



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

ASIGNACIÓN ÓPTIMA DE PRODUCTOS A OFRECER EN UNA  
CORREDORA DE SEGUROS, EN BASE A MODELOS DE  
PROPENSIÓN DE COMPRA POR ATRIBUTOS DE LOS  
PRODUCTOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL  
INDUSTRIAL

JOSÉ TOMÁS LAGOS AVID

PROFESOR GUÍA:  
PABLO MARÍN VICUÑA

MIEMBROS DE LA COMISIÓN:  
LUIS ABURTO LAFOURCADE  
CAROLINA SEGOVIA RIQUELME

SANTIAGO DE CHILE  
2019

**RESUMEN DE LA MEMORIA PARA OPTAR AL  
TÍTULO DE:** Ingeniero Civil Industrial  
**POR:** José Tomás Lagos Avid  
**FECHA:** 26/03/19  
**PROFESOR GUÍA:** Pablo Marín Vicuña

**ASIGNACIÓN ÓPTIMA DE PRODUCTOS A OFRECER EN UNA CORREDORA DE SEGUROS, EN BASE A MODELOS DE PROPENSIÓN DE COMPRA POR ATRIBUTOS DE LOS PRODUCTOS**

El presente proyecto se realiza en Seguros Falabella, corredora de seguros que pertenece al grupo Falabella. Está enfocado en el proceso de generación de bases para el canal Call-center, el cual durante el año 2017 concreto el 8% de las ventas de la compañía.

El objetivo principal de este informe es aumentar las ventas del Call-center, mediante la asignación óptima de los productos a ofrecer a los clientes. Para ello se evalúan distintos modelos de propensión con el fin de estimar la probabilidad de compra de los clientes a distintos productos. Posterior a ello, se elabora un modelo de optimización, que realiza la asignación según la propensión de compra de los clientes y las restricciones del canal, de tal manera de maximizar las ventas esperadas.

Se evalúan 4 metodologías para estimar la propensión de compra de los clientes. Se escogió un modelo Logit que incluye los atributos específicos de los productos, puesto que entrega mayor flexibilidad y un buen ajuste, capturando el 20% de las ventas en el primer decil. Se validó la selección de este modelo, con un producto con el cual no fue calibrado, en donde se observa que el primer decil es capaz de capturar el 19% de estas ventas.

Posterior a la selección del modelo de propensión, se construyó un modelo de optimización que realiza la asignación de cual producto ofrecer a cada cliente, el cual incorpora las restricciones del canal. Mediante una simulación se evalúa el impacto de la metodología propuesta, en donde se obtuvo un aumento en las ventas esperadas de un 8%. Además, se propuso una asignación más completa, que busque maximizar la rentabilidad del canal relajando sus restricciones. De esta manera, se comprueba que no es óptimo llamar a toda la base, sino que solo al 83% de esta.

La formulación de los modelos de propensiones presentados en este trabajo, permitieron evaluar el potencial impacto de nuevas configuraciones de productos con el fin de recomendar la generación de nuevos productos. Se determinó que todas las configuraciones posibles de atributos generaban un aumento en las ventas, y el producto que solo contaba con el atributo muerte natural es el que causa mayor impacto, aumentando las ventas mensuales esperadas en 1,1%.

Finalmente, para estudios futuros, se recomienda evaluar esta metodología de manera multi canal, de manera de buscar un mayor impacto al determinar el mejor canal por el cual se debe contactar al cliente.

# Agradecimientos

A mis padres, Eduardo y Suraia, y mi hermana, María Paz, por su incondicional apoyo en todo momento, acompañándome en las buenas y en las malas, para que este proyecto saliera adelante.

A mis amigos, por su preocupación y distracción en los momentos que lo necesitaba. En especial a Javier, con quien estuvimos estudiando juntos desde el colegio hasta el fin de la universidad

Al resto de mi familia, que, si bien puede pasar mucho tiempo sin que nos veamos, siempre han estado dispuesto a ayudarme.

A todas esas increíbles personas que he conocido en esta etapa de mi vida, aunque muchas de ellas no las conocí en la universidad, se convirtieron en pilares importantes para seguir adelante.

Al área de BI de Seguros Falabella, quienes no solo fueron compañeros, también se convirtieron en mis profesores laborales y en grandes amigos, ayudándome a realizar esta memoria.

A los profesores de la sección, por su paciencia y disposición en ayudarme y guiarme, durante esta etapa final. En especial a Pablo, quien no solo se preocupó de que guiarme en la memoria, también se preocupó de enseñarme a ser un buen profesional.

A todos ustedes y a muchos mas:

Gracias

# Tabla de contenido

<b>1</b>	<b>INTRODUCCIÓN</b>	<b>1</b>
1.1	CARACTERIZACIÓN DE LA EMPRESA	1
1.2	MERCADO	2
<b>2</b>	<b>JUSTIFICACIÓN DEL PROBLEMA</b>	<b>4</b>
2.1	IDENTIFICACIÓN DE LA OPORTUNIDAD DE MEJORA.	4
2.2	HIPÓTESIS Y ALTERNATIVA SOLUCIÓN	7
2.3	PROPUESTAS DE VALOR	9
<b>3</b>	<b>OBJETIVOS</b>	<b>11</b>
3.1	OBJETIVO GENERAL	11
3.2	OBJETIVOS ESPECÍFICOS	11
<b>4</b>	<b>ALCANCES</b>	<b>12</b>
<b>5</b>	<b>RESULTADOS ESPERADOS</b>	<b>13</b>
<b>6</b>	<b>MARCO CONCEPTUAL</b>	<b>14</b>
6.1	TÉCNICAS DE MINERA DE DATOS	14
6.1.1	<i>Regresión Logística</i>	14
6.1.2	<i>Arboles de Decisión</i>	15
6.2	MÉTRICAS DE COMPARACIÓN	16
6.2.1	<i>Matriz Confusión</i>	16
6.2.2	<i>Curva de Ganancia</i>	18
6.3	MODELOS DE OPTIMIZACIÓN	18
6.3.1	<i>Problema de la Mochila (Knapsack Problem)</i>	19
6.3.2	<i>Problema de la Mochila Múltiple (Multiple Knapsack Problem)</i>	20
6.4	NEXT BEST OFFER	20
6.5	DISEÑO EXPERIMENTAL	21
6.5.1	<i>Muestreo</i>	22
6.5.2	<i>Test de Proporciones</i>	22
6.6	METODOLOGÍAS DE BALANCEO DE CLASES.	23
6.7	METODOLOGÍA CRISP-DM	23
<b>7</b>	<b>METODOLOGÍA</b>	<b>26</b>
7.1	ENTENDIMIENTO DEL NEGOCIO:	26
7.2	COMPRESIÓN DE LOS DATOS:	26
7.3	PREPARACIÓN DE LOS DATOS	26
7.4	MINERÍA DE DATOS	27
7.5	MODELO DE ASIGNACIÓN	27
7.6	SIMULACIÓN DE POTENCIALES NUEVOS DE PRODUCTOS	27
7.7	DISEÑO EXPERIMENTAL	27
7.8	PRESENTACIÓN DE LAS CONCLUSIONES	27
<b>8</b>	<b>DESARROLLO DE LA METODOLOGÍA</b>	<b>28</b>
8.1	ENTENDIMIENTO DEL NEGOCIO	28

8.2	COMPRESIÓN DE LOS DATOS .....	28
8.2.1	<i>Productos:</i> .....	30
8.2.2	<i>Análisis Exploratorio</i> .....	31
8.2.3	<i>Conclusiones del análisis exploratorio</i> .....	42
8.3	PREPARACIÓN DE LOS DATOS .....	43
8.4	MINERÍA DE DATOS. ....	45
8.4.1	<i>Árbol de decisiones</i> .....	45
8.4.2	<i>Modelo Logit al ramo vida</i> .....	46
8.4.3	<i>Modelo Logit con interacción entre variables demográficas y atributos de los productos</i> .....	51
8.4.4	<i>Resultados</i> .....	52
8.4.5	<i>Elección del mejor Modelo</i> .....	53
8.4.6	<i>Validación del modelo seleccionado</i> .....	54
8.5	MODELO DE ASIGNACIÓN.....	55
8.5.2	<i>Modelo de asignación propuesto</i> .....	59
8.6	DISEÑO DE PRODUCTOS .....	62
8.7	DISEÑO EXPERIMENTAL .....	65
<b>9</b>	<b>CONCLUSIONES Y TRABAJOS FUTUROS.....</b>	<b>67</b>
<b>10</b>	<b>BIBLIOGRAFÍA .....</b>	<b>70</b>
<b>11</b>	<b>ANEXOS .....</b>	<b>72</b>
11.1	ANEXOS A.....	72
11.2	ANEXOS B.....	77
11.3	ANEXOS C .....	78
11.4	ANEXOS D .....	83
11.5	ANEXOS E.....	88

## Índice de tablas

TABLA 1: EVOLUCIÓN DE PRIMA INTERMEDIADA DE GRANDES TIENDAS.....	3
TABLA 2: POSIBLES IMPACTOS DEL PROYECTO (CONSIDERANDO LA PRIMA PROMEDIO DEL 2017).....	10
TABLA 3 : MATRIZ DE CONFUSIÓN. ....	17
TABLA 4: CANTIDAD DE DATOS A DISPOSICIÓN .....	29
TABLA 5 : SEGUROS OFRECIDOS POR EL CANAL CALL-CENTER .....	30
TABLA 6: ATRIBUTOS DE LOS SEGUROS OFRECIDOS POR EL CANAL CALL-CENTER.....	31
TABLA 7: ATRIBUTOS DEL NUEVO PRODUCTO QUE SE OFRECERÁ POR EL CANAL CALL-CENTER .....	31
TABLA 8: VENTAS DE CALL-CENTER .....	32
TABLA 9: SEGMENTACIÓN VARIABLE MONTO RENTA .....	37
TABLA 10: ANÁLISIS VARIABLES NUMÉRICAS .....	42
TABLA 11: VARIABLES QUE SE UTILIZARAN EN LA ETAPA DE MINERÍA DE DATOS .....	44
TABLA 12: RESULTADOS DE LOS ÁRBOLES DE DECISIONES.....	46
TABLA 13: COEFICIENTES MODELO LOGIT .....	47
TABLA 14: MATRIZ CONFUSIÓN MODELO LOGIT PARA EL RAMO VIDA CON CORTE DE 0,5 .	50
TABLA 15: MATRIZ CONFUSIÓN MODELO LOGIT CON INTERACCIONES CORTE 0,5.....	52
TABLA 16: RESULTADOS MÉTRICAS DE COMPARACIÓN .....	54
TABLA 17: DATOS UTILIZADOS PARA LA VALIDACIÓN DEL MODELO FINAL.....	54
TABLA 18: RESULTADOS DE LA VALIDACIÓN.....	54
TABLA 19: CORRELACIÓN DE LAS PROPENSIONES POR PRODUCTOS .....	55
TABLA 20: CAPACIDAD MÁXIMA POR PRODUCTO .....	58
TABLA 21: RESULTADOS DE LA SIMULACIÓN DEL MODELO APLICADO AL CASO REAL .....	58
TABLA 22: ANÁLISIS DE ROBUSTEZ DEL MODELO DE ASIGNACIÓN.....	59
TABLA 23: RESUMEN VALORES SIMULACIÓN CASO DE OPTIMIZACIÓN .....	61
TABLA 24: RESULTADOS SIMULACIÓN DEL MODELO DE ASIGNACIÓN PROPUESTO .....	62
TABLA 25: POTENCIALES NUEVOS PRODUCTOS SEGÚN MIX DE ATRIBUTOS.....	62
TABLA 26: CAPACIDAD DEL CANAL PARA EL CASO DE SIMULACIÓN DE DISEÑO DE PRODUCTOS .....	63
TABLA 27: RESULTADOS DE LA SIMULACIÓN DEL DISEÑO DE NUEVOS PRODUCTOS. ....	64
TABLA 28: FUENTES DE INFORMACIÓN RELEVANTES .....	72
TABLA 29: VARIABLES QUE APLICAN AL PROBLEMA .....	73
TABLA 30: MATRIZ DE CORRELACIÓN DE LAS VARIABLES .....	77
TABLA 31: MATRIZ CONFUSIÓN DEL ÁRBOL TIPO C.H.A.I.D CON CORTE 0,5 .....	78
TABLA 32: MATRIZ CONFUSIÓN DEL ÁRBOL TIPO C.H.A.I.D CON CORTE 0,5 .....	82
TABLA 33: COEFICIENTES DEL MODELO LOGIT CON ATRIBUTOS DE LOS PRODUCTOS .....	83
TABLA 34: MATRIZ CONFUSIÓN EN GRUPO DE VALIDACIÓN PARA EL PRODUCTO AP .....	86
TABLA 35: MATRIZ CONFUSIÓN EN GRUPO DE VALIDACIÓN PARA EL PRODUCTO VCB .....	86
TABLA 36: MATRIZ CONFUSIÓN EN GRUPO DE VALIDACIÓN PARA EL PRODUCTO VA5.....	87
TABLA 37: MATRIZ CONFUSIÓN EN GRUPO DE VALIDACIÓN PARA EL PRODUCTO APA .....	88
TABLA 38: EJEMPLO DEL OUTPUT MODELO DE ASIGNACIÓN .....	88

