen un costo por millar muy bajo.

[7]

Por todas las razones expuestas anteriormente este trabajo plantea la creación de una nueva base que haga posible una mejor explotación, un almacenamiento menor que sirva de soporte para realizar más y mejores estudios.

Capítulo 2

Definiciones

Considerando que el lector no está familiarizado con los elementos de la operación televisiva, en este capítulo se explicaran las definiciones necesarias para continuar con el desarrollo posterior.

En la primera sección describiremos la base de datos del consumo televisivo, por su parte en la segunda sección enunciaremos las definiciones que son utilizadas frecuentemente en este medio.

2.1 El Raw Data

El Raw Data, como ya se mencionó anteriormente, es una base que almacena todos los datos de consumo televisivo. Para comprender su composición es necesario entender el funcionamiento de la recolección

de los datos que lo forman, así como los sistemas y procedimientos que se utilizan para determinarlo y posteriormente, la forma como se utilizan dichos datos.

Designamos por **universo** al conjunto Ω formado por todos los Individuos u Hogares que cuentan con televisión. IBOPE-AGB México, es la empresa dedicada a la recolección y cálculo de los datos necesarios para la obtención de la Audiencia, considera un universo de individuos y hogares con un total de 47,894,995 personas y de 11,140,838 hogares que representan al 46.5 % y 47.5 %, respectivamente del total de la población, a cada uno de estos casos se les asigna un cierto *peso* dependiendo de la ubicación geográfica así como socio-económica de la misma, todo ello con objeto de representar al universo de la mejor manera posible. Debido a los hábitos de consumo televisivo, la muestra se toma en los hogares, teniendo como base principal las tres principales ciudades del país: D.F., Guadalajara y Monterrey, así como una muestra de otras ciudades que representan a todo el universo de estudio.[1]

Al principio la muestra estaba formada únicamente por hogares, pero después se extendió a las personas de esos hogares, formando

ahora “dos muestras”, que en realidad es la misma pero medida sobre individuos y sobre hogares. Al formar estas muestras se pretendía obtener los datos más relevantes del consumo televisivo, por ejemplo la cantidad de minutos vistos en un canal, así como la forma en que este consumo se distribuía a lo largo del día; o por ejemplo las características socio-demográficas del individuo u hogar. Posteriormente se fueron agregando características o se ampliaron algunas que ya se tenían, por ejemplo se extendieron las clasificaciones al demográfico para tener un refinamiento más específico de la población. Actualmente podemos saber hasta la posición que ocupa el televidente en la familia, esto es, podemos saber si el televidente es el jefe de la familia, la ama de casa, etcétera; así como también podemos conocer otros datos como su edad, nivel socio-económico, clasificación psicológica, etcétera. Es por esto que los datos obtenidos adoptaron una estructura en la cual se describía de manera sucesiva todos los datos anteriores, sin importar que esta estructura no fuera la mejor forma de acomodar los datos, complicando así los cálculos necesarios para la obtención de la Audiencia; a continuación describiremos exactamente como está conformada esta estructura.

La empresa IBOPE-AGB utiliza *medidores* de consumo televisivo, mejor conocidos como “People Meter”, con ellos logra estimar el consumo de televisión de la muestra y por consiguiente logra estimar el consumo de televisión del universo. El “People Meter” está conectado a la televisión y registra todos los movimientos que el televidente hace mientras ve la televisión, así como también registra los periodos en los cuales no hubo consumo alguno. Finalmente, para identificar al televidente se utiliza un método que consiste en lo siguiente: el “People Meter” tiene su propio control remoto, éste debe ser utilizado por cada miembro de la familia de forma similar, la diferencia radica en la forma de encenderlo, cada individuo cuenta con un botón especial para realizar esta función, con esto se conoce a la persona que está viendo la televisión, para poder lograr efectividad en su uso se da una capacitación previa a la familia en donde se va a instalar el “People Meter”.

[2]

Cada “People Meter” mide diariamente el consumo en minutos de cada televisor en la muestra; que al final de cada día envía los datos a la central de IBOPE-AGB (los días se miden de las 2:00 a.m. a la

1:59 a.m. del día siguiente), una vez ahí se concentran en un sistema llamado POLLUX donde se procesan y *limpian* mediante criterios de exclusión determinados de antemano, para finalmente acomodarse en la estructura que describimos a continuación.

El sistema POLLUX genera diariamente dos archivos, uno para hogares y otro para individuos, ambos en código ASCII. Cada archivo contiene varios registros, éstos aparecen tantas veces como individuos u hogares sean contemplados. Los registros inician con las letras **Z**, **W** y **D** que indican el **identificador**, **peso** y **demográfico** del hogar o individuo respectivamente. Luego de estos registros se encuentra la descripción del consumo de los canales por posición, ésta inicia con la letra **X** y termina con la letra **F**.

Cuando existe actividad en algún canal se iniciará con la forma XD0 y posteriormente se registrarán las letras **A** y **B** seguidas por la cantidad de minutos no vistos y minutos vistos respectivamente, de este modo, las letras **A** y **B** aparecen tantas veces según sea la cantidad de cambios que se ejecuten en el día, obsérvese que esto se hará hasta que la suma de minutos vistos y minutos no vistos sea igual a 1440. En caso que

no exista consumo de algún canal la forma de expresarlo será con las letras XF. A lo largo del presente trabajo supondremos que existe un número fijo y finito de canales.

Como puede observarse por la descripción anterior, la estructura del Raw Data contiene una gran cantidad de datos, muchos de ellos cuentan con una difícil utilización debido a que cada individuo u hogar contiene una descripción total de los canales existentes, no obstante que muchas de las personas u hogares no consumen todos los canales, ya que no todos cuentan con el sistema adecuado para poder disfrutarlos. Así por ejemplo, una persona común únicamente consume canales ofrecidos en televisión abierta, en consecuencia, el Raw Data, en la parte correspondiente a estas personas, podría reducirse sólo a estos canales, sin perder información, además con esta reducción podemos facilitar el almacenamiento del Raw Data.

El enfoque utilizado en este proyecto se basa en la idea de transformar los datos del Raw Data en un formato más compacto sin que este pierda información, es decir, en una base más simple, reducida, con mejor funcionalidad y mejor proceso de almacenamiento. La base

que construiremos comenzará por la reducción de todos los datos no representativos. Básicamente podemos decir que reduciremos los datos en donde el Individuo u Hogar no haya visto la televisión, contemplando sólo aquellos lapsos del día en donde se tuvo encendido el televisor. Posteriormente, arreglaremos los datos de manera compacta. A continuación mostramos un fragmento del Raw Data para ilustrarlo

20041003

Z9900004

W1883.32

D41101

XF

XF

XD0A694B50A696F

XF

XF

.

.

.

En la figura anterior podemos observar que el día 3 de octubre de 2004, la familia (Z) número 9900004, representa en la muestra a un total (W) de 1,883.32 familias del universo. Las características demográficas de esta familia se describen en el tercer renglón, la letra D indica el inicio de la descripción, los dígitos posteriores corresponden a las siguientes características:

- Nivel socio-económico: Se divide en dígitos del 1 al 4 los cuales representan los niveles Bajo, Medio Bajo, Medio y Alto, respectivamente.
- Ubicación geográfica: Se divide en números del 1 al 4, los cuales representan el D.F., Guadalajara, Monterrey y el interior del resto de las ciudades que se utilizan en la muestra, respectivamente.
- TV Paga: El 0 y el 1 representan si el Hogar cuenta o no cuenta con un sistema de paga para la televisión, respectivamente.
- Multivisión: Indica si la televisión cuenta con sistema de multivisión, nuevamente 0 y 1 son los estatus posibles y representan el tener o no tenerlo, respectivamente.

