



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL**

**CONFIGURACIÓN DE LA OFERTA DE PRODUCTOS HOGAR PARA CLIENTES
RESIDENCIALES DE UNA EMPRESA DE TELECOMUNICACIONES**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

DARÍO SIMÓN ROJAS STAPEL

**PROFESOR GUÍA:
MARCEL GOIC FIGUEROA**

**MIEMBROS DE LA COMISIÓN:
RICARDO MONTOYA MOREIRA
ALVARO OVANDO VASQUEZ**

**SANTIAGO DE CHILE
ENERO 2013**

**RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL
POR: DARIO ROJAS STAPEL
FECHA: 21/01/13
PROF.GUÍA: SR. MARCEL GOIC F**

**CONFIGURACIÓN DE LA OFERTA DE PRODUCTOS HOGAR PARA CLIENTES
RESIDENCIALES DE UNA EMPRESA DE TELECOMUNICACIONES**

La industria de las Telecomunicaciones destaca por su marcado dinamismo y su constante crecimiento de los últimos años. Adicionalmente, se observa un mayor empoderamiento de los clientes, reflejado en demandas de productos y servicios que se adapten a sus requerimientos. Lo anterior, genera incentivos importantes para que las empresas del rubro pongan especial atención en la comprensión de sus clientes para crear valor y satisfacer sus necesidades.

La industria posee dos grupos de usuarios; empresas y particulares. Estos últimos pueden ser divididos en dos tipos; clientes móviles y residenciales. Los clientes residenciales están asociados a una única locación y los principales productos ofrecidos son telefonía fija, televisión digital e internet fijo con sus variadas combinaciones y especificaciones. La memoria a desarrollar se enfoca en estos clientes y en los productos ofrecidos a ellos.

El proyecto busca desarrollar una metodología que permita profundizar el conocimiento de los clientes hogar, pudiendo comprender las dinámicas que se generan entre los consumidores y los productos ofrecidos por la empresa. En este sentido, el objetivo general de este proyecto consiste en asistir la venta a través de mecanismos estadísticos formales. Dichos mecanismos, estiman la propensión que tienen los diferentes perfiles de clientes a adquirir los diferentes productos ofrecidos por la empresa. Lo anterior permite direccionar los esfuerzos de venta, potenciando los actuales procesos de oferta de la compañía.

Para estimar la propensión que tienen los clientes a elegir las diferentes alternativas, se utilizan dos modelos independientes; un primer modelo Logit Binario para estimar la probabilidad que tiene un cliente a cambiar de producto, y un segundo modelo Logit Anidado, para estimar la propensión que tiene el cliente a elegir cada una de las alternativas, dado que ha realizado un cambio. El primer modelo es capaz de identificar el 52.7% de los cambios realizados por los clientes mientras que el modelo anidado logra predecir un 63.3% de las elecciones.

Al analizar los resultados obtenidos, se observa que la utilización de la metodología propuesta en este trabajo tiene el potencial de aumentar los ingresos asociados a los procesos de oferta de la empresa en casi un 500% en contraste con los resultados obtenidos actualmente.

AGRADECIMIENTOS

A mi madre.

TABLA DE CONTENIDO

1.	INTRODUCCIÓN.....	9
1.1	Antecedentes Generales	9
1.2	La Empresa y su Entorno	9
1.3	Planteamiento del Problema y Justificación.....	11
1.4	Objetivos.....	13
1.4.1	Objetivo General	13
1.4.2	Objetivos Específicos.....	14
1.5	Alcances	14
1.6	Resultados Esperados.....	14
2	MARCO CONCEPTUAL.....	15
2.1	Estimación de Probabilidad de Compra.....	15
2.1.1	Modelo Logit Binario	15
2.1.2	Modelo Logit Anidado (Nested Logit)	15
2.1.3	Heterogeneidad Observable	17
2.1.4	Heterogeneidad no Observable (Clases Latentes)	18
2.2	Función de Verosimilitud.....	18
2.3	Metodología KDD.....	19
3	METODOLOGÍA.....	20
3.1	Selección de Datos	20
3.2	Pre-Procesamiento de Datos.....	20
3.2.1	Eliminación de Observaciones	20

3.2.2	Descripción de Variables.....	20
3.2.3	Datos Faltantes	20
3.3	Transformación de Datos.....	21
3.3.1	Creación de Variables.....	21
3.4	Desarrollo de Modelos Económicos	21
3.4.1	Modelo sin Heterogeneidad	21
3.4.2	Modelo con Heterogeneidad Observable	21
3.4.3	Modelo con Heterogeneidad no Observable.....	21
3.5	Evaluación	22
3.5.1	Capacidad de Ajuste	22
3.5.2	Capacidad Predictiva	22
3.6	Propuestas de Acción	22
4	DESARROLLO	22
4.1	Selección de Datos	22
4.1.1	Características del Cliente	23
4.1.2	Comportamiento del Cliente.....	24
4.1.3	Productos Contratados.....	24
4.1.4	Atributos de los Productos	25
4.2	Pre-Procesamiento de Datos.....	26
4.2.1	Eliminación de Observaciones	26
4.2.2	Descripción de Variables.....	27
4.2.3	Datos Faltantes	38
4.3	Transformación de Datos.....	39
4.3.1	Creación de Variables.....	39

4.4	Desarrollo de Modelos Econométricos	39
4.4.1	Modelo de Cambio sin Heterogeneidad	41
4.4.2	Modelo de Cambio con Heterogeneidad Observable.....	41
4.4.3	Modelo de Cambio con Heterogeneidad no Observable.....	43
4.4.4	Modelo de Elección sin Heterogeneidad	50
4.4.5	Modelo de Elección con Heterogeneidad Observable	51
4.4.6	Modelo de Elección con Heterogeneidad no Observable.....	52
4.5	Evaluación	55
4.5.1	Capacidad Predictiva	55
4.5.2	Comparación con Procesos Actuales	59
4.6	Propuestas de Acción.	63
5.	CONCLUSIONES	65
5.1	Conclusiones del Trabajo	65
5.2	Trabajos Futuros.....	66
6.	BIBLIOGRAFÍA Y FUENTES DE INFORMACIÓN	68
7.	ANEXOS.....	69
7.1	Anexo A	69
7.2	Anexo B	70
7.3	Anexo C	76

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Esquematización de un modelo anidado	16
Ilustración 2: Metodología KDD	19
Ilustración 3: Esquematización decisiones anidadas	49
Ilustración 4: Árbol de decisión	50
Ilustración 5: Árbol de decisión Trio	78

ÍNDICE DE GRÁFICOS

Gráfico 1: Participación de mercado por categoría de producto	10
Gráfico 2: Evolución en penetración por hogar según categoría de producto	11
Gráfico 3: Velocidad de banda ancha promedio según edad	29
Gráfico 4: Penetración por categoría de canales adicionales	31
Gráfico 5: Plan de televisión contratado según grupo socioeconómico	31
Gráfico 6: Penetración canal adicional de deporte según antigüedad de servicio de televisión	32
Gráfico 7: Planes de telefonía.....	33
Gráfico 8: Penetración canal de telefonía ilimitado según segmentación preliminar	34
Gráfico 9: Penetración plan ilimitado de telefonía según minutos salientes	34
Gráfico 10: Penetración plan controlado de 150 o 250 minutos de telefonía según minutos salientes.....	35
Gráfico 11: Porcentaje de cambios según edad	36
Gráfico 12: Porcentaje de cambios según frecuencia de cambio	37
Gráfico 13: Porcentaje de cambios según grupo socioeconómico.....	37
Gráfico 14: Histograma y Boxplot de clases latentes.....	45
Gráfico 15: Variación en la probabilidad de cambio, según variación en gasto por servicio	46
Gráfico 16: Variación en probabilidad de cambio según tenencia de servicio televisivo.....	47
Gráfico 17: Variación de la probabilidad de cambio según canales adicionales contratados.....	47
Gráfico 18: Histograma y Boxplot de probabilidad de adquirir duo banda ancha-telefonía por clases	53
Gráfico 19: Probabilidad de migrar a Duo banda ancha-telefonía o Trio desde plan solo telefonía según años de telefonía	54

Gráfico 20: Probabilidad de adquirir planes de solo telefonía y Duo televisión-Banda Ancha según grupo etario	55
Gráfico 21: Capacidad predictiva según porcentaje de la población considerada .	56
Gráfico 22: Participación de mercado servicio de telefonía	69
Gráfico 23: Participación de mercado servicio de banda ancha	69
Gráfico 24: Participación de mercado servicio de televisión	70
Gráfico 25: Distribución Grupo Socioeconómico.....	70
Gráfico 26: Distribución Grupo Género	71
Gráfico 27: Distribución segmentación preliminar.....	71
Gráfico 28: Distribución velocidad de banda ancha	72
Gráfico 29: Distribución edad.....	72
Gráfico 30: Distribución gasto total por servicio (normalizado por 10,000)	73
Gráfico 31: Distribución antigüedad del cliente	73
Gráfico 32: Distribución numero de decodificadores.....	74
Gráfico 33: Penetración canales adicionales de cine según antigüedad del servicio de telefonía.....	75
Gráfico 34: Penetración canales adicionales de otras categorías según antigüedad del servicio de telefonía.....	75

INDICE DE TABLAS

Tabla 1: Datos eliminados del análisis.....	26
Tabla 2: Descripción variables numéricas	27
Tabla 3: Cambio de velocidad contratada de un periodo a otro.....	29
Tabla 4: Datos faltantes	38
Tabla 5: Resultados modelo de cambio sin heterogeneidad	41
Tabla 6: Comparación modelos de cambio con heterogeneidad observable.....	42
Tabla 7: Resumen resultados modelo cambio con heterogeneidad observable	43
Tabla 8: Comparación modelos de cambio con clases latentes	44
Tabla 9: Capacidad de ajuste modelo de cambio sin heterogeneidad.....	50
Tabla 10: Estadísticos de ajuste, modelos de elección con heterogeneidad	52
Tabla 11: Comparación modelos de elección con clases latentes.....	52
Tabla 12: Resumen capacidad predictiva modelo de cambio para puntos de corte considerados.....	57
Tabla 13: Capacidad predictiva modelo de elección para alternativas consideradas	58
Tabla 14: Capacidad predictiva modelo de elección a nivel de categorías para alternativas consideradas.....	59
Tabla 15: Comparación modelo de cambio con procesos actuales	59
Tabla 16: Capacidad predictiva según tipo de campaña	61
Tabla 17: Cambios con y sin campaña según tipo de campaña	61
Tabla 18: Beneficio promedio asociado a cada alternativa y beneficio actual	62
Tabla 19: Parámetros modelo de cambio con heterogeneidad no observable	77
Tabla 20: Parámetros modelos de elección sin heterogeneidad	79
Tabla 21: Parámetros modelos de elección con heterogeneidad observada.....	82
Tabla 22: Parámetros modelos de elección con heterogeneidad no observada....	87

1. INTRODUCCIÓN

1.1 Antecedentes Generales

En los últimos años, la Industria de las Telecomunicaciones ha presentado un marcado dinamismo y una creciente competitividad, reflejada tanto en el número de actores, como en las inversiones realizadas por los mismos. En el informe estadístico de Diciembre de 2011 elaborado por la Subsecretaria de Telecomunicaciones [1], se señala que el mercado ha mostrado fuertes cambios en respuesta a las nuevas exigencias legales, reflejado en la inversión sectorial más alta de los últimos 8 años, estableciéndose hoy como uno de los sectores más dinámicos del país.

En este contexto, ha existido un aumento en el interés de las empresas pertenecientes al rubro por mejorar su posición competitiva, derivando en un incremento de los esfuerzos en todas las áreas de negocio por crear valor y satisfacer las necesidades de los consumidores. En Marketing, estos nuevos intereses pueden responderse mediante un proceso denominado “Marketing Relacional” (RM) orientado a la generación, mantención y desarrollo de relaciones rentables con los clientes [2].

Las unidades de negocio de CRM (Customer Relationship Management) cuyo objetivo está orientado a la recopilación de información y análisis de clientes, juegan un importante papel en el contexto de la industria nacional y tiene directa relación con los procesos de RM, siendo parte fundamental en la generación de conocimiento de los consumidores, y por lo tanto, en la creación de valor para los mismos.

Desde el Marketing Relacional, la estimación de la propensión de compra de los consumidores adquiere vital importancia, determinando en gran medida la oferta por parte de las empresas. Dicha estimación, es relevante pues acota el número de productos que ofrece la compañía, focalizando los esfuerzos hacia los más relevantes para el cliente, pudiendo aumentar la satisfacción de los mismos y mejorar la rentabilidad de la firma.

El proyecto a desarrollar propone estimaciones de la propensión de compra de los clientes residenciales de Movistar, identificando las principales características que afecten su comportamiento de compra. Agrega valor al considerar modelos estadísticos robustos capaces de apoyar formalmente la oferta de productos hogar.

1.2 La Empresa y su Entorno

Movistar es el nombre comercial de una empresa de telecomunicaciones que ofrece productos fijos y móviles para clientes particulares y negocios. Por su parte, Telefónica corresponde a la marca institucional de la compañía [3]. El presente trabajo se desarrolla en el área de CRM de la firma, a cargo de la Subgerencia de Inteligencia de Negocios.

En la empresa, se distinguen dos grupos de clientes particulares, cliente hogar y cliente móvil. Los primeros son clientes residenciales cuyo servicio está asociado a un área física particular, por su parte, los clientes móviles pueden utilizar el servicio en diversos lugares, principalmente internet móvil y telefonía celular. Los principales servicios asociados a clientes hogar son telefonía fija, banda ancha y televisión digital, con sus diversas variaciones y combinaciones, además de servicios de valor agregado como protección y seguridad del computador, buzón de voz, canales adicionales, entre otros. La presente memoria se enfoca en los clientes residenciales y los principales productos asociados a ellos.

Actualmente, el mercado de servicios de telecomunicación fijos, está dominado por cinco principales actores, Entel, Movistar, VTR, Grupo Claro y Grupo GTD. Dicha participación difiere fuertemente según categoría de productos. Para finales del 2011, el claro dominador en Telefonía fija era Movistar con el 55% del mercado, en Televisión Digital VTR con un 44% y para Banda Ancha fija, Movistar con un 44% [1]. En el siguiente grafico se muestran las participaciones de mercado según categoría de los diferentes competidores.

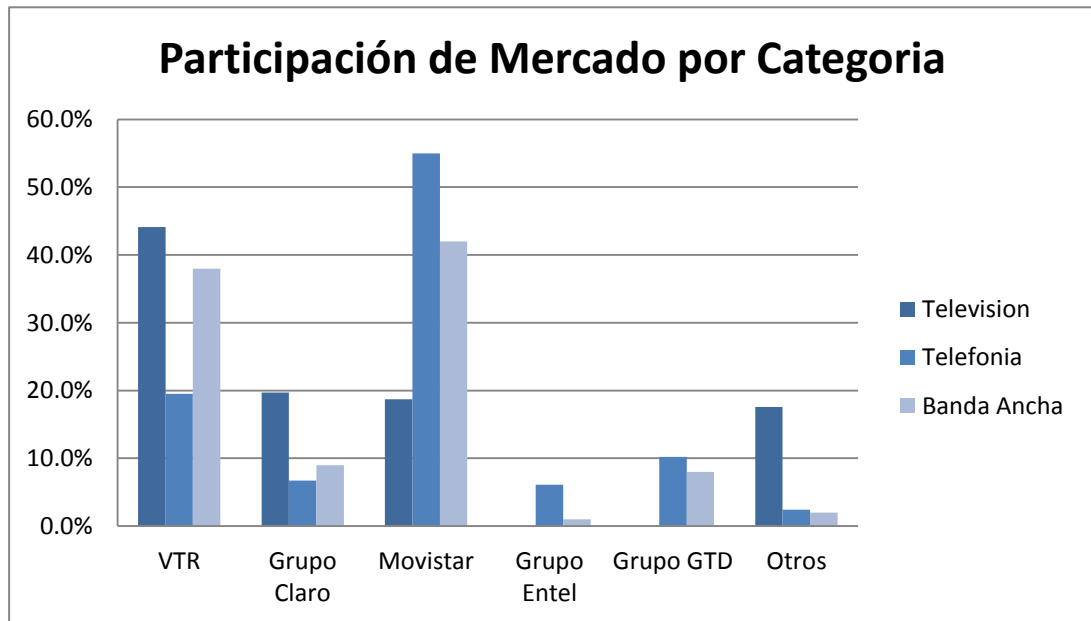


Gráfico 1: Participación de mercado por categoría de producto

Fuente: Elaboración Propia con datos de [1].

La demanda por servicios de telefonía ha disminuido en tanto que los servicios de internet y televisión han tenido una tendencia positiva en los últimos años. A continuación, se muestra la evolución de la penetración por hogar en diciembre desde 2006 a 2011 según categoría.

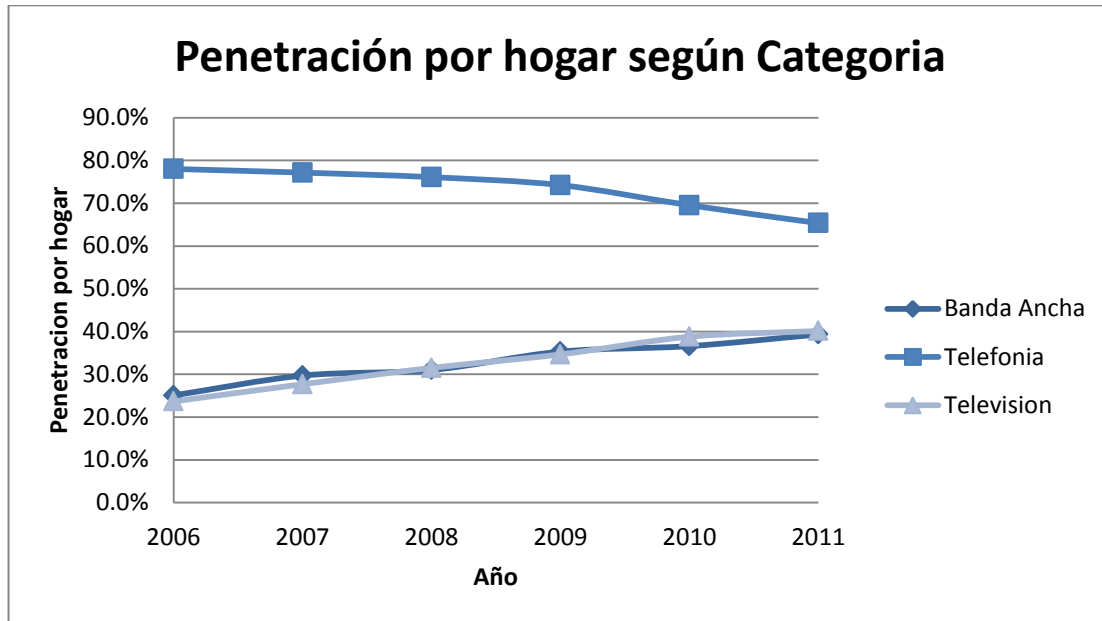


Gráfico 2: Evolución en penetración por hogar según categoría de producto

Fuente: Elaboración Propia con datos públicos de la Subsecretaría de Telecomunicaciones.

El gráfico anterior muestra la penetración por hogar de las diferentes categorías de productos, dicha penetración es calculada como el número de servicios contratados dividido el total de “hogares” en la población. Es posible observar una disminución en la penetración de los servicios de telefonía en contraste con el aumento en las otras categorías, reflejo de las nuevas tendencias presentes en la industria. Lo anterior es relevante para la empresa dado que su mayor poder de mercado en servicios fijos proviene de telefonía, y por lo tanto, existe una presión por generar políticas capaces de hacer frente a las nuevas tendencias de la industria.

En tanto el contexto regulatorio también ha cambiado en los últimos años. La compartición de infraestructura, la portabilidad numérica móvil y fija, entre otros cambios, han generado una disminución en las barreras de entrada que han tenido como resultado una mayor competitividad y dinamismo en la industria [2]. Debido a esto, el presente año han ingresado nuevos competidores al mercado tanto en servicios móviles como fijos, los que presentan una mayor presión para los actuales competidores por generar servicios capaces de entregar valor al cliente.

1.3 Planteamiento del Problema y Justificación

Las nuevas condiciones de la industria, hacen que el conocimiento de los clientes, en miras de ofrecer productos que se ajusten a sus necesidades, sea uno de los aspectos más relevantes para las compañías. En este contexto, no basta solo con la generación de productos y servicios interesantes para el cliente sino además, en la capacidad de ofrecer los productos adecuados a los consumidores indicados.

La estimación de la propensión de compra, es decir, de la correspondencia entre las características de un producto y los intereses de un cliente, es un aspecto relevante para las compañías pues permite dirigir la oferta de los servicios o productos considerando las valoraciones de cada cliente. Una estimación formal en este sentido permite disminuir esfuerzos y costos asociados a la venta, y aumentar los retornos asociados a cada cliente al ofrecer servicios de mayor rentabilidad que el cliente si este dispuesto a adquirir.

En la actualidad, la empresa cuenta con modelos estadísticos que describen la propensión de compra. Sin embargo, sus resultados son bastante agregados y poco interpretables en muchos casos. Específicamente, la empresa utiliza árboles de decisión y regresiones lineales que permiten asociar algunas características de clientes, tales como la edad, el grupo socioeconómico o una segmentación preliminar que determina tres grupos asociados al valor del cliente, a ciertos atributos de productos, tales como si posee internet, telefonía o televisión. Dichos modelos, tienen algunas limitaciones que se señalan a continuación.

I. No permiten describir la propensión de compra de un cliente particular, solo de grupos de clientes.

En efecto, los modelos utilizados por la compañía utilizan ciertas variables demográficas, como la edad, grupo socioeconómico o género para describir diferencias en la propensión de compra entre grupos de clientes.

Los arboles de decisión utilizados por la empresa discretizan las variables continuas generando grupos de clientes para evaluar la propensión de estos de adquirir banda ancha, telefonía o televisión. Dichos modelos determinan 10 grupos de clientes estimando la probabilidad de que estos contraten servicios de cada categoría. Dichos grupos separan a los clientes según algunas variables como la edad, mayores y menores de 50, o el grupo socioeconómico al que pertenece el individuo.

Los resultados obtenidos de estos modelos no permiten determinar la propensión asociada a un cliente particular, sino, a los grupos de clientes generados por el modelo. Este nivel de agregación no permite determinar o dirigir esfuerzos de oferta sino generar un análisis exploratorio de los diferentes tipos de clientes.

II. No permiten describir la propensión de compra por producto, solo por categoría de productos.

Los modelos desarrollados hasta el momento están asociados a categorías de producto, es decir, banda ancha, televisión o telefonía, por lo que no permiten describir el atractivo de un producto particular. Al igual que la limitante anterior, esto no permite apoyar directamente los esfuerzos de venta sino describir a grandes rasgos el comportamiento de los clientes.

III. Solo incluye patrones sociodemográficos de clientes por lo que no incluyen variables que pueden resultar relevantes, como patrones de consumo o características del “hogar”.

Los modelos desarrollados, tanto regresiones como arboles de decisión, solo utilizan características demográficas, no considerando la información de consumo en periodos anteriores, los años del cliente en la compañía, o variables asociadas a las características del grupo familiar.

Dadas estas limitaciones, el análisis derivado de dichos modelos no permite tomar decisiones concretas de oferta. En efecto, las decisiones relacionadas con a qué clientes deben ir dirigidas las ofertas, y que productos deben ser ofertados, responden a una sucesión de filtros basados en reglas de negocio e ideas generales del comportamiento de los clientes.

Considerando la falta de procedimientos formales para apoyar los procesos de oferta, la empresa, al no tener un foco claro de los clientes más valiosos o propensos a adquirir otro producto, realiza importantes esfuerzos para contactar a la mayor cantidad de clientes posible, lo que determina un aumento de los costos para la compañía y una probable fatiga por parte de los clientes. A lo anterior se suma el hecho de que los productos ofrecidos a cada persona no tienen relación con los intereses de las mismas, generando muchas veces una oferta poco atractiva para los clientes, perdiendo entonces la oportunidad de generar mayor valor para el cliente y un aumento de ingresos para la firma.

El presente trabajo, busca metodologías alineadas con las nuevas necesidades de Movistar, aportando modelos capaces de estimar la propensión de compra a un nivel desagregado, por lo que permiten apoyar la decisión de qué productos ofertar a cada perfil de cliente residencial, considerando la información disponible que se tiene respecto a las variables relevantes en el comportamiento de compra.

Específicamente, el proyecto caracteriza a los diferentes clientes de la empresa considerando el conjunto de variables que se consideren relevantes en cuanto a las decisiones de compra, se desarrollan modelos estadísticos que estimen la probabilidad de compra de cada producto que ofrece la compañía para cada perfil de cliente, se evalúan los diferentes modelos para finalmente proponer medidas que integren los resultados a los actuales procesos de venta de la firma.

1.4 Objetivos

1.4.1 Objetivo General

Diseñar y evaluar mecanismos formales de apoyo a los procesos de venta a clientes residenciales considerando las características particulares de cada cliente, el tipo de hogar y las dinámicas asociadas a su evolución de estilo de vida.

1.4.2 Objetivos Específicos

- I. Determinar el conjunto de variables que mejor describe el comportamiento de compra de clientes hogar.
- II. Diseñar e implementar modelos econométricos para describir el atractivo de cambiar a cada alternativa de oferta para cada perfil de cliente residencial.
- III. Evaluar la capacidad predictiva de los modelos para describir la propensión de compra de cada oferta.
- IV. Describir como los resultados de los modelos formales de análisis se integran a los procesos de negocios de la firma para asistir la venta.

1.5 Alcances

El proyecto a desarrollar, busca mejorar la oferta de servicios de Movistar desde el análisis del cliente y la relación de este con los servicios que ofrece la compañía. En este contexto se reconocen las siguientes limitaciones.

- I. Dentro de las variadas aristas que poseen los procesos de ventas, en el presente proyecto solo se analizarán las asociadas al *match* de productos con los diversos clientes. Por lo tanto, no se considerarán otras variables como los canales de venta utilizados, posibles promociones, entre otros.
- II. El desarrollo de los diversos modelos se realizará sobre una muestra de clientes tal que sea representativa de la base de datos completa. De todas formas, los modelos serán escalables y por tanto aplicables a la totalidad de clientes que posean los productos considerados en este trabajo.
- III. Los productos considerados serán únicamente los ofrecidos actualmente por la empresa en condiciones normales, no se consideraran por lo tanto, los productos ofrecidos bajo restricciones técnicas o a clientes antiguos que no han migrado a los nuevos productos. En este sentido, los modelos serán capaces de estimar la probabilidad de compra para la totalidad de clientes que posean alguno de estos productos.
- IV. Los datos utilizados en el proyecto considerarán solo los existentes actualmente en la empresa. En este sentido, no se considera el levantamiento de nueva información.
- V. El análisis será desarrollado solo sobre clientes residenciales particulares.
- VI. El presente proyecto analiza las migraciones de los clientes entre un producto y otro al interior de la compañía, utilizando información que posee la empresa del comportamiento de compra en los periodos anteriores, por lo tanto, es necesario contar con información de clientes activos en un periodo de al menos 4 meses.