# Índice de Ilustraciones

ILUSTRACIÓN 1: PARTICIPACIÓN POR PRIMA INTERMEDIADA GRANDES TIENDAS DICIEMBRE 2017.....	3
ILUSTRACIÓN 2: PROCESO DE GENERACIÓN DE BASES PARA CALL-CENTER.....	5
ILUSTRACIÓN 3: EVOLUCIÓN DE LA TASA DE VENTA DE CALL-CENTER.....	6
ILUSTRACIÓN 4: EVOLUCIÓN DE LA TASA DE FUGA A 3 MESES DE CALL-CENTER. ....	6
ILUSTRACIÓN 5: PROCESO DE GENERACIÓN DE BASES PARA CALL-CENTER PROPUESTO .	8
ILUSTRACIÓN 6: EVOLUCIÓN DE PRODUCTOS DISTINTOS OFRECIDOS EN CALL-CENTER. ...	9
ILUSTRACIÓN 7: CURVA DE GANANCIA ACUMULADA .....	18
ILUSTRACIÓN 8: METODOLOGÍA CRISP-DM .....	24
ILUSTRACIÓN 9: INTEGRACIÓN DE LOS DATOS .....	30
ILUSTRACIÓN 10: ANÁLISIS VARIABLE EDAD.....	32
ILUSTRACIÓN 11: ANÁLISIS VARIABLE CUPO .....	33
ILUSTRACIÓN 12: ANÁLISIS VARIABLE NÚMERO DE HIJOS.....	34
ILUSTRACIÓN 13 ANÁLISIS NÚMERO DE VEHÍCULOS.....	34
ILUSTRACIÓN 14: ANÁLISIS VARIABLE GÉNERO .....	35
ILUSTRACIÓN 15: ANÁLISIS VARIABLE ESTADO CIVIL.....	35
ILUSTRACIÓN 16: ANÁLISIS VARIABLE HABITACIÓN.....	36
ILUSTRACIÓN 17: ANÁLISIS VARIABLE ACTIVIDAD .....	37
ILUSTRACIÓN 18: ANÁLISIS VARIABLE MONTO RENTA .....	38
ILUSTRACIÓN 19: ANÁLISIS VARIABLE GSE .....	38
ILUSTRACIÓN 20: ANÁLISIS VARIABLE REGIÓN .....	39
ILUSTRACIÓN 21: ANÁLISIS VARIABLE RIGHT.....	40
ILUSTRACIÓN 22: ANÁLISIS VARIABLE LEFT.....	40
ILUSTRACIÓN 23: ANÁLISIS VARIABLE FREQUENCY .....	41
ILUSTRACIÓN 24: ANÁLISIS VARIABLE MONETARY ACTUAL.....	42
ILUSTRACIÓN 25 CURVA ACUMULADA DE GANANCIA DEL MODELO LOGIT PARA RAMO VIDA .....	50
ILUSTRACIÓN 26 CURVA DE GANANCIA ACUMULADA PARA EL MODELO LOGIT CON INTERACCIONES .....	53
ILUSTRACIÓN 27: CURVA DE GANANCIA ACUMULADA DEL ÁRBOL TIPO C.H.A.I.D .....	78
ILUSTRACIÓN 28: DIAGRAMA DEL ÁRBOL DE DECISIÓN TIPO C.H.A.I.D NODO 1 Y 2 .....	79
ILUSTRACIÓN 29: DIAGRAMA DEL ÁRBOL DE DECISIÓN TIPO C.H.A.I.D NODO 3. ....	80
ILUSTRACIÓN 30: DIAGRAMA DEL ÁRBOL DE DECISIÓN TIPO C.H.A.I.D NODO 4. ....	81
ILUSTRACIÓN 31: CURVA DE GANANCIA ACUMULADA DEL ÁRBOL TIPO C.A.R.T .....	82
ILUSTRACIÓN 32 CURVA DE GANANCIA ACUMULADA EN GRUPO DE VALIDACIÓN PARA EL PRODUCTO AP.....	85
ILUSTRACIÓN 33: CURVA DE GANANCIA ACUMULADA EN GRUPO DE VALIDACIÓN PARA EL PRODUCTO VCB.....	86
ILUSTRACIÓN 34: CURVA DE GANANCIA ACUMULADA EN GRUPO DE VALIDACIÓN PARA EL PRODUCTO VA5 .....	87
ILUSTRACIÓN 35: CURVA DE GANANCIA ACUMULADA EN GRUPO DE VALIDACIÓN PARA EL PRODUCTO APA .....	87

# 1 Introducción

En este capítulo, se hace una descripción de la empresa en donde se desarrolla el proyecto de memoria, describiendo el rubro y el mercado en el cual se desempeña.

## 1.1 Caracterización de la empresa

El presente trabajo de memoria se desarrolló en Seguros Falabella, compañía perteneciente al grupo Falabella, que se desempeña en el rubro de corredoras de seguros de grandes tiendas. Tiene presencia en 4 países: Chile, Argentina, Perú y Colombia.

Las corredoras de seguros funcionan como entidades intermediadoras entre las compañías de seguros y los clientes, ofreciendo y asesorando a estos últimos, sobre los productos con las coberturas más convenientes respecto a sus necesidades e intereses. Tienen además la responsabilidad de asistir a los clientes tanto en el proceso de compra, como al momento de producirse un siniestro.

Seguros Falabella ofrece seguros tanto para personas naturales como para empresas. Dentro de la cartera de productos que ofrece la compañía a sus clientes se encuentran:

- Seguros Automotrices.
- Seguros de Vida.
- Seguros de Salud.
- Seguros de Hogar.
- Seguros de viajes.
- Seguros Transaccionales.
- Seguros de Empresas (Patrimoniales, personas, pymes y vehículos).

Esta cartera de productos, es negociada por Seguros Falabella con las distintas compañías aseguradoras, con el fin de ofrecer los mejores productos de cara a los clientes. Dentro de las compañías aseguradoras relacionadas con la empresa



destacan: BCI seguros, BNP Paribas Cardif, Chilena Consolidada, Chubb, Consorcio, HDI seguros, Itaú compañía de seguros de vida, Liberty Seguros, MetLife, Zenit.

Seguros Falabella al pertenecer al grupo Falabella, hereda su prestigio y reconocimiento de marca, facilitando de esta manera su acercamiento con los clientes y permitiéndoles obtener los beneficios del grupo, por ejemplo, acumular o gastar puntos CMR. La compañía es considerada líder en el rubro de corredoras de seguros de grandes compañías, permitiéndoles negociar de mejor manera los productos con las compañías aseguradoras. Seguros Falabella ofrece sus servicios mediante varios canales, por ejemplo, página web, Call center y en todas las sucursales de Falabella y Homecenter. Esto le permite aprovechar el gran flujo de personas de las otras compañías del holding para captar nuevos clientes.

## **1.2 Mercado**

Tal como se mencionó anteriormente, Seguros Falabella pertenece al rubro de corredoras de seguros de grandes compañías, en donde los principales actores son:

- Compañías Aseguradoras: Empresas que ofrecen pólizas que cubren ciertos riesgos que podrían interesarles a los clientes.
- Clientes: Personas que buscan contratar algunos seguros para cubrir cierto riesgo.

De esta manera las corredoras funcionan como intermediadoras, ofreciendo distintos productos a los clientes y brindando asesoramiento en la contratación y asistencia en caso de un siniestro. Las corredoras de seguros son entidades independientes de las compañías aseguradoras, por lo tanto, cobran una comisión por el servicio que ofrecen, el cual corresponde a la diferencia de precio entre lo que cobra la corredora al cliente, con el precio que la compañía aseguradora ofrece a la corredora. [1]

Hasta el año 2017, tanto las aseguradoras como las corredoras eran reguladas por la Súper Intendencia de Valores y Seguros, entidad que desde enero del 2018 pasó a llamarse Comisión para el Mercado Financiero. Este ente público está encargado,

entre otras cosas, de supervisar, investigar y sancionar las distintas entidades financieras en Chile. De esta manera, busca el correcto funcionamiento, desarrollo y estabilidad del mercado financiero, en el cual se rigen tanto las aseguradoras como las corredoras. Esta misma entidad genera informes trimestrales sobre el mercado de corredoras de seguros. En el informe: “Corredoras de Seguros de Grandes Tiendas”, de diciembre del 2017, se señala que existen 2.559 corredores de seguros, de los cuales 8 son de grandes tiendas, con un total de ventas de aproximadamente 320 MM de pesos (Ventas por prima intermediada de seguros generales y seguros de vida).

Tabla 1: Evolución de prima intermediada de grandes tiendas.

	dic-13	dic-14	dic-15	dic-16	dic-17
<b>Crecimiento</b>	6,02%	15,51%	8,61%	18,03%	0,70%

Fuente: Informe Corredoras de Seguros Grandes Tiendas [2]

En la Tabla 1, se presenta el crecimiento de la industria en términos de prima intermediada, en ella se puede observar que el mercado de seguros venía experimentando altas tasas de crecimiento en los últimos 4 años, durante el 2017 se observó un crecimiento no tan marcado como en años anteriores. Por su parte Seguros Falabella, se muestra como líder en la participación de corredoras de grandes tiendas en la prima intermedia, con un 61% de participación, es seguido por Líder y La Polar con un 17% y 8% de participación respectivamente, tal y como se observa en la ilustración 1.

**Participación por Prima Intermediada Grandes Tiendas Diciembre 2017**

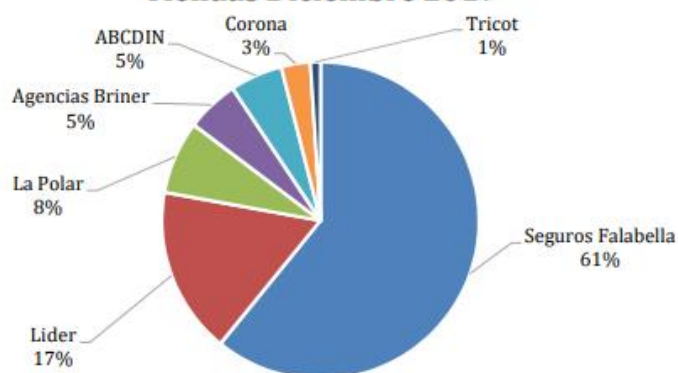


Ilustración 1: Participación por prima Intermediada Grandes tiendas Diciembre 2017.

Fuente: Informe Corredoras de Seguros Grandes Tiendas Diciembre 2017 [2]

## **2 Justificación del Problema**

A continuación, se identifica la oportunidad de mejora encontrada en la empresa, detallando evidencia de la relevancia de esta. Se identifican las hipótesis de la oportunidad de mejora, y se detallan las posibles soluciones y su impacto.

### **2.1 Identificación de la oportunidad de mejora.**

La corredora de seguros cuenta con múltiples canales de ventas, dentro de ellos destaca el canal Call-Center, el cual durante el año 2017 concreto el 8% de las ventas de la compañía, aportando en algunos productos con cerca del 50% de las ventas totales. Para el correcto funcionamiento del canal, existe un proceso de generación y cargas de bases. Este es un proceso mensual el cual tiene como finalidad el disponibilizar las bases de datos con la que los ejecutivos del canal contactan a los clientes, para ofrecerles los servicios y productos de la corredora. El proceso es desarrollado entre el canal Call-Center y el área de Inteligencia de Negocio de la compañía, y puede ser resumido en 3 pasos: recolección de la información, asignación de los productos a ofrecer a los clientes, y la generación final de bases y carga al sistema, para que estén disponibles para los ejecutivos del canal. La Ilustración 2 muestra el proceso descrito.