- Cablevisión: Funciona del mismo modo que el punto anterior.
- Niños de 0 a 12 años: Indica si el Hogar cuenta con niños de 0 a 12 años, el 1 y 2 representan tener o no tener niños de esa edad, respectivamente.
- Número de televisores: Los dígitos 1 y 2 indican si el Hogar cuenta con una o con más televisiones, respectivamente.
- Teléfono: El 1 y 2 representan el hecho de tener o no tener teléfono, respectivamente.
- Zona geográfica: Se divide en números del 1 al 6, los cuales representan a las zonas: Centro, Oeste, Noreste, Sur, Golfo y Noroeste de la República Mexicana, respectivamente.

La descripción del código D41101 de la figura anterior, indica que la familia es del D.F., cuenta con sistema de paga para la televisión, no tienen Multivisión, pero si tienen Cablevisión. [3] Del resto del código se puede determinar que:

- No vieron el primer canal
- No vieron el segundo canal
- El tercer canal lo vieron de la siguiente manera: no vieron los primeros 694 minutos del día, el día comienza a las 2:00 a.m., después lo vieron durante 50 minutos y finalmente lo dejaron de ver durante 696 minutos. Note que la suma de minutos es igual a los 1440 minutos del día.
- No vieron el cuarto canal
- No vieron el quinto canal
- ⋮

2.2 Definiciones Generales

Un **medio** es un canal de comunicación por el que se puede emitir un mensaje publicitario, por ejemplo, la prensa, la radio y la televisión. Consideraremos a la **publicidad** como el término que se utiliza para referirse a cualquier mensaje destinado a un cierto público, cuyo objetivo principal es fomentar la venta de bienes y servicios. Un **soporte publicitario** es un espacio físico y temporal del medio, utilizado para enviar un mensaje publicitario. [3]

El conjunto \mathcal{C} de canales estará compuesto por números naturales, esto es $\mathcal{C} = \{0, 1, \dots, m\}$, donde 0 representa el *canal* en donde la televisión está apagada y m la cantidad de canales disponibles. Nótese que cada canal por sí mismo es un soporte publicitario, a excepción del cero, el cual no satisface tal condición.

El **universo** (Ω) de personas que pueden entrar en contacto con la televisión estará representado por una muestra de televidentes, en nuestro modelo consideraremos a \mathcal{T} como el conjunto de esos televidentes. Denotaremos con t_z al z -ésimo televidente, así $\mathcal{T} = \{t_1, t_2, \dots, t_z\}$.

Además, cada televidente en \mathcal{T} podrá elegir ver cualquier canal de \mathcal{C} siempre y cuando tenga acceso a él.

Sea \mathcal{M} el conjunto de los 1440 minutos de un día, que los canales, denotaremos a este conjunto con números naturales, de esta forma $\mathcal{M} = \{1, 2, \dots, 1440\}$.

La **exposición** es una función $\phi : \mathcal{M} \rightarrow \mathcal{C}$ dada por $\phi(\mu) = c$ con $\mu \in \mathcal{M}$ y $c \in \mathcal{C}$, esta función indica el canal visto por el televidente en el μ -ésimo minuto.

La cantidad de gente que entra en contacto con la publicidad de un mensaje se considera como la **Audiencia**, la cual indica, en promedio por minuto, la cantidad de gente que fue alcanzada en un periodo de tiempo determinado.

Los **Encendidos** son un conjunto \mathbf{E} , formado por todas de personas que están viendo la televisión, en nuestro estudio nos enfocaremos en la cardinalidad de este conjunto para obtener otras medidas de interés. Definimos a los **GRP's** (Gross Rating Points), como la suma por minuto promedio de puntos de Audiencia acumulados en un periodo de

tiempo mayor o igual a una semana, por ejemplo: en periodos anuales, semestrales, mensuales o durante la vigencia de una campaña. [3]

El **Alcance** de un periodo o programa, será el porcentaje de personas expuestas a la televisión al menos un minuto del periodo o programa. La **Frecuencia** será el promedio de veces que la Audiencia está expuesta en un periodo determinado, es decir, es el cociente de los GRP's entre el Alcance. [3]

Designamos como **grupo objetivo** a los subconjuntos de \mathcal{T} , con una característica especial de estudio, por ejemplo, amas de casa, padres de familia, nivel socioeconómico, etcétera. Estos subconjuntos son de gran utilidad para analizar el éxito de una campaña publicitaria dirigida a un grupo objetivo determinado. [7]

Sea $t \in \mathcal{T}$ un televidente, diremos que t estuvo viendo la televisión si mantuvo encendido su televisor por más de un minuto y cumple ciertos criterios de selección, por ejemplo: la televisión no ha estado encendida en el mismo canal durante más de 12 horas consecutivas, estuvo viendo un canal por más de 30 segundos, etcétera. De este modo.

el **tiempo visto** será el conjunto \mathcal{V} formado por todos los minutos del día en los cuales el televidente estuvo viendo la televisión, es decir, $\mathcal{V} = \{\mu \in \mathcal{M} \mid \phi(\mu) \neq 0\}$.

La definición anterior sugiere, de forma natural construir conjuntos cuyos elementos se encuentren ordenados de forma sucesiva. Definimos a los **segmentos vistos**, como los conjuntos de minutos $s_{a_i}^{b_i}$, donde a_i representa el minuto del día en donde el televidente encendió la televisión por i -ésima ocasión y la estuvo viendo de manera sucesiva desde el minuto a_i hasta el minuto b_i . El minuto b_i representa el último minuto donde el televidente estuvo viendo la televisión, es decir, el minuto b_i satisface que $\phi(b_i + 1) = 0$; en caso de que el televidente haya visto la televisión durante todo el día consideraremos a b_i como el minuto 1440. Nótese que esta definición incluye todos los posibles segmentos del día, incluso aquéllos que constan de sólo un elemento, a los cuales llamaremos *segmentos triviales*, además, por definición de \mathcal{M} , los segmentos están ordenados de manera creciente, ya que heredan ese orden de \mathcal{M} .

Sea n el número de segmentos de un día, nótese que $n \in [0, 720]$,

ya que a lo largo de un día, a lo más, el televidente puede encender y apagar la televisión en cada minuto del día, de este modo el número de segmentos está acotado superiormente por 720; del mismo modo, puede suceder que no exista ningún segmento a lo largo del día, esto ocurre si $V=\emptyset$, por lo tanto n está acotado inferiormente por cero.

Sea \mathcal{S} el conjunto de segmentos del día, denotaremos a sus elementos de la siguiente manera $\mathcal{S} = \{s_{a_1}^{b_1}, s_{a_2}^{b_2}, \dots, s_{a_n}^{b_n}\}$, por simplificación los denotaremos como $\{s_1, s_2, \dots, s_n\}$.

Introduciremos una definición matemática general, para después poder aplicarla a nuestro problema.

Sea \mathcal{A} un conjunto no vacío, definimos una **partición** de \mathcal{A} como la familia $(A_i)_{i=1}^k$ que satisface:

1. $\cup_{i=1}^k A_i = \mathcal{A}$;
2. $A_i \cap A_j = \emptyset, \quad \forall i \neq j$.

De este modo, en nuestro caso tenemos varias particiones, la primera de ellas es la partición del conjunto \mathcal{M} , la cual está formada por la

familia $\mathcal{V}, \mathcal{V}^c$; a su vez $\mathcal{S}, \mathcal{S}^c$ es una partición de \mathcal{V} . Donde \mathcal{S}^c es el conjunto de minutos en los cuales el televidente no estuvo viendo la televisión, es decir, cuando estuvo viendo el canal cero; así $\mathcal{S}^c = \{\mu \in \mathcal{M} \mid \phi(\mu) = 0\} = \mathcal{M} \setminus \mathcal{S}$.