1.6 Resultados Esperados

Tras el desarrollo de este proyecto, se espera obtener los siguientes resultados:

- I. Identificar el conjunto de variables que mejor describen el comportamiento de compra de los clientes.

- II. Un conjunto de modelos estadísticos capaces de describir la propensión de compra de los diferentes clientes, cuyo error de predicción sea inferior a los desarrollados actualmente por la compañía. En particular, se espera que los modelos que consideran heterogeneidad no observable tengan una capacidad predictiva superior al del resto de modelos considerados.
- III. Comparaciones cualitativas y cuantitativas de la capacidad predictiva de los diversos modelos
- IV. Descripciones explícitas de cómo los resultados de los modelos pueden apoyar los procesos de ventas de la compañía.

2 MARCO CONCEPTUAL

La estimación de la propensión de compra es relevante pues permite a las empresas enfocar los recursos de oferta y venta en los productos que mejor correspondan a las necesidades de cada cliente. Este trabajo estima la dicha propensión mediante el cálculo de la probabilidad de compra a través de dos modelos independientes. Un primer modelo, asociado a la estimación de la probabilidad que tiene un cliente a cambiar de producto, y un segundo modelo que estima la probabilidad de que dado un cambio el cliente decida por cada una de las alternativas.

2.1 Estimación de Probabilidad de Compra

2.1.1 Modelo Logit Binario

Para estimar la propensión que tiene un cliente a cambiar de producto, se propone la utilización de un modelo Logit Binario. Dicho modelo corresponde a una alternativa para estimar la probabilidad de elección de un suceso {S} para un individuo {i} [4]. Esta probabilidad puede ser descrita mediante la siguiente ecuación.

$$P_i(S) = \frac{\exp(U_{Si})}{\exp(U_{Si}) + 1}$$

Donde U_{Si} representa la componente determinística de la utilidad que deriva un sujeto i al decidir el suceso S . Esta utilidad puede ser representada mediante la siguiente ecuación.

$$U_{Si} = X_i * \beta_i$$

Con X_i un vector que caracteriza al individuo i y el vector de parámetros β_i que representan la relación entre estas variables y la utilidad que obtiene el sujeto al elegir el suceso S .

2.1.2 Modelo Logit Anidado (Nested Logit)

Una vez que se ha decidido cambiar de producto, existen cerca de 200 variaciones entre las que puede elegir un cliente. El modelo Logit estándar considera

que los parámetros de sustitución entre un producto y otro son similares. En otras palabras, si se agrega (o elimina) una alternativa, la probabilidad estimada del resto disminuirá (aumentará) proporcionalmente [5]. Para los productos analizados en este trabajo, esto no es necesariamente cierto. En caso de remover una alternativa, por ejemplo Banda Ancha de 2mb de velocidad, la probabilidad de elegir algún plan de televisión no debiese verse afectada de la misma forma que la velocidad de 4mb.

Además, se supone que la decisión del producto que quiere el cliente no es necesariamente tan directa como se plantea en el modelo Logit estándar. Es decir, el cliente no decide el producto específico de una sola vez, sino que decide primero en un nivel más general para llegar luego a la decisión final. Por ejemplo, un cliente no decide inmediatamente si contratará un plan de 2mb de velocidad más televisión con adicionales de deporte, se supone, que este toma más bien una decisión secuencial, en la que decide primero tener solo banda ancha y televisión (excluyendo telefonía), luego decidirá un plan de televisión y finalmente si quiere o no algún canal adicional.

Otro problema del modelo Logit estándar es que su estimación puede ser compleja en términos numéricos frente a demasiadas alternativas, y por lo tanto, no es factible utilizarlo frente a más de 200 opciones.

Dadas las naturaleza del proceso de decisión más las limitaciones del modelo Logit estándar, se utilizará el modelo Logit Anidado (Nested Logit) para estimar la propensión que tiene un cliente a escoger cada una de las alternativas disponibles. Dicho modelo, separa las decisiones que realiza una persona en “nidos”. De esta forma, las elecciones que toma un individuo se consideran de forma secuencial. La siguiente ilustración esquematiza un modelo anidado.

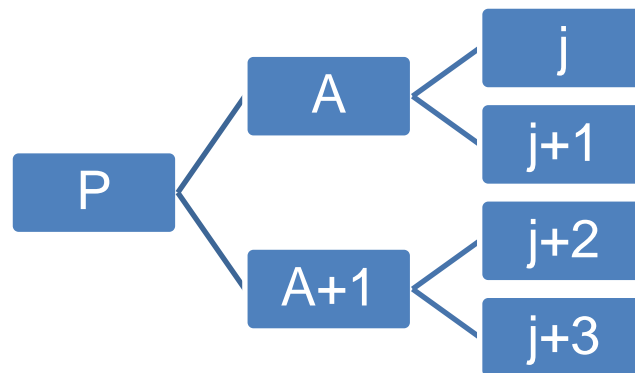


Ilustración 1: Esquematización de un modelo anidado

Fuente: Elaboración Propia.

Las decisiones finales condicionales a la elección del nido anterior en un modelo Logit Anidado pueden ser modeladas como un modelo Logit Estándar. De esta forma la probabilidad de elección de una alternativa j por un cliente i dada la elección de un nodo anterior A , dependen de la utilidad U_{ij} que deriva dicho cliente del producto j y del resto

de las alternativas que posee habiendo elegido A. Dicha probabilidad se calcula como se muestra en la siguiente ecuación.

$$P_i(j/A) = \frac{\exp(U_{ij})}{\sum_{l \in A} \exp(U_{il})}$$

Para los nodos interiores, la utilidad que representa para un cliente i elegir un nodo A queda representada en función de las utilidades que deriva de haber elegido dicho nodo. Dicha utilidad se denomina Valor Inclusivo del nodo. Este valor I_{iA} puede ser descrito de la siguiente forma.

$$I_{iA} = \ln\left(\sum_{l \in A} \exp(U_{il})\right)$$

La interpretación de este valor, está asociada a la máxima utilidad que puede obtener un individuo de los nodos sucesivos al nodo A. Es decir, la utilidad de un nodo interior está relacionada con la máxima utilidad que se puede derivar de los nodos posteriores habiendo escogido dicho nodo.

Luego, la probabilidad de elección del nodo A, dependerá de su valor Valor Inclusivo, en contraste al obtenido del resto de opciones con las que “compite” este nodo interior. En este sentido, la probabilidad de elección queda representada de la siguiente forma.

$$P_i(A) = \frac{\exp(I_{iA})}{\sum_{A \in P} \exp(I_{iA})}$$

Las componentes determinísticas de las utilidades U_{ij} asociadas a los nodos finales, serán representadas mediante ecuaciones lineales en función de las variables que mejor describan el comportamiento de compra de un cliente i . De esta forma la utilidad derivada de un producto j por un cliente i , será representada como muestra la siguiente ecuación.

$$U_{ij} = X_{ij} * \beta_i$$

El vector X_{ij} representa el valor que toman las variables explicativas de la utilidad obtenida tras elegir un producto j para el cliente i . Los parámetros del vector β_i representan la valoración relativa de dichas variables para ese individuo.

2.1.3 Heterogeneidad Observable

La probabilidad, tanto de cambiar de producto como de elección de una alternativa, puede variar entre un cliente y otro. En una primera instancia, dicha variación puede ser incluida en los modelos mediante la inclusión de heterogeneidad observable.

La no inclusión de heterogeneidad observable en la población, asume que los parámetros β_i son constantes e iguales para todos los individuos, es decir, $\beta_i = \beta \forall i$. Al incluir heterogeneidad observada, se asume que los parámetros β_i pueden variar entre un cliente y otro. Dicha variabilidad puede ser representada mediante la siguiente ecuación.

$$\beta_i = Y_i * \alpha$$

Donde Y_i representa las características de un cliente i y α representa la relación que existen entre el parámetro β_i y dichas características. Es importante notar que los parámetros α son fijos en este caso para todos los clientes.

2.1.4 Heterogeneidad no Observable (Clases Latentes)

Para capturar de mejor forma las variaciones entre un cliente y otro, es posible incluir heterogeneidad a un nivel no observable. En particular, en este trabajo se propone la utilización de clases latentes, que asumen la existencia de grupos al interior de la población, cuyo comportamiento difiere entre unos y otros [4].

La probabilidad asociada a cambiar de producto para el modelo Logit Binario puede ser escrita utilizando clases latentes de la siguiente forma.

$$P_i(S) = \sum_m^M S_m * \frac{\exp(U_{sim})}{\exp(U_{sim}) + 1}$$

Con M el número de clases latentes consideradas, S_m la probabilidad de pertenecer a la clase m y U_{sim} la utilidad que deriva un individuo de elegir el suceso S si perteneciera a la clase m . Para la probabilidad asociada a la elección de producto (Nested Logit), la probabilidad asociada a elegir un nodo final j , dada la elección de un nodo anterior A , puede ser escrita de la siguiente forma.

$$P_i(j/A) = \sum_m^M S_m * \frac{\exp(U_{ijm})}{\sum_{l \in A} \exp(U_{ilm})}$$

Donde U_{ijm} representa la utilidad que deriva un individuo i tras escoger un producto j si perteneciera a la clase m , los parámetros S_m y M representan la probabilidad de pertenecer a la clase m y el número de clases consideradas respectivamente.

2.2 Función de Verosimilitud

Al tener una muestra de datos $(x_1 \dots x_n)$ de una variable aleatoria X , cuya función de probabilidad es $f(X|\theta)$, es posible definir la función de verosimilitud como la

probabilidad de observar los datos que se poseen dada la función de distribución de la variable. En este sentido, la verosimilitud (L) puede ser escrita de la siguiente forma:

$$L = \prod_i^n f(x_i|\theta)$$

El estimador de máxima verosimilitud, es el parámetro θ estimado tal que maximiza la función de verosimilitud L [6].

Elementos útiles para la comparación de modelos son los estadísticos AIC (Akaike Information Criteria) y BIC (Bayesian Information Criteria) [4] que dependen de la verosimilitud L . A continuación se muestran las ecuaciones para calcular estos estadísticos.

$$AIC = -2 * \log(L) + 2 * k$$

$$BIC = -2 * \log(L) + k * \log(n)$$

Donde k representa el número de variables utilizadas y n el número de observaciones.

2.3 Metodología KDD.

Con el objetivo de ordenar los pasos metodológicos, se utilizará el esquema de procesamiento de datos KDD (Knowledge Discovery from Database). Dicha metodología consta de cinco pasos principales; Selección de datos, Pre-procesamiento de datos, Transformación de datos, Extracción de patrones y Evaluación [7].

La metodología KDD puede ser descrita mediante la siguiente ilustración.

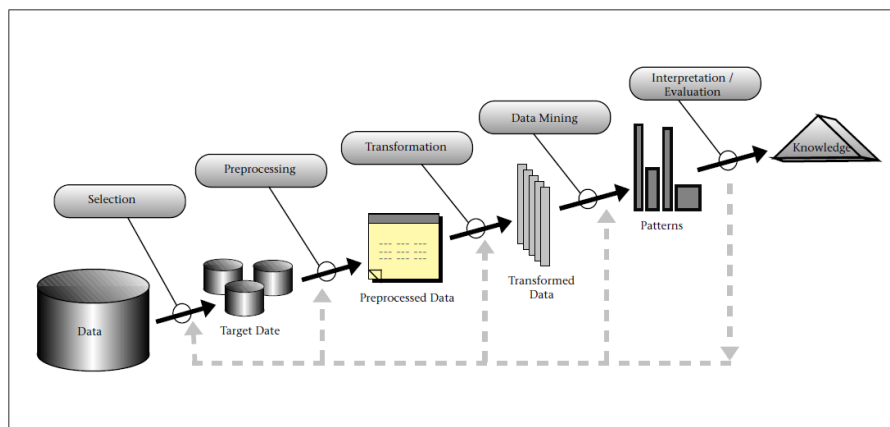


Ilustración 2: Metodología KDD

Fuente: Gregory Piatetsky-Shapiro, Padhraic Smyth Usama Fayyad [7]

3 METODOLOGÍA

Este trabajo, está estructurado bajo el marco metodológico KDD descrito en la sección 2.3. De esta forma se tendrán cinco etapas generales, selección de datos, pre-procesamiento de datos, transformación de datos, extracción de patrones y evaluación, además de un sexto paso correspondiente a las propuestas de acción posterior al análisis. A continuación se presenta una descripción de cada uno de estos pasos para este trabajo.

3.1 Selección de Datos

La selección de datos consiste en determinar con que datos se trabajará en el resto de la investigación. Esto se traduce en la selección del número de observaciones a utilizar, tanto para predicción como validación, las variables que se considerarán en los modelos y los periodos considerados.

3.2 Pre-Procesamiento de Datos

El Pre-Procesamiento de datos, será dividido en tres partes principales, la eliminación de datos irrelevantes, una descripción exploratoria de las variables a utilizar y el tratamiento de datos faltantes.

3.2.1 Eliminación de Observaciones

Tras la selección de los datos a utilizar en los modelos, se eliminarán del análisis aquellos datos que produzcan ruido o errores en las estimaciones. Esto se traduce en la eliminación de las observaciones cuyo valor en alguna de las variables relevantes sea muy superior o inferior a los valores corrientes (outliers), o la eliminación de casos particulares que no se consideran dado los alcances del proyecto.

3.2.2 Descripción de Variables

Para comprender las características y relaciones entre los datos, es necesaria una descripción cuantitativa de las observaciones y variables consideradas en la selección de datos. Para esto, se utilizarán análisis de frecuencias, correlaciones y tendencias presentes en los datos.

3.2.3 Datos Faltantes

El tratamiento de los valores faltantes (Missing Values) consiste en eliminar o estimar las observaciones en las que se carezca de datos en algunas de las variables analizadas. Dicho tratamiento dependerá de la cantidad de faltantes en cada variable y de la capacidad de predecir un valor estimado de los mismos. En cuyo caso serán removidos de la data, reemplazados por otro valor (valor promedio o estimación realizada) o considerados como un conjunto de datos aparte.

3.3 Transformación de Datos

3.3.1 Creación de Variables

Los modelos estadísticos propuestos, requieren la utilización de variables que se determinan directamente de los datos. Es decir, es necesaria la creación y/o transformación de las variables para poder realizar el modelamiento propuesto. Lo anterior se traduce en la creación de variables *dummy* y la normalización de variables.

3.4 Desarrollo de Modelos Económicos

La propensión de compra, será estimada mediante dos modelos independientes, un modelo Logit estándar para estimar la propensión a cambiar de producto y un modelo Logit Anidado, para estimar la probabilidad de elección de cada alternativa. Dichos modelos serán estimados con diferentes niveles de heterogeneidad, los cuales se describen a continuación.

3.4.1 Modelo sin Heterogeneidad

En primer lugar, se realizarán estimaciones sin incluir heterogeneidad, es decir, sin consideración de variaciones entre los individuos. Dicha estimación puede ser utilizada como una primera aproximación del comportamiento promedio de los clientes. En este contexto, esta primera versión de los modelos será estimada utilizando exclusivamente la información de los productos y las elecciones realizadas por los clientes.

3.4.2 Modelo con Heterogeneidad Observable

Para describir la propensión a cambiar de producto considerando heterogeneidad a un nivel observable, se desarrollará un modelo Logit donde las variables explicativas estén asociadas a las características del cliente.

Para describir la probabilidad de elección de los diferentes productos, se utiliza un modelo Logit Anidado en el que se incluyen los atributos de las diferentes alternativas, y en el que la sensibilidad asociada a estos dependerá de las características de los clientes.

3.4.3 Modelo con Heterogeneidad no Observable

Una vez incluida heterogeneidad observable en los modelos, se incluirá heterogeneidad no observable a través de la utilización de clases latentes. En concreto, se evaluará la mejor especificación, tanto para el modelo de cambio como de elección, y se incluirán clases latentes pudiendo mejorar el ajuste y capacidad predictiva de estos.

Los tres modelos serán desarrollados en el software R y los parámetros serán estimados maximizando la función de verosimilitud.

3.5 Evaluación

3.5.1 Capacidad de Ajuste

El desarrollo y evolución de las diversas especificaciones, estará ligada al análisis de la capacidad de ajuste de las mismas, además de la significatividad individual de los diversos parámetros. Dicha capacidad será evaluada a través de los estadísticos AIC, BIC y el valor del R cuadrado ajustado de McFadden.

Además, se realizarán evaluaciones cualitativas de la interpretabilidad de los modelos, y la capacidad de generar medidas de acción en base a los resultados de los mismos.

3.5.2 Capacidad Predictiva

Además de la capacidad de ajuste, es relevante analizar la capacidad predictiva de los modelos. Para esto se utilizará una base de datos de testeo, distinta a la utilizada para realizar el ajuste, pudiendo evaluar la capacidad que tiene el modelo de predecir el comportamiento de los clientes sobre los cuales no se ha realizado ningún análisis ni modelamiento.

Por otro lado, se generarán mediciones cuantitativas del aporte potencial que tienen los modelos en contraste con los actuales procedimientos que utiliza la compañía. Es decir, se generarán medidas explícitas para aplicar los modelos, y se evaluarán las diferencias con los actuales procesos de oferta.

3.6 Propuestas de Acción

Tras la evaluación y comparación de los resultados de los diversos modelos, se propondrán líneas de acción que integren dichos resultados a los procesos de venta de la compañía. En este sentido, se espera generar una interpretación adecuada de los resultados que permita a la empresa justificar sus acciones de oferta a los diferentes tipos de cliente.

4 DESARROLLO

4.1 Selección de Datos

Actualmente la empresa tiene algo más de 1.3 millones de clientes residenciales. Lo anterior, se traduce en la existencia de una base de datos con más de un millón de observaciones por mes, asociadas a cada cliente y el producto que tiene contratado en dicho periodo. Por motivos de simplicidad, todos los análisis desarrollados posteriormente, se realizan sobre una muestra aleatoria simple correspondiente al 20% de los datos originales, es decir, aproximadamente 260,000 clientes.

El análisis será desarrollado sobre los clientes activos en el periodo Agosto 2012, que hayan tenido un plan con la empresa en los 3 periodos anteriores, es decir, Julio, Junio y Mayo de 2012. En este sentido, los modelos serán desarrollados para estimar el comportamiento de Agosto, en función de la información que se posee del cliente en los 3 meses anteriores a este.

Es importante notar, sin embargo, que la empresa posee la información referente a un mes cualquiera, recién el día 15 de dicho periodo. Lo anterior implica la necesidad de generar las políticas de oferta con un mes de desfase en cuanto a la información disponible. En otras palabras, los procesos de oferta realizados en un periodo t , cuyo cambio (de ser exitoso) es generado en un periodo $t+1$, deben utilizar la información disponible a lo más de $t-1$, pues la información del periodo t no estará disponible hasta mediados de dicho periodo. En el proyecto, lo anterior se traduce en que las estimaciones utilizan información de Junio y Mayo para estimar la propensión de cambio que tiene el cliente desde el periodo Julio a Agosto.

Se utilizarán solo 3 meses de historia, dado que la mayoría de las variables consideradas no aporta información relevante en periodos más antiguos. En efecto, las variables asociadas a las características del cliente, por ejemplo edad o género, no varían de forma importante entre un periodo y otro. Por otro lado las variables asociadas al consumo, como el plan contratado en periodos anteriores, no aportan información relevante en periodos más antiguos de la que es posible obtener en los periodos más actuales.

Lo anterior es especialmente cierto para el modelo de elección, que intenta determinar la propensión que tiene el cliente a adquirir cada producto de la parrilla ofrecida por la empresa. Para el modelo que estima la propensión que tiene un cliente a cambiar de producto, sin embargo, puede ser relevante mirar cierta información de periodos más antiguos, como la última vez que el cliente ha cambiado de producto o que tan frecuente es dicho cambio. Dado lo anterior, en este modelo se considerarán algunas variables asociadas al comportamiento histórico del cliente en un periodo de 18 meses hacia el pasado, es decir, hasta enero del 2011.

Las variables consideradas, pueden ser divididas en 3 grupos principales, características del cliente, información asociada al comportamiento del cliente y productos contratados en cada periodo. A continuación, se detallan las variables utilizadas en este proyecto.

4.1.1 Características del Cliente

Las variables asociadas al cliente consideradas a priori en los diferentes modelos serán:

- I. **Comuna asociada a la cuenta:** Comuna en la cual se presta el servicio.
- II. **Segmentación del valor del cliente realizada por la compañía:** Segmentación en función de características inherentes del cliente (reconocido, político, entre

otras) y el valor promedio de transacciones con la compañía. Dicha segmentación se divide a los consumidores en tres grupos Alto Valor, Preferente o Tradicional.

III. **Grupo socioeconómico:** Segmentación de 5 Grupos socioeconómico del cliente, se dividen en ABC1, C2, C3, D y E.

IV. **Género**

V. **Edad del cliente**

4.1.2 Comportamiento del Cliente

Las variables asociadas al comportamiento consideradas a priori en los modelos serán:

- VI. **Gasto total por el servicio adquirido en los últimos 4 meses:** Gasto total por la utilización de todos los servicios residenciales ofrecidos por la compañía.
- VII. **Numero de meses en la compañía:** Numero de meses que el cliente ha estado asociado a la empresa.
- VIII. **Antigüedad de servicio de internet:** Numero de meses que el cliente ha tenido contratado el servicio de internet con la empresa.
- IX. **Antigüedad de servicio de Telefonía:** Numero de meses que el cliente ha tenido contratado el servicio de Telefonía con la empresa.
- X. **Antigüedad de servicio de televisión:** Numero de meses que el cliente ha tenido contratado el servicio de Televisión con la empresa.
- XI. **Reclamos:** Total de reclamos realizado por el cliente.
- XII. **Cambios de producto:** Si ha realizado o no un cambio de producto en cada periodo considerado.
- XIII. **Total de minutos salientes:** Número de minutos salientes realizados por el cliente.

4.1.3 Productos Contratados

El producto contratado en cada periodo puede ser definido a través de las siguientes variables:

- I. **Precio del Producto**
- II. **Plan de Telefonía:** Tipo de plan de Telefonía, de minutos (150, 250 o 350), ilimitado o prepago (de \$5,000 y \$6,000)
- III. **Plan de Internet:** Tipo de plan de internet, sus características y velocidad.
- IV. **Plan de Televisión:** Tipo de plan de televisión, canales incluidos, decodificadores, etc.
- V. **Número de decodificadores:** Número de decodificadores que posee el cliente.
- VI. **Canales adicionales de Televisión:** Canales adicionales en televisión, los cuales pueden ser de cine, deporte, para adultos, entre otros.

El presente trabajo, analiza los productos en base a sus atributos, es decir, a un conjunto de variables que determinan el atractivo de cada alternativa. De las variables

asociadas a los productos contratados es posible determinar 12 atributos que caracterizan cada producto los cuales se detallan a continuación.

4.1.4 Atributos de los Productos

- I. **Precio del Producto**
- II. **Velocidad de Internet**
- III. **Si posee plan de televisión Preferente:** Dummy si posee plan de televisión preferente.
- IV. **Si posee plan de televisión TOP:** Dummy si posee plan de televisión TOP.
- V. **Si posee decodificador HD:** Dummy de si posee o no decodificador de HD.
- VI. **Si posee canales adicionales de Deporte:** Dummy de si posee canales adicionales de deporte.
- VII. **Si posee canales adicionales de Cine:** Dummy de si posee canales adicionales de cine.
- VIII. **Si posee otros canales adicionales:** Dummy de si posee canales adicionales de otras categorías a deporte y cine.
- IX. **Si posee plan de telefonía ilimitada:** Dummy si posee plan de telefonía ilimitada.
- X. **Si posee plan de telefonía controlada de 350 min:** Dummy si posee plan de telefonía controlada de 350 minutos.
- XI. **Si posee plan de telefonía controlada de 150 o 250 min:** Dummy si posee plan de telefonía controlada de 150 o 250 minutos.
- XII. **Si posee plan de telefonía prepago:** Dummy si posee plan de telefonía prepago de \$5.000 o \$6.000.

Es importante notar que estos atributos determinan la unicidad de un producto, es decir, las combinaciones factibles (de oferta por parte de la empresa) entre los diferentes atributos, determinan un producto. En otras palabras, un producto estará determinado por una combinación única de los atributos mencionados anteriormente.

Dado que el objetivo de este trabajo es apoyar los actuales procesos de venta, en el análisis de modelos se consideraran solo las observaciones en las que se haya adquirido alguno de los productos que ofrece actualmente la compañía. Estos productos representan un 62.4% de la data considerada, se excluyen de esta forma productos antiguos que representan una pequeña parte de la data o productos exclusivos para localidades particulares.

Los atributos mencionados, asociadas a un producto, serán los que definen el conjunto de productos a considerar en los modelos, por otro lado, la heterogeneidad estará dada por las variables asociadas al cliente y a las observaciones individuales por consumidor.

Actualmente, algunas de las variables mencionadas anteriormente son analizadas con el fin de comprender mejor el comportamiento de compra de los consumidores. En efecto, los arboles de decisión desarrollados por la empresa permiten analizar variables

como la edad, el grupo socioeconómico o el género del cliente, dejando de lado otras variables asociadas a su ciclo de vida al interior de la compañía, como la antigüedad de los servicios, o de consumo, como el gasto realizado en periodos anteriores.

Dichos análisis consideran ciertas características de los productos como si tiene o no banda ancha, o posee o no algún plan de televisión, pero no considera atributos más desagregados que permiten analizar la propensión de compra de los productos ofrecidos por la empresa como la velocidad de conexión, el tipo de plan de telefonía, o los canales adicionales contratados.

4.2 Pre-Procesamiento de Datos

4.2.1 Eliminación de Observaciones

Una vez seleccionada la muestra, se han eliminado los datos que tengan valores muy distintos a los datos regulares y por tanto pueden producir ruido y afectar de forma negativa el modelamiento. Específicamente se han eliminado aquellas personas mayores de 80 y menores de 17 años, datos faltantes en la variable “antigüedad cliente”, valores de gasto total por servicio superiores a \$100.000 u observaciones contradictorias como la contratación de canales adicionales de televisión sin poseer un plan de televisión. El siguiente cuadro resume las eliminaciones de observaciones efectuadas.