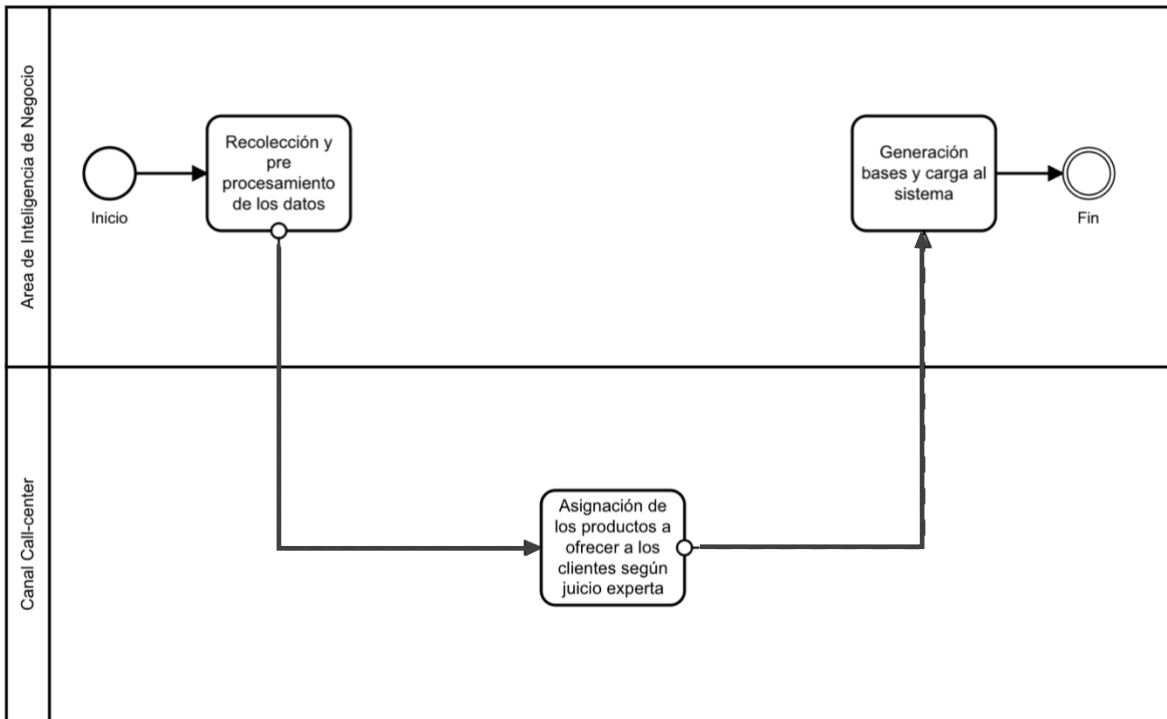
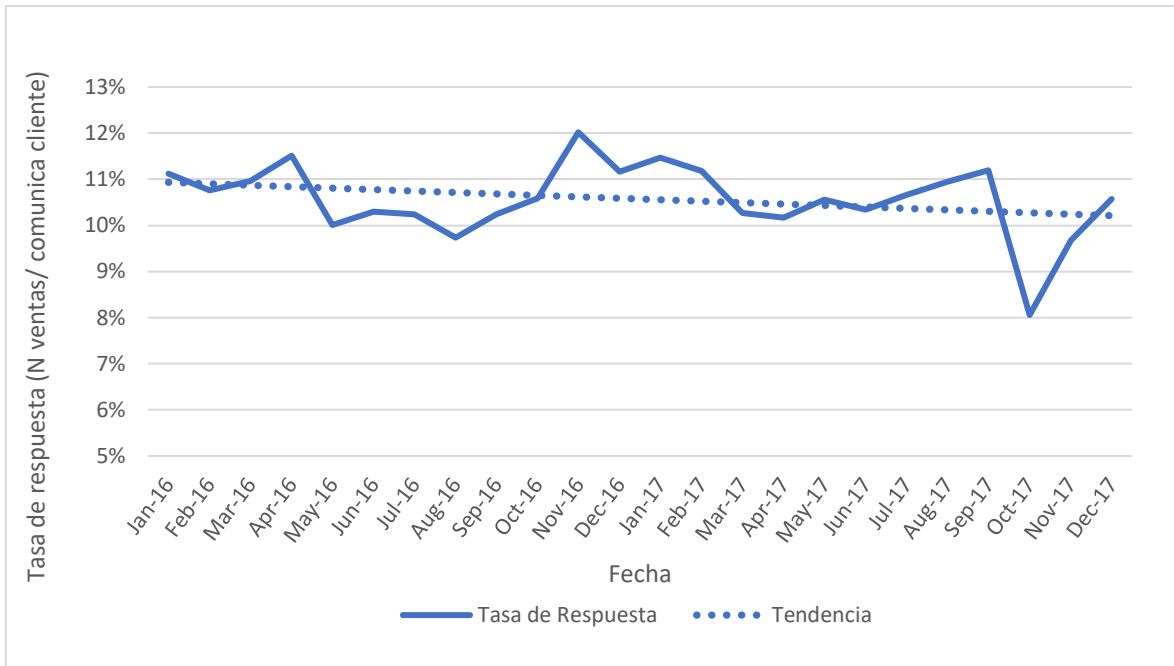


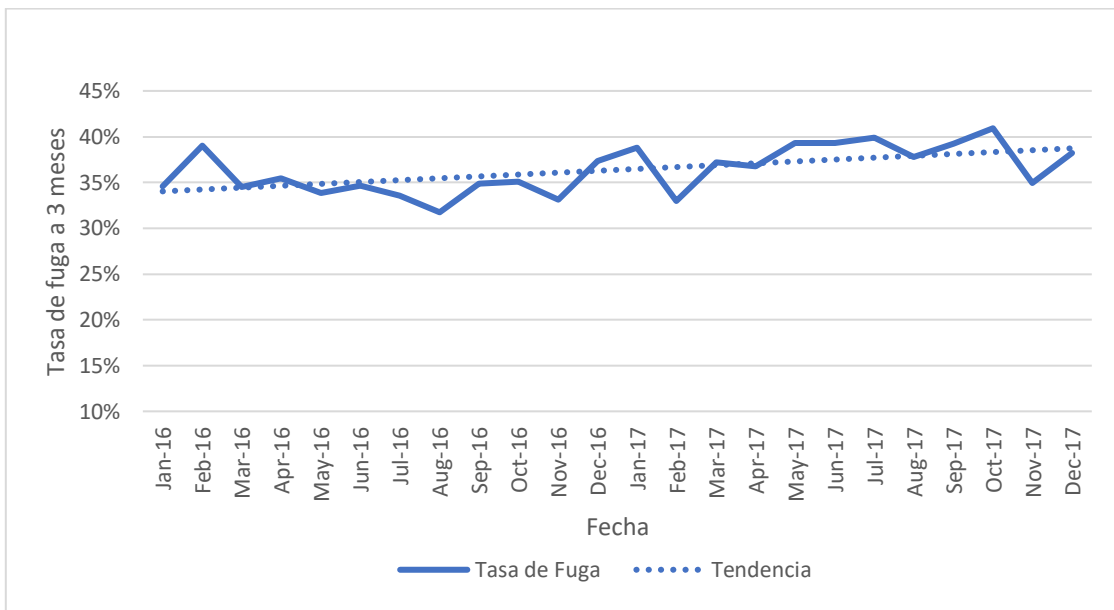
Ilustración 2: Proceso de generación de bases para Call-Center  
Fuente: Elaboración Propia

Llama la atención del proceso que son los supervisores del canal quienes asignan que productos ofrecer a los clientes mediante su juicio experto, lo que puede estar generando dos problemas importantes:

1. No necesariamente están ofreciendo el mejor producto a cada cliente, lo que conlleva a que se estén perdiendo ventas. Se puede apreciar en la Ilustración 3, que la tasa de ventas del canal se ha mantenido estable entre un 10% y 12%.
2. Se están realizando ventas no sostenibles en el tiempo, generando que los clientes anulen las suscripciones a los productos. La Ilustración 4 muestra la tasa de fuga del canal durante los primeros 3 meses después de realizada la venta, se puede ver una pequeña alza es este indicador durante el último año.



**Ilustración 3: Evolución de la tasa de venta de Call-Center**  
Fuente: Elaboración propia



**Ilustración 4: Evolución de la tasa de fuga a 3 meses de Call-center.**  
Fuente: Elaboración propia

Dado estos indicadores, se observa que la tasa de ventas que perduran más de 3 meses ha bajado levemente con respecto al año anterior. Este indicador es relevante dado el rubro en el cual se desempeña la empresa, en donde los pagos de los clientes son de manera mensual, por lo que se busca aumentar el número de clientes activos mes a mes en la compañía, o el “stock” de clientes. Para aumentar este “stock” de clientes se pueden aumentar las ventas o disminuir la fuga. Es debido a esto último que la fuga a 3 meses es un indicador relevante para la empresa. Además, el proceso al tener un traspaso de información entre dos áreas que no se encuentran físicamente juntas genera lentitud en el proceso, generando tiempos muertos mientras se espera la acción de la otra área.

## **2.2 Hipótesis y alternativa solución**

Al revisar el proceso mencionado anteriormente, se tiene como hipótesis que la manera actual de realizar la asignación de productos a ofrecer a los clientes, no es la óptima. El presente trabajo de memoria, propone un rediseño de este proceso utilizando una metodología que realice la asignación de productos a ofrecer a los clientes, mediante la resolución de un problema de optimización. Para ello se prueban distintos modelos de propensión de compra, los cuales son utilizados como inputs, para un modelo de optimización que busque maximizar las ventas que perduran más de 3 meses, respetando las restricciones propias del canal y la compañía. La Ilustración 5 resume el nuevo proceso de generación de bases, el cual ya no necesita el traspaso de información entre las áreas, sino que solo la entrega de las restricciones que deben ser consideradas para la asignación.

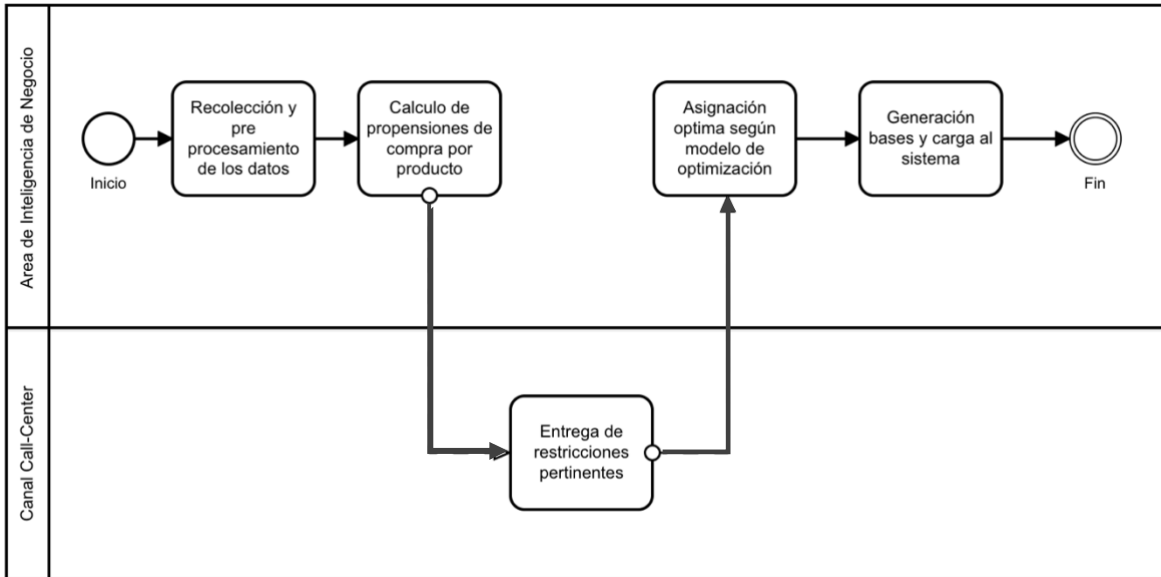


Ilustración 5: Proceso de generación de bases para Call-Center propuesto  
Fuente: Elaboración propia

Es importante señalar, que, dado que el trabajo se contextualiza en una corredora de seguros, esta se encuentra constantemente negociando con las compañías aseguradoras. Esto genera que su cartera de productos no sea constante en el tiempo, y que los productos ofrecidos varíen. La Ilustración 6 muestra cómo ha ido cambiando el número de productos ofrecidos por Call-center en los últimos dos años.

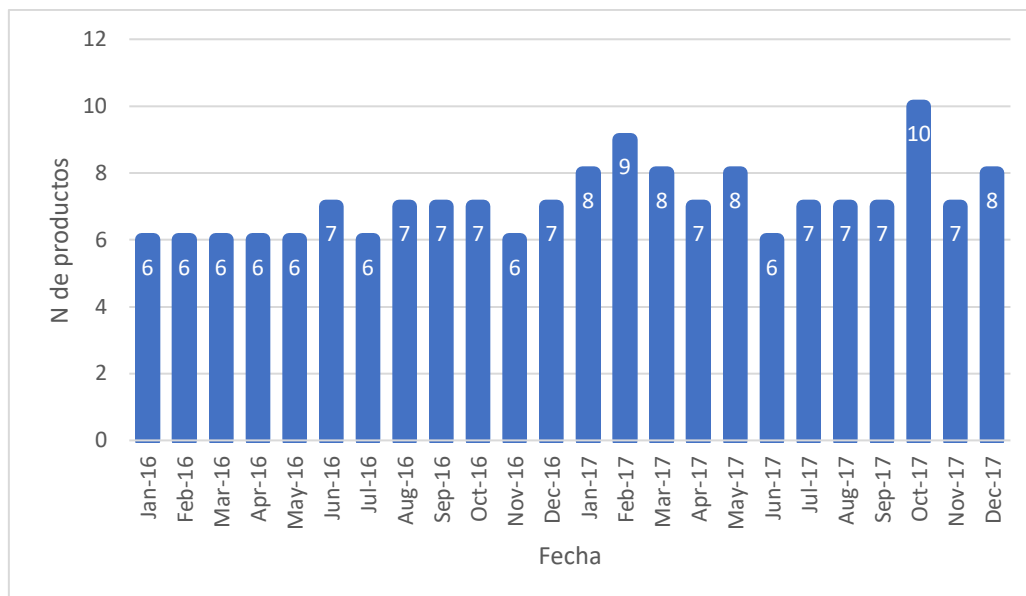


Ilustración 6: Evolución de productos distintos ofrecidos en Call-center.  
Fuente: Elaboración propia.

Dada esta variabilidad en los productos que son ofrecidos, es relevante generar una solución que sea perdurable en el tiempo. Es por ello que se propone generar modelos de propensión que incorporen como variables explicativas, los atributos relevantes de los productos. Esto adicionalmente, permitirá recomendar el diseño de nuevos productos con la combinación de estos atributos, de tal manera de aumentar las ventas perdurables del canal.

## 2.3 Propuestas de Valor

**La solución propuesta pretende impactar en varios niveles al problema:**

- Permitir evaluar distintos perfiles de clientes y su respuesta tanto en nivel de compra como de fuga a los distintos productos, permitiendo ofrecer el producto más conveniente a cada cliente.
- Agilizar los procesos de generación de bases del área, disminuyendo el tiempo perdido por el constante traspaso de información entre las áreas.
- Al incorporar como variables los atributos de los productos, se evalúan los impactos de estos en las tasas de venta, lo que entrega flexibilidad al modelo, permitiendo evaluar nuevos productos.



- Se busca, además, generar un aumento en las ventas del canal, mediante el uso de un modelo de optimización que utilice los modelos de propensión. En la Tabla 2 se puede ver un análisis del posible impacto del modelo final.
- Evaluar el impacto que de potenciales nuevos productos en las ventas del canal.

Tabla 2: Posibles impactos del proyecto (Considerando la prima promedio del 2017).