Sea c un canal visto en un segmento no trivial $s_i \subset \mathcal{S}$. Definimos a los **periodos de c** en s_i como los conjuntos $c_i = \{\mu \in s_i \mid \phi(\mu) = c\}$. Es importante resaltar lo siguiente:

$$c_i \cup (c_i)^c = s_i \quad \forall c \in \{1, 2, \dots, m\}; \quad i \in \{1, 2, \dots, n\} \quad (2.1)$$

y

$$\bigcup_{c \in \mathcal{C}} c_i = s_i \quad \forall i \in \{1, 2, \dots, n\} \quad (2.2)$$

Así, los periodos de los canales en un segmento dado, forman una partición de éste, es decir, el i -ésimo segmento tiene una partición dada por $(c_i)_{c=1}^m$, de hecho puede ser que podamos reducir esta familia ya que no necesariamente todos los canales se ven en un periodo dado, o incluso pueden existir canales que el televidente no vea durante todo el día, de este modo existen $c \in \mathcal{C}$ tales que $c_i = \emptyset$ para algunas $c \in \{1, \dots, m\}$, no obstante la familia completa garantiza una partición de s_i .

Sean μ_1 y μ_2 dos minutos de c_i , diremos que μ_1 y μ_2 son conexos si y sólo si $|\mu_1 - \mu_2| \leq 1$, de esta manera, denotaremos a $c_{i,j}$ como el j -ésimo conjunto conexo de minutos del segmento i -ésimo, donde el televidente estuvo viendo el canal c . Observe que en los periodos de un canal no necesariamente existen elementos conexos, de hecho pueden existir casos en donde cada elemento no sea conexo con cualquier otro del conjunto.

Estas definiciones se ilustrarán posteriormente con objeto de dejar clara la idea de las mismas.

Gracias a las definiciones anteriores podremos modificar de manera fácil el modelo con fin de adaptarlo a cualquier cambio que pudiese presentarse, ya sea en el tamaño de la muestra, la cantidad de canales, etcétera. Esto resulta útil debido a que estos cambios son frecuentes.

Capítulo 3

Un Nuevo Modelo de Explotación

Para transformar el Raw Data en una base de mejor explotación construiremos, para cada televidente, la matriz asociada a cada uno de ellos, el objetivo de la matriz de consumo es almacenar, de forma más reducida, la misma información que presenta el Raw Data para cada televidente. La existencia de la matriz de consumo estará en función del consumo que haya tenido el televidente, es decir, existirá si el televidente estuvo viendo la televisión durante el día. Su construcción es importante debido a que cada renglón de ella representará la cantidad de minutos que el televidente estuvo viendo el canal que corresponda a dicho renglón, por su parte, en las columnas de la matriz de consumo podremos conocer el total de minutos vistos por el televidente cada

ocasión que encendió la televisión así como el número de ocasiones que lo hizo el cual es equivalente al número de columnas existentes.

3.1 Hipótesis y Metodología del Modelo

Para nuestro modelo supondremos que cada televidente puede ver, ya sean cero minutos o bien un total de 1440 minutos, o cualquier número natural contenido entre estos dos. El día empezará a las 2:00 *hrs.* y terminará a la 1:59 *hrs.* del día siguiente.

Con objeto de poder reconocer a cada televidente conservaremos su identificador, su peso en la muestra, su demográfico y su matriz de consumo a la cual denotaremos como mc , en caso de que no exista se conservarán sólo los primeros parámetros. Esto servirá para la reconstrucción total del Raw Data.

Sea $t_z \in \mathcal{T}$ un televidente fijo, supondremos que t_z vió m canales en el día, durante n segmentos, con $m \in \mathcal{C}$ y $0 \leq n \leq 720$. Además, en cada minuto t_z puede ver un sólo canal, esto es, si en el minuto x está viendo el canal q y el canal r , entonces $q = r$.

Inicialmente contemplaremos el caso en el que sólo existe un televidente, él cual puede ver sólo un canal, comentaremos las simplificaciones que este caso presenta y aprovecharemos su estructura para iniciar nuestro desarrollo posterior, nos daremos una idea del almacenamiento así como la manera de extracción de los datos en la nueva base.

Posteriormente desarrollamos el caso de un televidente con dos canales, esta vez el problema necesitará de gran parte de nuestras definiciones, con las cuales podremos desarrollar de mejor manera su estructura y trataremos de dejar clara la idea de los segmentos así como su partición.

Una vez comprendido el caso anterior, nos daremos a la tarea de explicar su generalización, es aquí donde desarrollaremos gran parte del trabajo ya que, una vez completado este caso procederemos a explicar el problema del Raw Data en su totalidad, es decir, donde existen z televidentes y m canales.

La idea central a lo largo de todo el trabajo es la partición del día en

segmentos que representen los intervalos de tiempo que el televidente estuvo viendo la televisión. Una vez identificado este tiempo debemos analizar el comportamiento del televidente, con objeto de conocer el consumo de cada canal.

En todos los casos anteriores, definiremos una función que comprima cada segmento en un número, dando las explicaciones necesarias para poder hacer esto, así mismo seguiremos la misma idea para proceder a la extracción de estos datos. Desafortunadamente nuestra compresión no identificará el minuto inicial de cada segmento, motivo por el cual almacenaremos esta información para cada televidente por separado, en el desarrollo posterior observaremos que este almacenamiento es de gran utilidad ya que sirve de base para reconstruir el consumo televisivo.

3.2 El caso inicial

Sean t un televidente y 1 el único canal que puede ver t , como se puede apreciar desde el principio este caso se reduce a ver o no ver la televisión, es decir, el problema se reduce a un problema binario, donde cero y uno representan no ver y ver la televisión, respectivamente.

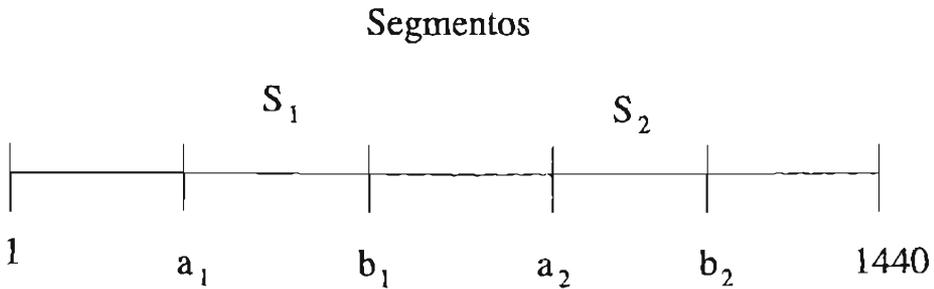
Note que aquí, como sólo puede ver un canal, los segmentos vistos coinciden con los periodos del canal uno. por este motivo la extracción es simple ya que el almacenamiento guardará el mismo número tanto para el segmento como para los periodos del canal uno en cualquier segmento dado, además la matriz de consumo de t es en realidad un vector en \mathfrak{R}^n , donde n es la cantidad de segmentos que t vio en el día.

Finalmente, nótese que este caso sugiere de forma natural una generalización a través de los canales, ya que intuitivamente es más fácil agregar renglones para construir una matriz y después hacer lo propio con cada televidente.

Este caso sirve para dar una idea del comportamiento inicial del problema que estamos tratando, en la siguiente sección pretendemos

interpretar totalmente las definiciones establecidas anteriormente.

A continuación ilustramos los segmentos, mismos que coinciden con los periodos del canal, por ello no marcamos diferencia alguna, las diferencias se verán en los casos posteriores.