Datos Removidos del análisis	
Datos iniciales	100.0%
Productos ofrecidos actualmente	62.4%
Datos Outliers	15.6%
Total datos considerados	46.8%

Tabla 1: Datos eliminados del análisis

Fuente: Elaboración Propia.

Como es posible observar en la tabla 1, del 100% de las observaciones iniciales, solo un 62% están asociadas a los productos considerados en este trabajo (productos principales ofrecidos actualmente), el restante 38%, corresponde a productos específicos para algunas localidades, o servicios antiguos que los clientes aún no han cambiado. Por otro lado, casi un 16% de las observaciones son consideradas datos outlier, y por lo tanto, se trabajará con un 46.8% de la base de datos original, es decir, aproximadamente 130.000 observaciones.

4.2.2 Descripción de Variables

Para poder observar la composición de los datos, se genera un análisis de frecuencias de las diferentes variables consideradas, dividiendo dicho análisis en las variables numéricas y nominales.

- **Variables numéricas**

La tabla 2, resume las características de las principales variables numéricas consideradas en el análisis, mostrando el promedio, desviación estándar, máximo y mínimo valor.

	Promedio	Desviación Estándar	Máximo Valor	Mínimo Valor
Velocidad de Banda Ancha (kb)	2972.22	2431.66	6144	0
Edad (años)	47.40	14.34	80	20
Gasto Total por Servicio (pesos)	20680	10290	100000	0
Antigüedad del cliente (meses)	129.35	119.32	530	1
Antigüedad del servicio de telefonía (meses)	88.56	103.81	530	0
Antigüedad del servicio de Internet (meses)	21.07	27.52	138	0
Antigüedad del servicio de Televisión (meses)	7.10	15.39	73	0
Minutos salientes	256.30	406.37	60772	0
Numero de decodificadores	.88	1.41	11	0
total reclamos realizados	.04	.26	8	0
recencia de cambio (tiempo desde el último cambio) (meses)	10.01	6.23	17	0
Frecuencia de cambio (total cambios sobre el número de meses activo)	.07	.13	1	0

Tabla 2: Descripción variables numéricas

Fuente: Elaboración Propia.

Para la velocidad de banda ancha, los valores oscilan entre 0 y 6144, dado que 6144 kb (6 mb) es la máxima velocidad considerada en este trabajo y 0 significa no poseer internet. El gasto total por servicio oscila entre 0 y 100.000 pesos, dadas las restricciones impuestas en la eliminación de outliers.

En cuanto a la antigüedad del cliente, es posible observar que esta oscila entre 1 y 530 meses, es decir, entre 0 y 44 años. Dicha variable puede poseer una fuerte correlación con la variable edad, y en efecto, el coeficiente de correlación de Spearman, que oscila entre -1 y 1 tiene un valor 0.54. Dicha correlación, no es relevante al

comparar con las variables de antigüedad de los servicios de televisión e internet, aunque si lo es con la antigüedad de telefonía, cuyo coeficiente tiene un valor de 0.42.

Las variables recencia de cambio, asociada al número de meses transcurridos desde que el cliente ha realizado un último cambio, oscila entre 0 y 17 meses. Cero en este caso, significa que el cliente no ha realizado un cambio en los últimos 17 meses (considerando un total de historia de 18 meses hacia el pasado), mientras que 0, significa que el último cambio fue realizado justamente en el periodo anterior. La frecuencia toma valores entre 0 y 1. Cero en este caso representa un cliente que no ha cambiado en los últimos 18 periodos, mientras que un valor de 1 significa que el cliente ha cambiado en todos los periodos desde que está activo.

- **Variables nominales**

Las variables nominales consideradas son básicamente cuatro. La comuna de residencia, el género, el grupo socioeconómico y la segmentación preliminar del cliente.

Existen un total de 291 comunas distintas en la base de datos distribuidas en 12 regiones del país, 5 grupos socioeconómicos, ABC1, C2, C3, D y E, el género masculino o femenino y 3 segmentos asociados a la segmentación preliminar, los cuales son clientes de “alto valor”, “preferentes” y “tradicionales”.

Los clientes de movistar están concentrados en los grupos socioeconómicos C2, C3 y D, representando un 87.3% del total de clientes. En cuanto al género, la base de datos está compuesta por un porcentaje algo mayor de mujeres que de hombres, 49% y 42% respectivamente. En cuanto a la segmentación preliminar, se observa que la gran mayoría de clientes corresponden al grupo tradicional, representado por más de un 84% de la data.

En cuanto a los productos hogar, estos pueden ser definidos a través de 3 categorías principales, banda ancha, televisión y telefonía. Los productos banda ancha se diferencian unos de otros básicamente por la velocidad de conexión. Los servicios de televisión pueden ser o bien “Preferente” que corresponde a un plan de 80 canales o “Top” que corresponde a 127 canales incluidos de alta definición, además es posible contratar canales adicionales de deporte, cine u otras categorías en alta o baja definición. En telefonía existen planes ilimitados, controlados de 350, 250 o 150 minutos o prepagos de \$5,000 o \$6,000.

Un cliente puede decidir una combinación de estos atributos, que sujeta a restricciones de la empresa, definen la parrilla de productos de la compañía. A continuación se muestra un análisis exploratorio de las decisiones realizadas por los clientes en cada una de las categorías mencionadas y la relación de estas con las variables numéricas y nominales descritas con anterioridad.

- **Banda Ancha**

Aproximadamente un 69% de los clientes analizados tiene contratado un plan que incluye banda ancha. La decisión, tanto de poseer un servicio de internet, como la velocidad del mismo, depende de múltiples factores. En efecto, es posible observar que la velocidad de banda ancha contratada tiene una tendencia a disminuir en función de la edad del cliente. El gráfico 3 muestra dicha relación.

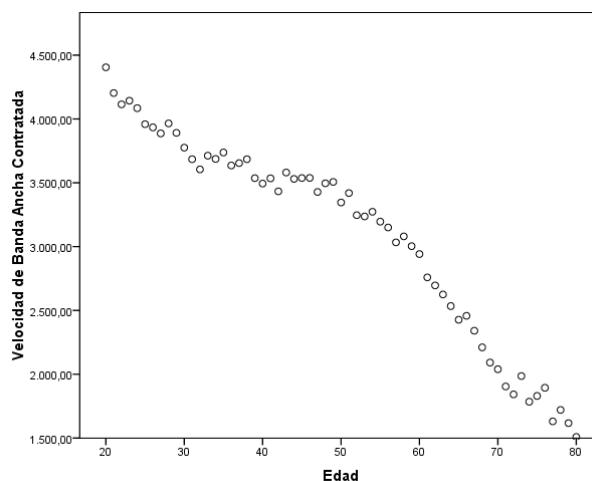


Gráfico 3: Velocidad de banda ancha promedio según edad

Fuente: Elaboración Propia.

Una segunda variable que se relaciona con la velocidad de banda ancha contratada, tiene relación con el plan de internet que tiene el cliente en el periodo anterior. En efecto, un 99% de los clientes no cambia la velocidad contratada de un periodo a otro. Sin embargo, en los clientes que si cambian, es posible observar una tendencia a aumentar la velocidad del plan actual. La tabla 3 muestra la relación entre el plan contratado en un periodo anterior t-1 y el plan elegido (dado que el cliente cambia la velocidad) en un periodo t.

		Velocidad en Periodo t				TOTAL CAMBIOS
		SIN BA	2 MB	4 MB	6 MB	
Velocidad en Periodo t-1	SIN BA	0.0%	19.4%	31.3%	49.3%	22.6%
	2 MB	8.7%	0.0%	64.4%	26.9%	23.4%
	4 MB	7.5%	22.3%	0.0%	70.2%	38.9%
	6 MB	20.7%	20.0%	59.3%	0.0%	15.2%

Tabla 3: Cambio de velocidad contratada de un periodo a otro

Fuente: Elaboración Propia.

En la tabla anterior, es posible observar que los clientes que poseían un plan de 2 MB y deciden realizar un cambio de velocidad, tienden a aumentar a un plan de 4MB en un 64% de los casos. Lo mismo ocurre para los clientes que poseían una velocidad de 4 MB contratada, donde en el 70% de los casos aumenta a 6 MB. Lo anterior sugiere que los clientes tienden a aumentar la rapidez de conexión contratando el “siguiente” plan más rápido disponible, del que poseen actualmente.

La cuarta columna de la tabla, muestra el total de cambios asociados al plan del periodo t-1, como porcentaje del total de cambios de velocidad realizados en ese periodo. Es relevante notar, que solo un 15% del total de cambios realizados está asociado a los clientes que poseen un plan de 6 MB. Este hecho refleja una suerte de estancamiento en el cambio de velocidad cuando se ha alcanzado un plan de una rapidez mayor.

En Agosto del 2012, el precio de banda ancha de 2 megas era de \$14.490, el precio de 4 megas de \$18.990 y el de 6 megas de \$19.990. Como puede apreciarse, aunque el precio aumenta conforme la velocidad, cada vez lo hace en menor medida. Lo anterior, puede ser interpretado como el reconocimiento por parte de la empresa de que la utilidad marginal que reportan planes de mayor velocidad es decreciente, este aspecto debe ser considerado en los modelos y será realizado mediante la inclusión del logaritmo de la velocidad como variable explicativa.

- **Televisión**

Aproximadamente un 39% de los clientes de la empresa posee un plan que incluye televisión. Dichos planes pueden ser de dos tipos, “preferente” o “top”. El plan “preferente” es el plan más básico disponible y posee 80 canales. Por su parte, el plan “top” posee 127 canales incluyendo de alta definición.

Un 97% de los clientes que deciden tener un plan de televisión optan por el plan “preferente”, dicho plan tiene un costo de \$10.000 menos que el plan “top”, por lo que es probable que dicho valor determine en gran medida las diferencias de elección. Además de los planes mencionados, es posible contratar paquetes adicionales de canales, los cuales pueden dividirse en 5 tipos, canales de cine, deporte, otros, cine alta definición y deporte alta definición. El porcentaje de cliente que contrata cada una de estas categorías se muestra en el gráfico 4.

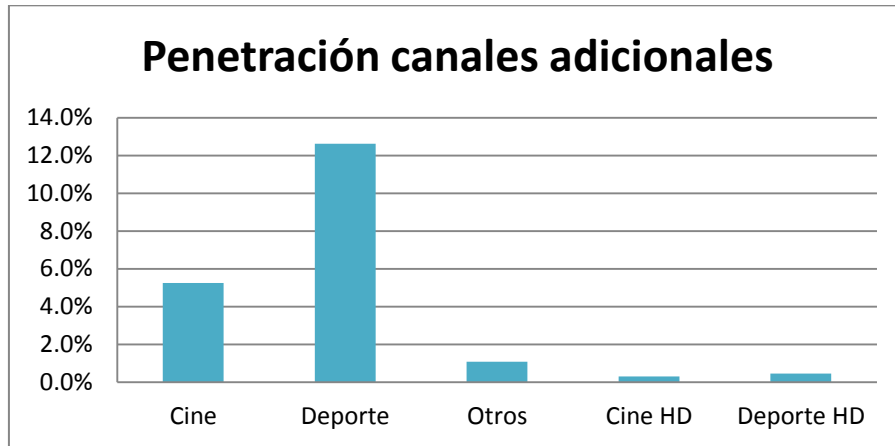


Gráfico 4: Penetración por categoría de canales adicionales

Fuente: Elaboración Propia.

En el gráfico anterior es posible observar que los clientes suelen contratar más canales de deporte que de cine, tanto en alta como en baja definición. Lo anterior, aun considerando que los canales de deporte tienen un precio superior que los adicionales de cine.

Dadas las fuertes diferencias de precio y la naturaleza de cada uno de los planes de televisión, se espera que personas con un mayor poder adquisitivo tiendan a adquirir más frecuentemente un plan "top" que un plan preferente. En efecto, el grupo socioeconómico asociado al cliente determina en gran medida el plan de televisión que el cliente escoge. El gráfico 5 muestra dicha relación, mostrando el porcentaje de elecciones de los diferentes planes en función del GSE del cliente.

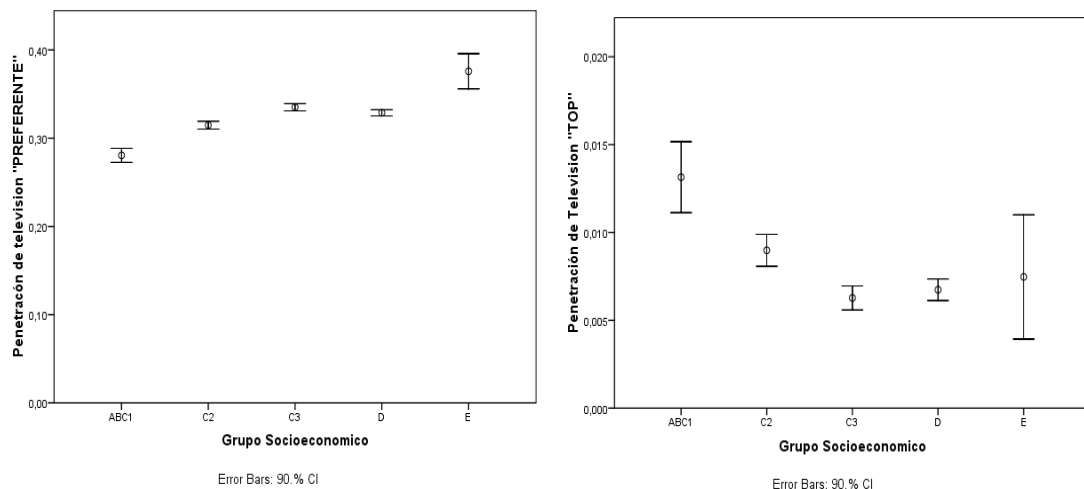


Gráfico 5: Plan de televisión contratado según grupo socioeconómico

Fuente: Elaboración Propia.

Como es posible observar en el gráfico anterior, los clientes pertenecientes al grupo socioeconómico ABC1, tienden a adquirir con mayor frecuencia un plan “top”, en contraste con los otros grupos. Dicha relación es contraria a lo que ocurre con los planes de televisión “preferente”, donde se puede observar que los clientes ABC1 son los que poseen la mínima penetración.

En cuanto a los canales adicionales, es posible observar que la antigüedad del servicio de televisión, es decir, el tiempo que el cliente ha tenido contratado un plan de televisión está relacionado positivamente con la contratación de canales adicionales. El gráfico 6 muestra esta relación para los canales adicionales de deporte.

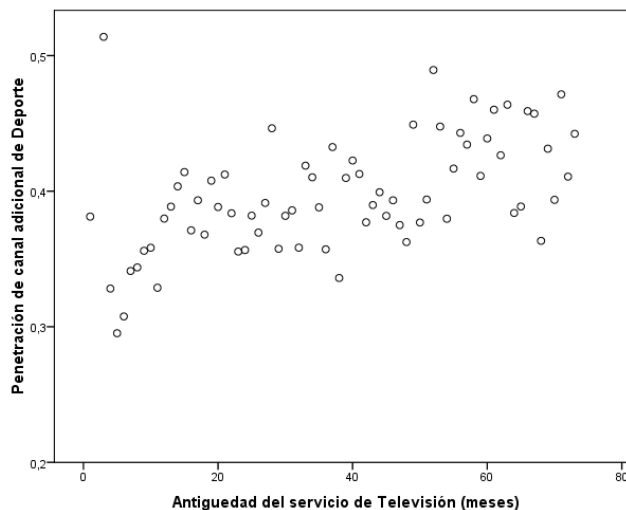


Gráfico 6: Penetración canal adicional de deporte según antigüedad de servicio de televisión

Fuente: Elaboración Propia.

En el gráfico anterior, es posible observar que a medida que aumenta el tiempo que un cliente posee un plan de televisión contratado, existe una tendencia a aumentar la proporción de clientes que posee un canal adicional de deporte. Lo anterior puede responder a una suerte de acostumbramiento por parte de las personas a los canales contratados, derivando en la necesidad de adquirir adicionales.

- **Telefonía**

Casi un 87% de la población analizada posee algún plan de telefonía contratado, dichos planes pueden dividirse en 6 tipos, plan ilimitado, plan controlado de 350, 250 y 150 minutos y planes pre-pagados de \$5.000 y \$6.000. En el siguiente gráfico se muestran las diferencias de contratación de los diferentes planes.

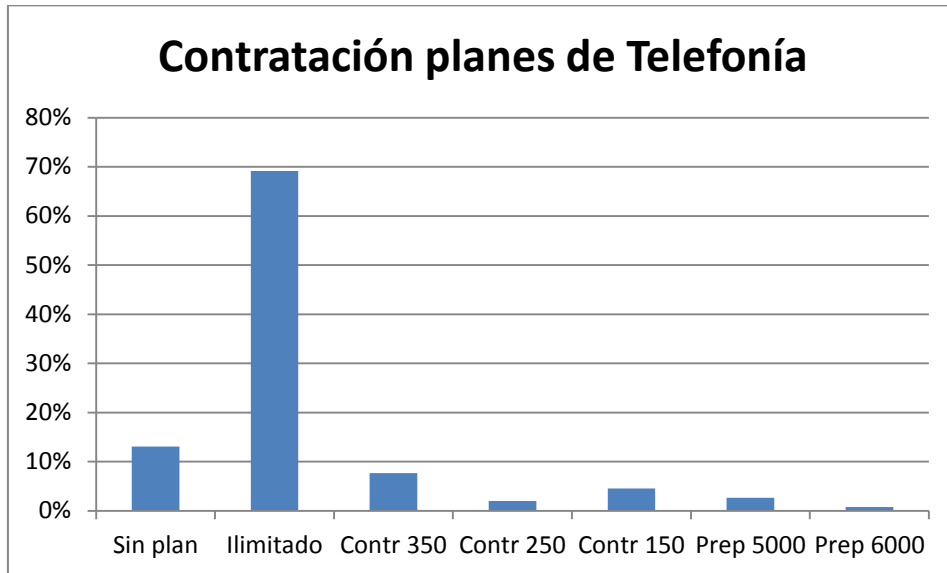


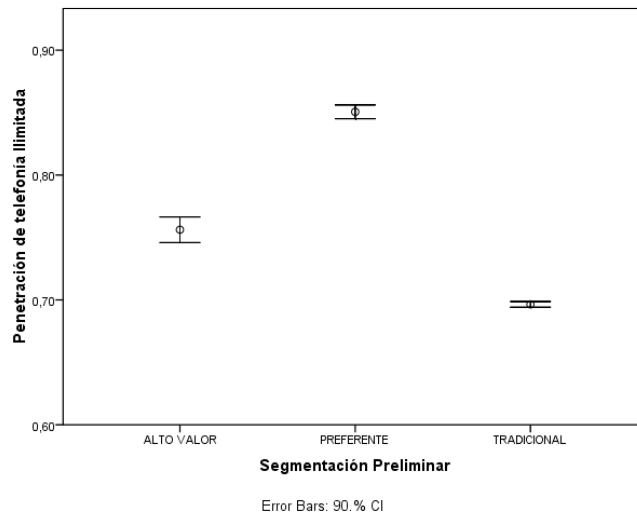
Gráfico 7: Planes de telefonía

Fuente: Elaboración Propia.

El plan ilimitado es sin lugar a dudas el preferido por los clientes. Por su parte, los planes controlados de 250 minutos y prepago de \$6.000 parecieran no ser muy atractivos para los usuarios, en contraste con los planes de 150 minutos y de \$5.000.

Para este trabajo se opta por considerar como similares los planes de 250 y 150 minutos dado que no difieren demasiado en sus características, y dado que el número de elecciones es relativamente bajo. Lo mismo ocurre con los planes prepagos de \$5.000 y \$6.000. Por lo tanto, se agruparan en un solo grupo los planes controlados de 150 y 250 minutos, y en otro, los planes de \$5.000 y \$6.000.

Una segmentación preliminar realizada por la empresa, que divide a los clientes en tres tipos, alto valor, preferente y tradicional, puede tener relación con la elección del plan de telefonía que realizan los clientes. El siguiente grafico muestra las diferencias entre cada tipo de cliente y la contratación de planes de telefonía ilimitada.



Fuente: Elaboración Propia.

Gráfico 8: Penetración canal de telefonía ilimitado según segmentación preliminar

En efecto, es posible observar que los clientes “preferentes” tienen una tendencia mayor a poseer telefonía ilimitada, en contraste con los otros dos grupos de clientes. Otra variable que puede ser relevante para explicar la elección de plan de telefonía es el número de minutos salientes. Es de esperarse que en tanto un cliente llama más desde su teléfono, prefiera un plan con más minutos o simplemente ilimitado. El siguiente gráfico, muestra la relación entre la penetración de un plan de telefonía ilimitado, en función del número de minutos salientes.

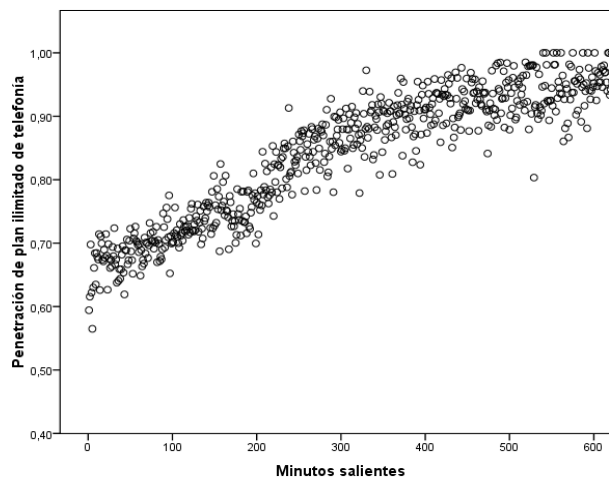


Gráfico 9: Penetración plan ilimitado de telefonía según minutos salientes

Fuente: Elaboración Propia.

Tal y como se espera, a medida que las personas realizan más llamadas desde su teléfono, la proporción de gente que posee un plan ilimitado es mayor. Lo contrario debiese ocurrir con un plan controlado o prepago, ya que estos tienen una cantidad limitada de minutos, debiendo pagar una suma de dinero extra tras sobrepasar el límite establecido. El gráfico 10 corrobora dicha hipótesis, mostrando la relación entre los minutos salientes y un plan controlado de 150 o 250 minutos.

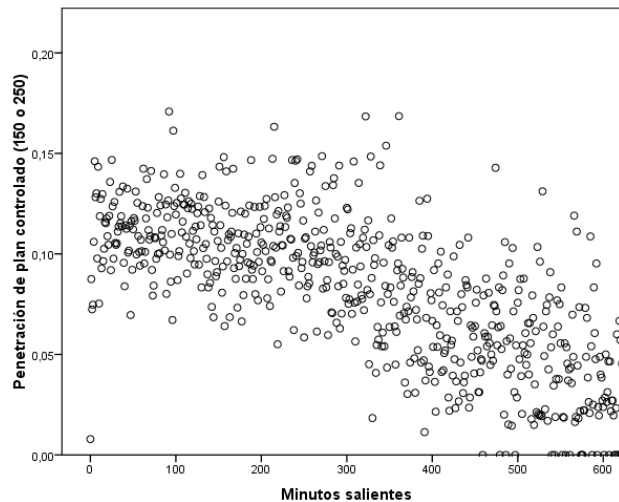


Gráfico 10: Penetración plan controlado de 150 o 250 minutos de telefonía según minutos salientes

Fuente: Elaboración Propia.

El presente trabajo, no solo pretende estimar la propensión de los diferentes clientes a adquirir los diferentes productos, sino además, la probabilidad asociada a cambiar el servicio actual, es decir, a contratar un producto distinto al que se posee en un periodo. Dado esto, puede ser relevante explorar la relación entre las diferentes variables, y la proporción de clientes que realiza un cambio. A continuación, se presenta un análisis exploratorio relacionado con la propensión a cambiar de los diferentes clientes.

- **Propensión de cambio**

Una primera variable relacionada con la propensión de cambio es la edad del cliente, al parecer, clientes más jóvenes tienden a ser más propensos a realizar cambios que clientes de mayor edad. La relación entre estas variables se muestra en la siguiente gráfica.

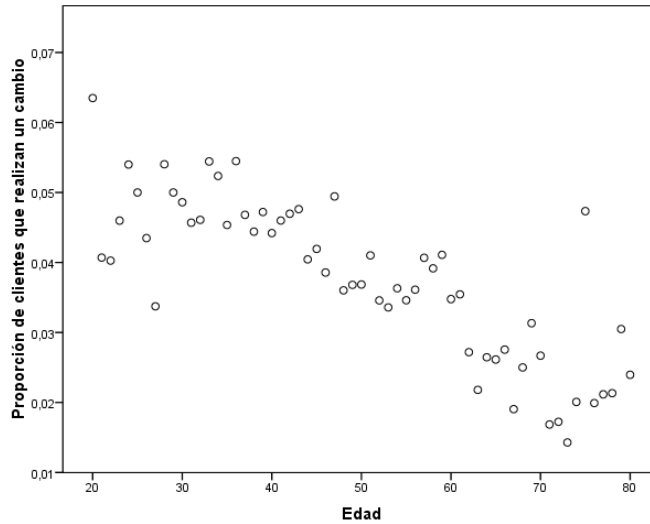


Gráfico 11: Porcentaje de cambios según edad

Fuente: Elaboración Propia.

Como es posible observar en la gráfica anterior, los clientes más jóvenes tienen un promedio de cambio cercano a un 5%, en contraste con los clientes de mayor edad, cuyo promedio bordea el 2%. Lo anterior puede significar que los clientes más jóvenes son más propensos a cambiar de plan, mientras que los de mayor edad tienden a permanecer con el mismo servicio por más tiempo.

Otras variables interesantes de observar, tienen relación con la historia de cambio del cliente. Asociado a lo anterior, una variable que puede ser relevante es la frecuencia con la que el cliente ha cambiado de productos en el pasado. Es de esperar, que un cliente que ha cambiado en un mayor número de ocasiones, sea más propenso a cambiar en la actualidad. El gráfico número 12, muestra la relación entre la frecuencia de cambio, medida como el total de cambios dividido el total de periodos en que el cliente estuvo activo en el total de periodos considerados, y la proporción de clientes que ha realizado un cambio en el periodo en el que se realiza la estimación.