<b>Incremento porcentual</b>	<b>Pólizas anuales</b>	<b>Ventas Anuales</b>
1%	850	\$ 34 MM
2%	1.700	\$ 68 MM
3%	2.550	\$ 102 MM
4%	3.400	\$ 136 MM
5%	4.250	\$ 170 MM

## **3 Objetivos**

Dado lo expuesto anteriormente, se definen los objetivos a desarrollar en el proyecto

### **3.1 Objetivo General**

Aumentar las ventas perdurables del Call-center de una corredora de seguros, mediante una metodología de asignación óptima que considere la propensión de compra de los clientes y las restricciones del canal.

### **3.2 Objetivos Específicos**

- Estimar la probabilidad de compra de los clientes a los productos ofrecidos al canal, considerando los atributos de los productos y sus interacciones con variables de los clientes, mediante la calibración y construcción de modelos de propensión.
- Diseñar un problema de optimización que resuelva el problema de asignación óptima, considerando las propensiones de compra y las restricciones del canal.
- Generar recomendaciones al negocio, con respecto a la asignación producto-cliente, y con respecto al mix de atributos de los productos.

## 4 Alcances

Los alcances del presente trabajo de memoria se limitarán en los siguientes aspectos:

- El trabajo se realizó Seguros Falabella, concretamente, en el proceso de generación de bases para el canal Call-Center.
- Se trabajó con un segmento de clientes categorizados como “aperturas”, este segmento de potenciales clientes tiene la particularidad de que son personas que abrieron una tarjeta de crédito en el holding en el último mes, y por lo tanto no tienen necesariamente un vínculo con la corredora. Este segmento de clientes es considerado el más importante para el canal de Call-center, pues durante el año 2017 aportó cerca del 70% de las ventas del canal.
- Se utilizó los datos de enero 2017 hasta mayo del 2018 para entrenar los modelos, puesto que se tiene información completa con respecto al estado de las ventas. Los datos de Junio y Julio, fueron utilizados como validación del modelo seleccionado.
- Dada la categoría de clientes con la que se trabajara y el periodo de tiempo que se estudiara, se entrenaran los modelos en base a 4 productos que estuvieron activos en este escenario: Vida con Bonificación, Accidentes personales con bonificación y muerte accidental tipo 4 y 5. Se modelara en base a sus atributos, los cuales serán levantados en conjunto con el área de producto de la compañía.
- Solo se planteará el diseño experimental, debido a que las condiciones en el momento de evaluar la metodología no son adecuadas para obtener resultados estadísticamente significativos.

## 5 Resultados Esperados

El trabajo busca entregar los siguientes resultados a la empresa:

- Calibrar y disponibilizar a la organización modelos de propensión de compra por productos, que tenga la flexibilidad necesaria para ser utilizables en un contexto donde los productos ofrecidos varían en el tiempo. Para ello se deberán determinar las características relevantes de los clientes, además de realizar un levantamiento previo de los atributos significantes de los productos, con el fin de incluir todos estos elementos en los modelos de propensión.
- Diseñar y disponibilizar a la organización modelos de optimización que realicen una asignación óptima de los productos a ofrecer a los clientes. Se espera por lo tanto realizar un levantamiento de las restricciones del canal de Call-center para ser considerados en la asignación.
- Probar mediante simulación que la metodología propuesta puede generar un aumento en las ventas perdurables del canal, y evaluar el diseño de potenciales nuevos productos utilizando la metodología estudiada.

## 6 Marco Conceptual

En este capítulo se detalla la teoría y modelos conceptuales existentes respecto al problema a desarrollar, y se seleccionan las metodologías que estén más relacionados con la solución que se propone.

### 6.1 Técnicas de minería de Datos

Las técnicas de minería de datos frecuentemente pueden ser divididas en 3 categorías, métodos de clasificación, métodos de agrupación y métodos de detección de anomalías. Los métodos de clasificación son usualmente utilizados para responder la principal pregunta del marketing directo, ¿cómo va reaccionar cada cliente a la oferta de un producto o servicio? Los métodos de clasificación permiten calcular la probabilidad de que el cliente reaccione de manera positiva o compre el producto ofrecido, para ello generalmente utiliza tanto el comportamiento histórico del cliente como sus características demográficas. Dentro de los métodos de clasificación destacan [3]:

- Regresión logística binaria y multinomial.
- Árbol de Decisiones.
- Support Vector Machine.
- Redes Neuronales.
- Random Forest.
- Método del Bayesiano.

En este trabajo se utiliza el método de la regresión logística y los arboles decisiones, debido a que han sido utilizados para resolver problemas similares con buenos resultados, además de que tienen la particularidad de ser de fácil interpretación. [4], [5], [6], [7], [8].

#### 6.1.1 Regresión Logística

La regresión logística es un modelo no lineal supervisado, que se utiliza para determinar la probabilidad que ocurra un evento, por lo tanto, es importante definir cuál será la variable a predecir y cuando será considerado un evento exitoso (valor

1) y cuando no (valor 0).

Este modelo calcula la probabilidad de ocurrencia a partir de variables explicativas, la esperanza de que ocurra un suceso se puede ver en la formula 1.

$$E(Y_i) = P_i = \frac{e^{(\beta_0 + \beta_1 * X_{1i} + \dots + \beta_n * X_{ni})}}{1 + e^{(\beta_0 + \beta_1 * X_{1i} + \dots + \beta_n * X_{ni})}} \quad (1)$$

En donde:

$X_{ni}$  = Valor o categoría de la variable N del individuo i

$\beta_n$  = Coeficiente de la variable N.

La regresión logística tiene la ventaja de que es simple y de fácil interpretación, permitiendo conocer el impacto de cada variable sobre la clasificación. En el presente trabajo se utiliza la regresión logística para calcular la propensión de compra de los clientes en los distintos productos, además aportara señalando el impacto que tienen los atributos de los productos sobre esta propensión.

### 6.1.2 Árboles de Decisión

Son modelos supervisados, en donde se genera un grafo formado por un nodo inicial con toda la población a estudiar y que posteriormente se comienza a ramificar dependiendo de las variables observables. De esta manera se va clasificando a la población y se va calculando la probabilidad de que ocurra el evento estudiado. Dentro de las ventajas de los árboles de decisión es que permiten tener variables tanto numéricas como categóricas, junto con esto, los arboles de decisión tienen la particularidad de que son de fácil interpretación, permitiendo encontrar de manera simple grupos de clientes que cumplen condiciones similares y tienen una respuesta similar sobre la variable estudiada

Existen varios métodos para ir ramificando el árbol, en la bibliografía destacan particularmente el método C.H.A.I.D y C.A.R.T

- C.H.A.I.D: consiste en ir ramificando mediante las variables con mayor significancia según el estadístico  $\chi^2$  hasta que se cumpla algún criterio de

parada que se le entrega al modelo. Este tipo de modelación tiene como ventaja el entregar reglas de decisión valiosas para el negocio de manera gráfica. La formula 2 muestra el calculo de  $\chi^2$ :

$$\chi^2 = \sum_{i=1} \frac{(OBSERVADA_i - TEORICO_i)^2}{TEORICO_i} \quad (2)$$

- C.A.R.T: Consiste en ir ramificando en base al índice de Gini, es decir, cada ramificación busca aumentar la diferencia en las proporciones de las clases. La formula 3 muestra el calculo del índice de Gini:

$$Gini(D) = 1 - \sum_{i=1}^m \rho_i^2 \quad (3)$$

Donde:

$\rho_i$  = Probabilidad de ocurrencia de i

En este trabajo se utilizaron los arboles de decisión para calcular la propensión de compra de los clientes. Lo que permitirá encontrar segmentación de clientes más propensos que otras.

## 6.2 Métricas de Comparación

### 6.2.1 Matriz Confusión

La matriz de confusión es una herramienta que permite evaluar el los resultados de una predicción, clasificando la predicción en 4 categorías:

- Verdaderos Positivos (VP): Valores que el modelo de predicción considero positivos y en la realidad también lo fueron.
- Falsos Positivos (FP): Valores que el modelo de predicción considero positivas, pero en la realidad eran negativos.

- Falsos Negativos (FN): Valores que el modelo de predicción considero negativos pero que en realidad eran positivos.
- Verdaderos Negativos (VN): Valores que el modelo de predicción considero negativos y en la realidad también eran negativos

Tabla 3 : Matriz de confusión.

		REAL	
		POSITIVO	NEGATIVO
PREDICHO	POSITIVO	VP	FP
	NEGATIVO	FN	VN

Dada esta matriz se pueden calcular las siguientes métricas de desempeño de los modelos, que permitirán seleccionar los modelos más adecuados:

- Sensibilidad: Se mide la proporción de las observaciones verdaderas positivos, sobre todos los positivos reales. La formula 4 muestra el calculo de la sensibilidad.

$$Sensibilidad = \frac{VP}{VP + FN} \quad (4)$$

- Especificidad: Se mide la proporción de las observaciones verdaderos negativos sobre todas las negativas reales. La formula 5 muestra el calculo de la especificidad.

$$Especificidad = \frac{VN}{VN + FP} \quad (5)$$

- Exactitud: Se mide la proporción de los valores correctamente calificados. La formula 6 muestra el calculo de la exactitud.

$$Exactitud = \frac{VP + VN}{VP + FP + VN + FN} \quad (6)$$



## 6.2.2 Curva de Ganancia

La curva de ganancia, es una manera gráfica de representar la efectividad del modelo, entregando cuanto porcentaje de casos de éxito se están clasificando correctamente por el criterio. Se construyen ordenando las propensiones y generando deciles de estas, para luego capturar el porcentaje de casos de éxitos obtenidos en estos deciles. De esta manera se puede comparar distintos modelos, frente al caso base o sin modelo, en donde entre mayor es el porcentaje de éxitos capturado por los primeros deciles, mejor será la clasificación del modelo. Este modelo tiene la ventaja de que funciona de igual manera para bases tanto balanceadas como no, y no depende de un criterio de corte.

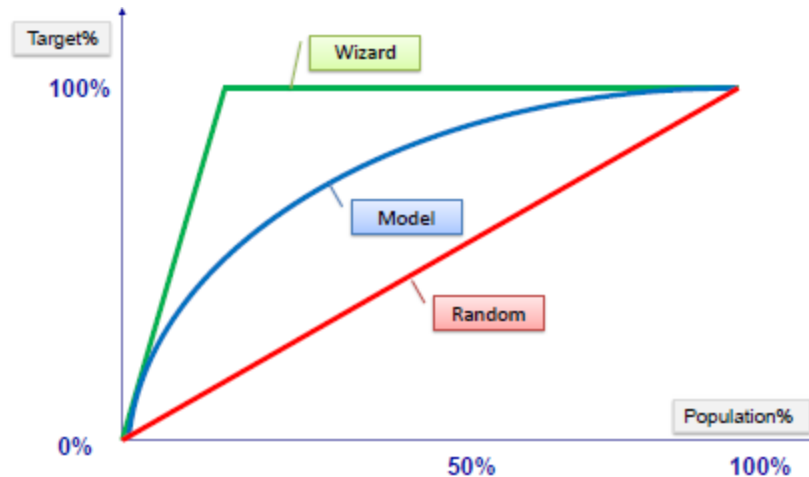


Ilustración 7: Curva de ganancia acumulada

Fuente: [http://www.saedsayad.com/model\\_evaluation\\_c.htm](http://www.saedsayad.com/model_evaluation_c.htm)

## 6.3 Modelos de Optimización

Un modelo de optimización busca encontrar el valor máximo o mínimo de una función, definida en un dominio o sujeto a algunas restricciones. En general un problema de optimización puede ser descrito por 5 elementos:

1. Conjuntos: Son las agrupaciones que se utilizarán durante la modelación.  
Ejemplo: Productos = {VIDA CON BONIFICACIÓN, ACCIDENTES PERSONALES, MUERTE POR ACCIDENTE}
2. Parámetros: Son aquellos valores que no dependen de alguna decisión del modelo.

Ejemplo:  $v_i = \text{valor de la prima del producto } i$

3. Variables Decisión: Son todas aquellas variables que el modelo debe determinar sus valores mediante la resolución del problema.

Ejemplo:  $x_{ki} = 1$  *si al cliente k le ofrezco el producto i*

4. Restricciones: Características que la solución final deba cumplir.

Ejemplo:  $\sum x_{ki} = 1$  *Todo cliente debe tener asignado un producto*

5. Función Objetivo: Corresponde el valor que se quiere optimizar

Ejemplo: Maximizar  $\sum x_{ki} * v_i$

### 6.3.1 Problema de la Mochila (Knapsack Problem)

El problema de la mochila busca decidir qué productos se deben seleccionar en la mochila con el fin de maximizar la utilidad, de tal manera de no sobrepasar la capacidad dada de la mochila [9]. De esta manera el modelo puede ser descrito de la siguiente manera:

- $x_j$  es la variable decisión binaria en donde toma el valor 1 si el producto j es incluido en la mochila y 0 si no.
- $p_j$  es la utilidad obtenida por incluir el producto j a la mochila
- $w_j$  es el peso del producto j.
- C es la capacidad de la mochila

De esta manera, el problema de optimización puede ser descrito tal y como muestra la formula 7.

$$\text{Max} : \sum_{j=1}^n p_j x_j \quad (7)$$

Sujeto a:

$$\sum_{j=1}^n w_j x_j \leq c$$

### 6.3.2 Problema de la Mochila Múltiple (Multiple Knapsack Problem)

Este problema es la generalización del problema de la mochila [10], en donde se tienen múltiples mochilas, y se debe decidir si incorporar cada producto a una de las mochilas, con el fin de maximizar la utilidad y asegurar de no superar la capacidad de ninguna de las mochilas. De esta manera el modelo puede ser descrito de la siguiente manera:

- $x_{ij}$  la variable decisión binaria en donde toma el valor 1 si el producto  $j$  es incluido en la mochila  $i$  y 0 si no.
- $p_j$  es la utilidad obtenida por incluir el producto  $j$  a la mochila
- $w_j$  es el peso del producto  $j$ .
- $c_j$  es la capacidad de la mochila  $i$

De esta manera, el problema de optimización puede ser descrito tal y como muestra la formula 8.

$$Max : \sum_{j=1}^n p_j x_{ij} \quad (8)$$

Sujeto a:

$$\sum_{j=1}^n w_j x_{ij} \leq c_i$$

### 6.4 Next Best Offer

Tal como su nombre lo dice, esta metodología busca determina la mejor oferta que puede recibir el cliente de tal manera de maximizar la probabilidad de respuesta. El enfoque tradicional del Next Best offer cuenta con los siguientes pasos: [11]

1. Para cada producto a ofertar se desarrolla un modelo de propensión.
2. Cada Cliente tendrá un score para cada oferta.
3. Se asigna a cada cliente las ofertas con mayor Score.