3.3 Un televidente con dos canales

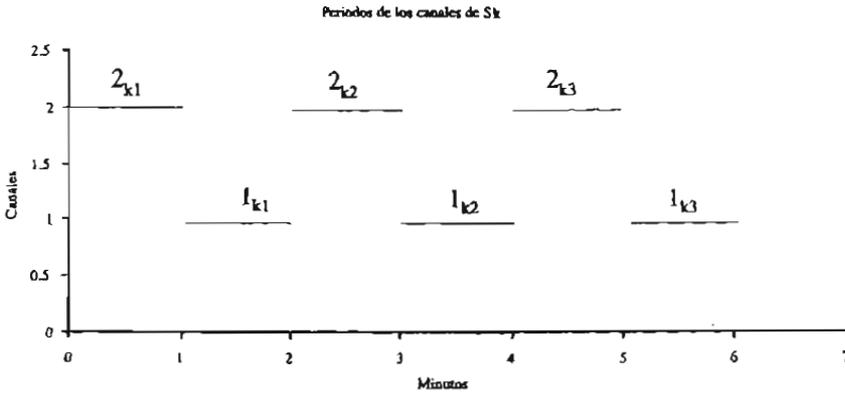
Sean t un televidente de \mathcal{T} y 1 y 2 dos canales de \mathcal{C} .

A diferencia del caso anterior, cuando t encienda el televisor, podrá elegir ver entre los canales 1 y 2, claramente esto hace que los periodos de los canales 1 y 2 sean de cardinalidad menor o igual a la de los segmentos correspondientes.

Nuestra labor en este caso es dejar claro el proceso de separación de los segmentos en los periodos de los canales, para ello usaremos la función *exposición* (ϕ), ya que con la imagen de esta función sabemos el canal visto por t en cada minuto μ del día.

Sea $\mathcal{S} = \{s_1, s_2, \dots, s_n\}$ el conjunto de los segmentos vistos por t durante el día, el subíndice $i \in \{1, 2, \dots, n\}$ representa los minutos a_i y b_i donde t encendió y apagó respectivamente el televisor por i -ésima vez en el día.

A continuación mostramos los periodos de los canales en este caso:



Para transformar estos n segmentos en números, definimos la siguiente función:

Sea $s_i \subset \mathcal{S}$, definimos: $\varphi : P(\mathcal{S}) \rightarrow \mathcal{N}$ dada por

$$\varphi(s_i) = \varphi(s_{a_i}^{b_i}) = \sum_{\mu \in s_{a_i}^{b_i}} 2^{b_i - \mu}$$

Donde $P(\mathcal{S})$ representa el conjunto potencia de \mathcal{S}

A continuación enunciamos una Proposición con las propiedades de la función φ .

Proposición 1 Sean s_i el i -ésimo segmento del día, n el total de segmentos del día y c_i los periodos del canal c en el i -ésimo segmento, se tienen las siguientes propiedades:

1. $\varphi(s_i) = \varphi(s_{a_i}^{b_i}) = 2^{b_i - a_i + 1} - 1$
2. $\varphi(c_i) \leq \varphi(s_i) \quad \forall i \in \{1, 2, \dots, n\}$ y $\forall c \in C$
3. $\varphi(s_i) = \sum_{c=1}^m \varphi(c_i) \quad \forall i \in \{1, 2, \dots, n\}$

La igualdad en la segunda propiedad se cumple si y sólo si t vió el canal c durante todos los minutos del segmento s_i , es decir, en caso de que $c_i = s_i$.

En la última propiedad, la igualdad se cumple para cualquier segmento ya que estamos sumando sobre todos los canales.

Demostración:

1. Por definición $s_i = s_{a_i}^{b_i} = \{a_i, a_i + 1, \dots, b_i\}$, de este modo

$$\varphi(s_{a_i}^{b_i}) = \sum_{\mu | \mu \in s_{a_i}^{b_i}} 2^{b_i - \mu} = \sum_{\mu=a_i}^{b_i} 2^{b_i - \mu} = \sum_{l=0}^{b_i - a_i} 2^l = 2^{b_i - a_i + 1} - 1$$

2. Nuevamente, por definición $c_i \subseteq s_i$, de este modo $|c_i| \leq |s_i|$, entonces $\sum_{\mu|\mu \in c_i} 2^{b-\mu} \leq \sum_{\mu|\mu \in s_i} 2^{b-\mu}$, por tanto $\varphi(c_i) \leq \varphi(s_i) \quad \forall i \in \{1, 2, \dots, n\}$.

3. Sean c, d dos canales distintos, note que $c_i \cap d_i = \emptyset$, ya que si existiera $\mu \in \mathcal{M}$ tal que $\mu \in c_i \cap d_i$ entonces t vió, en el minuto μ , el canal c y el canal d . lo cual es imposible ya que por hipótesis $c \neq d$. Por lo tanto $c_i \cap d_i = \emptyset$, ahora basta notar que en s_i , t vio alguno de los m canales, por lo tanto para cualquier $\mu \in s_i$ existe un único $c \in \mathcal{C}$ tal que $\phi(\mu) = c$, por lo tanto $\varphi(s_i) = \sum_{c=1}^m \varphi(c_i)$

■

Observe que los periodos de los canales no se intersectan entre sí.

Además, por la Ecuación 2.2 tenemos que

$$\varphi(s_i) = \varphi(\cup_{c=1}^m c_i) = \sum_{c=1}^m \varphi(c_i)$$

Aplicando la función φ a cada segmento de \mathcal{S} obtendremos n números que describen perfectamente el consumo del televidente. Estos números son distintos de cero ya que en caso de tener un sólo elemento φ toma el valor uno. Además, cada número almacenado es diferente

para cada canal ya que la función φ es inyectiva, lo cual nos es de utilidad para proceder a la extracción de los datos una vez que éstos sean almacenados, asegurando la unicidad del canal visto.

Es importante mencionar que aunque los segmentos estén asociados por la función φ a números distintos de cero, pueden existir $c \in \mathcal{C}$ para los cuales $\varphi(c) = 0$, es decir, $c_i = \emptyset$, $\forall i \in \{1, 2, \dots, n\}$. Si este es el caso, omitiremos el canal c ya que no contiene información relevante, salvo la ausencia de consumo de ellos, lo cual puede caracterizarse mediante la eliminación de su registro en la matriz de consumo.

Con estos datos podemos formar la matriz de consumo de t . El siguiente esquema ilustra la forma de cada una de estas matrices, donde A_{ij} representa el j -ésimo conjunto de minutos conexos, contenidos en el segmento i -ésimo, en que el televidente estuvo viendo el canal A

$$\begin{pmatrix} 1_{1,1} & 1_{1,2} & \dots & 1_{1,n} & \dots & 1_{k,1} & 1_{k,2} & \dots & 1_{k,n} \\ 2_{1,1} & 2_{1,2} & \dots & 2_{1,n} & \dots & 2_{k,1} & 2_{k,2} & \dots & 2_{k,n} \end{pmatrix}$$

Una vez que tenemos ordenados los datos en la matriz de consumo debemos explicar y justificar su extracción, sin embargo, esto lo tratare-

mos en la siguiente sección, con el propósito de dar una idea más general, ya que en ese caso tendremos m posibles canales.

La propiedad anterior establece que:

$$\varphi(s_i) = 2^{b_i - a_i + 1} - 1$$

Despejando a b_i tenemos:

$$b_i = \frac{\ln(\varphi(s_i) + 1)}{\ln 2} + a_i - 1$$

Lo que reduce considerablemente el almacenamiento de esta parte del Raw Data, ahora continuaremos con la descripción del proceso de extracción, pero antes enunciaremos la siguiente proposición.