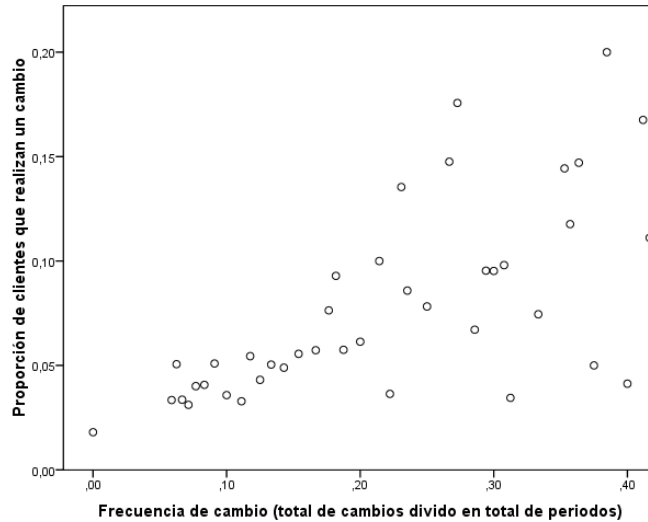


Gráfico 12: Porcentaje de cambios según frecuencia de cambio

Fuente: Elaboración Propia.

En el gráfico anterior es posible observar que en efecto, existe una tendencia de los clientes que han cambiado con mayor frecuencia en el pasado, a realizar un mayor número de cambios en el presente. Otras variables que pueden ser relevantes para estimar la propensión de cambio, tienen relación con el grupo socioeconómico de los clientes. Al parecer, los clientes ABC1, poseen una tendencia a realizar menos cambios, en contraste con los clientes pertenecientes al grupo C3 o D. Dicha observación, se muestra gráficamente en el gráfico 13.

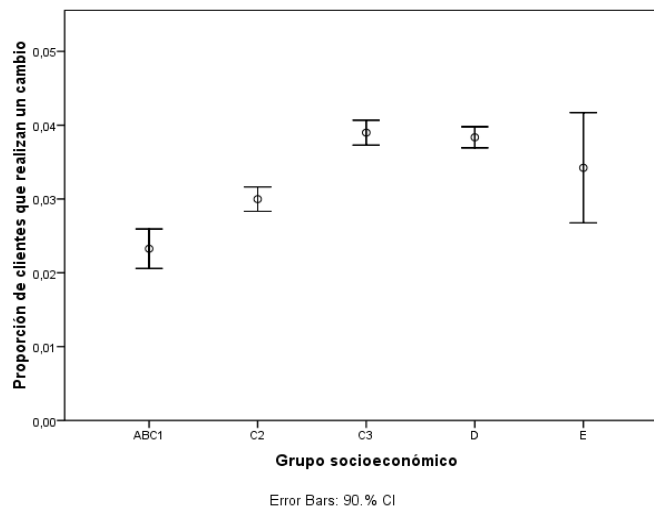


Gráfico 13: Porcentaje de cambios según grupo socioeconómico

Fuente: Elaboración Propia.

4.2.3 Datos Faltantes

El análisis descriptivo desarrollado, corresponde a un primer acercamiento para comprender los datos disponibles y la relación entre las diferentes variables. Habiendo desarrollado este punto, es posible generar un análisis de los datos faltantes y determinar la forma de proceder en cada uno de estos casos. En la tabla 4 se presenta una descripción del número de observaciones en las cuales se carece de información para cada una de las variables consideradas.

	Datos Válidos	Datos Faltantes
Comuna	99.1%	0.9%
Segmentación preliminar	97.0%	3.0%
GSE	95.1%	4.9%
Género	91.4%	8.6%
Edad	56.5%	43.5%

Tabla 4: Datos faltantes

Fuente: Elaboración Propia.

Tras analizar los resultados se observa que para la mayoría de las variables, los valores faltantes representan un pequeño porcentaje del total de datos. En particular, las variables “Comuna”, “Segmentación Preliminar” y “GSE”, poseen menos de un 5% de faltantes en el total de datos. Para la variable “Genero”, dicho porcentaje es algo mayor aunque aún no supera el 9% de los datos.

Considerando lo anterior y dado que las variables son nominales, es decir, cuyos valores no son numéricos sino que representan una característica particular. Se generará un nuevo grupo de clientes asociado a aquellos en los que no se posee el valor de la observación. En otras palabras, se generará un nuevo grupo asociado a los clientes que poseen un valor nulo para cada una de las variables. Lo anterior, es relevante pues es posible generar modelos escalables a la mayoría de los clientes de la empresa, y los problemas asociados a la capacidad predictiva de los mismos no debiesen ser demasiado importantes dada la baja cantidad de casos en los que no se poseen datos.

En cuanto a la variable “Edad”, en este caso se posee un 43% de datos faltantes. Una posibilidad es eliminar aquellos datos en los que no se tenga el valor, lo que significaría reducir la escalabilidad de los modelos a prácticamente la mitad de los clientes. Otra posibilidad es estimar los datos faltantes a través de un modelo de imputación de datos, el problema de esta alternativa es que se utilizan variables que se utilizarán luego en los modelos, quedando la edad del cliente altamente correlacionada con el resto de las variables, además, las estimaciones realizadas no presentan una

buena capacidad predictiva. Una tercera opción, es nominalizar la variable edad, es decir, generar grupos etarios con un grupo extra asociado a los datos faltantes.

Finalmente, se opta por esta última opción, dado que es posible incluir la variable edad sin necesidad de eliminar datos, y es posible utilizar la información real disponible para ajustar los modelos. En este contexto se generan 4 grupos etarios asociados a los clientes que poseen entre 20 y 40, 40 y 60 y más de 60 años, además de un cuarto grupo asociado a los clientes en los que no se posee la información.

4.3 Transformación de Datos

4.3.1 Creación de Variables

Antes de desarrollar los modelos propuestos, es necesario transformar los datos de forma de poder utilizar las diferentes variables. Para esto se han creado las siguientes variables

- I. **Variable “Choice”:** 210 variables Dummy correspondiente a cada uno de los productos que puede elegir un cliente. Dicha variable tiene un uno en caso de que el cliente haya escogido ese producto y cero en caso contrario.
- II. **Variable “Cambio”:** Variable que contiene un 1 si el cliente ha realizado un cambio de producto en dicho periodo, cero en caso contrario.
- III. **Variable “Precio”:** 210 variables numéricas que contienen el precio de cada una de las alternativas en cada mes. Existe una variable “precio” por cada variable “choice” creada.
- IV. **Variable “Velocidad”:** 210 variables numéricas que contienen la velocidad de internet de cada una de las alternativas en cada mes. Existe una variable “velocidad” por cada variable “choice” creada.
- V. **Atributos de los productos:** Variables dummy que contienen un 1 si el producto elegido tiene esa característica, por ejemplo, “dummy internet”, tiene un 1 si el producto posee internet, cero en caso contrario. Lo mismo ocurre para las variables, “tv preferente”, “tv top”, “telefonía ilimitada”, “telefonía de 350 minutos”, entre otras.
- VI. **Variables binarias:** Transformación de variables nominales a variables binarias, esto es por ejemplo la transformación de la variable “GSE”, a 5 variables Dummy correspondientes a cada grupo. Lo mismo para la variable “Segmentación Preliminar o “Género”.
- VII. **Normalización de variables:** Por simplicidad numérica, se normalizarán las variables “precio” y “gasto en periodos anteriores”, la primera dividiéndola por 1.000 y la segunda por 10.000.

4.4 Desarrollo de Modelos Económicos

Una vez realizada la selección, pre-procesamiento y la transformación de datos, se desarrollan los modelos predictivos propuestos. Lo anterior significa el desarrollo de dos modelos independientes, un primer modelo para estimar la propensión que tiene un

cliente a cambiar de producto, y un segundo, para estimar la propensión que tiene el cliente a elegir cada alternativa disponible, dado que realizará un cambio.

Dicho esto, para estimar la propensión que tiene un cliente a adquirir un producto cualquiera, los modelos descritos deben ser utilizados de forma sucesiva, pudiendo estimar en primer lugar la propensión que tiene un cliente a migrar de producto y luego, dado que realizará un cambio, se estima cual es la alternativa que mejor calza dado su perfil particular.

De esta forma, es posible responder de forma sucesiva a las preguntas relevantes asociadas al proceso de oferta, es decir, a quién debe ir dirigida la oferta, cuya respuesta estará relacionada a la propensión que tiene un cliente a migrar de producto, y luego de cuál debe ser la oferta, asociada a la preferencia o propensión que tiene un cliente por cada una de las alternativas.

Para la calibración del modelo de cambio, se utilizara una muestra aleatoria simple de 100.000 clientes activos en los periodos Agosto, Julio, Junio y Mayo de 2012. Para el modelo de elección, es decir, de la propensión a adquirir los diferentes productos, se utilizará una muestra aleatoria simple de 5,000 clientes activos en los mismos periodos, que hayan realizado un cambio en el último mes. Las diferencias en el tamaño de la muestra hacen relación con la naturaleza de cada uno de los modelos, tanto por la velocidad de procesamiento como por las características intrínsecas de las estimaciones.

La elección de 4 periodos se justifica dado el desfase de un mes necesario para generar las políticas de oferta y dado que los modelos utilizan información de dos periodos de actividad del cliente, es decir, dado que para estimar el comportamiento de Agosto, es necesario utilizar los datos disponibles de dos meses hacia atrás, vale decir, Junio del mismo año, dadas las restricciones de la empresa, y además los modelos utilizan información de la actividad del cliente en el periodo anterior a este (es decir Mayo).

Por simplicidad, la descripción del desarrollo será separada en dos partes, una asociada al modelo de cambio, y otra relacionada con el modelo de elección. A continuación se presenta el desarrollo del modelo Logit Binario, cuyo objetivo es estimar la probabilidad de migración de producto de los clientes.

- **Modelo de Propensión de Cambio**

La estimación de la propensión de cambio será desarrollada a través de un modelo Logit binario. La variable dependiente de dicho modelo corresponderá a un 1 si el cliente ha realizado un cambio, 0 en caso contrario. Se han diferenciado los modelos en tres tipos, modelo sin heterogeneidad, con heterogeneidad observable y con heterogeneidad no observable, a continuación se muestran los resultados de cada uno de ellos.

4.4.1 Modelo de Cambio sin Heterogeneidad

El modelo sin heterogeneidad no incluye diferencias entre los distintos perfiles de clientes. En el modelo de cambio, esto se traduce en la utilización de una única constante para estimar los modelos. La siguiente tabla resume los resultados obtenidos para este modelo.

Variable	Valor	Desviación Estándar	P-valor
Constante	-3.367	0.176	0.00

Tabla 5: Resultados modelo de cambio sin heterogeneidad

Fuente: Elaboración Propia.

La única variable utilizada en este modelo es una constante cuyo valor es de -3,367. Lo anterior significa que los clientes otorgan una valoración negativa a cambiar, en contraste con una valoración nula a no realizar ningún cambio. El modelo deriva en una probabilidad de cambio homogénea para todos los clientes de 3.33%, es decir, los clientes tienen una probabilidad promedio a cambiar de 3.33%.

Aunque los resultados descritos pueden parecer simples y poco relevantes, es importante notar que la capacidad descrita a través del valor del pseudo R cuadrado de Mc Fadden, asciende a 78.92%. Lo anterior ocurre porque la proporción de clientes que cambia cada periodo es bastante baja, y por lo tanto, un modelo básico que no dijera absolutamente nada, es decir, que todos los parámetros fueran iguales a cero, no aporta ninguna información verosímil. Luego, para evaluar la capacidad de ajuste de los modelos que se desarrollan a continuación, será relevante compararlos con este primer modelo, que entrega una noción al menos básica de la realidad, traducida en que los clientes son más propensos a no cambiar que a hacerlo.

4.4.2 Modelo de Cambio con Heterogeneidad Observable

Un segundo modelo desarrollado incluye heterogeneidad a un nivel observable, traducida en variaciones entre un cliente y otro derivadas directamente de los datos. La complejidad en este punto radica en la identificación del conjunto de variables que mejor describan el comportamiento de los clientes.

Para determinar la mejor especificación que describa la propensión de cambio de los clientes se han desarrollado diversos modelos, y se ha discriminado utilizando los criterios de Akaike, Bayes y Pseudo R cuadrado ajustado de Mc Fadden, además de las desviaciones estándar de cada variable para determinar la significancia individual de las mismas. La tabla 6 resume las diversas especificaciones desarrolladas con los respectivos estadísticos de comparación.

Especificación	Número de variables	R cuadrado Ajustado ¹	AIC	BIC
Modelo s/heterogen.	1	0.00%	29228	29237.51
Modelo c/heterogen 1	12	1.04%	28922	29036.16
Modelo c/heterogen 2	17	8.02%	26882	27043.72
Modelo c/heterogen 3	19	8.22%	26824	27004.75
Modelo c/heterogen 4	24	10.31%	26214	26442.31
Modelo c/heterogen 5	27	11.34%	25912	26168.85
Modelo c/heterogen 6	33	11.52%	25860	26173.93
Modelo c/heterogen 7	26	11.53%	25856	26103.34

Tabla 6: Comparación modelos de cambio con heterogeneidad observable

Fuente: Elaboración Propia.

En la tabla anterior, es posible observar que el mínimo valor para el criterio de Akaike (AIC), está asociado al modelo número 7, con 26 variables explicativas. Dicho modelo posee además el mayor valor del R cuadrado ajustado (en contraste con el modelo sin heterogeneidad) y uno de los menores valores para el criterio de Bayes.

Lo anterior determina la elección del modelo número 7, como la mejor especificación posible para describir el comportamiento de cambio de los clientes. En la siguiente tabla, se muestran los valores de los diferentes parámetros para este modelo, junto a sus desviaciones estándar.

Variable	Valor	Desviación Estándar	P-valor
Constante	- 3.8622	0.0872	0.00
Edad entre 20 y 40	0.1580	0.0491	0.00
Edad entre 40 y 60	0.0971	0.0436	0.03
GSE ABC1	- 0.1919	0.0928	0.04
GSE C3	0.1851	0.0497	0.00
GSE D	0.1724	0.0463	0.00
Antigüedad Telefonía	- 0.0007	0.0003	0.01
Antigüedad Televisión	- 0.0027	0.0012	0.03
Antigüedad Telefonía / antigüedad cliente	0.1727	0.0667	0.01
Antigüedad Televisión / antigüedad cliente	- 0.1518	0.0726	0.04
Numero de decodificadores	0.1409	0.0267	0.00
Telefonía prepago antes	- 0.5263	0.2227	0.02
Telefonía ilimitada antes * minutos salientes	0.0001	0.0000	0.01

¹ R cuadrado ajustado en contraste con un modelo base asociado a la especificación que incluye una sola variable explicativa

Adicional de deporte antes	1.2735	0.1044	0.00
Adicional de cine antes	0.4374	0.0536	0.00
Adicional de deporte antes * número de decodificadores	- 0.2076	0.0351	0.00
Frecuencia de cambio	1.8786	0.1989	0.00
Recencia último cambio	- 0.0290	0.0046	0.00
Recencia último cambio * frecuencia de cambio	- 0.2537	0.0445	0.00
Velocidad de 4MB antes	0.0643	0.0404	0.11
Plan trio antes	0.4392	0.0804	0.00
Duo banda ancha-televisión antes	0.5054	0.1429	0.00
Duo telefonía-televisión antes	0.5229	0.1357	0.00
Solo Banda Ancha antes	- 0.7436	0.1373	0.00
Solo telefonía antes	- 0.4540	0.0766	0.00
Diferencia gasto periodos t-1 y t-2	0.0734	0.0314	0.02

Tabla 7: Resumen resultados modelo cambio con heterogeneidad observable

Fuente: Elaboración Propia.

Es posible observar que todas las variables consideradas son significativas a un 90% de confianza, a excepción de la velocidad de 4MB en el periodo anterior (que si es significativa a un 89% de confianza). Como se ha descrito en la tabla 6, dicho modelo tiene un valor de un R cuadrado ajustado de McFadden de 11.53%, representando la contribución relativa entre este modelo y el que utiliza como único insumo una constante.

Es posible observar en la tabla anterior que, tal y como se observaba en la descripción de variables, los clientes más jóvenes tienden a ser más propensos a realizar cambios que los clientes de mayor edad cuyos parámetros son menores que los asociados al grupo más joven. De la misma forma, el modelo rescata la relación existente entre el grupo socioeconómico del cliente y su propensión de cambio. Al parecer, los clientes ABC1, tienden a realizar menos cambios en contraste con los clientes pertenecientes a los grupos C3 o D.

Se puede observar que los clientes que contratan canales adicionales de cine o de deportes, tienden a cambiar con más frecuencia que el resto de clientes. Lo mismo ocurre con aquellos clientes que poseen una velocidad de 4 MB contratada en contraste con los que no poseen internet, o poseen velocidades de 2 o 4 MB.

4.4.3 Modelo de Cambio con Heterogeneidad no Observable

Los resultados del modelo descrito con anterioridad, sirven como base para desarrollar un último nivel de heterogeneidad no observable, a través de la utilización de clases latentes. Para determinar el número de segmentos o clases a utilizar se utilizan nuevamente los criterios de Akaike, Bayes y pseudo R cuadrado. La tabla 8 resume dichos criterios utilizando 1, 2, y 3 clases latentes.

Especificación	Número de variables	R cuadrado Ajustado	AIC	BIC
Modelo 7 con 1 clase latente	26	11.53%	25856	26103.34
Modelo 8 con 2 clases latentes	53	11.91%	25746	26250.19
Modelo 9 con 3 clases latentes	80	11.82%	25772	26533.03

Tabla 8: Comparación modelos de cambio con clases latentes

Fuente: Elaboración Propia.

El modelo descrito en la sección anterior, corresponde al modelo 7 con una sola clase latente. Es posible observar en la tabla anterior, que al incluir una segunda clase el R cuadrado ajustado aumenta al mismo tiempo que disminuye el valor del estadístico de Akaike. Lo anterior sugiere que el aumento en la capacidad de ajuste es relevante, aun contrastando con la utilización de una mayor cantidad de variables explicativas. Al incluir una tercera clase, sin embargo, el aumento de la capacidad de ajuste no es suficiente en contraste con la inclusión de más variables, y por lo tanto, se decide por la utilización del modelo 8 que incluye 2 clases latentes.

Luego, el modelo que mejor describe la propensión a cambiar de los clientes es un modelo que utiliza 53 variables explicativas, posee un R cuadrado ajustado de 11.91% (en contraste al modelo sin heterogeneidad) y considera la utilización de dos clases latentes. Es importante notar, que las variables utilizadas en este modelo son similares a las utilizadas en el modelo 7 con heterogeneidad observable, solo que en este caso cada variable estará repetida una vez en cada clase.

Las probabilidades asociadas a cada una de las clases son, 94.82% para la primera, y 5.18% para la segunda. Dichos resultados sugieren la existencia de dos grupos de clientes, uno que representa a la gran mayoría, y un segundo grupo asociado a un pequeño porcentaje de la población. Como se ha mencionado, la mayoría de los clientes son poco propensos a cambiar, y por lo tanto, una primera interpretación de estos grupos puede sugerir que la segunda clase latente corresponde al grupo de clientes más propensos a realizar cambios. Para corroborar dicha hipótesis, es posible graficar las probabilidades asociadas a cada una de las clases encontradas. En el gráfico 14, es posible apreciar los histogramas y *Boxplots* de la probabilidad de cambiar de producto, dada cada una de las clases.

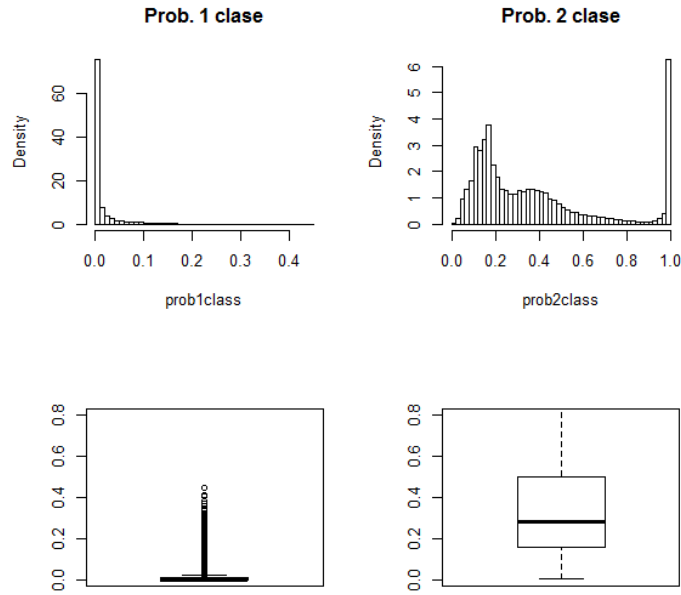


Gráfico 14: Histograma y Boxplot de clases latentes

Fuente: Elaboración Propia.

Como es posible observar en la gráfica anterior, las probabilidades asociadas a la primera clase están mucho más cercanas a cero que las de la segunda. En efecto, si miramos el *boxplot* de cada una de ellas, se puede apreciar que la media de la primera clase es cercana a cero, mientras que la de la segunda tiene un valor que bordea el 0.3. Sumado a esto, la segunda clase pareciera tener un importante grupo de clientes a los que asigna una probabilidad de cambio cercana a 1, mientras que la primera, no pareciera tener valor más allá de 0.2.

Luego, efectivamente las clases representan al grupo de clientes poco propensos a cambiar, en contraste con aquellos cuya probabilidad de migrar a otro producto es mayor. Las variables asociadas a este grupo, ponen especial énfasis en el comportamiento pasado de los clientes, es decir, a las variables asociadas a frecuencia o recencia de cambio.

Es relevante indagar en la relación que tienen las diferentes variables respecto a la propensión de cambio de los clientes. Al parecer, los clientes que varían su gasto por servicios en periodos anteriores, suelen ser más propensos a cambiar que los clientes que mantienen el gasto de un periodo a otro. Lo anterior se puede apreciar en el gráfico 15.

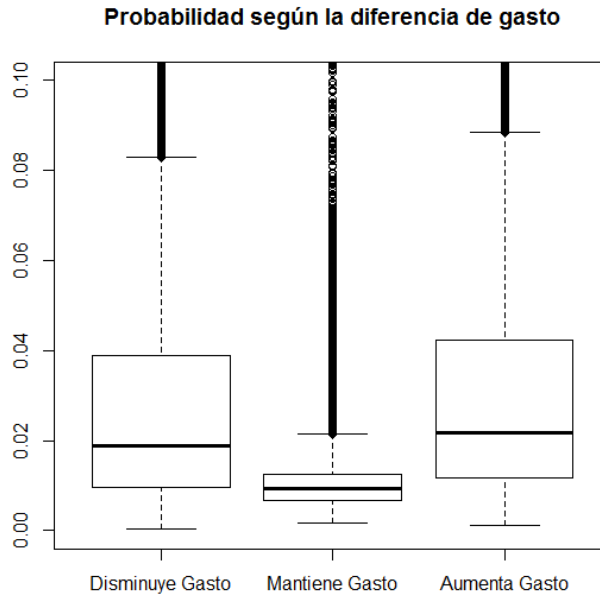


Gráfico 15: Variación en la probabilidad de cambio, según variación en gasto por servicio

Fuente: Elaboración Propia.

El gráfico anterior muestra que los clientes que han variado su gasto en servicios en los periodos anteriores suelen cambiar con más frecuencia su plan actual, en contraste con quienes han mantenido su gasto. Esto es consecuente con el hecho de que un cliente que no cambia el monto gastado, estará relativamente conforme con su plan actual en comparación con otro que ha cambiado sus patrones de consumo y por tanto ha tenido que variar sus gastos.

Por otro lado, tal y como se observa en la descripción de variables, el modelo logra rescatar el hecho de que los clientes más jóvenes tienden a cambiar con mayor frecuencia que aquellos de mayor edad. El modelo desarrollado es capaz de capturar dicha tendencia otorgando una mayor propensión de compra a aquellos clientes de grupos etarios más bajos.

Otro conjunto de variables sumamente relevante para explicar la propensión de cambio, tienen relación con las características de consumo televisivo del cliente. Al parecer, los clientes que poseen televisión, tienen una propensión a cambiar de producto mayor que los clientes que no han contratado dicho servicio. El gráfico 16 muestra esta diferencia.

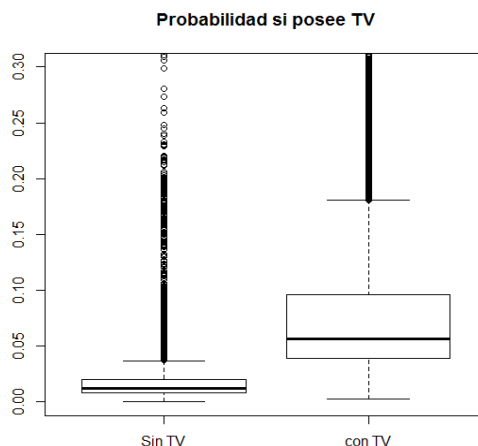


Gráfico 16: Variación en probabilidad de cambio según tenencia de servicio televisivo

Fuente: Elaboración Propia.

Sumado a lo anterior, es posible observar que de los clientes que poseen televisión, aquellos con mayor tendencia a cambiar son aquellos que poseen o bien canales de cine o de deporte adicionales contratados. La siguiente grafica muestra las diferencias en la probabilidad de cambiar, para los clientes que poseen planes de televisión y poseen o no canales adicionales contratados.

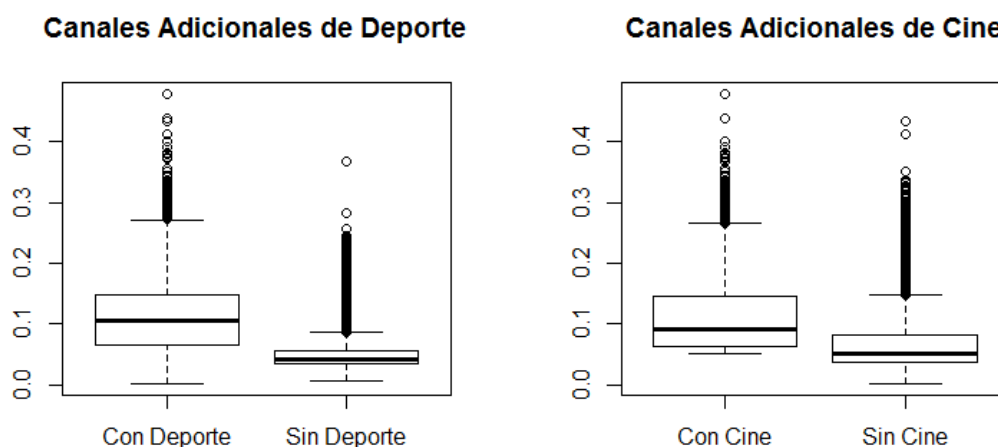


Gráfico 17: Variación de la probabilidad de cambio según canales adicionales contratados

Fuente: Elaboración Propia.