El problema de este enfoque, es que no toma en consideración las restricciones del negocio, como por ejemplo presupuesto de campañas, capacidad de ejecutivos, retorno esperado, etc. Dado esto se realizará un enfoque que incorpore la resolución de un problema de optimización [8] [11], [12], [13]. De esta manera la secuencia será:

1. Para cada producto a ofertar se desarrolla un modelo de propensión de compra perdurable.
2. Cada Cliente tendrá un score para cada oferta.
3. Resolver problema de optimización que asigne la mejor oferta al cliente, sujeto a las restricciones del negocio.
4. Se asigna a cada cliente las ofertas correspondientes según los resultados de la optimización.

De esta manera la optimización buscara maximizar las ventas perdurables del canal asignando a cada cliente una oferta, pero tomando en consideración las restricciones propias del negocio y del canal.

## **6.5 Diseño Experimental**

Se propondrá un diseño experimental con el fin de evaluar la metodología propuesta, y como afecta a la tasa de ventas perdurables estudiada. Para ello se utilizará la metodología de grupos de control, la cual consiste en contrastar entre conjuntos de individuos. Por lo general se crean dos grupos, un grupo de tratamiento, que será afectado por el trabajo realizado y un grupo de control, el cual no será afectado [14]. Es importante que el diseño de los grupos cumpla una asignación aleatoria entre los grupos. (Un cliente debe tener igual probabilidad de

quedar en alguno de los grupos) y la independencia entre los grupos (el resultado de un grupo no debe afectar o intervenir al otro). De esta manera se busca asegurar la causalidad en la experimentación. Dentro de un diseño experimental es importante señalar que métrica se quiere evaluar. Durante este trabajo se propone una metodología que busque aumentar la tasa de ventas perdurables.

### 6.5.1 Muestreo

Para determinar el tamaño de la muestra de una población estadísticamente representativa, se debe calcular mediante la formula 9.

$$n = \frac{X^2 * N * P * (1 - P)}{d^2 * (N - 1) + X^2 * P * (1 - P)} \quad (9)$$

Dónde:

n = tamaño de la muestra.

$X^2$  = Valor de la tabla chi-cuadrado para un nivel de confianza necesario.

N= Tamaño de la población.

P = proporción de la población.

d=error deseado.

En donde para N lo suficientemente grande, se puede escribir la formula 10.

$$n = \frac{X^2 * P * (1 - P)}{d^2} \quad (10)$$

### 6.5.2 Test de Proporciones

El test de proporciones busca probar si existe diferencia significativa en las poblaciones de estudio, en este trabajo se evaluará si existe diferencia significativa en la tasa de ventas perdurable. Por lo tanto, el test evaluara las siguientes hipótesis.

Hipótesis nula  $H_0: p_1 = p_2$ . No existe diferencia entre las tasas de ventas perdurables del grupo de tratamiento y el grupo de control

Hipótesis alternativa:  $H_1: p_1 \neq p_2$  Si existe diferencia entre las tasas de ventas perdurables.

Para muestras grandes el estadístico de prueba se puede calcular como muestra la formula 11.

$$Z = \frac{p_1 - p_2}{\sqrt{p * q * (\frac{1}{n_1} + \frac{1}{n_2})}} \quad (11)$$

Con:  $p = \frac{x_1 + x_2}{n_1 + n_2}$ .

$q = 1 - p$ .

Dónde:  $x_i$ = Casos favorables en la muestra i.

$n_i$ = Casos totales en la muestra i.

En donde z sigue una normal de tal manera que la hipótesis nula se rechaza si el valor  $z > \frac{z\alpha}{2}$  en donde  $\alpha$  corresponde al intervalo de confianza deseado.

## 6.6 Metodologías de balanceo de clases.

Al analizar los resultados de las campañas de una compañía, es común encontrarse con el problema de que existen pocas respuestas positivas de los clientes, esto genera que, al aplicar metodologías de minería de datos, se entrenen los modelos con datos desbalanceados. Esto es comúnmente resultado con las metodologías de over sampling o under sampling. En este proyecto se utiliza la metodología de random over sampling, debido a que se ha estudiado que se obtienen mejores resultados al ajustar modelos, mientras que al utilizar el under sampling se puede estar perdiendo datos importantes. [14]

## 6.7 Metodología Crisp-DM

El trabajo de esta memoria necesita extraer información de grandes volúmenes de data para convertirlos en conocimiento agregado para las empresas. La metodología Crisp-DM [15] es un proceso iterativo de minería de datos que permite descubrir patrones de tal manera de entender la información y contribuir en la toma de decisiones del negocio. Esta Metodología pone énfasis en la etapa de entendimiento del negocio que otras metodologías como el KDD no consideran. Esto permite tener mayor conocimiento sobre los procesos internos, de tal manera de evaluar la situación actual y proponer mejoras pertinentes.

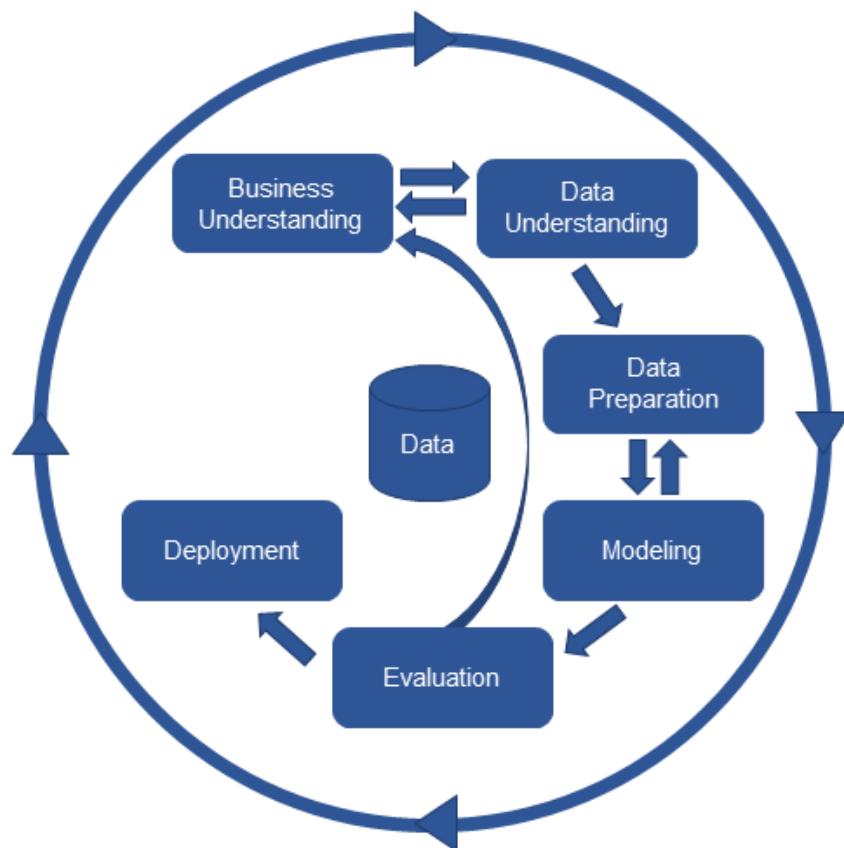


Ilustración 8: Metodología Crisp-DM  
Fuente: CRISP DM 1.0 Step-by-Step data mining guide

Las fases del proceso metodológico son las siguientes:

- Entendimiento del negocio: Etapa inicial del proceso que busca entender y definir los objetivos desde una perspectiva del negocio.

- **Comprensión de los datos:** En esta fase se busca conocer y estudiar la data. Realizar una exploración de esta de tal manera, de identificar problemas en ella y obtener los primeros insight sobre el problema que se está abordando.
- **Preparación de la data:** En esta etapa se busca mejorar la calidad de la sabana de datos que será utilizada en fases posteriores. Para ello se realiza la limpieza y selección de los datos. Además, se transforman y se crean variables con el fin de obtener mejores resultados.
- **Modelar:** Se seleccionan y se aplican distintas técnicas de modelamiento que puedan aplicar al objetico inicial.
- **Evaluación:** Se evalúan y comparan los modelos generados, así como todas las etapas anteriores, en caso de encontrar algún problema en alguna fase, se deben volver a ella e iterar el proceso nuevamente, en caso contrario se puede avanzar a la siguiente fase.
- **Desarrollo:** En esta etapa se busca presentar el conocimiento obtenido a la empresa, esta entrega debe ser entendible y de fácil implementación. Se busca además entregar recomendaciones generales al negocio sobre los insights conseguidos en las etapas de modelación y evaluación.



## **7 Metodología**

Se utiliza como base la metodología Crisp-DM, y es adaptada al caso de tal manera de estar alineado con los objetivos plateado. De esta manera la metodología constara de los siguientes pasos:

### **7.1 Entendimiento del negocio:**

Durante esta etapa se busca conocer el negocio, realizar los primeros levantamientos de la situación actual y definir los objetivos y alcances del problema explicado anteriormente. Se realizó la revisión bibliográfica con el fin de determinar el marco conceptual y metodologías que podrían ser usados para solucionar el problema propuesto. Además, se estudia los distintos softwares que servirán para la aplicación de los distintos métodos de modelamiento.

### **7.2 Comprensión de los datos:**

Dada la gran cantidad de información que puede almacenar una empresa, se requiere familiarizarse con las distintas fuentes de información y los datos que almacena la compañía, de esta manera se obtienen una primera aproximación de cómo es la sabana de datos necesaria para los modelos de propensión. Se realiza además un estudio exploratorio de los datos de tal manera de tener noción de cómo afecta a la variable que interesa predecir.

### **7.3 Preparación de los datos**

Una vez conocidas las tablas relevantes se seleccionarán los datos relevantes para el problema al que se le busca dar solución, para posteriormente realizar un pre-procesamiento de estos, en donde se pretende limpiar los datos mal ingresados y outliers con el fin de tener una data de mejor calidad para trabajar. Además, se transformar y se crean nuevas variables que puedan entregar un mejor ajuste a los modelos.

## **7.4 Minería de Datos**

Se construyen los modelos predictivos seleccionados para predecir la propensión de compra de los clientes para cada producto estudiado. Se evalúan los modelos mediante métricas de comparación de tal manera de seleccionar el que tenga mejor aporte al contexto estudiado.

## **7.5 Modelo de Asignación**

Se diseña un modelo de optimización que realice la asignación óptima de tal manera de maximizar las ventas del canal. Para ello se utilizan los resultados de los modelos de propensión como input y se consideraran restricciones propias del canal y de la compañía. Además, se realiza una simulación del caso de tal manera de evaluar el potencial incremento de la solución.

## **7.6 Simulación de potenciales nuevos de productos**

Dada la metodología estudiada, se evalúan los posibles nuevos productos que podría ofrecer el canal, mediante una simulación se determina que producto aumentaría las ventas esperadas.

## **7.7 Diseño Experimental**

Se plantea un diseño experimental con la finalidad de determinar el efecto de la nueva metodología en las tasas de ventas perdurables del canal.

## **7.8 Presentación de las Conclusiones**

Se presentan los principales resultados de cada una de las etapas de la metodología, con el fin de generar recomendaciones a la organización respecto al modelamiento. Además, se presentarán posibles trabajos futuros.

## **8 Desarrollo de la metodología**

### **8.1 Entendimiento del Negocio**

El presente informe tiene como propósito el aportar en la toma de decisiones del canal de Call-center, específicamente en la manera de asignar que producto ofrecer a los clientes, de tal manera de pasar de una asignación por opinión experta, a una impulsada por los datos. (Data-Driven).

En la actualidad Call-center trabaja bases de datos de 5 fuentes distintas. Dentro de ellas están las catalogadas como aperturas, esta fuente tiene la particularidad de que solo contiene clientes que abrieron una tarjeta dentro del holding durante el mes anterior. Por lo que se consideran clientes nuevos del holding. Durante el 2017, el 70% de las ventas de Call-center, procedieron de este segmento de clientes, siendo por lo tanto el más relevante para el canal.

Los productos que se ofrecen en este segmento son principalmente seguros de vida, los cuales con el paso del tiempo van variando, tanto en las coberturas, como los precios que se les ofrecen a los clientes, esto dificulta la creación de un modelo de propensión de compra de un producto.

Es por esto, que el presente proyecto busca asignar el mejor producto a cada cliente según su propensión de compra, que será estimada mediante un modelo de propensión por los atributos de los productos, de tal manera de entregar flexibilidad al cambio continuo de productos que ofrece el canal.