Proposición 2 Sea $w \in \mathbb{N}$, tal que $w \geq 2$, entonces:

$$2^w > \sum_{i=1}^{w-1} 2^i$$

Demostración:

Obsérvese que $2^w = 2 * 2^{w-1}$, entonces

$$2^w = 2^{w-1} + 2^{w-1}$$

$$2^w = 2^{w-1} + 2 * 2^{w-2}$$

$$2^w = 2^{w-1} + 2^{w-2} + 2^{w-2}$$

⋮

$$2^w = 2^{w-1} + 2^{w-2} + \dots + 2^2 + 2^2$$

$$2^w = 2^{w-1} + 2^{w-2} + \dots + 2^2 + 2 + 2$$

$$2^w = \sum_{i=1}^{w-1} 2^i + 2$$

dado que $2 > 0$, tenemos que

$$2^w > \sum_{i=1}^{w-1} 2^i$$

■

Esta proposición será de gran utilidad ya que gracias a ella podremos establecer criterios para la extracción de datos. El primero de ellos se basa en el siguiente lema:

Denotaremos a $\{ \cdot \}$ como la función menor entero.

Lema 1 Sea μ un minuto de un segmento fijo s , y sea c un canal fijo contenido en la matriz de consumo, entonces:

$$\left[\frac{\varphi(c_i)}{2^{b_i - \mu}} \right]$$

es impar si y sólo si t vio el canal c en el minuto μ , en otro caso es par.

Demostración:

Note que existen dos posibles casos:

1. $\mu \in c_i$, ó
2. $\mu \notin c_i$.

Sin pérdida de generalidad, en el primer caso, supongamos que $c_i = \{\mu_1, \mu_2, \dots, \mu_x\}$. Entonces $\mu = \mu_\xi$ para algún $\xi \in \{1, 2, \dots, x\}$.

Además, por definición de φ , tenemos que

$$\begin{aligned} \frac{\varphi(c_i)}{2^{b_i - \mu_\xi}} &= \frac{\sum_{\mu \in c_i} 2^{b_i - \mu}}{2^{b_i - \mu_\xi}} = \frac{2^{b_i - \mu_1} + 2^{b_i - \mu_2} + \dots + 2^{b_i - \mu_\xi} + \dots + 2^{b_i - \mu_x}}{2^{b_i - \mu_\xi}} \\ &= \underbrace{\frac{2^{b_i - \mu_1} + 2^{b_i - \mu_2} + \dots + 2^{b_i - \mu_{\xi-1}}}{2^{b_i - \mu_\xi}}}_I + \underbrace{\frac{2^{b_i - \mu_\xi}}{2^{b_i - \mu_\xi}}}_{II} + \underbrace{\frac{2^{b_i - \mu_{\xi+1}} + \dots + 2^{b_i - \mu_x}}{2^{b_i - \mu_\xi}}}_{III} \end{aligned}$$

Nótese que para cualquier $\xi \in \{1, 2, \dots, x\}$, se cumple que $\left\lceil \frac{\varphi(c_i)}{2^{b_i - \mu}} \right\rceil$ es impar ya que:

- El sumando I siempre es par, pudiendo ser cero si ξ es el primer minuto del segmento
- El sumando III siempre es uno y

- El sumando III siempre es menor que uno por la proposición anterior.

Análogamente, en el segundo caso, II no existe, porque t no está viendo ese canal, es decir, $\mu \notin \{\mu_1, \mu_2, \dots, \mu_T\}$, sin embargo, el sumando I es par y el sumando III es menor que uno.

Por lo tanto el término $\left[\frac{\varphi(c_t)}{2^{b_t - \mu}} \right]$ es impar si y sólo si t vio el canal c en el minuto μ ; será par cuando t no vió el canal c en el minuto μ .



Con ayuda de lo anterior describiremos el proceso de extracción de los datos almacenados en la matriz de consumo.

Sea Λ la matriz de consumo de t . El proceso a seguir para la extracción de la Audiencia de un cierto canal fijo c , en un periodo arbitrario del día, $P = [u, v] \subset \mathcal{M}$, se describe a continuación:

Debemos verificar que c esté en Λ , ya que si no está presente, la Audiencia de c en cualquier periodo del día es cero, en especial en P .

Si el canal c se encuentra en Λ debemos empezar por reconstruir los

segmentos, es decir, obtener los extremos finales de cada segmento. Una vez hecho esto, debemos buscar los segmentos para los cuales $s_i \cap P \neq \emptyset$, obsérvese que si no existen segmentos que tengan minutos en común con P indicará que la Audiencia de cualquier canal en P es cero.

Nótese que si existen segmentos de \mathcal{S} que coincidan con P , entonces:

- P está contenido en segmentos sucesivos de \mathcal{S} , en caso de estar contenido en más de un segmento; y
- Pueden existir minutos $\mu \in P$ tales que $\mu \notin s_i$ para ninguna $i \in \{1, 2, \dots, n\}$.

Por estas observaciones supondremos, sin pérdida de generalidad, que s_i, s_{i+1}, \dots, s_k son los segmentos que coinciden con P , donde $1 \leq i \leq k \leq n$.

Si c esta presente en Λ , debemos buscar sobre el renglón correspondiente las entradas c_i, c_{i+1}, \dots, c_k y extraer aquellas que sean distintas de cero. Nuevamente, si todas las entradas son idénticamente cero implicará que la Audiencia de c en P es cero. Además, nótese que si $\varphi(c_\zeta) = \varphi(s_\zeta)$ para algún $\zeta \in \{i, i+1, \dots, k\}$ indicará que el canal c

se vio durante todos los minutos de s_c y la extracción de este canal se reduce a conocer la duración de este segmento.

Sin pérdida de generalidad, supongamos que P se intersecta con los periodos c_i, c_{i+1}, \dots, c_k y que las entradas de ese renglón son diferentes de cero, también supongamos que $|c_j| < |s_j| \quad \forall j \in \{i, i+1, \dots, k\}$. Es importante destacar que:

- $P \subseteq s_i$, o
- $P \supseteq s_i$

Obsérvese que si $P \supseteq s_i$ tendremos que considerar a todos los minutos de s_i , por el contrario, si $P \subseteq s_i$ debemos encontrar aquellos minutos que coincidan en ambos, ya que el primer supuesto es más simple nos enfocaremos primero al segundo caso.

Sea s_i un segmento y sea c un canal visto en este segmento. Por las Ecuaciones 2.1 y 2.2 tenemos que:

$$\varphi(c_i) + \varphi((c_i)^c) = \varphi(s_i) \quad (3.1)$$

ya que $c_i \cap (c_i)^c = \emptyset$.

Para obtener la Audiencia de c en P debemos saber cuántos minutos se estuvo viendo este canal en este periodo, para conocer esto debemos reconstruir el consumo de c en s_i .

La forma más fácil de hacer esto es conocer los minutos de s_i para los cuales se empezó y se dejó de ver c respectivamente, para ello definimos lo siguiente:

Sean s_i un segmento y c un canal visto en s_i , es decir, $s_i \cap c_i \neq \emptyset$, y sea $c_i = \mu_1, \mu_2, \dots, \mu_p$, definimos a los minutos de s_i en los que se empezó y dejó de ver, el canal c como los minutos $a_{i,j}$ y $b_{i,j}$, respectivamente, en donde

$$a_{i,j} = \mu_j \text{ con } j = \{1, 3, 5, \dots, p-1\} \text{ y}$$

$$b_{i,j} = \mu_j \text{ con } j = \{2, 4, 6, \dots, p\}$$

Nuestro objetivo es obtener a μ_j en un segmento, para lo cual usaremos la siguiente proposición.

Proposición 3 Sea $s_i \subset \mathcal{S}$ un segmento y sea c un canal visto en este segmento, tal que $\varphi(c_i) \neq \varphi(s_i)$.