Como es posible observar en los gráficos 16 y 17, los clientes que tienen televisión son más propensos a cambiar de plan que aquellos que no poseen. Más aun,

aquellos clientes que poseen televisión y algún canal adicional, ya sea cine o deporte, tienen una mayor propensión a cambiar de plan que aquellos que no poseen adicionales contratados, aun teniendo contratado el servicios de televisión.

Lo anterior puede deberse a que los canales adicionales de televisión pueden contratarse y des-contratarse con mayor facilidad que el resto de los planes, probablemente porque corresponden más bien un anexo a un plan principal, que un plan en sí mismo. Más aún, existen clientes que solo contratan canales adicionales por algunos meses, generando así una constante dinámica de cambio.

- **Modelo de Elección, o propensión a cambiar a cada alternativa**

Una vez desarrollado el modelo de cambio, se muestran en esta sección los resultados obtenidos tras desarrollar un modelo Logit Anidado para describir la propensión que tiene un cliente a elegir cada alternativa. Como se ha mencionado, este modelo es desarrollado en una muestra aleatoria simple de 10.000 clientes que han realizado un cambio en el último periodo considerado.

Para describir el atractivo de cada alternativa, el modelo considera la especificación de un producto a nivel de atributos, es decir, cada producto estará determinado por una combinación única de 12 atributos, los cuales se describen a continuación.

- **Precio:** Precio del Producto
- **Velocidad:** Velocidad de internet del producto, cero en caso de que el producto no tenga internet.
- **Televisión Top:** Variable dummy que posee un 1 si el producto posee Televisión Top
- **Televisión Preferente:** Variable dummy que posee un 1 si el producto posee Televisión Preferente
- **Decodificador HD:** Variable dummy que posee un 1 si el producto posee Decodificador de Alta definición
- **Canal adicional de Deporte:** Variable dummy que posee un 1 si el producto posee Canales adicionales de Deporte no HD
- **Canal adicional de Cine:** Variable dummy que posee un 1 si el producto posee Canales adicionales de Cine no HD
- **Canal adicional de Otra categoría:** Variable dummy que posee un 1 si el producto posee Canales adicionales de Otra categoría (Adultos o tv Perú)
- **Plan Ilimitado:** Variable dummy que posee un 1 si el producto posee Telefonía Ilimitada
- **Plan de 350 minutos:** Variable dummy que posee un 1 si el producto posee Telefonía controlada de 350 minutos
- **Plan de 150 o 250 minutos:** Variable dummy que posee un 1 si el producto posee Telefonía controlada de 150 o 250 minutos

- **Plan prepago de 5,000 o 6,000 pesos:** Variable dummy que posee un 1 si el producto posee Telefonía prepago de \$5,000 o \$6,000

Estas variables definen un total de 210 combinaciones únicas, que representan las principales alternativas que posee un cliente al momento de contratar un producto. En otras palabras, existen 210 combinaciones de atributos que describen los principales productos hogar que ofrece la compañía.

Al ser un modelo anidado, los 210 productos están representados mediante los nodos finales del modelo. Sumando los nodos intermedios existe un total de 250 nodos. Dichos nodos hacen referencia a las decisiones sucesivas que realiza un individuo para tomar una decisión final. Se asume, que los individuos toman las decisiones primero, escogiendo las categorías que desea (televisión, banda ancha y/o telefonía), luego decide la velocidad de banda ancha, luego el cliente escoge el plan de telefonía que quiere contratar, el plan de televisión y finalmente la combinación de canales adicionales (cine, deporte y/u otros) acorde con las decisiones que haya tomado antes (por ejemplo, no decide velocidad si no ha escogido un plan con internet antes). La siguiente figura esquematiza dicha decisión.

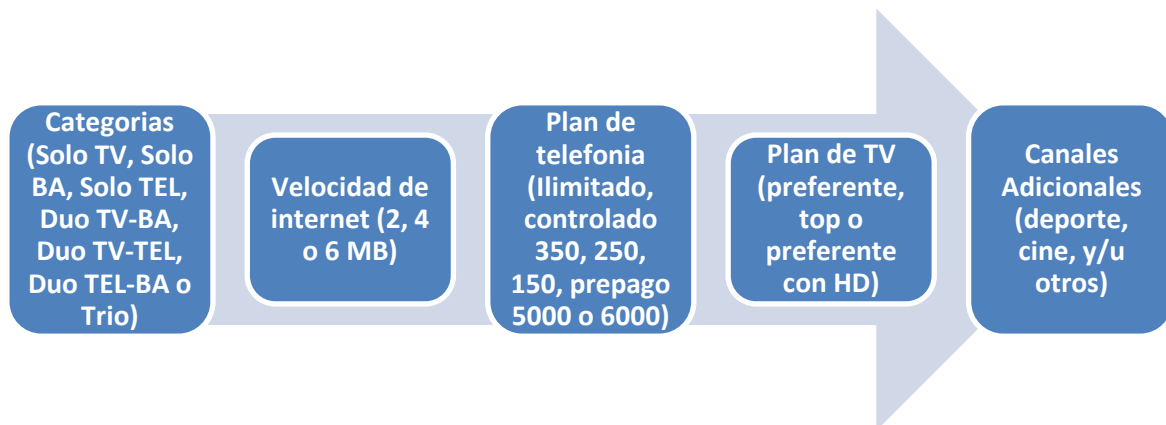


Ilustración 3: Esquematización decisiones anidadas

Fuente: Elaboración Propia.

De esta forma, el cliente toma las decisiones más generales en un comienzo, y comienza a agregar decisiones más específicas en los nodos posteriores. Se asume que la primera decisión es a nivel de categorías, es decir, los servicios generales que quiere contratar, luego, si el cliente ha decidido tener internet decidirá una velocidad, en caso contrario, decidirá el plan de telefonía (en caso de que haya escogido una categoría con telefonía). Finalmente el cliente elige un plan de televisión y junto con esto, si tendrá o no un decodificador de alta definición, luego de esto, podrá decidir si quisiera además agregar algún tipo de canal adicional. La ilustración 4 esquematiza el árbol de decisiones que toma el cliente considerado en los modelos mostrando dos posibles categorías y algunas opciones posteriores que puede realizar el cliente.

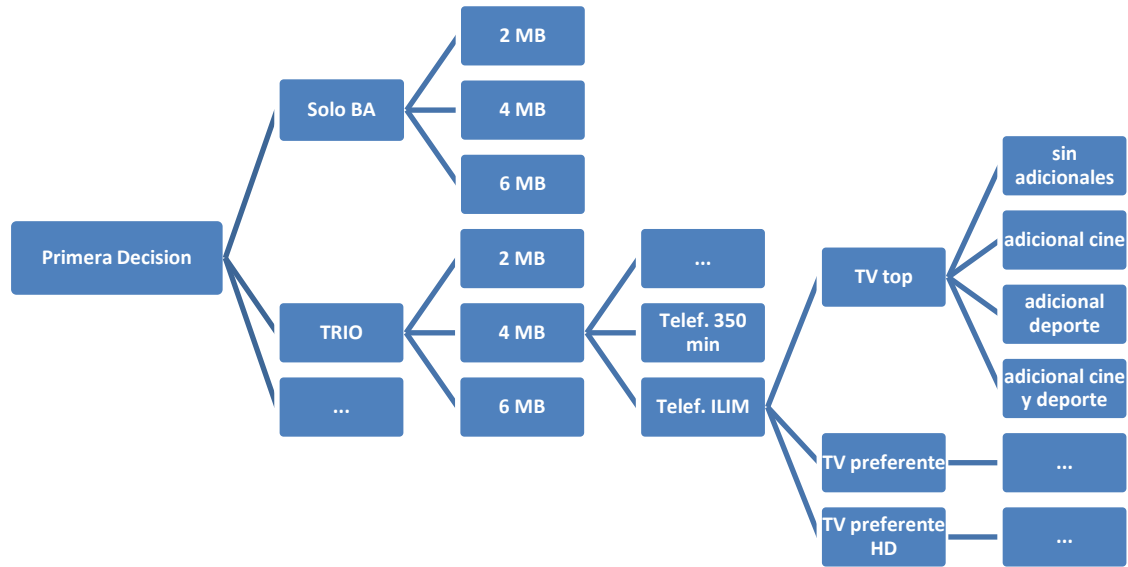


Ilustración 4: Árbol de decisión

Fuente: Elaboración Propia.

La ilustración anterior, muestra la sucesión de decisiones que toma una persona. De esta forma, si un cliente escoge tener solo banda ancha, estará limitado a las velocidades disponibles de la empresa. Por otro lado, si el cliente decide tener un “trío”, elegirá en primer lugar la velocidad, pero luego deberá seguir decidiendo a niveles más específicos hasta llegar al producto final.

4.4.4 Modelo de Elección sin Heterogeneidad

El primer modelo desarrollado es un modelo logit anidado que considera los parámetros beta constantes entre un cliente y otro. En este sentido se obtiene un total de 12 parámetros que ponderan la utilidad de un individuo en función de los atributos que describen un producto.

Todos los parámetros son significativos al 95% de confianza, dado el p-valor asociado a cada variable es menor que 0.05, a excepción del parámetros asociado a telefonía prepago de \$5.000 y \$6.000 que si es significativo al 85% de confianza. La tabla número 9, muestra la capacidad de ajuste del modelo a través de los estadísticos de Akaike, Bayes y R cuadrado ajustado de McFadden

Especificación	Número de variables	R cuadrado Ajustado	AIC	BIC
Modelo s/heterogen.	12	34.52%	35048	35126.21

Tabla 9: Capacidad de ajuste modelo de cambio sin heterogeneidad

Fuente: Elaboración Propia.

Utilizando los resultados obtenidos en este primer modelo es posible generar un árbol de decisión con las probabilidades asociadas a cada cliente, de elegir cada nodo, condicional a la elección de nodos anteriores. Es importante notar, sin embargo, que las probabilidades en este caso serán similares para todos los clientes dado que este modelo no incluye heterogeneidad de ningún tipo, y por tanto su interpretación hace relación con la sensibilidad promedio de los clientes para cada uno de los atributos considerados.

4.4.5 Modelo de Elección con Heterogeneidad Observable

Para incluir heterogeneidad en los modelos, los parámetros calculados en el punto anterior dependerán de las características del cliente. De esta forma, un conjunto de características definen la sensibilidad de los clientes a los diferentes atributos de los productos.

Para identificar el mejor conjunto de variables que describan la propensión a adquirir los diferentes productos, se utilizan los estadísticos de Akaike, Bayes y R cuadrado ajustado de McFadden, además de la significatividad individual de los diferentes parámetros. La tabla 10 resume el número de variables utilizadas y los estadísticos pseudo R cuadrado ajustado, Akaike y Bayes para los diferentes modelos desarrollados.

Especificación	Número de variables	R cuadrado Ajustado	AIC	BIC
Modelo 1 s/heterog.	12	34.45%	35048	35126.21
Modelo 2	68	35.42%	34534	34977.17
Modelo 3	68	42.15%	30934	31377.17
Modelo 4	52	44.30%	29784	30122.89
Modelo 5	42	40.48%	31826	32099.72
Modelo 6	44	39.45%	32376	32662.76
Modelo 7	45	38.66%	32798	33091.27
Modelo 8	71	44.62%	29612	30074.72
Modelo 9	25	36.94%	33718	33880.93
Modelo 10	25	37.24%	33560	33722.93
Modelo 11	81	42.89%	30536	31063.89
Modelo 12	81	38.64%	32808	33335.89
Modelo 13	85	40.44%	31846	32399.96
Modelo 14	120	48.20%	27696	28478.06
Modelo 15	46	34.85%	34836	35135.79
Modelo 16	112	47.62%	28006	28735.93
Modelo 17	105	48.23%	27682	28366.31
Modelo 18	101	48.24%	27676	28334.24

Modelo 19	99	48.25%	27672	28317.20
-----------	----	--------	-------	----------

Tabla 10: Estadísticos de ajuste, modelos de elección con heterogeneidad

Fuente: Elaboración Propia.

De la tabla anterior se deriva que la mejor especificación, en cuanto a menores valores de los criterios AIC, BIC y mayor valor del R cuadrado ajustado de McFadden es el modelo 19. Dicho modelo incluye 99 variables explicativas, y tiene un pseudo R cuadrado de 48.25%, en contraste con un modelo base que no incluye ninguna variable explicativa (o todos los parámetros son iguales a cero).

4.4.6 Modelo de Elección con Heterogeneidad no Observable

Tras desarrollar el modelo con heterogeneidad observable, se genera un último modelo que incluye heterogeneidad no observable utilizando clases latentes. A continuación se presentan los estadísticos de Akaike, Bayes y R cuadrado ajustado de McFadden para los modelos con 1 y 2 clases latentes.

Especificación	Número de variables	R cuadrado Ajustado	AIC	BIC
Modelo 19 con 1 clase latente	99	48.25%	27672	28317.2
Modelo 20 con 2 clases latentes	199	51.91%	25712	27008.92

Tabla 11: Comparación modelos de elección con clases latentes

Fuente: Elaboración Propia.

Como se puede observar en la tabla anterior, los estadísticos tanto de Akaike como de Bayes tienen menores valores cuando se utilizan dos clases latentes, al compararlos con el modelo de una única clase. Consecuente con lo anterior, el valor del estadístico R cuadrado tiene un mayor valor para el modelo número 20. Dado lo anterior, el modelo que mejor describe la propensión a cambiar a cada una de las alternativas, es el modelo 20, que considera 199 variables y dos clases latentes cuyo valor para el estadístico R cuadrado ajustado asciende a 51.91%.

Aunque la inclusión de una tercera clase para este modelo pudiera ser en teoría factible, la complejidad computacional asociada es sumamente alta y por lo tanto poco viable en la práctica. En concreto, dicho modelo incluiría 299 incrementando de forma exponencial el tiempo computacional necesario para maximizar la función de verosimilitud.

La probabilidad de pertenecer a la primera clase es de un 9.01%, en contraste con la probabilidad de pertenecer a la segunda de un 90.99%. Es posible apreciar diferencias entre estos grupos a nivel de categorías, donde las principales

desigualdades se encuentran en la categoría “Duo Banda Ancha-telefonía”. El gráfico número 18 muestra las diferencias en la probabilidad de adquirir un producto en dicha categoría, dada la pertenencia a cada una de las clases.

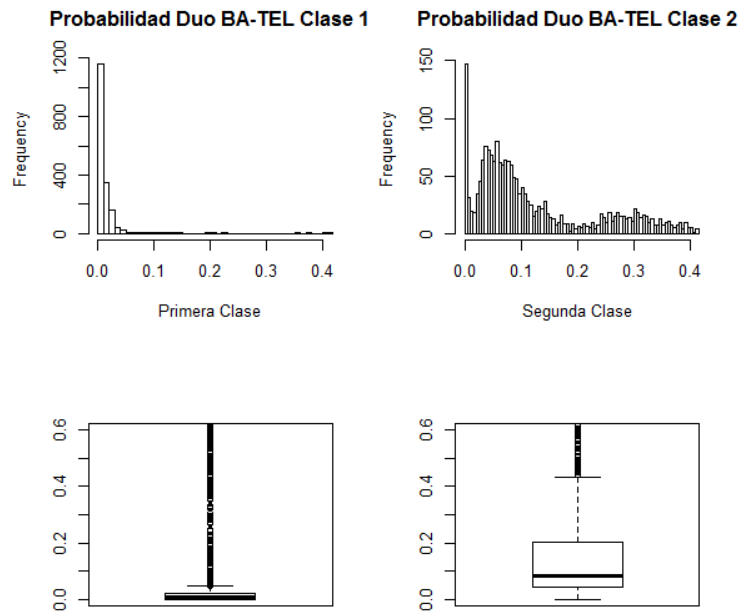


Gráfico 18: Histograma y Boxplot de probabilidad de adquirir duo banda ancha-telefonía por clases

Fuente: Elaboración Propia.

En la gráfica anterior es posible observar que en efecto, la primera clase tiene probabilidades de escoger un duo banda ancha y telefonía mucho más cercanas a cero que para el caso de la segunda. La diferencia aquí apreciada, es similar para la categoría “Solo telefonía”, solo que para este caso, la segunda clase tiene un comportamiento más cercano a cero, en contraste con la primera.

En cuanto al comportamiento de las diferentes probabilidades, en función de las variables explicativas, es posible apreciar algunas relaciones que pueden ser interesantes. La probabilidad de migrar a un plan más completo, es decir, de un plan con más atributos que el actual, parece depender de múltiples factores entre los cuales destaca el plan que se poseen en el periodo anterior, y el número de periodos en los cuales se ha poseído.

Al parecer un cliente tiene una propensión más baja de modificar su plan a medida que lo posee por más tiempo. Aún más, pareciera que los clientes tienden a realizar cambios más abruptos cuando han tenido un plan por menos tiempo, en contraste con los que realizan cuando ha pasado un periodo mayor desde que tienen contratado el plan.

En efecto, los clientes que poseen un plan de solo telefonía, tienden con alta probabilidad a cambiar a un Duo banda ancha-telefonía o a un Trio. Dicha probabilidad cambia drásticamente en función de hace cuánto tiempo lleva el cliente con su plan de telefonía. El gráfico 19 muestra el cambio de probabilidad de adquirir un plan Duo de internet más telefonía, un trio, o mantener el plan de solo telefonía, en función del tiempo que ha tenido el cliente contratado un servicio de telefonía.

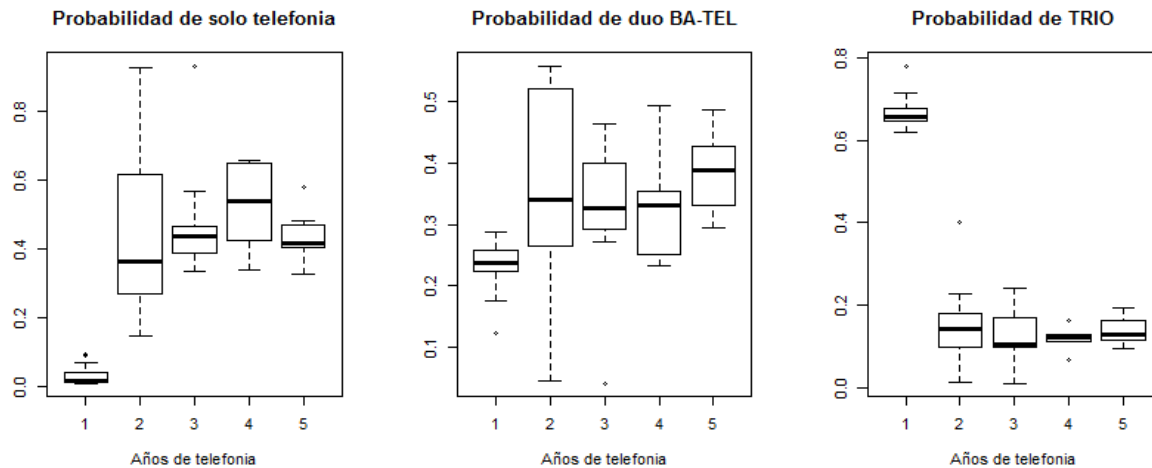


Gráfico 19: Probabilidad de migrar a Duo banda ancha-telefonía o Trio desde plan solo telefonía según años de telefonía

Fuente: Elaboración Propia.

En el gráfico 19, se puede observar que los clientes que poseen un plan solo de telefonía, tienden a migrar con mucha mayor probabilidad, tanto a un TRIO como a un DUO Banda ancha más telefonía, en el primer año de antigüedad del servicio. Este hecho cambia drásticamente al segundo año donde la probabilidad de mantener un plan de solo telefonía aumenta mientras que la probabilidad de migrar a un TRIO disminuye prácticamente a un cuarto de la probabilidad del primer año. Por su parte, la probabilidad de migrar al plan Duo, aumenta levemente en los años posteriores.

Lo anterior refleja el hecho de que un cliente que tiene por más tiempo un tipo de plan, tiende a mantenerlo (o a cambiar a planes muy similares), y de cambiar de categoría, lo hace de forma no demasiado drástica, por ejemplo pasando de un plan de un solo servicio a un plan “duo” y no directamente a un “trio”. Dicho comportamiento es similar para otros tipos de clientes, por ejemplo aquellos que poseen un plan de solo banda ancha o solo televisión.

La edad del cliente parece tener directa relación con el plan que este escoge, en efecto, los clientes de más edad tienden a cambiar a planes más simples y en general que tengan telefonía. Los clientes más jóvenes en tanto, son menos propensos a elegir planes con telefonía y más a aquellos que poseen Banda Ancha. El gráfico 20 muestra las diferencias entre grupos etarios en cuanto a la propensión de adquirir planes de solo

telefonía (izquierda) o a adquirir planes de Duo de televisión más banda ancha (derecha).

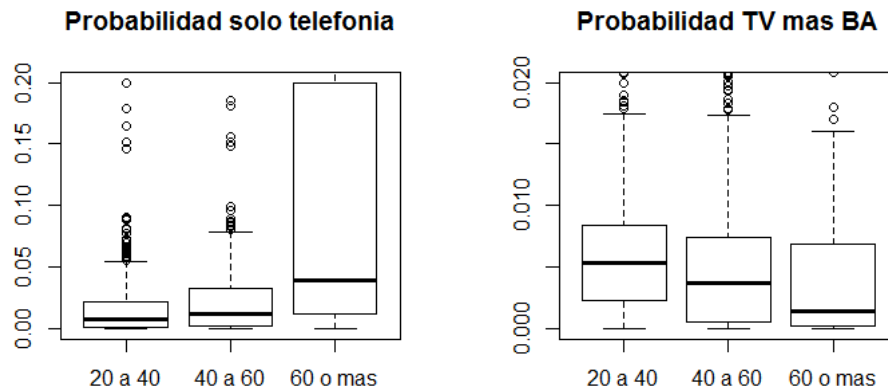


Gráfico 20: Probabilidad de adquirir planes de solo telefonía y Duo televisión-Banda Ancha según grupo etario

Fuente: Elaboración Propia.

Como muestra la gráfica anterior, los clientes de mayor edad tienden con mayor probabilidad a migrar a planes de solo telefonía en contraste con los clientes más jóvenes, en tanto que los grupos de con más edad tienden con baja probabilidad a migrar a planes sin telefonía como un Duo de televisión más banda ancha en comparación a los grupos etarios más jóvenes.

Luego, ambos modelos desarrollados, tanto el para estimar la propensión de cambio, como la probabilidad de elegir los diferentes productos, son capaces de ajustar con bastante precisión a los datos, reflejada en valores relativamente altos de los estadísticos R cuadrado ajustado. Además, los modelos son capaces de reconocer el conocimiento que se posee a priori analizado en la descripción inicial de variables.

Sin embargo, no solo es importante evaluar en cuanto a la capacidad de ajuste y la relación entre las variables y resultados, sino además, la capacidad de predecir el comportamiento real de los clientes. La siguiente sección analiza la capacidad predictiva de los modelos en cuanto y la capacidad de mejorar los actuales procesos de oferta de la compañía

4.5 Evaluación

4.5.1 Capacidad Predictiva

La evaluación de la capacidad predictiva de los modelos se realiza en una muestra de 20.000 clientes para el modelo de cambio y 2.000 clientes para el modelo de elección, ambos independientes a las muestras utilizadas para ajustar los modelos.

Para evaluar la capacidad predictiva, es importante traducir los resultados obtenidos a decisiones concretas respecto a lo que decidirá un cliente. Es decir, es necesario transformar las probabilidades entregadas por los modelos a decisiones concretas de si un cliente cambiará o no de producto y a qué servicio específico lo hará.

- **Modelo de cambio**

Para el modelo de cambio, lo anterior se traduce en establecer un punto de corte para decidir si el cliente realizará o no un cambio. En otras palabras, es necesario decidir por sobre que probabilidad entregada por el modelo, será traducida como que el cliente realiza un cambio.

Es importante notar, que lo relevante es encontrar a aquellos clientes que efectivamente realizaron un cambio, pero reconociendo que al realizar la oferta se estará incurriendo en un costo, por lo que se debe buscar también considerar a la menor masa de clientes posibles.

Si el punto de corte, es cercano a cero, se señala que todos los clientes realizarán un cambio, dado que todos los clientes tienen probabilidad mayor a cero entregada por el modelo y por lo tanto, se acertará en todos los casos en los que ocurra un cambio. Por el contrario, si el punto de cambio es cercano a uno, se estará diciendo que ningún cliente realizará un cambio y por lo tanto, no se acertará en ningún caso a los cambios reales realizados. Lo relevante entonces, es encontrar un punto interior en el cual se considere que cierto porcentaje de la población realizará un cambio, y se acierte a la mayor cantidad de cambios posibles. El gráfico 21, muestra la cantidad de aciertos entregados por el modelo como porcentaje de los cambios reales ocurridos en el periodo estimado, en función del porcentaje total de la población que se considera cambiará en dicho periodo.

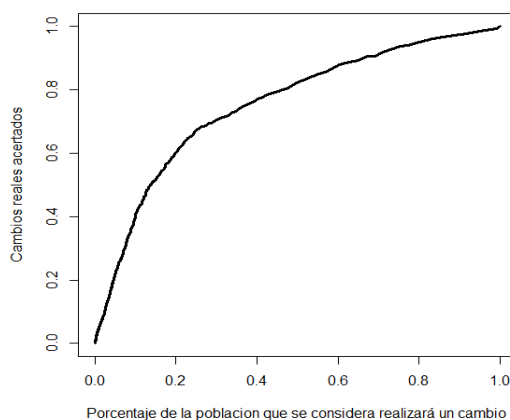


Gráfico 21: Capacidad predictiva según porcentaje de la población considerada

Fuente: Elaboración Propia.

La línea trazada en la gráfica anterior muestra como el porcentaje de aciertos aumenta en función de la población considerada, cuyo aumento está relacionado con una disminución del punto de corte. Si se considera un corte que incluya al 20% de la población, se observa que se acierta aproximadamente al 60% de los cambios reales ocurridos.

Una posible métrica para definir el punto de corte, es la diferencia entre el porcentaje de aciertos y el porcentaje de la población considerada. Al maximizar dicho valor, se obtiene que el punto óptimo de corte es 5.41%, es decir, se considerará que el cliente cambiará de producto a todos aquellos que tengan una probabilidad de cambio superior a este punto entregada por el modelo. Lo anterior se traduce en la consideración de un 23.49% de la población y un acierto de un 65.27% de los casos reales de cambio ocurridos en dicho periodo.