Para resolver la asignación, se utilizará un modelo de optimización, que busque maximizar las ventas de canal, tomando en consideración las restricciones del negocio.

### **8.2 Comprensión de los Datos**

Se trabajó directamente con la base de datos de la empresa para recopilar información. Dentro de las tablas relevantes se encuentra:

- **TMP\_BI\_CMI\_CC\_FINAL**: tabla que contiene la gestión del canal Call-center, en donde se encuentra el estado final de todos los registros. Contiene información sobre: a quien se llamó, que se le ofreció, cuando se llamó y cuál fue la respuesta del cliente.
- **Casar\_SF**: Tabla con información demográfica y transaccional de los clientes.
- **TMK\_BI\_RFM**: Tabla con variables RFM de los clientes, según un periodo de tiempo determinado.

Las variables con las que cuenta cada tabla se encuentran en anexos A. La tabla que se utiliza como base para la integración es **TMP\_BI\_CMI\_CC\_FINAL**, que contiene la gestión total del Call-center desde enero del año 2017 hasta mayo del año 2018, se encuentran cerca de 1 millón de datos.

De las gestiones del canal, dado el alcance definido, solo interesan las que fueron de la categoría de aperturas, lo que significa un 28% de los datos. Además, como es gestión de un Call-center, se debe tomar en cuenta solo aquellos llamados que fueron contestados por el cliente, de tal manera de tener respuestas. De esta manera se obtiene un total de 192 mil registros aproximadamente para el periodo estudiado. En la Tabla 4 se muestra un resumen de la limpieza de los registros.

Tabla 4: Cantidad de datos a disposición

	<b>N de datos</b>	<b>% de Datos</b>
Gestión Total de Call-center	1.025.546	100%
Gestión Segmento Aperturas	288.410	28%
Conecta Cliente	192.583	19%

Tal como se mencionó anteriormente, en la tabla **Casar\_SF**, se encuentran tanto datos demográficos como transaccionales de los clientes, dado que se trabajará con clientes nuevos muchas variables no pueden ser utilizadas. En Anexos A se detalla que variables si son utilizables y cuáles no.

Por último, la tabla **TMK\_BI\_RFM**, que contiene las variables RFM de los clientes, si bien son datos transaccionales, estos no están asociadas a la tarjeta del holding. De esta manera, si un cliente fue alguna vez cliente de la compañía, pero no tenía su tarjeta del holding abierta, igual será considerado dentro del segmento de clientes que interesa.

La integración de los datos queda de la siguiente manera:

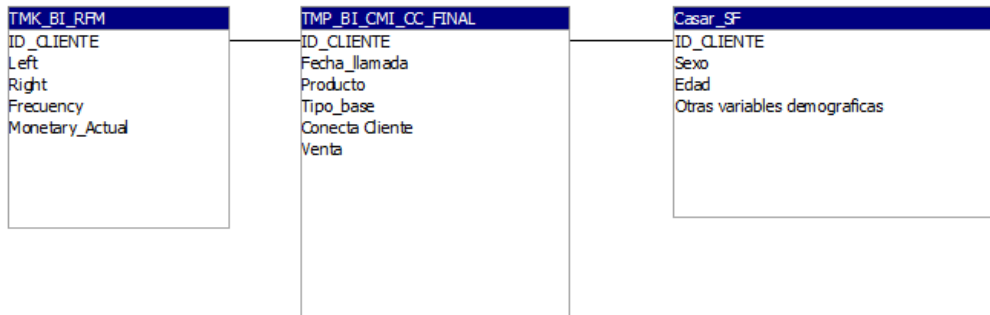


Ilustración 9: Integración de los datos  
Fuente: Elaboración Propia

### 8.2.1 Productos:

Los seguros ofrecidos por el canal Call-center durante el periodo de estudio fueron 4, los cuales se detallan en la Tabla 5.

Tabla 5 : Seguros ofrecidos por el canal Call-center

Producto	Sigla	N de Registros
Vida con Bonificación	VCB	94.790
Accidentes Personales con bonificación	AP	83.858
Accidentes Personales V4	VA4	3.369
Accidentes Personales V5	VA5	10.566
<b>Total</b>		<b>192.583</b>

Uno de los aspectos relevantes del proyecto es el considerar los atributos de los productos para realizar los modelos de propensión. Para el levantamiento de estos atributos, se conversó con el área de productos de la empresa. De esta manera se consideró que los atributos relevantes de los productos son las coberturas y los precios medios de estos. A continuación, se detallan estos atributos.

Tabla 6: Atributos de los seguros ofrecidos por el canal Call-center

Producto	Muerte Natural	Muerte Accidental	Bonificación	ITP por Accidentes	Descuento en Farmacia
AP	0	1	1	0	1
VA5	0	1	0	1	1
VA4	0	1	0	1	0
VCB	1	1	1	0	1
APA	0	0	1	0	0

Es importante señalar que los productos AP, VA5, VA4 y VCB fueron ofrecidos durante todo el 2017 y el primer semestre del 2018. Mientras que, para el segundo semestre del 2018, la compañía solo ofreció los productos APA y VA5

## 8.2.2 Análisis Exploratorio

Para tener un mayor conocimiento de cómo afectan las distintas variables a la probabilidad de compra perdurable, se realiza un análisis exploratorio de estas.

### 8.2.2.1 Variable Dependiente

Se utilizará como variable dependiente en los modelos, las ventas perdurables. Estas son definidas como aquellas ventas que tuvieron una permanencia mayor a 3 meses, estas ventas son las relevantes para la compañía, debido a su estructura de cobros por mensualidades, y no cobros únicos por ventas.

En la Tabla 7 se encuentra las ventas del canal Call-center y las ventas perdurables (aquellas que tuvieron una permanencia mayor a 3 meses). Se observa que tan solo el 8% de los registros del canal son ventas perdurables, muy por debajo de las 27 mil ventas totales, dando por lo tanto mayor relevancia a utilizar este dato como variable dependiente.

Tabla 7: Ventas de Call-center durante el 2017

	<b>Cantidad</b>	<b>%</b>
Total de Clientes	192.583	100%
Ventas	27.033	14%
Ventas Perdurables	15.509	8%

### 8.2.2.2 Descripción de las variables

A continuación, se detallan todas las variables demográficas que en principio pueden ser utilizadas para calibrar los modelos. Es importante recalcar, que variables transaccionales y muchas otras no pueden ser utilizadas debido al segmento de clientes con el que se trabaja, según lo explicado en los alcances del proyecto, ver Anexos A. Tal como se menciona anteriormente, se trabajó con una población de 192.583 clientes, lo que representan el 100% de la muestra.

#### 8.2.2.2.1 Variables Demográficas:

**Edad:** De la Ilustración 10, se observa los clientes con edades entre 18-25 años. Son aquellos con mayor tasa de ventas perdurables con un 9,1%, además son la segunda categoría más poblada con 31% de los clientes. Por otro lado, la categoría con más registros (34%), es la con menor tasa de venta perdurable (5,8%).

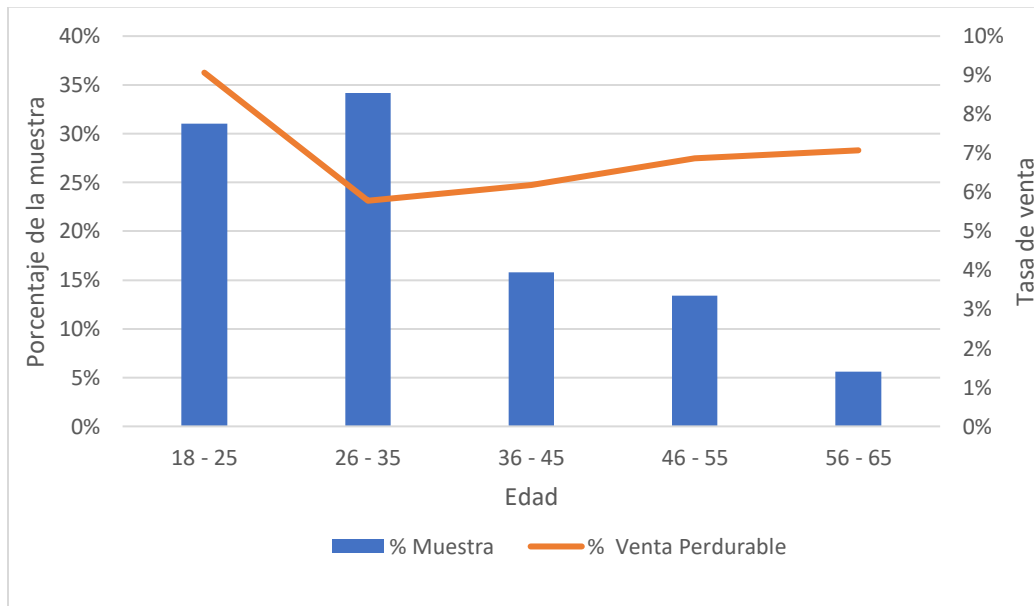


Ilustración 10: Análisis variable Edad

**Cupo:** Como se ha mencionado anteriormente, el trabajo está enfocado a un segmento de clientes denominado aperturas (aquellos que abrieron una tarjeta dentro del holding hace poco). El cupo se refiere a la capacidad de compra que tienen con la tarjeta. Se puede observar, de la Ilustración 11, que la gran mayoría recibe cupos bajo el millón de pesos, y que son estos además los que tienen la mayor tasa de ventas perdurable.

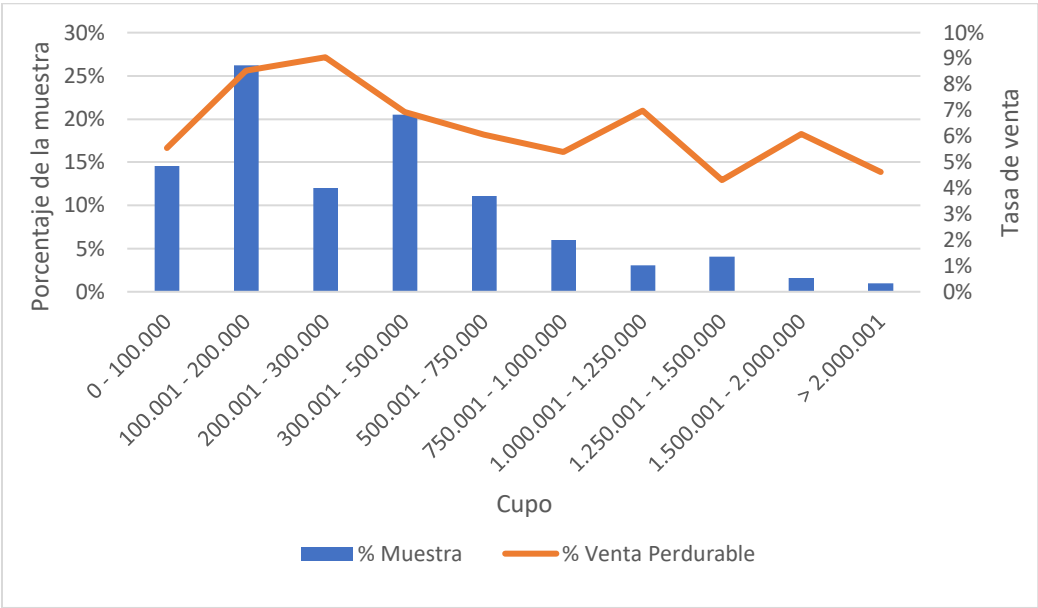


Ilustración 11: Análisis variable Cupo

**N de Hijos:** Se observa, de la Ilustración 12, que a medida que los clientes tienen más hijos el porcentaje de clientes que han sido venta perdurable aumenta.



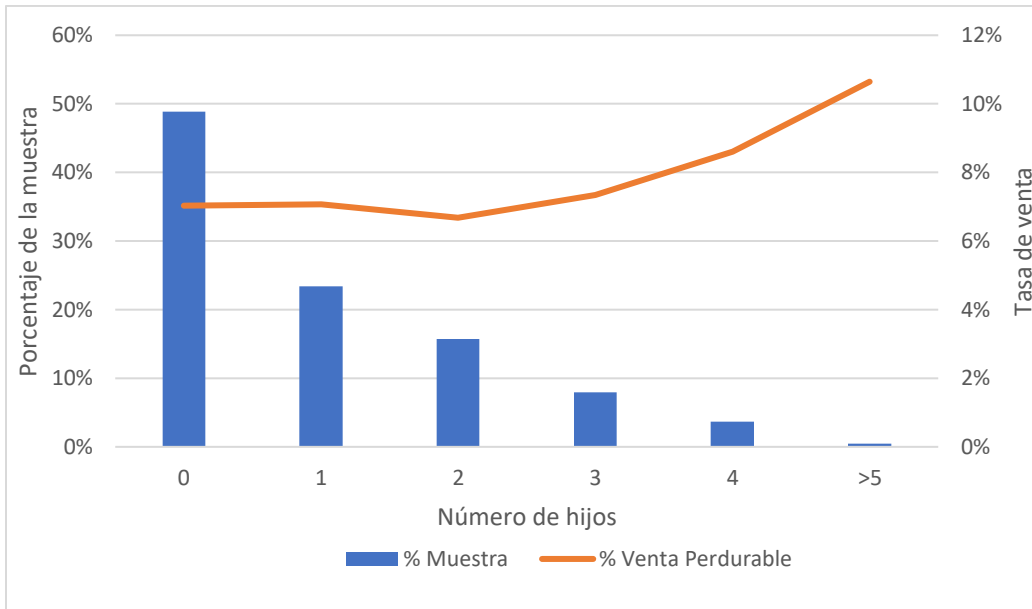


Ilustración 12: Análisis variable Número de Hijos

**N de Vehículos:** Se observa de la Ilustración 13, que más del 60% de los clientes no cuenta con un automóvil, y que son el segundo segmento de clientes con más porcentaje de ventas perdurables, superado solo por aquellos clientes que tienen más de 3 autos, pero estos representan solo el 1% de la muestra estudiada.