Entonces, el primer minuto de s_i en el que se vió el canal c esta dado por:

$$b_i - \left\lceil \frac{\ln(\varphi(c_i))}{\ln 2} \right\rceil$$

Demostración:

Sea $c_i = \{\mu_1, \mu_2, \dots, \mu_p\}$ un conjunto de minutos, de este modo $\mu_j = a_{i,j}$ para alguna $j \in \{1, 3, 5, \dots, p-1\}$ y $\mu_j = b_{i,j}$ para alguna $j \in \{2, 4, 6, \dots, p\}$ y sea $s_i = \{a_i, a_{i+1}, \dots, b_i\}$; tomando los elementos de c_i tenemos que $s_i = \{a_i, a_{i+1}, \dots, \mu_1, \dots, \mu_p, \dots, b_i\}$, en donde $a_i \leq \mu_1 \leq \dots \leq \mu_p \leq \dots b_i$

Como

$$\varphi(c_i) = \sum_{\mu \in c_i} 2^{b_i - \mu}$$

Entonces $\varphi(c_i) \geq 2^{b_i - \mu_1}$, despejando a μ_1 tenemos que

$$\frac{\varphi(c_i)}{2^{b_i - \mu_1}} \geq 1$$

$$\ln \left(\frac{\varphi(c_i)}{2^{b_i - \mu_1}} \right) \geq \ln(1)$$

$$\ln(\varphi(c_i)) - \ln(2^{b_i - \mu_1}) \geq 0$$

$$\ln(\varphi(c_i)) \geq (b_i - \mu_1) * \ln(2)$$

$$\frac{\ln(\varphi(c_i))}{\ln 2} \geq b_i - \mu_1$$

ya que $\ln(2) > 0$

$$\mu_1 \geq b_i - \frac{\ln(\varphi(c_i))}{\ln 2}$$

Nótese que el término $\frac{\ln(\varphi(c_i))}{\ln 2}$ no necesariamente es un entero. sin embargo b_i y μ_1 si lo son, por tanto necesitamos que $\frac{\ln(\varphi(c_i))}{\ln 2}$ sea entero, para hacer esto usamos la función entero ya que necesitamos el mínimo entero que satisfaga:

$$\frac{\ln(\varphi(c_i))}{\ln 2} \geq b_i - \mu_1$$

Por lo tanto

$$\mu_1 = b_i - \left\lceil \frac{\ln(\varphi(c_i))}{\ln 2} \right\rceil$$



De esta manera podemos encontrar el primer minuto (μ_1) en el que se vio un canal en un segmento. Ahora debemos conocer cual es el minuto en el que se deja de ver este canal por primera vez, es decir μ_2 para ello usaremos la Ecuación 3.1 así como las propiedades de la función φ .

Sin pérdida de generalidad, supongamos que $\mu_1 > a_i$, es decir, el canal c se empezó a ver después $\mu_1 - a_i$ minutos del segmento s_i . Por lo tanto los minutos $\{a_i, a_i + 1, \dots, \mu_1 - 1\} \in (c_i)^c$. De este modo, restando los sumandos correspondientes a estos minutos encontraremos el minuto de s_i para el cual $(c)^c$ se empezó a ver por segunda vez, la primera vez fue en a_i . Para encontrar estos sumandos es suficiente ver que son de la forma:

$$2^{b_i - a_i} + 2^{b_i - (a_i + 1)} + \dots + 2^{b_i - (\mu_1 - 1)}$$

que podemos simplificar mediante la ecuación

$$2^{b_i - a_i} + 2^{b_i - (a_i + 1)} + \dots + 2^{b_i - (\mu_1 - 1)} = 2^{b_i - a_i + 1} - 2^{b_i - \mu_1 + 1}$$

De esta forma, podemos usar la Proposición 3 para encontrar el segundo minuto inicial de $(c_i)^c$, por lo tanto, el minuto en el que se

dejó el canal c es un minuto antes, esto es, si en el minuto $\hat{\mu}$ se empezó a ver el canal $(c)^c$, entonces en el minuto $\hat{\mu}-1$ se dejó de ver el canal c , por tanto $\mu_2 = \hat{\mu} - 1$.

El método de extracción de un minuto μ_j con $j = \{1, 2, 3, \dots, p\}$ esta basado en:

Dado que conocemos a μ_1 y μ_2 , para encontrar μ_3 debemos proceder de forma similar, sólo debemos restar a $\varphi(c_i)$ los sumandos correspondientes a los minutos $\{\mu_1, \mu_1+1, \dots, \mu_2\}$, de igual forma, éstos se pueden representar con la ecuación:

$$2^{b_i-\mu_1} + 2^{b_i-(\mu_1+1)} + \dots + 2^{b_i-(\mu_2-1)} = 2^{b_i-\mu_1+1} - 2^{b_i-\mu_2}$$

Usando la proposición anterior podemos encontrar μ_3 . En general, definimos a:

$$\mu_1 = b_i - \left[\frac{\ln[\varphi(c_i)]}{\ln 2} \right]$$

$$\alpha_1 = 2^{b_i-\mu_1+1} - 2^{b_i-(\mu_1-1)}$$

$$\mu_2 = b_i - \left[\frac{\ln[\varphi((c_i)^c) - \alpha_1]}{\ln 2} \right]$$

$$\alpha_2 = 2^{b_i - (\mu_1 - 1)} - 2^{b_i - \mu_2}$$

$$\mu_3 = b_i - \left[\frac{\ln[\varphi(c_i) - \alpha_2]}{\ln 2} \right]$$

$$\alpha_3 = 2^{b_i - \mu_2} - 2^{b_i - (\mu_3 - 1)}$$

$$\mu_4 = b_i - \left[\frac{\ln[\varphi((c_i)^c) - \alpha_1 - \alpha_3]}{\ln 2} \right]$$

$$\alpha_4 = 2^{b_i - (\mu_3 - 1)} - 2^{b_i - \mu_4}$$

$$\vdots$$

$$\mu_{2p-1} = b_i - \left[\frac{\ln[\varphi(c_i) - \alpha_2 - \dots - \alpha_{2p-2}]}{\ln 2} \right]$$

$$\alpha_{2p-1} = 2^{b_i - \mu_{2p-2}} - 2^{b_i - (\mu_{2p-1} - 1)}$$

$$\mu_{2p} = b_k - \left[\frac{\ln[\varphi((c_i)^c) - \alpha_1 - \dots - \alpha_{2p-1}]}{\ln 2} \right]$$

$$\alpha_{2p} = 2^{b_k - (\mu_{2p-1} - 1)} - 2^{b_k - \mu_{2p}}$$

Seguimos este procedimiento hasta que llegemos a alguno de los dos casos citados anteriormente, en ese momento habremos reconstruido el consumo del canal ν en el segmento s_k . Cabe mencionar que aseguramos llegar a cualquiera de los dos casos anteriores ya que los números que utilizamos están acotados y por ende el proceso es finito.

Para $P = [u, v]$ y $P \subseteq s_i$ ó $P \supseteq s_i$ tomaremos, en ambos casos a μ_0 como $\max\{a_i, u\}$ y como minuto final al minuto μ_f , dado como $\min\{b_i, v\}$ debido a que el consumo se reduce al segmento $[\mu_0, \mu_f]$.

Ya que tenemos reconstruido el consumo en el segmento, debemos conocer el consumo en $[\mu_0, \mu_f]$, para hacer esto basta con fijar el minuto μ_0 y conocer si t vió o no vió el canal c en este minuto (véase p. 55).

Sin pérdida de generalidad, supongamos que t vió el canal c en el minuto μ_0 , hay que encontrar el menor índice δ para el cual $\mu_0 < \mu_\delta$, con $\mu_\delta \in \{\mu_1, \dots, \mu_p\} = c_i$.

Nótese que existen dos posibilidades:

- a) $\mu_\delta \geq \mu_f$, es decir, t vió el canal c durante todo el periodo P .
- b) $\mu_\delta < \mu_f$, entonces, t no vió el canal c durante todo el periodo P .