Otra posible métrica, corresponde a un punto de corte tal que se ofrezca a la misma cantidad de personas que se ofrece actualmente en la compañía. Para esto se consideran las principales ofertas realizadas en julio del 2012 que corresponden a up-selling de banda ancha, up-selling de televisión, cross-selling de banda ancha y cross-selling de televisión. En estas campañas se ha ofertado a un 15.69% del total de clientes. El punto de corte asociado a este porcentaje de clientes es del 8.32% y se acierta a un total de 52.72% de los cambios reales. La tabla numero 12 resume los puntos de corte considerados y la capacidad predictiva de los modelos en cuanto a la cantidad de cambios reales predichos.

Punto de Corte	Clientes considerados	Cambios predichos
5.41%	23.49%	65.27%
8.32%	15.69%	52.72%

Tabla 12: Resumen capacidad predictiva modelo de cambio para puntos de corte considerados

Fuente: Elaboración Propia.

- **Modelo de Elección**

Para el modelo de elección, lo relevante es decidir hacia que producto migrará el cliente, dado que ha realizado un cambio. Para esto, es necesario definir cómo interpretar los resultados del modelo tomando una decisión de cambio. Los resultados del modelo de elección es una probabilidad de elección asociada a cada uno de los 210 productos considerados para cada cliente.

Una primera alternativa, es decir que el cliente cambiará al producto cuya probabilidad de elección entregada por el modelo sea máxima entre todas las decisiones que puede tomar. Al considerar dicho criterio, se acierta en el producto al que ha migrado el cliente en un 25.45% de los casos.

Es importante notar sin embargo, que se está considerando una parrilla de 210 alternativas y por lo tanto, las diferencias entre un producto y otro en muchos casos es pequeña. Luego, considerar como única alternativa de elección al máximo de las probabilidades del modelo puede ser demasiado exigente. Dado esto, se considera un grupo de productos como output del modelo, correspondientes a las probabilidades más altas entregadas por el mismo.

En este sentido, lo relevante es encontrar el número de productos que se consideraran como posibles para la migración del cliente. Dado que el objetivo de este trabajo es apoyar los procesos de oferta, la pregunta puede ser respondida al considerar el máximo número de productos que la empresa puede considerar para ofrecer a un cliente, es decir, el máximo número de productos que pueden ser ofrecidos a un cliente en el proceso de oferta.

Dado lo anterior, se considerarán 4 alternativas de migración del cliente, correspondientes a las cuatro máximas probabilidades de elección entregadas por el modelo. La tabla 13, resume la capacidad predictiva asociada a cada una de estas alternativas.

Alternativa	Aciertos de Elección	Acierto Agregado
1	25.45%	25.45%
2	17.90%	43.35%
3	13.00%	56.35%
4	7.00%	63.35%

Tabla 13: Capacidad predictiva modelo de elección para alternativas consideradas

Fuente: Elaboración Propia.

Como se aprecia en la tabla anterior, se acierta a la elección realizada por el cliente en alguna de las 4 alternativas entregadas por el modelo en un 63.35% de los casos.

Es importante notar, que los procesos de oferta no siempre son tan específicos, y en algunos casos puede ser útil ofrecer productos más generales o categorías de productos a los clientes, antes de ofrecer productos más particulares. Dado esto, puede ser relevante considerar la capacidad predictiva de los modelos en cuanto a la categoría que decidirá el cliente, entendiéndose esta como la primera decisión que realiza el cliente, asociada a la primera elección del árbol de decisiones anidadas del modelo. En otras palabras, puede ser relevante conocer la propensión del cliente en respecto a la combinación de categorías de producto prefiere, entre internet, televisión y/o telefonía. La tabla 14, muestra la capacidad predictiva del modelo considerando dos alternativas de categorías.

Alternativa	Aciertos de Elección	Acierto Agregado
1	77.85%	77.85%
2	15.90%	93.75%

Tabla 14: Capacidad predictiva modelo de elección a nivel de categorías para alternativas consideradas

Fuente: Elaboración Propia.

Es posible observar en la tabla anterior, que al considerar dos alternativas de categorías entregadas por el modelo, se acierta a la elegida por el cliente en un 93.75% de los casos. Dada la alta capacidad predictiva en este sentido, dicho resultado puede ser utilizado para guiar el proceso de oferta de un nivel más desagregado hasta llegar al producto final.

4.5.2 Comparación con Procesos Actuales

Tras analizar la capacidad predictiva de los modelos, es posible comparar los resultados obtenidos con los procesos que actualmente utiliza la compañía. Para esto, se considerarán las principales campañas de oferta de productos fijos que realiza la empresa. Dicha comparación, permite analizar en qué casos la utilización de las metodologías aquí desarrolladas pueden tener un mayor impacto en la firma.

Los procesos actuales de oferta de la empresa, utilizan una sucesión de filtros, para determinar el conjunto de clientes al que debe ser enviada la oferta en cada periodo. Dicha sucesión, considera información del producto que tiene contratado actualmente el cliente, el gasto que tiene por servicio, si posee o no deudas, entre otros.

Para evaluar la efectividad de los procesos actuales, es relevante analizar los cambios reales producidos por dichos procesos, en contraste con los cambios predichos por el modelo de cambio, considerando los puntos de corte mencionados anteriormente. Es importante evaluar además, las diferencias en el porcentaje de la población llamada en los procesos actuales, versus la considerada por los modelos. La tabla número 15 resume las diferencias entre las campañas actuales y los resultados del modelo de cambio.

Situación	Cambios predichos	Clientes llamados
Actual	16.86%	15.69%
Modelo corte 5.41%	65.27%	23.52%
Modelo corte 8.32%	52.72%	15.69%

Tabla 15: Comparación modelo de cambio con procesos actuales

Fuente: Elaboración Propia.

Es posible observar que los procesos actuales consideran a un 15.69% de la población total, acertando en un 16.86% de los cambios realizados por los clientes. Lo anterior sugiere que las metodologías utilizadas en los procesos actuales para decidir a qué clientes debe ir dirigida la oferta tiene una componente aleatoria importante, dado que se está acertando a casi al mismo porcentaje predicho de cambios reales que los clientes que son considerados. En otras palabras, la sucesión de filtros utilizada por la compañía, no parece tener una capacidad predictiva real para con los clientes propensos a realizar cambios, dado que se observa que la capacidad de acierto de los cambios reales producidos, es prácticamente igual al porcentaje de clientes escogidos para realizar la oferta.

Los modelos considerados, sin embargo, son capaces de dirigir la oferta a aquellos clientes más propensos a realizar cambios, teniendo que llamar a un porcentaje de clientes considerablemente menor, pudiendo generar un efecto similar en cuanto al número de clientes que realizan finalmente un cambio. En efecto, el modelo que utiliza un punto de corte de 5.41%, es capaz de predecir un 65% de los cambios, considerando a un 23.52% de los clientes. Comparado con el proceso actual, esto significa considerar aproximadamente la mitad de clientes en los procesos de oferta, acertando aproximadamente cuatro veces más que los procesos actuales. Por su parte, al comparar el modelo cuyo corte es de 8.32%, se observa que es necesario considerar a la misma cantidad de personas, acertando aproximadamente 3 veces más que los procesos actuales.

Lo anterior sugiere que la utilización de los modelos puede ser de gran utilidad para la firma, pudiendo disminuir los costos asociados a los procesos de oferta, al considerar a menor cantidad de personas, y pudiendo aumentar además, la efectividad de dichos procesos. En particular, es posible observar que el punto de corte de 8.32% es Pareto superior a la situación actual, dado que se considera el mismo volumen de gente (y por tanto los mismos costos), pero una mayor efectividad.

Es relevante notar que la comparación aquí realizada, considera los procesos a un nivel agregado, y por lo tanto, es relevante analizar que ocurre para cada una de las campañas realizadas por la empresa. Dichas campañas pueden dividirse en cuatro grupos, up-selling de banda ancha y televisión, y cross-selling de banda ancha y televisión, es decir, campañas que apuntan a aumentar la velocidad contratada de internet del cliente o mejorar el plan de televisión, y campañas que apuntan a ofrecer banda ancha o televisión a clientes que no tienen actualmente este servicio. La tabla número 16 muestra las diferencias en la capacidad predictiva para cada una de estas campañas, para los procesos actuales y los modelos desarrollados. Dicha tabla, resume los cambios reales producidos por los procesos actuales, y los cambios genuinos que los modelos son capaces de predecir.

	Cambios predichos			
	Campaña up BA	Campaña up TV	Campaña cross BA	Campaña cross TV
Actual	12.31%	12.17%	33.33%	43.24%
Modelo corte 5.41%	28.21%	69.43%	21.43%	0.00%
Modelo corte 8.32%	15.38%	46.50%	7.14%	0.00%

Tabla 16: Capacidad predictiva según tipo de campaña

Fuente: Elaboración Propia.

Es posible observar en la tabla anterior, que los modelos desarrollados son capaces de estimar la propensión de cambiar de producto de mejor forma que los procesos actuales para las campañas de up-selling de la empresa. Esto no ocurre al comparar los resultados con las campañas de cross-selling realizadas por la compañía. Lo anterior puede ser debido a que el modelo no es capaz de reconocer la probabilidad de cambio de un cliente que cambiará la categoría del producto, pero si lo hace para aquellos que mantendrán la categoría cambiando las características del plan.

Sin embargo, es importante preguntarse cuántos de los cambios realizados por los clientes son realmente asociados a una propensión a cambiar, o cuantos son en realidad inducidos por el proceso de oferta. En otras palabras, puede ser relevante preguntar qué porcentaje de los cambios son producidos al estar en campaña y cuantos son en realidad espontáneos, es decir, por motivaciones propias del cliente. La tabla 17 resume dichos porcentajes por tipo de campaña realizada.

	Cambios por campaña	cambios sin campaña
Campaña up BA	12.31%	87.69%
Campaña up TV	12.17%	87.83%
Campaña cross BA	33.33%	66.67%
Campaña cross TV	43.24%	56.76%

Tabla 17: Cambios con y sin campaña según tipo de campaña

Fuente: Elaboración Propia.

Es posible observar que para las campañas de cross selling, el porcentaje de clientes que cambia de forma genuina es bastante menor que los que cambian tras haber pasado por un proceso de oferta, comparado con las campañas de up selling. Lo anterior tiene relación con la efectividad del proceso de oferta y puede explicar en cierta medida porque los modelos predicen de mejor forma en uno u otro caso. Lo anterior, dado que los modelos estiman la propensión de cambio de un cliente habiendo o no pasado por campañas y si la propensión de un cliente cambia demasiado al pasar por

un proceso de oferta, como es el caso de las campañas de cross selling, el modelo subestima los clientes que cambiarán, dado que no reconoce que estos pasaron por el procesos de venta.

Para el modelo de elección, la comparación con los procesos actuales puede ser realizada mediante la comparación del delta precio asociado a la oferta realizada al cliente. De esta forma, es posible tener una comparación de los procesos actuales, versus las ofertas realizadas a través de los modelos actuales.

Obtener el beneficio promedio estimado para cada alternativa no es suficiente dado que se sabe que no todos los clientes aceptarán las alternativas propuestas por los modelos. Si consideramos entonces la capacidad predictiva del modelo de elección, es posible calcular el aporte promedio asociada a cada una de las alternativas. Utilizando entonces los resultados obtenidos en la tabla 13 asociados a la capacidad predictiva de cada alternativa y multiplicándolos por los beneficios promedio asociado a cada una de ellas, se obtiene una estimación razonable del aporte promedio total al utilizar el modelo de elección propuesto.

La tabla 18 muestra el beneficio promedio en pesos que producen las campañas actuales, y el beneficio promedio que producen las diferentes alternativas mencionadas en la sección anterior, expuestas en la tabla 13, describiendo el beneficio neto estimado de cada una de las alternativas.

Alternativa	Aciertos de Elección	Beneficio promedio	Beneficio estimado
ACTUAL	-	\$ 1,210	\$ 1,210
1	25.45%	\$ 2,424	\$ 617
2	17.90%	\$ 3,367	\$ 603
3	13.00%	\$ 3,223	\$ 419
4	7.00%	\$ 4,247	\$ 297

Tabla 18: Beneficio promedio asociado a cada alternativa y beneficio actual

Fuente: Elaboración Propia.

Al sumar todos los beneficios estimados, asociados a las alternativas propuestas por el modelo de elección, ponderados por la probabilidad de acertar en cada caso, se obtiene que el beneficio promedio asociado a la utilización de los modelos para guiar los procesos de oferta sea de \$1,936. En contraste con los \$1,210 actuales por persona. Lo anterior significa un aumento de los beneficios asociados a las campañas cercanos a un 60%.

Luego, si se considera contactar a la misma cantidad de personas que en la situación actual, es decir, cerca del 15.69% de la base de datos, se obtiene que los ingresos asociados a la utilización de los modelos bordean los \$31.000.000 mensuales,

en contraste con los casi \$6.400.000 asociados a los procesos actuales. Por lo tanto, la utilización de la metodología propuesta en este trabajo tiene el potencial de aumentar los ingresos asociados a los procesos de oferta de la empresa en casi un 500% de la situación actual.

4.6 Propuestas de Acción.

Tras evaluar la capacidad predictiva de los modelos, es posible proponer líneas de acción asociadas a la utilización de los mismos. Por tanto, el objetivo de esta sección es generar un conjunto de pautas claras, respecto a la forma en que deben ser utilizados los modelos.

En primer lugar, es importante notar que la utilización de las metodologías propuestas en este trabajo, están sujetas a limitaciones que es relevante tener en cuenta al momento de generar líneas de acción. Luego, se reconoce la importancia de tomar en cuenta las siguientes condiciones:

1. Los clientes deben estar a lo menos dos meses en la compañía, de forma de tener suficientes observaciones para la funcionalidad de los modelos
2. Los clientes deben tener contratado alguno de los productos principales fijos de la empresa, o servicios similares que tengan características comparables con dichos productos.

Teniendo en cuenta dichas limitaciones, se debe seleccionar un grupo de clientes a los cuales sea factible incluir en los procesos de oferta, esto es, clientes que cumplan con las reglas de negocio establecidas por la compañía y que cumplan las condiciones asociadas a la usabilidad de los modelos. Sobre el grupo de clientes seleccionados, se deben utilizar los parámetros de los modelos obteniendo la probabilidad asociada al cambio de producto, y la probabilidad de elección de cada una de las alternativas.

Es importante destacar, que en una primera instancia se debe mantener las prácticas actuales que utilizan los procesos de oferta de la empresa, y por lo tanto, es necesario generar estructuras de acción cuyas características sean similares a las utilizadas actualmente. Dicha estructura, se basa en la definición a priori de la cantidad de ofertas por cada tipo de producto, es decir, en definir y fijar un número de prospectos para cada tipo de campaña.

En este sentido, es relevante que la utilización de los modelos aquí descritos sea capaz de definir un número fijo de campañas de up selling de televisión y banda ancha, y un número fijo de campañas de cross selling de las mismas categorías, o las que sean necesarias en cada periodo.

Se ha definido en secciones anteriores, que el modelo de elección entrega cuatro alternativas de oferta para los clientes, reflejado en los cuatro productos cuya probabilidad de migración sea más alta. Dado lo anterior, se propone que se utilicen estas alternativas para definir la oferta que se hará a cada cliente. De esta forma, si se

necesitan enviar 100,000 ofertas de aumento de la velocidad de internet, se propone escoger aquellos clientes que en la primera alternativa arrojada por los modelos tengan un producto que cumpla dicha condición (un aumento de la velocidad de internet respecto el plan actual). Si considerando todas las primeras alternativas aún no se tiene la cantidad necesaria, se debe mirar la segunda, luego tercera y finalmente cuarta alternativa.

Luego, se propone enviar la oferta a los clientes más propensos a realizar un cambio, mirando los resultados entregados por el modelo de cambio. De esta forma, el modelo de elección será utilizado para definir el grupo de clientes a los que se enviara la oferta, y el segundo modelo, para priorizar entre un prospecto y otro.

Las acciones propuestas hasta este punto, pueden ser interesantes dado que permiten utilizar los modelos propuestos de forma sencilla, sin necesidad de reestructurar los actuales procesos de oferta, sino que simplemente, definir de forma distinta los clientes incluidos en los procesos.

Sin embargo, se debe notar que al mantener la estructura actual de los procesos, se están dejando de lado el resto de alternativas entregadas por el modelo de elección a las que el cliente puede ser también propenso a cambiar. Por ejemplo, si se toma un cliente al que se le ofrece un aumento en la velocidad de internet, y este no acepta la oferta, sería prudente darle algunas otras opciones a las cuales es propenso también a cambiar, representadas por las alternativas posteriores entregadas por los modelos.

Este hecho ocurre porque los procesos de oferta actuales no son lo suficientemente flexibles para ofrecer varios productos diferentes a un mismo cliente, en caso de que no acepte la primera oferta que se la hace. En este sentido, se considera relevante generar una flexibilización de los procesos de oferta, pudiendo utilizar todo el potencial de los modelos. Si este fuera el caso, la utilización de los modelos puede ser algo distinta, yendo en búsqueda de aumentar la rentabilidad asociada al cambio de productos.

Es importante notar que dicha flexibilización de los procesos implica un costo no considerado en la evaluación presentada anteriormente. Dicho costo puede ser importante al reconocer que implica la reestructuración de parte importante de los procesos de oferta, las dinámicas entre los equipos de planificación y quienes realizan la oferta y la estructuración de incentivos para estos últimos. De todas formas, dicho costo es principalmente a corto plazo, en tanto que los beneficios asociados a la reestructuración se mantienen en el tiempo.

En este contexto, se propone la utilización de los dos modelos desarrollados de forma conjunta, en miras de estimar la rentabilidad asociada a la oferta de productos para cada cliente. La forma de utilización propuesta de los modelos para este caso, se describe a continuación.

En primer lugar, se debe calcular cual es el aumento en el precio que el cliente paga por el producto, para cada una de las alternativas derivadas del modelo de elección. Hecho esto, se deben multiplicar dichos incrementos por la probabilidad de que el cliente efectivamente cambie a ese producto, entregadas por el mismo modelo. La suma de estos valores, corresponde a una estimación de la rentabilidad esperada de la oferta, dado que el cliente ha realizado el cambio.

Para la estimación de la rentabilidad esperada final, se debe multiplicar dicho valor por la probabilidad obtenida del modelo de cambio, que corresponde a la probabilidad estimada de que el cliente efectivamente realice un cambio. Este valor final, corresponde a la rentabilidad esperada asociada al proceso de oferta para ese cliente particular.

Luego, el proceso de oferta debe priorizar a los clientes cuya rentabilidad esperada sea mayor, y debe ofrecer los productos derivados de la utilización del modelo de elección, sin importar si hay más o menos ofertas de uno u otro tipo.

5. CONCLUSIONES

5.1 Conclusiones del Trabajo

El objetivo general de este trabajo es diseñar y evaluar mecanismos formales capaces de apoyar los actuales procesos de oferta a clientes residenciales de Movistar. En este sentido, se busca proponer metodologías formales en términos teóricos para dar respuesta a un problema aplicado en una empresa. Dicho objetivo ha sido resuelto con éxito tras elaborar modelos econométricos congruentes con las necesidades de la compañía, evaluar en términos numéricos la efectividad de los resultados obtenidos en contraste con las metodologías actuales y proponer medidas directas de cómo se adaptan dichos resultados a la realidad de la firma.

La elaboración de este trabajo es relevante, pues da solución a un problema latente en la empresa, asociado a la comprensión del cliente y su relación con los diferentes productos que ofrece la compañía, en miras de mejorar la efectividad de los procesos de venta. En este contexto, este trabajo propone solución a un problema relevante no solo para la firma, sino además, para las empresas del rubro interesadas en generar diferenciación mediante el análisis de sus clientes para crear valor y satisfacer sus necesidades.

Entre las principales complejidades de este trabajo, destacan en primer lugar la incorporación de heterogeneidad en los modelos, y por lo tanto, la búsqueda de resultados específicos para cada perfil de cliente. En segundo lugar, la gran cantidad de clientes y de alternativas disponibles, hacen que el modelamiento en términos tanto numéricos como de procesamiento computacional sea sumamente complejo. Por último, al ser un problema aplicado, una de las principales complejidades radica en la

evaluación de los modelos, y principalmente en la comparación con los actuales procesos existentes en la compañía.

En cuanto a las estimaciones realizadas, se observa que los modelos que mejor describen el comportamiento de los clientes son los que incluyen heterogeneidad no observable. Para el modelo que estima la probabilidad de que un cliente realice un cambio de producto, se observa que este incluye un 52% de los cambios totales, al considerar un 15.7% de los clientes de la empresa. Lo anterior, es sumamente positivo si se considera que los actuales procesos de oferta, contactan al mismo 15.7% de los clientes, de los cuales solo un 16.8% realiza algún cambio. En cuanto al modelo de elección, se observa que el modelo acierta en un 63% de las elecciones realizadas por los clientes, lo que determina un aumento potencial de un 60% de la rentabilidad asociada a los procesos de oferta.

Sin embargo, es importante notar que los modelos son capaces de predecir mejor el comportamiento de los clientes bajo ciertas condiciones. En particular, se observa que los modelos de cambio no entregan buenos resultados para los clientes que han cambiado la categoría de productos, en contraste con los que han cambiado de producto dentro de una misma categoría.

Aun considerando lo anterior, los modelos son capaces de apoyar positivamente los actuales procesos de oferta de la empresa, teniendo por una parte el potencial de disminuir costos tanto para la compañía como para los clientes, reflejados en la posibilidad de reducir la cantidad de contactos realizados, y por otro, la posibilidad de aumentar la rentabilidad asociada a los actuales procesos de oferta, al considerar productos que tengan relación con las necesidades del cliente, aumentando la probabilidad de que estos reaccionen de forma positiva a la oferta.

Por último, se observa que los resultados de los modelos son aplicables a la realidad de la empresa, pudiendo mantener la estructura actual de los procesos de venta. De todas formas, se propone una reestructuración de los procesos actuales, traducida en una flexibilización de la oferta, pudiendo aprovechar todo el potencial de los resultados derivados de este trabajo, que determina finalmente un aumento de los ingresos de la compañía.

5.2 Trabajos Futuros

Dadas las limitaciones y alcances de este proyecto, se reconocen diferentes aristas mediante las cuales se puede complementar el trabajo desarrollado. En este sentido, se proponen posibles trabajos a realizar en el futuro utilizando los resultados o metodologías aquí expuestas.

Un posible trabajo futuro es utilizar variables explicativas que no fueron analizadas en este proyecto, que pueden tener directa relación con la propensión de elección de un cliente. Entre estas se destacan, las variables asociadas a las promociones pasadas

realizadas al cliente, el análisis de los canales de venta utilizados y una caracterización más completa del grupo familiar.

Dado que este proyecto corresponde a una propuesta particular para resolver el problema, se propone la utilización de metodologías distintas, tales como modelos Probit, modelos que incluyan jerarquía bayesiana, redes neuronales, entre otros. Pudiendo analizar las diferencias en cuanto a la capacidad predictiva y velocidad de procesamiento de las diversas alternativas.

Otro trabajo futuro propuesto, es la realización de un análisis similar utilizando los productos que en este proyecto han sido excluidos. Además, puede ser interesante evaluar los modelos en muestras de clientes más grandes, lo que corresponde a un importante desafío en cuanto al tiempo de procesamiento de los modelos.

6. BIBLIOGRAFÍA Y FUENTES DE INFORMACIÓN

Ministerio de Transportes y Telecomunicaciones (SUBTEL), "Sector
1] Telecomunicaciones," Santiago de Chile, Diciembre 2011.

Shelby D. Hunt Robert M. Morgan, "The Commitment-Trust Theory of
2] Relationship Marketing," *Journal of Marketing*, vol. 58, Julio 1994.

Subgerencia de Inteligencia Competitiva, "Presentación Inducción
3] Telcomunicaciones y Movistar," Telefonica Chile, Santiago de Chile, 2012.

Byung-Do Kim, Scott A. Neslin Robert C. Blattberg, *Database Marketing :
4] Analyzing And Managing Customers.:* Springer, 2008.

Kenneth E. Train, "Discrete Choice Methods with Simulation," University of
5] California, Berkeley, 2003.

Nancy Lacourly, *Apuntes de Estadística.* Santiago, Chile: Escuela de Ingeniería
6] y Ciencias, Universidad de Chile, 1996.

Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth , "From Data
7] Mining to Knowledge Discovery in Database," *AI Magazine*, vol. 17, 1996.

7. ANEXOS

7.1 Anexo A

La participación de mercado de la industria, está claramente diferenciada según categoría de productos. Los siguientes gráficos muestran las diferencias en la participación de mercado de servicios de banda ancha, telefonía y televisión, en 2011.

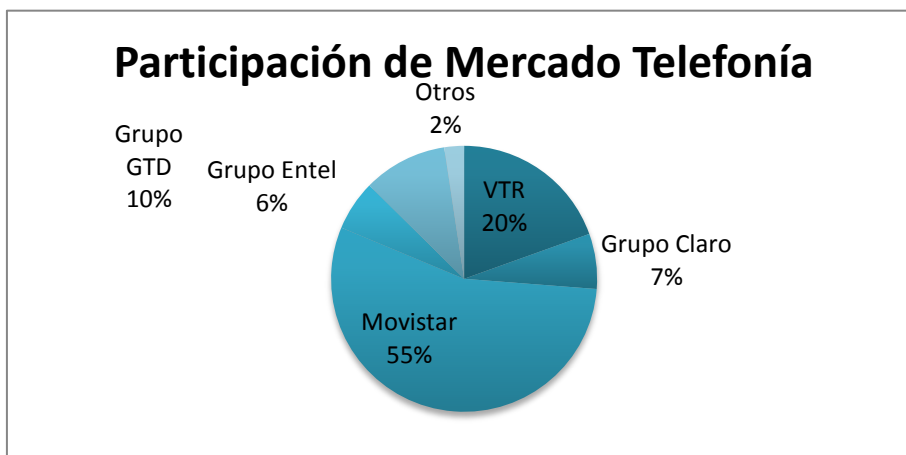


Gráfico 22: Participación de mercado servicio de telefonía

Fuente: Elaboración Propia con datos del Ministerio de Transportes y Telecomunicaciones (SUBTEL) [1].