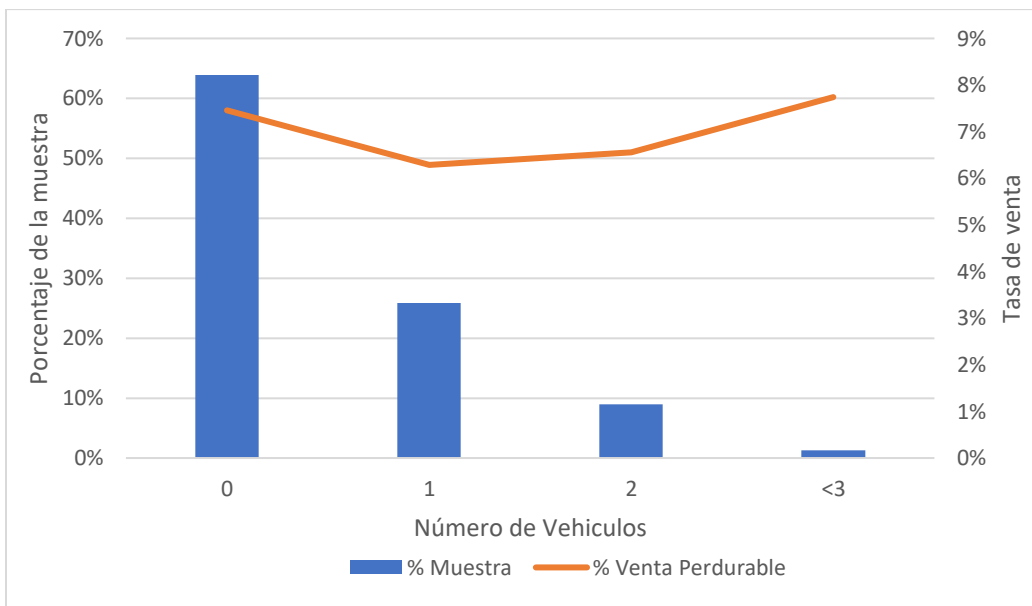


Ilustración 13 Análisis número de vehículos

**Género:** Con respecto al género, la Ilustración 14 muestra que mayor parte de la población es masculina, además, estos tienen una tasa de venta de 2,3% mayor respecto a las mujeres.

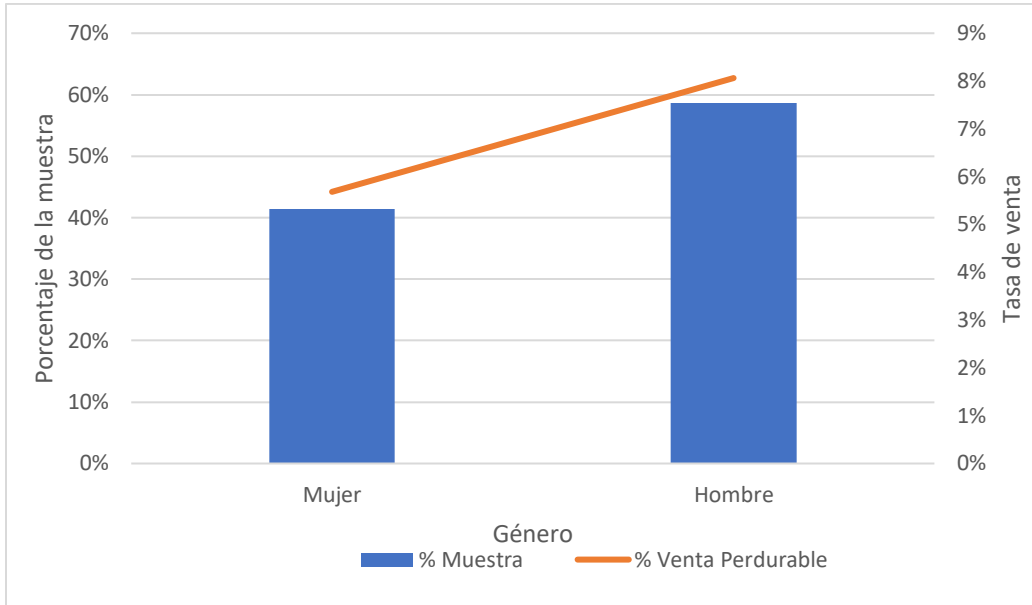


Ilustración 14: Análisis Variable Género

**Estado Civil:** Al analizar el estado civil de los clientes en la Ilustración 15, se observa que cerca de un 80% de los clientes de la muestra son solteros, y estos tienen una tasa de venta de casi 1% mayor que los casados.

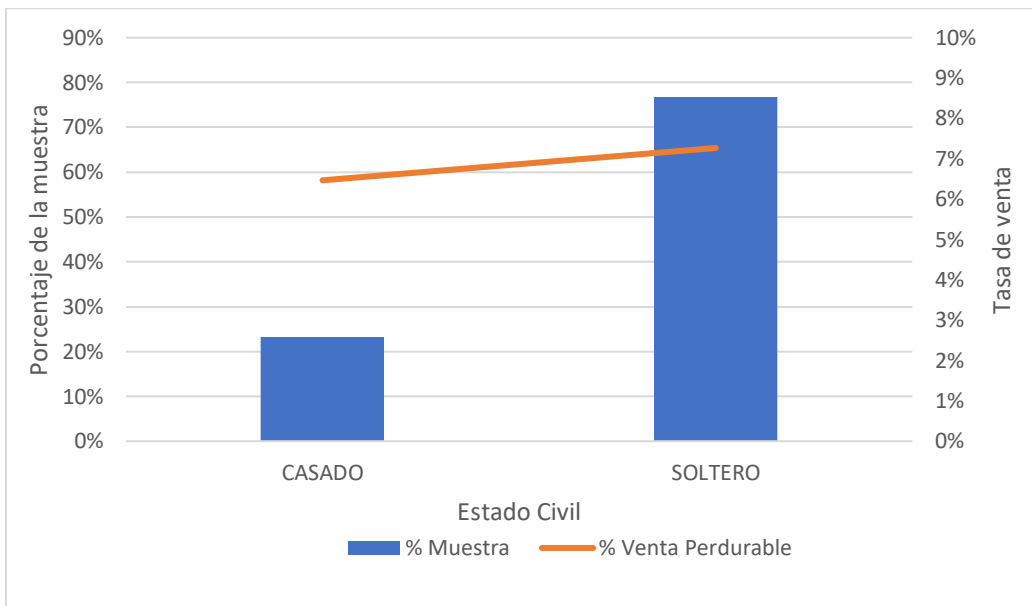


Ilustración 15: Análisis variable Estado Civil

**Habitación:** Otra variable interesante con la que se cuenta es la habitación de los clientes, tipo de domicilio en el que viven. Lamentablemente se observa de la Ilustración 16, que cerca de un 60% de los datos no se ha informado el tipo de habitación, por lo que se decide no utilizar en los modelos.

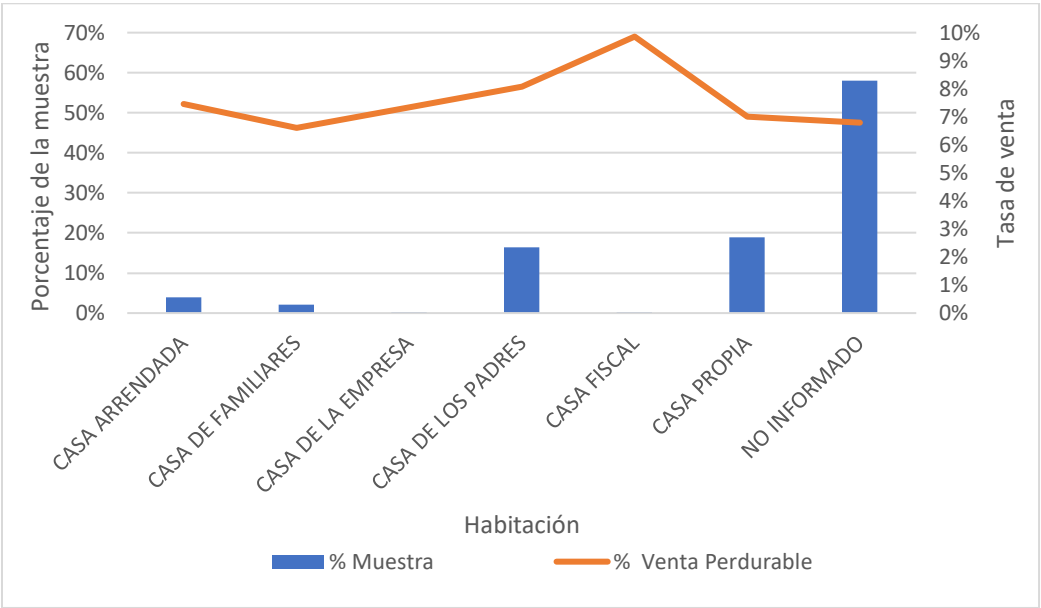


Ilustración 16: Análisis variable Habitación

**Actividad:** Con respecto a la actividad de los clientes, se observa que 5 actividades representan cerca del 98% de la muestra. Dentro de estas 5 categorías, desatacan los obreros y auxiliares con la tasa de ventas perdurables más alta de un 10% tal y como se observa en la Ilustración 17. Mientras que los profesionales universitarios empleados e independientes son los que tienen las tasas más bajas. Fuera de las 5 categorías más pobladas, se observa que los clientes con actividad de personal de seguridad tienen una tasa de venta perdurable de un 12%.

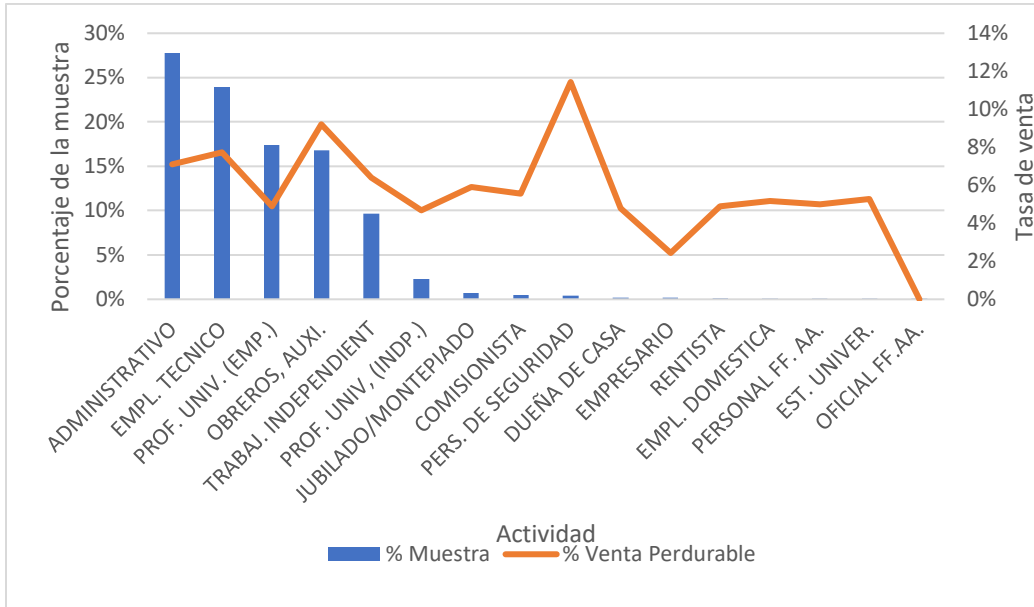


Ilustración 17: Análisis variable Actividad

**Monto renta:** Dentro de las variables con las que cuenta la empresa se encuentra el monto renta. Esta variable se encuentra segmentada en 9 categorías, en donde la categoría 1 son los clientes con renta más bajas, mientras que los que se encuentran en la categoría 9 son los clientes con la renta más alta. La Tabla 8 detalla la construcción de estos segmentos.

Tabla 8: Segmentación variable Monto Renta

INICIO	FIN	RANGO
0	250.000	1
250.001	500.000	2
500.001	750.000	3
750.001	1.000.000	4
1.000.001	1.250.000	5
1.250.001	1.500.000	6
1.500.001	1.750.000	7
1.750.001	2.000.000	8
> 2.000.000		9

De la Ilustración 18, se observa que los clientes que se encuentran en los segmentos medios bajos son los más poblados. Además, se observa que los clientes del primer segmento son los que tienen menor probabilidad de compra.