Para (a) basta conocer el minuto μ_f para concluir con el consumo de c ; por su parte, en (b) se incluye la posibilidad de contar con más segmentos para los cuales haya existido consumo del canal c en P , por ello describiremos este último.

Tomamos a μ_1 como el primer minuto en el que se vio el canal c , por hipótesis $\mu_\delta < \mu_f$, por lo tanto existe un primer índice γ para el cual $\mu_\gamma \geq \mu_f$. Para obtener el consumo del canal c en $[\mu_0, \mu_f]$ basta con sumar la longitud de los segmentos a partir del minuto μ_1 hasta el minuto μ_f si γ es par, es decir, si $\mu_\gamma = \mu_j$ con $\mu_j \in \{\mu_2, \mu_4, \dots, \mu_p\}$, ó hasta el minuto $\mu_{\gamma-1}$ si $\mu_\gamma = \mu_j$ con $\mu_j \in \{\mu_1, \mu_3, \dots, \mu_{p-1}\}$.

Con esto obtenemos el consumo de c en $[\mu_0, \mu_f]$, por lo tanto obtenemos el consumo de c en cualquiera de los casos citados.

Hasta este punto hemos reconstruido el consumo del canal c en un segmento, sin embargo, por hipótesis las entradas c_i, c_{i+1}, \dots, c_k del renglón c en Λ son las que coinciden con P , por tanto bastará con realizar el mismo procedimiento para los segmentos s_i, s_{i+1}, \dots, s_k , de esta forma obtendremos el consumo del canal c en cualquier periodo.

Finalmente, para calcular la Audiencia del canal c en el periodo P , hacemos el cociente:

$$\text{Audiencia} = \frac{\text{Minutos vistos en el canal } c \text{ durante } P}{\text{Duración de } P} \quad (3.2)$$

Como se mencionó en las definiciones, existen otras medidas de gran uso en este medio, entre ellas están las siguientes:

$$\text{GRP}'s = \sum_{i \in D} \text{Audiencia}_i \quad (3.3)$$

donde D es un periodo de tiempo determinado mayor a una semana y Audiencia_i es la Audiencia del i -ésimo día de D .

Con los GRP's medimos la cantidad de impactos de un cierto producto, en este caso medimos la cantidad de veces que t es alcanzado.

La fórmula para calcular el alcance es:

$$\text{Alcance} = \frac{\text{Personas expuestas al menos un minuto}}{\text{Universo}} \quad (3.4)$$

Ya que sólo tenemos un televidente, el alcance es 100% ó 0% en caso de que t haya visto ó no haya visto la televisión, respectivamente.

Por último, la frecuencia es el número promedio de veces en la que se impacta a una persona.

$$Frecuencia = \frac{GRP's}{Alcance} \quad (3.5)$$

Con esto concluimos esta sección, antes de pasar a nuestro último caso haremos algunas observaciones importantes.

Hasta este momento no hemos utilizado el peso del individuo en la muestra porque sólo hemos trabajado con un individuo, por lo tanto su peso nos es indiferente ya que él representa a toda la muestra, por ende a todo el universo. Además no hemos podido ver el ahorro en el almacenamiento ya que anteriormente suponíamos que el televidente t siempre veía la televisión, sin embargo, en el Raw Data existen muchos televidentes que no presentan consumo alguno, para éstos podremos ahorrar memoria considerablemente ya que únicamente almacenaremos la información del identificar, peso y demográfico de los mismos.

3.5 El caso general

En esta sección contemplaremos a todos los televidentes de \mathcal{T} , es decir, la existencia de z televidentes, diremos que cada t_z vió m_z canales durante s_z segmentos. De esta manera nuestra nueva base consistirá de una sucesión de matrices, cada una de ellas de distinto orden y además independiente de las demás, independiente en el sentido de que no tienen nada en común, salvo que representan el consumo televisivo de un individuo.

Gracias al modo de generalización, en este caso es suficiente hacer notar que para calcular la Audiencia de un periodo deseado $P = [u, v]$, basta con seguir con el procedimiento de la sección anterior en cada individuo, la ventaja que tendremos esta vez es que muchos individuos no poseen matriz de consumo y por lo tanto su aportación a la Audiencia es únicamente su representación en la muestra, es decir, su identificador, peso y demográfico, ya que su consumo televisivo es nulo, además seguramente tendremos muchos otros individuos en los cuales no haya consumo para ese periodo determinado, o si lo hay, muy probablemente no haya consumo en todos los canales. por tanto la extracción

de los datos se reduce considerablemente. Esto nos ha ahorrado tanto el almacenamiento como el método de cálculo de la Audiencia.

Con base en esto, la formula para calcular la Audiencia de un periodo $P = [u, v]$, con $1 \leq u \leq v \leq 1440$, en el canal c es:

$$Audiencia = \frac{\sum_{t \in \mathcal{T}} \mathcal{T}(\text{Minutos vistos en } c \text{ durante } P \text{ por } t) \times W_t}{\sum_{t \in \mathcal{T}} \mathcal{T}(\text{Duración de } P) \times W_t} \quad (3.6)$$

donde W_t es el peso del televidente t en la muestra, es decir, su expansor.

Los GRP's en este caso, quedan expresados por la ecuación:

$$GRP's = \sum_{i \in D} Audiencia_i \quad (3.7)$$

donde D es un periodo de tiempo determinado mayor a una semana y $Audiencia_i$ es la Audiencia del i -ésimo día de D , calculada con la Ecuación 3.6.

En este caso, el alcance toma mayor sentido, ya que anteriormente tomaba los valores $\{0, 100\}$, debido a que sólo contabamos con un televidente, sin embargo ahora cobra mayor sentido. Además, gracias a

nuestra construcción si un televidente no cuenta con matriz de consumo sabemos que no podrá aportar algo a la Audiencia y para aquellos que cuenten con una matriz de consumo bastará con conocer si en P estuvieron viendo la televisión. El cálculo del Alcance queda expresado por:

$$\text{Alcance} = \frac{\text{Personas expuestas al menos un minuto}}{\text{Universo}} \quad (3.8)$$

De igual forma, la frecuencia toma un sentido más representativo y nuevamente se define como:

$$\text{Frecuencia} = \frac{\text{GRP's}}{\text{Alcance}} \quad (3.9)$$

Como observaciones finales a este caso podemos resaltar la importancia de trabajar con datos acotados ya que así garantizamos la convergencia del proceso, además estas cotas se alcanzan con muy poca frecuencia haciendo del método descrito un proceso eficiente.

Con esto hemos logrado reducir el Raw Data, pero desafortunadamente en la práctica los métodos de cálculo para reducir y exportar los datos almacenados en la matriz de consumo sufren de imperfec-

ciones por los propios algoritmos de los sistemas computacionales, por lo tanto, en el siguiente capítulo explicaremos las modificaciones pertinentes para ajustar el modelo.

Capítulo 4

Consideraciones Numéricas

Uno de los principales problemas entre desarrollar una teoría numérica e implementarla es que en algunas ocasiones la aplicación sufre de problemas computacionales, desafortunadamente nuestro trabajo cae en esta categoría. La implementación numérica de nuestro modelo sufre de serios problemas, el principal de ellos es el almacenamiento de cantidades demasiado grandes que ocasionan diferencias sustanciales en la extracción.