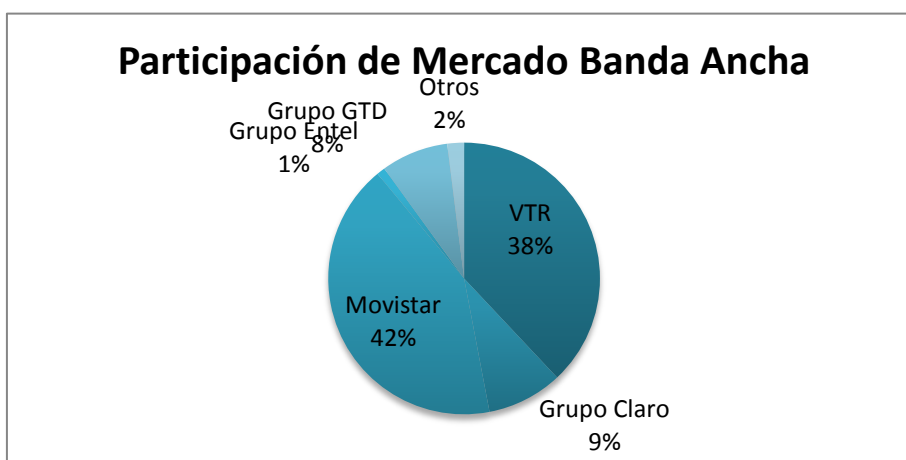


Gráfico 23: Participación de mercado servicio de banda ancha

Fuente: Elaboración Propia con datos del Ministerio de Transportes y Telecomunicaciones (SUBTEL) [1].

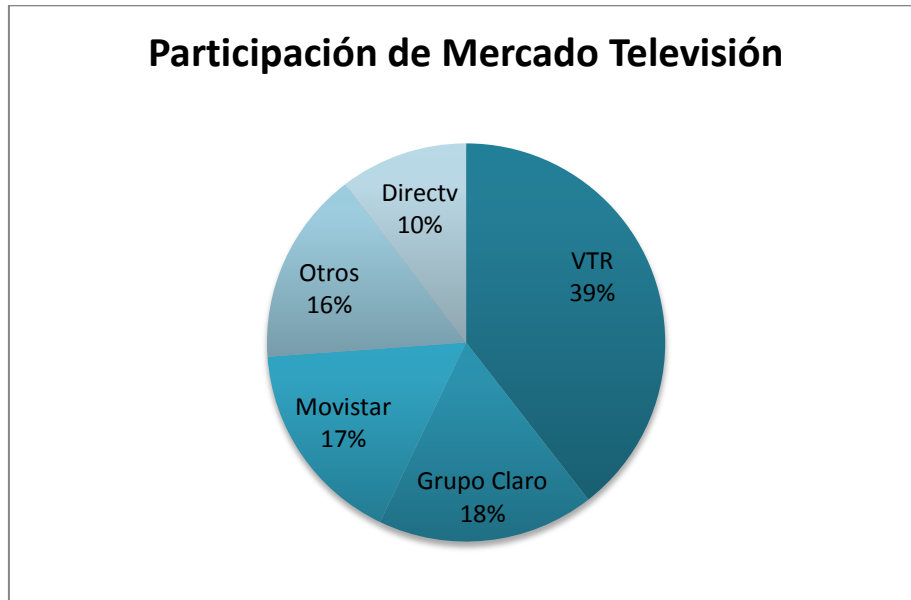


Gráfico 24: Participación de mercado servicio de televisión

Fuente: Elaboración Propia con datos del Ministerio de Transportes y Telecomunicaciones (SUBTEL) [1].

7.2 Anexo B

En los gráficos 25, 26 y 27 se muestra la distribución en términos porcentuales, asociada a las variables nominales, Grupo socioeconómico, género y segmentación preliminar respectivamente.

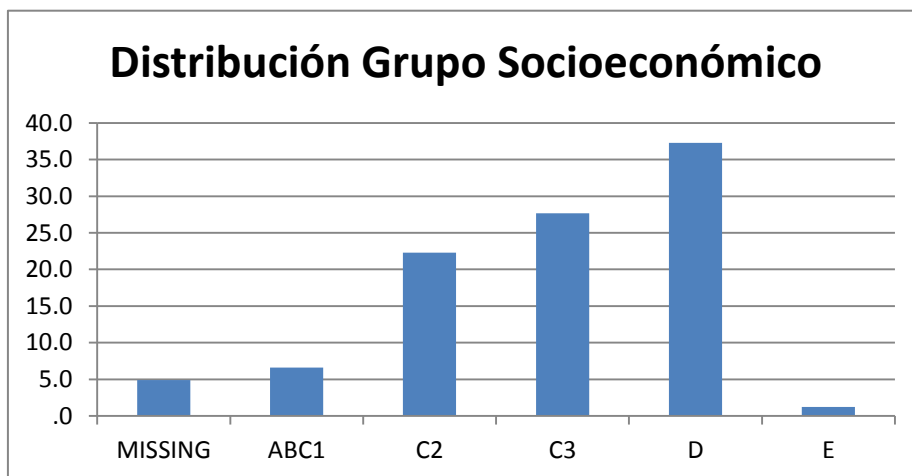


Gráfico 25: Distribución Grupo Socioeconómico

Fuente: Elaboración Propia.

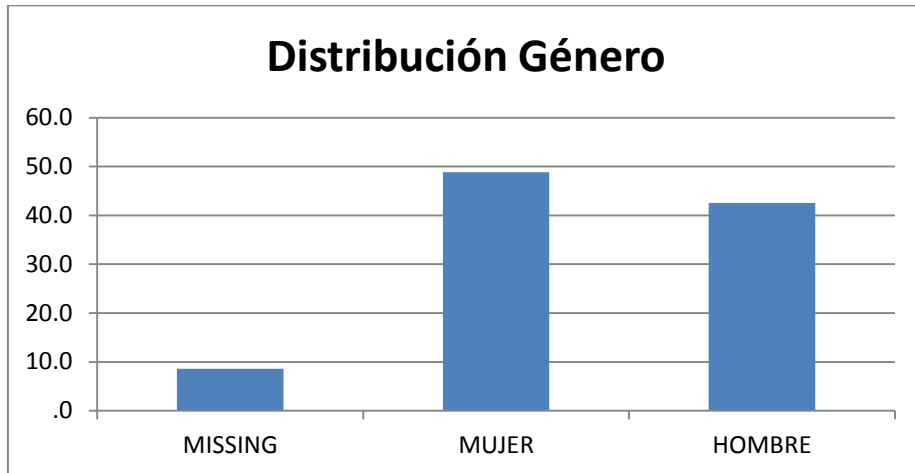


Gráfico 26: Distribución Grupo Género

Fuente: Elaboración Propia.

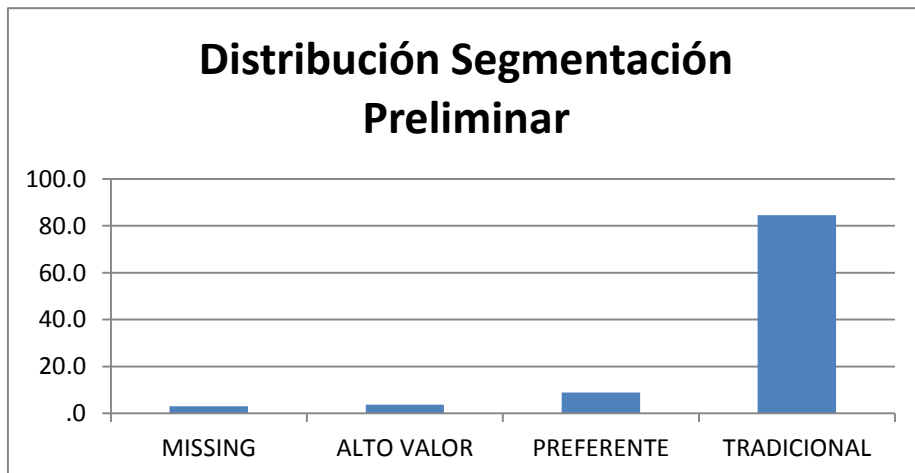


Gráfico 27: Distribución segmentación preliminar

Fuente: Elaboración Propia.

Es posible observar que existen casi un 50% de mujeres y un 42% de hombres en la data. En cuanto a la segmentación preliminar, pareciera que la gran mayoría de clientes son tradicionales, y solo algunos, corresponderían a preferentes o de alto valor.

En los siguientes gráficos, se muestra la distribución de las variables numéricas, velocidad de banda ancha, edad, gasto total por servicio, antigüedad del cliente y número de decodificadores.

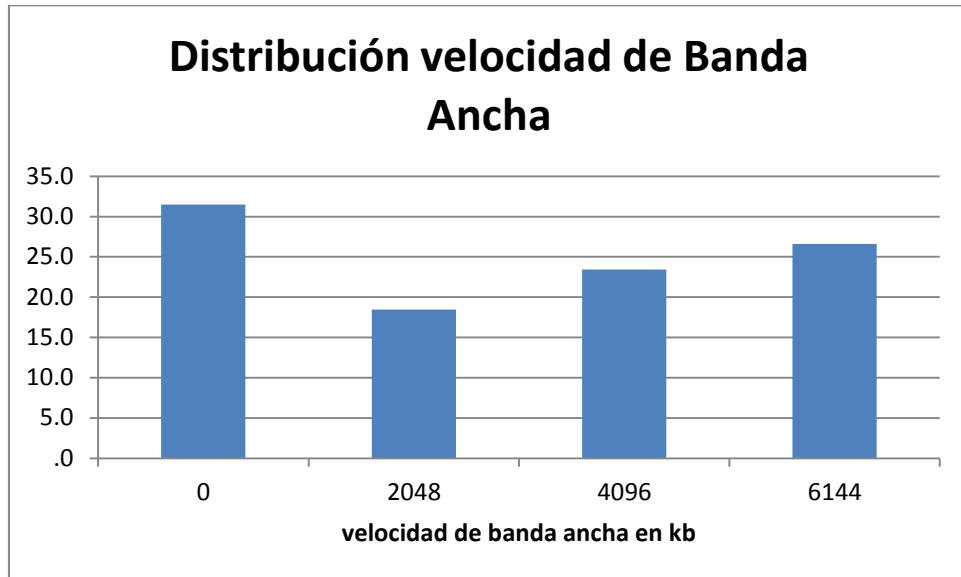


Gráfico 28: Distribución velocidad de banda ancha

Fuente: Elaboración Propia.

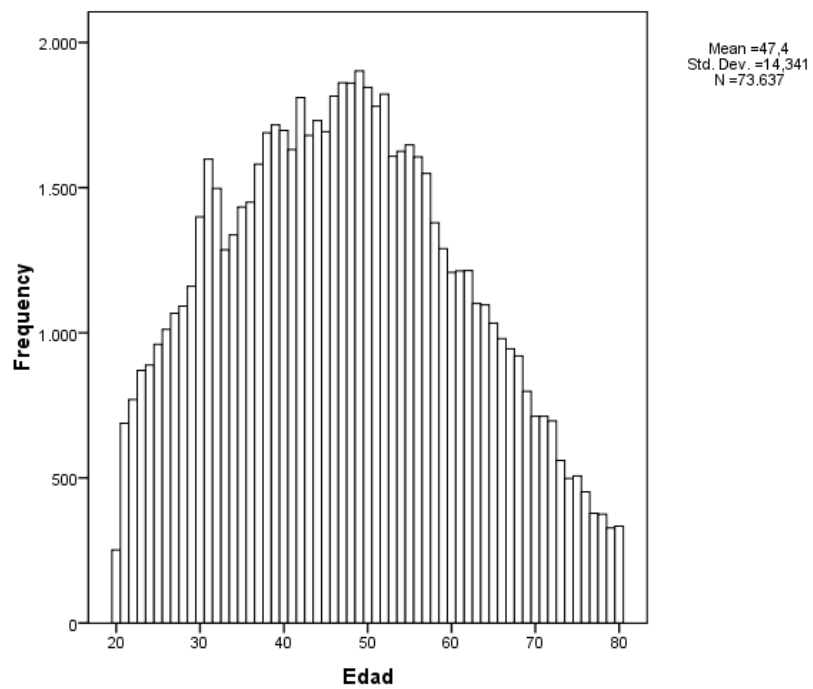


Gráfico 29: Distribución edad

Fuente: Elaboración Propia.

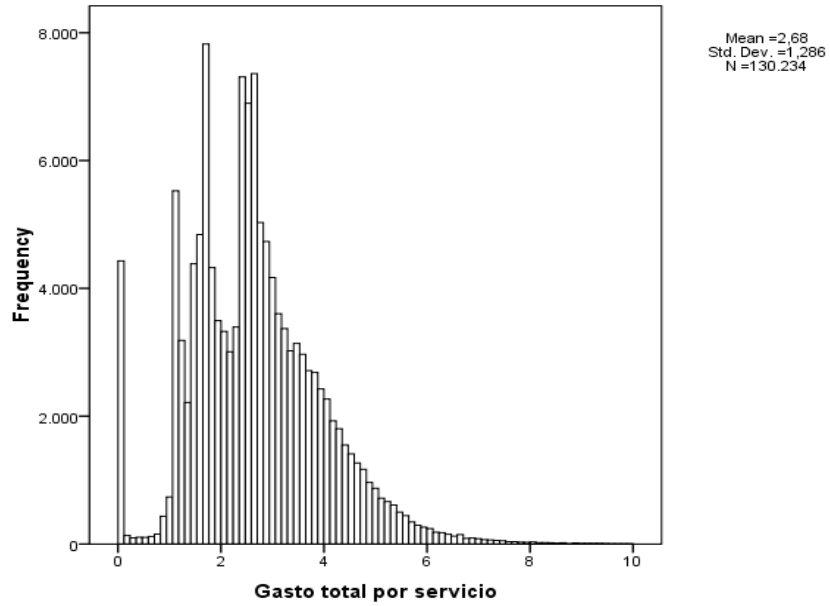


Gráfico 30: Distribución gasto total por servicio (normalizado por 10,000)

Fuente: Elaboración Propia.

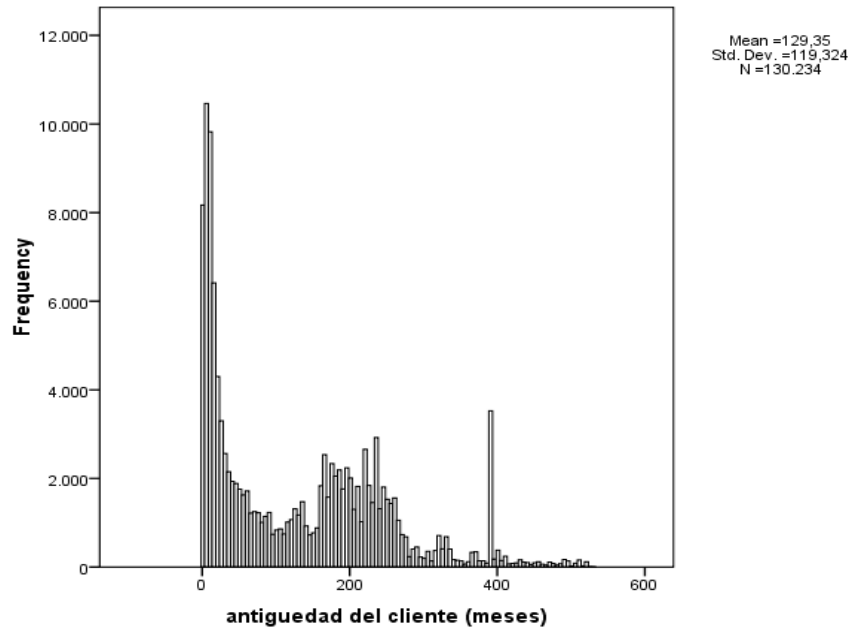


Gráfico 31: Distribución antigüedad del cliente

Fuente: Elaboración Propia.

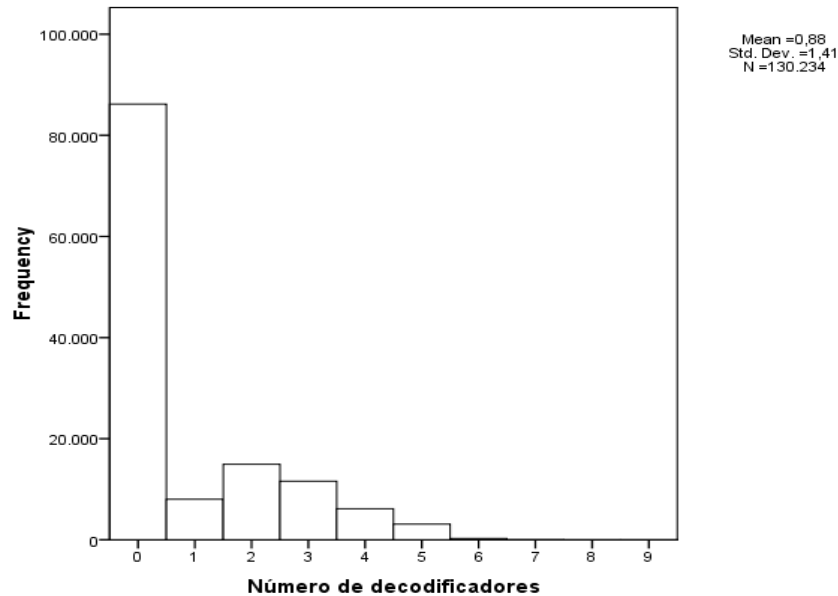


Gráfico 32: Distribución numero de decodificadores

Fuente: Elaboración Propia.

Es posible observar que la velocidad de 6 MB es la más contratada por los clientes, seguida por lo de 4, y finalmente la de 2 MB. En cuanto a la edad, esta parece estar distribuida casi de forma normal, centrada en los 47 años, los cortes en las edades de 20 y 80 años, son producidos debido a las restricciones impuestas en la eliminación de datos outliers.

En cuanto al gasto total por servicio, es posible observar que existe una gran masa de clientes que gasta hasta \$30,000, y luego la frecuencia comienza a disminuir notoriamente para los gastos superiores. En cuanto a la antigüedad del cliente parece tener una distribución bimodal, donde existe una gran masa de clientes “jóvenes” en la empresa y otra masa de clientes que están hace ya varios periodos en la compañía. Esto puede deberse a que existe un masa de clientes que está recién entrando a la compañía, y otra masa completamente fidelizada que lleva varios años al interior de la empresa.

El número de decodificadores tiene una alta frecuencia en cero, correspondiente a clientes que no tienen internet, y luego parece tener un máximo en dos decodificadores. Al parecer, las personas que tienen televisión, suelen poseer más de un decodificador, aun considerando que los planes de la empresa, vienen con un único decodificador incluido.

Como se muestra en el cuerpo del informe, la frecuencia asociada a la adquisición de canales adicionales, está directamente relacionada con la antigüedad del servicio de

telefonía que tiene el cliente. En los gráficos 33 y 34, se muestra la dependencia de los canales adicionales de cine y otros, según dicha antigüedad.

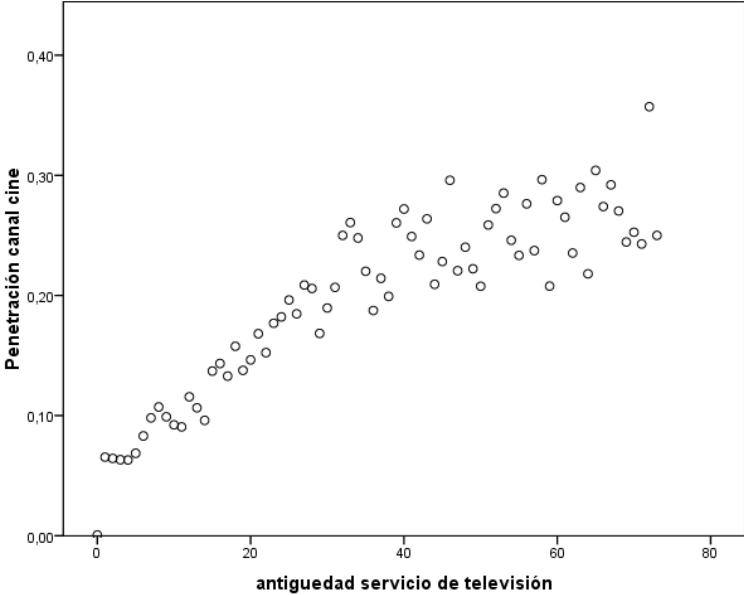


Gráfico 33: Penetración canales adicionales de cine según antigüedad del servicio de telefonía

Fuente: Elaboración Propia.

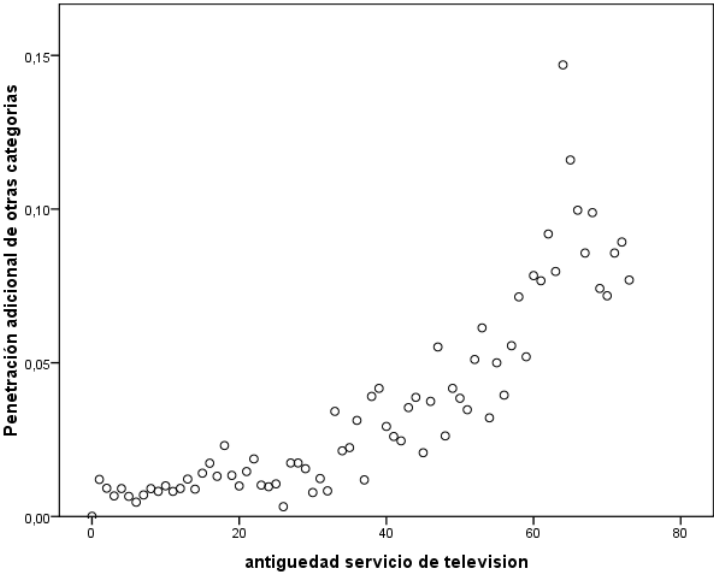


Gráfico 34: Penetración canales adicionales de otras categorías según antigüedad del servicio de telefonía

Fuente: Elaboración Propia.

Es posible observar, que al igual que ocurre con los canales de deportes, descritos en el cuerpo del informe, la antigüedad de telefonía afecta positivamente en la proporción de personas que poseen un canal adicional en las categorías de cine y otras.

7.3 Anexo C

En el cuerpo del informe se muestra los parámetros asociados al modelo de cambio (Logit Binomial) sin heterogeneidad y con heterogeneidad observada. En la siguiente tabla, se muestran los parámetros asociados al modelo final con heterogeneidad no observable introducida mediante la utilización de dos clases latentes.

Variable	Valor	Clase
Constante	-4.576484	Primera
Edad entre 20 y 40	0.197483	Primera
Edad entre 40 y 60	0.175674	Primera
GSE ABC1	-0.112019	Primera
GSE C3	0.170488	Primera
GSE D	0.101975	Primera
Antigüedad Telefonía	-0.000371	Primera
Antigüedad Televisión	-0.006891	Primera
Antigüedad Telefonía / antigüedad cliente	0.078167	Primera
Antigüedad Televisión / antigüedad cliente	0.008906	Primera
Numero de decodificadores	0.222462	Primera
Telefonía prepago antes	-1.216686	Primera
Telefonía ilimitada antes * minutos salientes	0.000075	Primera
Adicional de deporte antes	1.532279	Primera
Adicional de cine antes	0.427322	Primera
Adicional de deporte antes * número de decod.	-0.271333	Primera
Frecuencia de cambio	1.828359	Primera
Recencia último cambio	-0.201162	Primera
Recencia último cambio * frecuencia de cambio	-0.227552	Primera
Velocidad de 4MB antes	0.080743	Primera
Plan trio antes	0.833409	Primera
Duo banda ancha televisión antes	0.997664	Primera
Duo telefonía televisión antes	0.886382	Primera
Solo Banda Ancha antes	-16.709901	Primera
Solo telefonía antes	0.998208	Primera
Diferencia gasto periodos t y t-1	0.085401	Primera
Constante	-2.162576	Segunda

Edad entre 20 y 40	0.235036	Segunda
Edad entre 40 y 60	0.061320	Segunda
GSE ABC1	-0.649830	Segunda
GSE C3	0.302125	Segunda
GSE D	0.488207	Segunda
Antigüedad Telefonía	-0.001445	Segunda
Antigüedad Televisión	0.010035	Segunda
Antigüedad Telefonía / antigüedad cliente	0.333208	Segunda
Antigüedad Televisión / antigüedad cliente	0.033186	Segunda
Numero de decodificadores	0.034642	Segunda
Telefonía prepago antes	-0.541878	Segunda
Telefonía ilimitada antes * minutos salientes	0.000345	Segunda
Adicional de deporte antes	9.596096	Segunda
Adicional de cine antes	14.538167	Segunda
Adicional de deporte antes * número de decod.	-1.449319	Segunda
Frecuencia de cambio	-2.221110	Segunda
Recencia último cambio	0.059743	Segunda
Recencia último cambio * frecuencia de cambio	0.763336	Segunda
Velocidad de 4MB antes	0.258830	Segunda
Plan trio antes	0.585192	Segunda
Duo banda ancha televisión antes	0.366119	Segunda
Duo telefonía televisión antes	1.991786	Segunda
Solo Banda Ancha antes	-0.392915	Segunda
Solo telefonía antes	-0.976876	Segunda
Diferencia gasto periodos t y t-1	0.247090	Segunda
Parámetro asociada a probabilidad primera clase	2.907091	-

Tabla 19: Parámetros modelo de cambio con heterogeneidad no observable

Fuente: Elaboración Propia.

Como se ha discutido en el cuerpo del informe, los modelos de elección asumen la toma de decisiones de forma anidada, la siguiente ilustración muestra la sucesión de decisiones que realiza un cliente para la categoría TRIO.

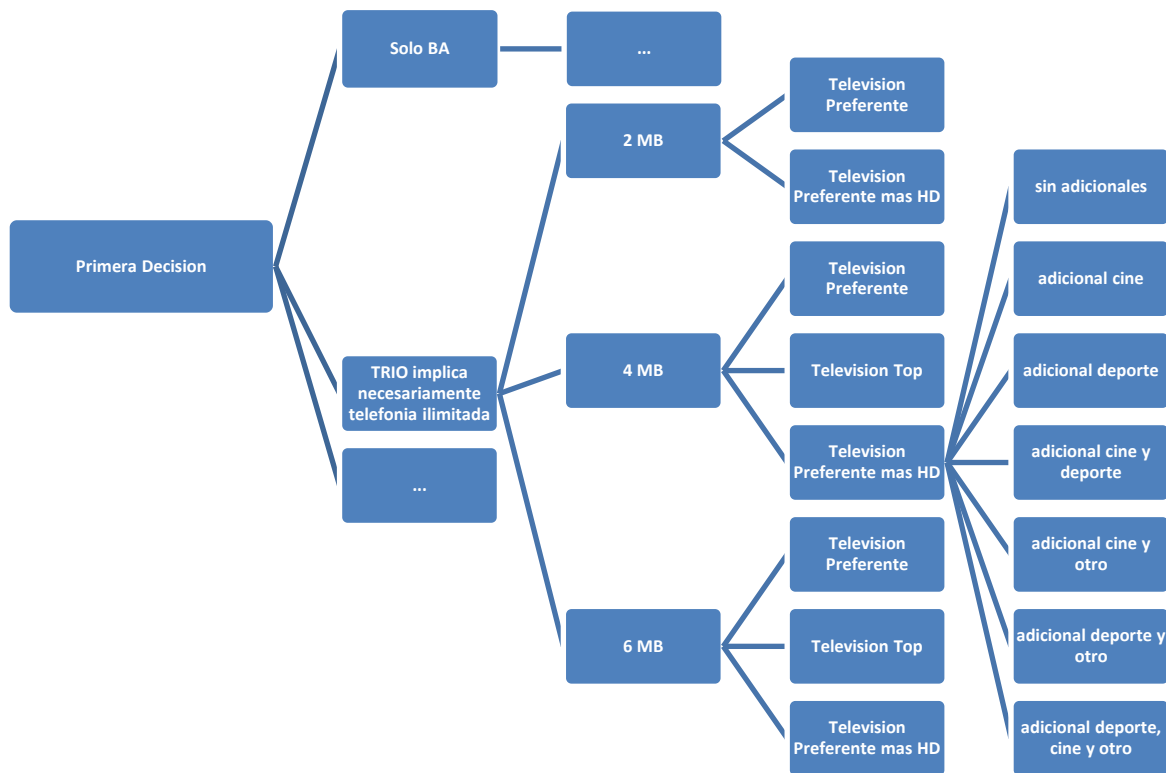


Ilustración 5: Árbol de decisión Trio

Fuente: Elaboración Propia.

Se observa en la figura anterior, las sucesiones de decisión que se asume realiza un cliente. En particular se muestra las sucesiones factibles para la categoría Trio. Es importante notar por ejemplo, que un cliente que decide una velocidad de 2 mb no podrá elegir televisión top, dadas las restricciones impuestas en la parrilla de productos por parte de la empresa.

El último nivel del árbol no contiene los canales adicionales para todos los planes de televisión, por motivos de presentación, sin embargo, estos son aplicables para todos los planes que se muestran en la ilustración.

Es relevante notar la forma en que están definidos los productos, y por lo tanto el nivel de granularidad entregados por lo modelos. Por ejemplo, se asume que una persona que posee el canal de deportes “CDF”, es exactamente igual a otra que posee un canal de deportes “FOZ SPORT”. Dado que ambos clientes habrán elegido el mismo nodo.

Dada la gran cantidad de parámetros utilizadas en el modelo de elección, no se presentan los parámetros en el cuerpo del informe para ninguno de los niveles de

heterogeneidad desarrollados, las tablas 20, 21 y 22, muestran respectivamente los parámetros para los modelos de elección sin heterogeneidad, con heterogeneidad observable y con heterogeneidad no observable utilizando 2 clases latentes.

Variable	Valor	Desviación Estándar	P-valor
Precio	-0.3824	0.0169	0.00
Logaritmo de la Velocidad	0.5634	0.0184	0.00
Televisión Top	5.5727	0.2694	0.00
Televisión Preferente	5.0719	0.2078	0.00
Decodificador HD	0.4923	0.1409	0.00
Adicional de deporte	1.1703	0.0970	0.00
Adicional de cine	0.8854	0.0949	0.00
Adicional de otra categoría	-2.3067	0.1003	0.00
Plan ilimitado de minutos	5.5215	0.1654	0.00
Plan controlado de 350 min	3.3267	0.2133	0.00
Plan controlado de 150 o 250 min	3.8897	0.1438	0.00
Plan prepago de 5000 o 6000	-0.2801	0.1887	0.14

Tabla 20: Parámetros modelos de elección sin heterogeneidad

Fuente: Elaboración Propia.

Variable	Valor	Desviación Estándar	P-valor
Precio	-0.5191	0.0280	0.00
Precio * diferencia de gasto t-2 a t-1	-0.0053	0.0032	0.10
Precio * gasto total en t -1	0.0128	0.0024	0.00
Precio * gse C2	-0.0191	0.0063	0.00
Precio * gse C3	-0.0076	0.0049	0.12
Precio * edad entre 40 y 60	0.0079	0.0045	0.08
Precio * reclamos en periodos anteriores	0.0024	0.0007	0.00
Logaritmo de la Velocidad	0.2799	0.0829	0.00
Logaritmo de la Velocidad * antigüedad de internet (meses)	0.0011	0.0005	0.04
Logaritmo de la Velocidad * antigüedad de telefonía (meses)	-0.0004	0.0001	0.00
Logaritmo de la Velocidad * antigüedad de telefonía / antigüedad cliente	0.1410	0.0321	0.00
Logaritmo de la Velocidad * velocidad de banda ancha antes	-1.0275	0.1850	0.00
Logaritmo de la Velocidad * velocidad de banda ancha antes al cuadrado	0.1928	0.0272	0.00
Logaritmo de la Velocidad * telefonía ilimitada antes	0.3111	0.0767	0.00
Logaritmo de la Velocidad * telefonía controlada antes	0.2319	0.0747	0.00
Logaritmo de la Velocidad * televisión preferente antes	-0.5285	0.0390	0.00

Logaritmo de la Velocidad * television top antes	-0.6387	0.1118	0.00
Logaritmo de la Velocidad * trio antes	1.6862	0.2696	0.00
Logaritmo de la Velocidad * duo ba-tel antes	1.2869	0.2706	0.00
Logaritmo de la Velocidad * duo ba-tv antes	2.1863	0.2795	0.00
Logaritmo de la Velocidad * solo ba antes	2.1217	0.3101	0.00
Logaritmo de la Velocidad * gse C2	0.0295	0.0250	0.24
Television Top	7.3055	0.4775	0.00
Television Top * antigüedad de television / antigüedad cliente	0.5254	0.2781	0.06
Television Top * antigüedad de telefonía / antigüedad cliente	-1.2477	0.2734	0.00
Television Top * numero de decodificadores al cuadrado	0.0961	0.0199	0.00
Television Top * television preferente antes	-0.4540	0.0988	0.00
Television Top * decodificador HD	1.6779	0.2537	0.00
Television Preferente	4.2745	0.3602	0.00
Television Preferente * antigüedad de banda ancha (meses)	-0.0035	0.0019	0.07
Television Preferente * antigüedad de telefonía (meses)	-0.0014	0.0006	0.01
Television Preferente * numero de decodificadores	0.2053	0.0516	0.00
Television Preferente * television preferente antes	2.7813	0.3294	0.00
Television Preferente * television top antes	1.0647	0.4458	0.02
Television Preferente * canal deporte anes	0.5936	0.1046	0.00
Television Preferente * gse ABC1	0.2165	0.2029	0.29
Television Preferente * trio antes	-1.0119	0.2774	0.00
Television Preferente * duo ba-tv antes	-0.6803	0.3040	0.03
Television Preferente * duo ba-tel antes	1.6570	0.1985	0.00
Television Preferente * solo ba antes	1.4620	0.2677	0.00
Decodificador HD	1.5092	0.2557	0.00
Decodificador HD * antigüedad de television (meses)	0.0113	0.0025	0.00
Decodificador HD * telefonía controlada antes	-1.2242	0.7221	0.09
Decodificador HD * canal cine antes * numero de decodificadores	0.1756	0.0350	0.00
Decodificador HD * gse D	-1.1175	0.1632	0.00
Decodificador HD * gse C3	-0.6983	0.1720	0.00
Decodificador HD * gse C2	-0.3855	0.1864	0.04
Decodificador HD * gse E	-1.7141	0.7398	0.02
Adicional de deporte	1.5318	0.1577	0.00
Adicional de deporte * antigüedad de television (meses)	0.0101	0.0020	0.00
Adicional de deporte * antigüedad de telefonía (meses)	-0.0012	0.0005	0.02
Adicional de deporte * canal otro antes	0.4596	0.1948	0.02
Adicional de deporte * solo television antes	0.4854	0.2066	0.02
Adicional de deporte * duo ba-tv antes	0.4954	0.1662	0.00
Adicional de deporte * duo ba-tel antes	-0.5330	0.1373	0.00
Adicional de cine	-1.1761	0.2978	0.00

Adicional de cine * canal cine antes	0.5669	0.0880	0.00
Adicional de cine * canal deporte antes * numero de decodificadores	-0.0771	0.0236	0.00
Adicional de cine * canal otro antes * numero de decodificadores	0.1484	0.0587	0.01
Adicional de cine * television top antes	2.4095	0.4451	0.00
Adicional de cine * television preferente antes	2.6761	0.3735	0.00
Adicional de cine * edad entre 20 y 40	0.1145	0.0848	0.18
Adicional de cine * solo television antes	0.3296	0.3370	0.33
Adicional de cine * duo ba-tv antes	-0.5481	0.3241	0.09
Adicional de otra categoria	-3.4881	0.3847	0.00
Adicional de otra categoria * antigüedad de television (meses)	0.0084	0.0046	0.07
Adicional de otra categoria * canal otro antes	3.6101	0.2339	0.00
Adicional de otra categoria * canal deporte antes * numero de decodificadores	0.0999	0.0580	0.09
Adicional de otra categoria * solo television antes	1.5336	0.5700	0.01
Adicional de otra categoria * trio antes	0.5395	0.3998	0.18
Adicional de otra categoria * duo ba-tv antes	1.3014	0.5617	0.02
Plan ilimitado de minutos	6.2341	0.4007	0.00
Plan ilimitado de minutos * antigüedad telefonia / antigüedad cliente	1.2632	0.3142	0.00
Plan ilimitado de minutos * antigüedad de banda ancha (meses)	0.0253	0.0060	0.00
Plan ilimitado de minutos * antigüedad de telefonia (meses)	0.0081	0.0015	0.00
Plan ilimitado de minutos * antigüedad de banda ancha / antigüedad de cliente	-0.8599	0.2792	0.00
Plan ilimitado de minutos * antigüedad de television (meses)	-0.0394	0.0061	0.00
Plan ilimitado de minutos * velocidad de banda ancha antes	-1.0615	0.2133	0.00
Plan ilimitado de minutos * velocidad de banda ancha antes al cuadrado	0.1004	0.0284	0.00
Plan ilimitado de minutos * telefonia ilimitada antes	2.6263	0.2883	0.00
Plan ilimitado de minutos * duo ba-tv antes	-1.9024	0.3444	0.00
Plan ilimitado de minutos * solo television antes	-4.6529	1.0609	0.00
Plan ilimitado de minutos * trio antes	1.1524	0.2510	0.00
Plan controlado de 350 min	3.2126	0.5165	0.00
Plan controlado de 350 min * edad entre 20 y 40	-1.2433	0.5676	0.03
Plan controlado de 350 min * antigüedad telefonia / antigüedad cliente	4.5709	0.5179	0.00
Plan controlado de 350 min * antigüedad television / antigüedad cliente	-2.5629	0.5957	0.00
Plan controlado de 150 o 250 min	2.4894	0.6136	0.00
Plan controlado de 150 o 250 min * antigüedad telefonia / antigüedad cliente	3.5524	0.5381	0.00
Plan controlado de 150 o 250 min * antigüedad cliente (meses)	0.0071	0.0015	0.00
Plan controlado de 150 o 250 min * antigüedad television / antigüedad cliente	-0.2519	0.0661	0.00
Plan controlado de 150 o 250 min * velocidad de banda ancha antes	-6.1624	0.9984	0.00
Plan controlado de 150 o 250 min * velocidad de banda ancha al cuadrado	1.3025	0.2070	0.00
Plan controlado de 150 o 250 min * minutos salientes antes	-0.0027	0.0007	0.00
Plan controlado de 150 o 250 min * telefonia ilimitada antes	4.3561	0.4129	0.00
Plan prepago de 5000 o 6000	-0.0164	0.4803	0.97

Plan prepago de 5000 o 6000 * antigüedad de telefonía / antigüedad cliente	2.4917	0.5775	0.00
Plan prepago de 5000 o 6000 * televisión preferente antes	-3.3514	0.8232	0.00
Plan prepago de 5000 o 6000 * minutos salientes antes	-0.0025	0.0015	0.10

Tabla 21: Parámetros modelos de elección con heterogeneidad observada

Fuente: Elaboración Propia.

Variable	Valor	Clase
Precio	-1.1169	Primera
Precio * diferencia de gasto t-2 a t-1	-0.0073	Primera
Precio * gasto total en t-1	-0.0137	Primera
Precio * gse C2	-0.0115	Primera
Precio * gse C3	-0.0072	Primera
Precio * edad entre 40 y 60	0.0002	Primera
Precio * reclamos en periodos anteriores	-0.0001	Primera
Logaritmo de la Velocidad	0.6712	Primera
Logaritmo de la Velocidad * antigüedad de internet (meses)	0.0024	Primera
Logaritmo de la Velocidad * antigüedad de telefonía (meses)	-0.0016	Primera
Logaritmo de la Velocidad * antigüedad de telefonía / antigüedad cliente	0.3263	Primera
Logaritmo de la Velocidad * velocidad de banda ancha antes	-0.3182	Primera
Logaritmo de la Velocidad * velocidad de banda ancha antes al cuadrado	0.0392	Primera
Logaritmo de la Velocidad * telefonía ilimitada antes	0.2202	Primera
Logaritmo de la Velocidad * telefonía controlada antes	0.8225	Primera
Logaritmo de la Velocidad * televisión preferente antes	-0.2460	Primera
Logaritmo de la Velocidad * televisión top antes	-0.3614	Primera
Logaritmo de la Velocidad * trio antes	0.9001	Primera
Logaritmo de la Velocidad * duo ba-tel antes	0.9777	Primera
Logaritmo de la Velocidad * duo ba-tv antes	1.4832	Primera
Logaritmo de la Velocidad * solo ba antes	2.1209	Primera
Logaritmo de la Velocidad * gse C2	0.0209	Primera
Televisión Top	18.4181	Primera
Televisión Top * antigüedad de televisión / antigüedad cliente	0.0802	Primera
Televisión Top * antigüedad de telefonía / antigüedad cliente	0.0279	Primera
Televisión Top * número de decodificadores al cuadrado	0.0894	Primera
Televisión Top * televisión preferente antes	0.2164	Primera
Televisión Top * decodificador HD	12.7071	Primera
Televisión Preferente	4.3561	Primera
Televisión Preferente * antigüedad de banda ancha (meses)	0.0008	Primera
Televisión Preferente * antigüedad de telefonía (meses)	-0.0042	Primera

Television Preferente * numero de decodificadores	0.0164	Primera
Television Preferente * television preferente antes	3.4470	Primera
Television Preferente * television top antes	0.1402	Primera
Television Preferente * canal deporte anes	-0.2454	Primera
Television Preferente * gse ABC1	0.2956	Primera
Television Preferente * trio antes	0.8975	Primera
Television Preferente * duo ba-tv antes	0.8283	Primera
Television Preferente * duo ba-tel antes	1.7460	Primera
Television Preferente * solo ba antes	3.6115	Primera
Decodificador HD	-1.4990	Primera
Decodificador HD * antigüedad de television (meses)	0.0215	Primera
Decodificador HD * telefonia controlada antes	-1.6334	Primera
Decodificador HD * canal cine antes * numero de decodificadores	0.0089	Primera
Decodificador HD * gse D	-0.8543	Primera
Decodificador HD * gse C3	0.0994	Primera
Decodificador HD * gse C2	-0.5478	Primera
Decodificador HD * gse E	-3.4456	Primera
Adicional de deporte	13.2003	Primera
Adicional de deporte * antigüedad de television (meses)	0.0115	Primera
Adicional de deporte * antigüedad de telefonia (meses)	0.0002	Primera
Adicional de deporte * canal otro antes	0.7787	Primera
Adicional de deporte * solo television antes	0.7154	Primera
Adicional de deporte * duo ba-tv antes	0.4903	Primera
Adicional de deporte * duo ba-tel antes	-2.1912	Primera
Adicional de cine	3.9050	Primera
Adicional de cine * canal cine antes	4.7722	Primera
Adicional de cine * canal deporte antes * numero de decodificadores	0.8025	Primera
Adicional de cine * canal otro antes * numero de decodificadores	0.4899	Primera
Adicional de cine * television top antes	12.9204	Primera
Adicional de cine * television preferente antes	1.8875	Primera
Adicional de cine * edad entre 20 y 40	0.1156	Primera
Adicional de cine * solo television antes	0.2266	Primera
Adicional de cine * duo ba-tv antes	-0.7886	Primera
Adicional de otra categoria	-0.7133	Primera
Adicional de otra categoria * antigüedad de television (meses)	-0.0181	Primera
Adicional de otra categoria * canal otro antes	3.1996	Primera
Adicional de otra categoria * canal deporte antes * numero de decodificadores	0.0545	Primera
Adicional de otra categoria * solo television antes	3.1260	Primera
Adicional de otra categoria * trio antes	1.6229	Primera
Adicional de otra categoria * duo ba-tv antes	0.6818	Primera

Plan ilimitado de minutos	18.3576	Primera
Plan ilimitado de minutos * antigüedad telefonía / antigüedad cliente	-0.0993	Primera
Plan ilimitado de minutos * antigüedad de banda ancha (meses)	0.0689	Primera
Plan ilimitado de minutos * antigüedad de telefonía (meses)	0.0164	Primera
Plan ilimitado de minutos * antigüedad de banda ancha / antigüedad de cliente	-1.8678	Primera
Plan ilimitado de minutos * antigüedad de televisión (meses)	-0.1458	Primera
Plan ilimitado de minutos * velocidad de banda ancha antes	-3.2567	Primera
Plan ilimitado de minutos * velocidad de banda ancha antes al cuadrado	-0.0106	Primera
Plan ilimitado de minutos * telefonía ilimitada antes	7.3129	Primera
Plan ilimitado de minutos * duo ba-tv antes	7.7478	Primera
Plan ilimitado de minutos * solo televisión antes	-8.4290	Primera
Plan ilimitado de minutos * trio antes	8.2443	Primera
Plan controlado de 350 min	2.3689	Primera
Plan controlado de 350 min * edad entre 20 y 40	-1.3505	Primera
Plan controlado de 350 min * antigüedad telefonía / antigüedad cliente	9.1505	Primera
Plan controlado de 350 min * antigüedad televisión / antigüedad cliente	-0.4376	Primera
Plan controlado de 150 o 250 min	-0.8115	Primera
Plan controlado de 150 o 250 min * antigüedad telefonía / antigüedad cliente	8.3149	Primera
Plan controlado de 150 o 250 min * antigüedad cliente (meses)	0.0150	Primera
Plan controlado de 150 o 250 min * antigüedad televisión / antigüedad cliente	-0.0093	Primera
Plan controlado de 150 o 250 min * velocidad de banda ancha antes	-8.2297	Primera
Plan controlado de 150 o 250 min * velocidad de banda ancha al cuadrado	0.7562	Primera
Plan controlado de 150 o 250 min * minutos salientes antes	0.0014	Primera
Plan controlado de 150 o 250 min * telefonía ilimitada antes	24.3040	Primera
Plan prepago de 5000 o 6000	0.0004	Primera
Plan prepago de 5000 o 6000 * antigüedad de telefonía / antigüedad cliente	-0.1974	Primera
Plan prepago de 5000 o 6000 * televisión preferente antes	-13.1695	Primera
Plan prepago de 5000 o 6000 * minutos salientes antes	0.0003	Primera
Precio	-0.5015	Segunda
Precio * diferencia de gasto t-2 a t-1	-0.0018	Segunda
Precio * gasto total en t -1	0.0131	Segunda
Precio * gse C2	-0.0189	Segunda
Precio * gse C3	-0.0081	Segunda
Precio * edad entre 40 y 60	0.0111	Segunda
Precio * reclamos en periodos anteriores	0.0027	Segunda
Logaritmo de la Velocidad	0.3035	Segunda
Logaritmo de la Velocidad * antigüedad de internet (meses)	0.0005	Segunda
Logaritmo de la Velocidad * antigüedad de telefonía (meses)	-0.0002	Segunda
Logaritmo de la Velocidad * antigüedad de telefonía / antigüedad cliente	0.0921	Segunda
Logaritmo de la Velocidad * velocidad de banda ancha antes	-6.5476	Segunda

Logaritmo de la Velocidad * velocidad de banda ancha antes al cuadrado	1.1327	Segunda
Logaritmo de la Velocidad * telefonia ilimitada antes	0.3065	Segunda
Logaritmo de la Velocidad * telefonia controlada antes	0.2226	Segunda
Logaritmo de la Velocidad * television preferente antes	-0.5755	Segunda
Logaritmo de la Velocidad * television top antes	-0.7119	Segunda
Logaritmo de la Velocidad * trio antes	9.0072	Segunda
Logaritmo de la Velocidad * duo ba-tel antes	8.8150	Segunda
Logaritmo de la Velocidad * duo ba-tv antes	9.4326	Segunda
Logaritmo de la Velocidad * solo ba antes	9.4974	Segunda
Logaritmo de la Velocidad * gse C2	0.0299	Segunda
Television Top	6.6677	Segunda
Television Top * antigüedad de television / antigüedad cliente	0.3832	Segunda
Television Top * antigüedad de telefonia / antigüedad cliente	-1.7319	Segunda
Television Top * numero de decodificadores al cuadrado	0.0486	Segunda
Television Top * television preferente antes	-0.8196	Segunda
Television Top * decodificador HD	3.0227	Segunda
Television Preferente	3.9581	Segunda
Television Preferente * antigüedad de banda ancha (meses)	-0.0052	Segunda
Television Preferente * antigüedad de telefonia (meses)	-0.0010	Segunda
Television Preferente * numero de decodificadores	0.2428	Segunda
Television Preferente * television preferente antes	3.3231	Segunda
Television Preferente * television top antes	1.3389	Segunda
Television Preferente * canal deporte anes	0.7798	Segunda
Television Preferente * gse ABC1	0.4012	Segunda
Television Preferente * trio antes	-1.4395	Segunda
Television Preferente * duo ba-tv antes	-1.0552	Segunda
Television Preferente * duo ba-tel antes	2.0917	Segunda
Television Preferente * solo ba antes	1.4175	Segunda
Decodificador HD	1.4248	Segunda
Decodificador HD * antigüedad de television (meses)	0.0109	Segunda
Decodificador HD * telefonia controlada antes	-1.3350	Segunda
Decodificador HD * canal cine antes * numero de decodificadores	0.1794	Segunda
Decodificador HD * gse D	-1.1350	Segunda
Decodificador HD * gse C3	-0.7364	Segunda
Decodificador HD * gse C2	-0.2859	Segunda
Decodificador HD * gse E	-1.5226	Segunda
Adicional de deporte	1.0038	Segunda
Adicional de deporte * antigüedad de television (meses)	0.0133	Segunda
Adicional de deporte * antigüedad de telefonia (meses)	-0.0018	Segunda
Adicional de deporte * canal otro antes	0.5560	Segunda

Adicional de deporte * solo television antes	0.5620	Segunda
Adicional de deporte * duo ba-tv antes	0.6653	Segunda
Adicional de deporte * duo ba-tel antes	-0.7539	Segunda
Adicional de cine	-1.8741	Segunda
Adicional de cine * canal cine antes	0.5994	Segunda
Adicional de cine * canal deporte antes * numero de decodificadores	-0.2161	Segunda
Adicional de cine * canal otro antes * numero de decodificadores	0.2464	Segunda
Adicional de cine * television top antes	2.6906	Segunda
Adicional de cine * television preferente antes	3.1669	Segunda
Adicional de cine * edad entre 20 y 40	0.0887	Segunda
Adicional de cine * solo television antes	0.4146	Segunda
Adicional de cine * duo ba-tv antes	-0.6317	Segunda
Adicional de otra categoria	-3.9955	Segunda
Adicional de otra categoria * antigüedad de television (meses)	0.0154	Segunda
Adicional de otra categoria * canal otro antes	4.1954	Segunda
Adicional de otra categoria * canal deporte antes * numero de decodificadores	0.0350	Segunda
Adicional de otra categoria * solo television antes	1.4692	Segunda
Adicional de otra categoria * trio antes	0.5264	Segunda
Adicional de otra categoria * duo ba-tv antes	1.5777	Segunda
Plan ilimitado de minutos	5.4868	Segunda
Plan ilimitado de minutos * antigüedad telefonia / antigüedad cliente	2.6025	Segunda
Plan ilimitado de minutos * antigüedad de banda ancha (meses)	0.0280	Segunda
Plan ilimitado de minutos * antigüedad de telefonia (meses)	0.0074	Segunda
Plan ilimitado de minutos * antigüedad de banda ancha / antigüedad de cliente	-1.3959	Segunda
Plan ilimitado de minutos * antigüedad de television (meses)	-0.0403	Segunda
Plan ilimitado de minutos * velocidad de banda ancha antes	-1.2835	Segunda
Plan ilimitado de minutos * velocidad de banda ancha antes al cuadrado	0.1830	Segunda
Plan ilimitado de minutos * telefonia ilimitada antes	3.6468	Segunda
Plan ilimitado de minutos * duo ba-tv antes	-2.5120	Segunda
Plan ilimitado de minutos * solo television antes	-3.7875	Segunda
Plan ilimitado de minutos * trio antes	0.1284	Segunda
Plan controlado de 350 min	3.3890	Segunda
Plan controlado de 350 min * edad entre 20 y 40	-1.1832	Segunda
Plan controlado de 350 min * antigüedad telefonia / antigüedad cliente	5.2577	Segunda
Plan controlado de 350 min * antigüedad television / antigüedad cliente	-3.2254	Segunda
Plan controlado de 150 o 250 min	2.3643	Segunda
Plan controlado de 150 o 250 min * antigüedad telefonia / antigüedad cliente	4.7133	Segunda
Plan controlado de 150 o 250 min * antigüedad cliente (meses)	0.0072	Segunda
Plan controlado de 150 o 250 min * antigüedad television / antigüedad cliente	-2.1551	Segunda
Plan controlado de 150 o 250 min * velocidad de banda ancha antes	-26.9549	Segunda

Plan controlado de 150 o 250 min * velocidad de banda ancha al cuadrado	-4.2079	Segunda
Plan controlado de 150 o 250 min * minutos salientes antes	-0.0039	Segunda
Plan controlado de 150 o 250 min * telefonía ilimitada antes	5.2444	Segunda
Plan prepago de 5000 o 6000	0.0034	Segunda
Plan prepago de 5000 o 6000 * antigüedad de telefonía / antigüedad cliente	3.4609	Segunda
Plan prepago de 5000 o 6000 * televisión preferente antes	-3.3152	Segunda
Plan prepago de 5000 o 6000 * minutos salientes antes	-0.0034	Segunda
Parametro asociado a probabilidad de primera clase	-2.3083	-

Tabla 22: Parámetros modelos de elección con heterogeneidad no observada

Fuente: Elaboración Propia.