Como es bien sabido, los sistemas computacionales tienen un límite de capacidad numérico, este problema no se basa en una limitante de espacio, más bien es un problema de operación, los sistemas pueden hacer operaciones numéricas hasta un determinado límite. en general

los sistemas utilizan distintos tipos de números, el tipo más completo de ellos es el sistema *float*, el cual tiene capacidad de trabajar con números de hasta 58 dígitos, una vez sobrepasada esta cantidad el sistema carece de precisión operacional, por si esto no fuera poco, no todos los sistemas cuentan con esta capacidad, de hecho la mayoría posee una capacidad menor, por tanto es necesaria una revisión de las capacidades de los sistemas comunes. [8]

Como es lógico de esperarse, nuestro almacenamiento maneja cantidades hasta del tamaño $2^{1440-1+1} - 1 \approx 2^{1440}$ ya que este es el caso en el que se haya visto la televisión durante todo el día, evidentemente esta cantidad sobrepasa la capacidad de cualquier sistema, por estas razones debemos modificar nuestro proceso con fin de apegarlo a las limitantes de los computadores convencionales. Afortunadamente este nuevo proceso, aunque menos compacto que el desarrollado previamente, sigue siendo más dinámico y con la mismas características que el original.

Las modificaciones realizadas consisten en la partición de los segmentos, recuerdese que éstos son los que almacenan las cantidades más grandes y, además, si dividimos los segmentos el consumo de los canales

vistos en ellos se distribuye de forma adecuada. Inicialmente pensamos en una división por segmentos de una hora de duración, desafortunadamente éstos seguían presentando problemas computacionales, no fue sino hasta segmentos de 49 minutos de duración, donde ya no se presentaban estos problemas, sin embargo en el ámbito televisivo, esta cantidad no es manejada frecuentemente, en este medio son frecuentes los intervalos de media hora, una hora, dos horas, etcétera, por esta razón decidimos dividir los segmentos en intervalos de 30 minutos de duración como máximo, en los cuales conservamos toda la teoría desarrollada anteriormente y además libramos los problemas computacionales de las computadoras de cualquier empresa enfocada a este medio.

De este modo los cambios en el método son los siguientes:

Sea $t \in \mathcal{T}$ fijo y $s_i \in \mathcal{S}$ tal que t vió la televisión en el segmento s_i , debemos verificar que:

$$|s_i| \leq 30 \tag{4.1}$$

Si esto se satisface para toda i no tendremos ningún problema, sin embargo, si existen segmentos $s_l, \dots, s_k \in \mathcal{S}$ que no cumplan la condición

de la Ecuación 4.1 debemos particionar cada segmento en subsegmentos de 30 minutos de duración.

Sea s_l un segmento fijo, tal que $|s_l| = L > 30$, entonces tendremos $p = \left\lfloor \frac{L}{30} \right\rfloor$ sub-segmentos de treinta minutos de duración y a lo más un segmento adicional de duración $r = 30 * \left[\frac{L}{30} - \left\lfloor \frac{L}{30} \right\rfloor \right]$.

En este caso sólo reordenamos los segmentos del día y en lugar de tener n segmentos tendremos $n + p$ segmentos, conservando las mismas características que los originales.

Haciendo esto a cada segmento que no satisfaga la Ecuación 4.1, construiremos una partición de los segmentos del día en subsegmentos de a lo más treinta minutos de duración, con los cuales expresamos todo el Raw Data.

Capítulo 5

Conclusiones

El medio televisivo tiene cada vez más importancia, no obstante de la aparición de nuevos medios de comunicación, tales como Internet, de hecho este último es el que cuenta con mayores oportunidades de crecimiento publicitario, en el sentido de que la creatividad es la única limitante de expansión y que además cuenta con la preferencia de la gente como medio de entretenimiento, atributo que se ha ganado por ser económico y de fácil acceso para la mayor parte de la población mundial.

Gracias a los avances tecnológicos, la televisión cobrará aún más fuerza de la que posee actualmente, además su cobertura se ha ampliado de forma tal que casi cualquier persona, por más lejana que

esta se encuentre, tienen la posibilidad de captar la señal de un cadena televisora. Finalmente, el proceso de globalización ha otorgado a la televisión una expansión que hace algunos años no se creía posible alcanzar. Por todas estas razones la televisión es el medio de comunicación masivo más importante que existe en nuestros tiempos. Por esta misma razón, el cálculo de la Audiencia es sumamente importante ya que de él depende el desarrollo del medio televisivo, específicamente de sus patrocinadores. En este caso particular, la Audiencia toma un singular importancia ya que TV-Azteca basa sus ventas en el concepto del CPA.

Desafortunadamente, la base utilizada para calcular la Audiencia (El Raw Data), es una base de difícil aprovechamiento, por esto surge la necesidad de crear una nueva base coherente con su explotación. En este trabajo presentamos una posibilidad que satisface estas carencias, sin embargo, deben buscarse formas alternativas a ésta, que sustituyan o complementen a la presentada aquí, buscando optimizar su rendimiento cada vez más. Con las mejoras propuestas en el presente trabajo intentamos dejar claro que pueden realizarse análisis más completos, que

reflejen fielmente todo lo que esta pasando con el consumo televisivo y para los cuales se puedan encontrar más y mejores conclusiones, este es el primer paso en el avance del Raw Data, esperamos sinceramente que el medio evolucione con conceptos e ideas que vengan a fortalecerlo y que las empresas que trabajen en él sean capaces de asimilar dicho desarrollo y sepan utilizarlo para su beneficio.

A lo largo de todo el trabajo desarrollamos la metodología para la muestra de individuos, no obstante los mismos resultados son aplicables a la muestra de hogares, en cuyo caso podemos reducir aún más el almacenamiento ya que un hogar comunmente cuenta con más de un televisor, de este modo, registraremos una sucesión de matrices, en donde cada elemento puede constar de cero, uno, dos, tres, o más elementos.

Deseamos resaltar la importancia en el buen manejo de los datos ya que es de ahí de donde parte el buen funcionamiento del análisis, que sirve para la oportuna toma de decisiones, lo que lleva a un funcionamiento de la empresa. Como hemos visto, los datos forman parte de la base necesaria para el buen desempeño de la compañía, por ende

debemos buscar la mejor forma de aprovechamiento de los mismos y darles la importancia que merecen.

Quizá el punto más importante es el hecho de motivar a los participantes del medio para que cada vez se vean más envueltos en el análisis de variables que permitan a las agencias llevar a cabo una planeación más a la medida de cada anunciante que requiere de sus servicios y que esto mismo sirva para que las cadenas televisivas sean más exhaustivas y dediquen más empeño al área de investigación, provocando una reforma en el medio que sea de conveniencia para todos sus participantes. Todo esto traerá como consecuencia una mayor demanda de información, misma que debe ser satisfecha por la empresa IBOPE-AGB.

En resumen, esperamos que en un futuro el medio sea capaz de evolucionar para contar con una mejor publicidad, ya que actualmente lo más importante en la televisión es ésto, sin ella no sería posible mantener a estas empresas que día a día llevan entretenimiento a gran parte de la población y que implícitamente visto desde el punto de vista del telespectador, se sustentan mediante la venta de espacios dentro de sus programas, sin embargo, desde el punto de vista de las cadenas

televisivas, el medio es una oportunidad para ofrecer a los anunciantes la posibilidad de exhibir su producto a un gran número de personas ya que este es precisamente su negocio.

Bibliografía

- [1] Establishment Survey, 2004, IBOPE AGB
- [2] <http://www.ibope.com.mx>
- [3] Manual de Conceptos Básicos, 2004, IBOPE AGB
- [4] <http://www.tvazteca.com.mx>
- [5] <http://www.esmas.com/televisahome>
- [6] Robert W. Bly, 1999, Advertising Manager's Hand Book
- [7] Emilio Martínez Ramos, 1992, El uso de los medios de Comunicación en Marketing y Publicidad
- [8] Richard L. Burden, 1985, Análisis Numérico
- [9] <http://www.maths.tcd.ie/~dwilkins/LaTeXPrimer>