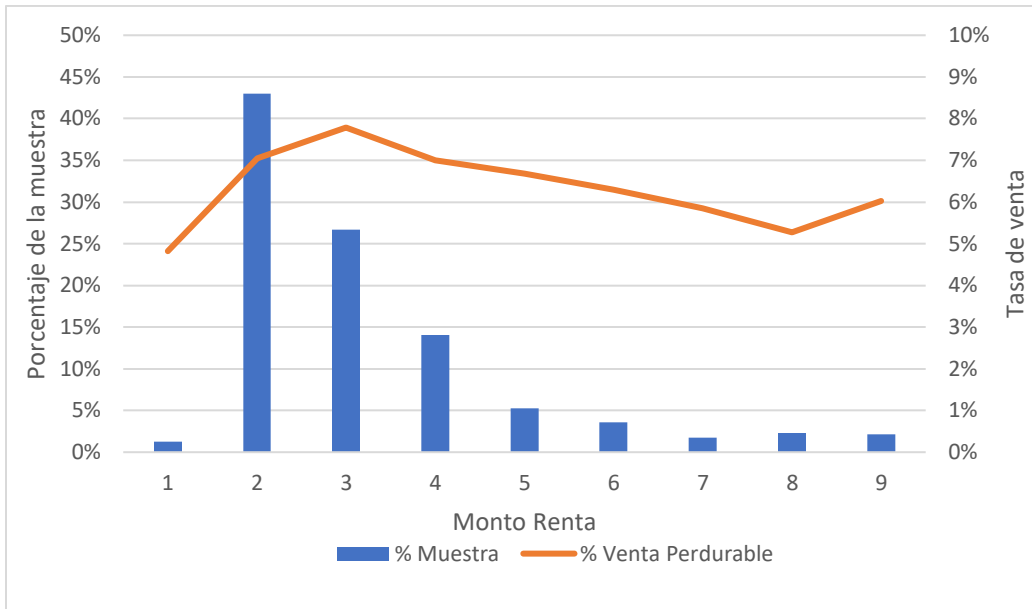


Ilustración 18: Análisis variable Monto Renta

**Gse:** Sobre la variable GSE se observa en la Ilustración 19, que para más del 95% de los clientes no se tienen información, dado esto la variable no será utilizada en los modelos.

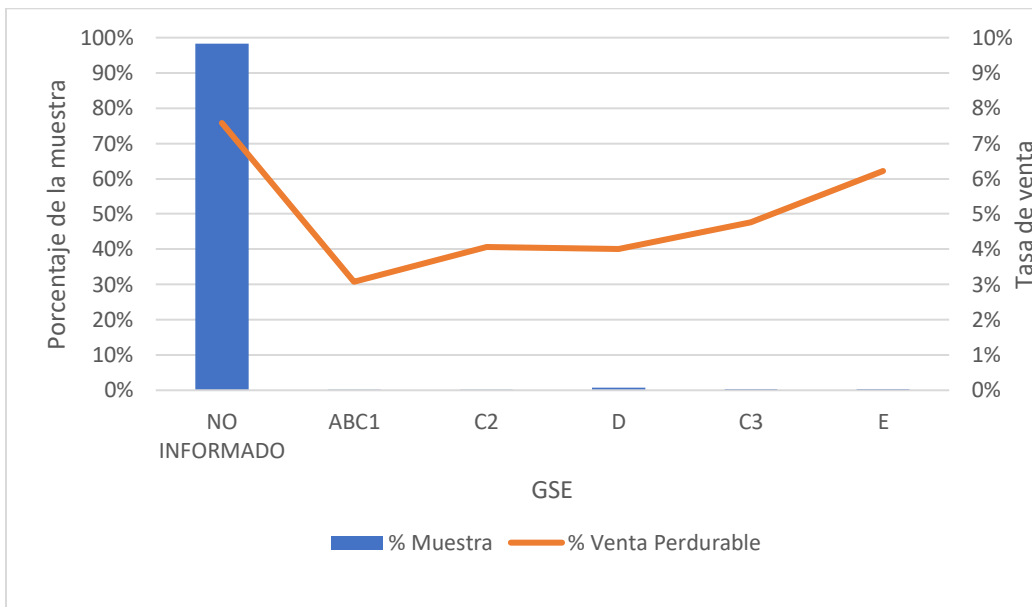


Ilustración 19: Análisis variable Gse

**Región:** En la Ilustración 20 ,se observa una gran concentración de clientes en la Región Metropolitana, pese a esto, en las regiones del norte y sur del país se observa una tasa de venta perdurable más alta.

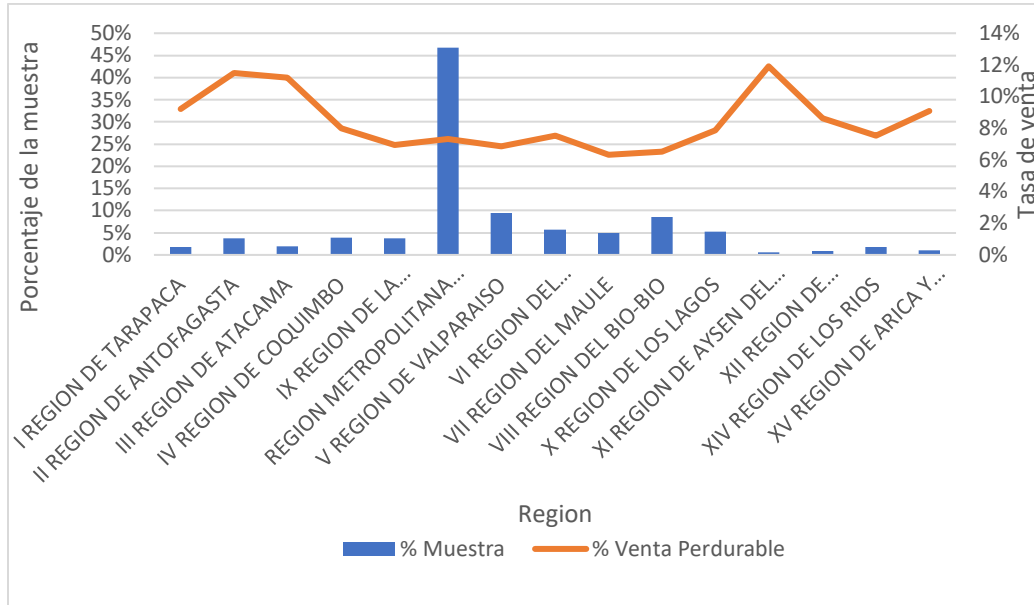


Ilustración 20: Análisis variable Región

#### 8.2.2.2.2 Variables RFM:

Tal como se mencionó anteriormente, se cuenta con variables RFM, las cuales serán explicadas a continuación:

- **Right:** Esta variable está definida como cuánto tiempo ha pasado desde la última contratación hasta el periodo de generación de bases. Se observa de la Ilustración 21, que la mayoría de los clientes estudiados han comprado un seguro en el último mes. Mientras que la muestra restante no tiene compra anterior en Seguros Falabella.

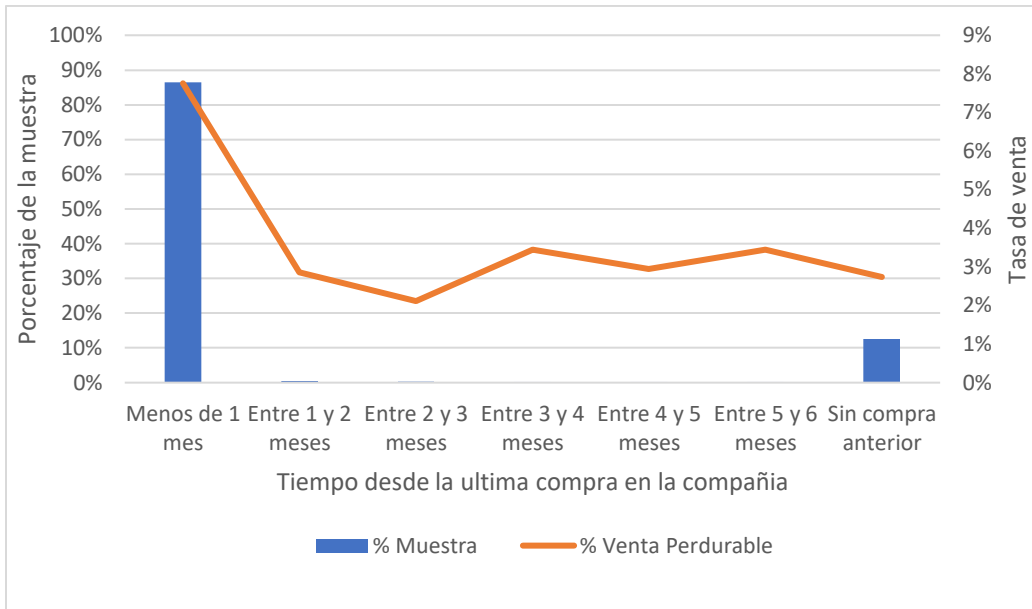


Ilustración 21: Análisis variable Right

- Left: Esta variable está definida como el tiempo que ha pasado desde la primera contratación hasta el periodo de generación de bases. Con respecto a esta variable se observa algo similar al caso del Right, en donde la mayoría de los clientes contrataron su primer producto hace menos de un mes tal y como se ve en la Ilustración 22.

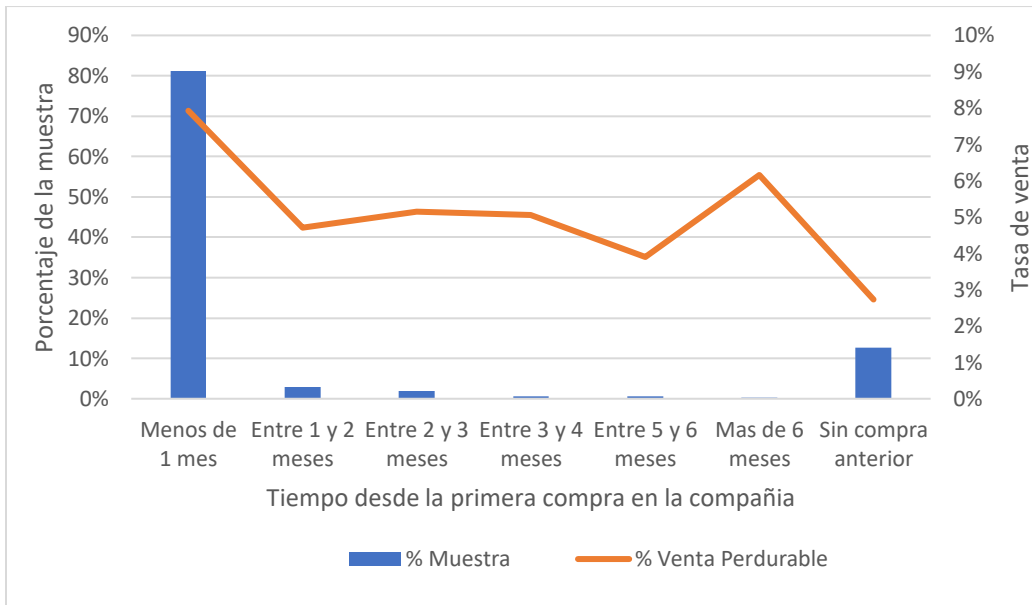


Ilustración 22: Análisis variable Left

- **Frequency:** Esta variable está definida como la cantidad de seguros que tiene el cliente antes del periodo de generación de bases. Se observa de la Ilustración 23, que la mayoría de los clientes tienen algún seguro contratado antes del periodo de generación de variables, esto es debido a que al momento de que el cliente abre su tarjeta con la compañía, se le ofrecen seguros asociados a la tarjeta. Se observa que los clientes que tienen una mayor cantidad de seguros contratados previamente a ser llamados son más propensos a comprar los productos.

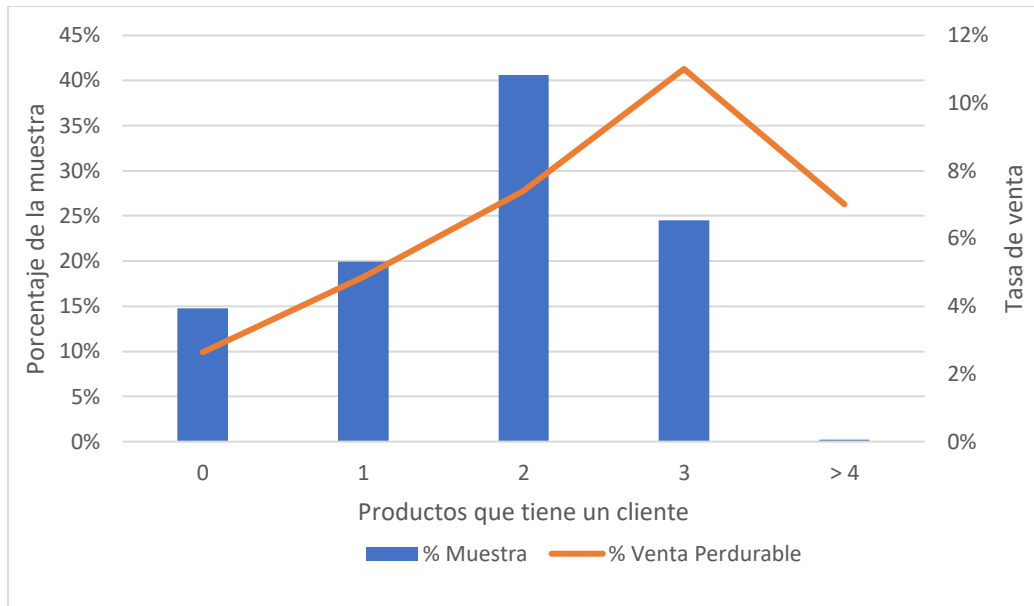


Ilustración 23: Análisis variable Frequency

- **Monetary Actual:** Esta variable está definida como el monto en UF gastadas por los clientes en seguros en la compañía, al momento previos de ser contactados por Call-center. Se observa de la Ilustración 24 que cerca del 50% no tiene gastos asociados, esto puede deberse a que los seguros contratados durante el primer mes de la apertura de la tarjeta tienen costo 0 por un par de periodos. En general se observa que, a mayor gasto, mayor es el porcentaje de ventas perdurables salvo por dos puntos.



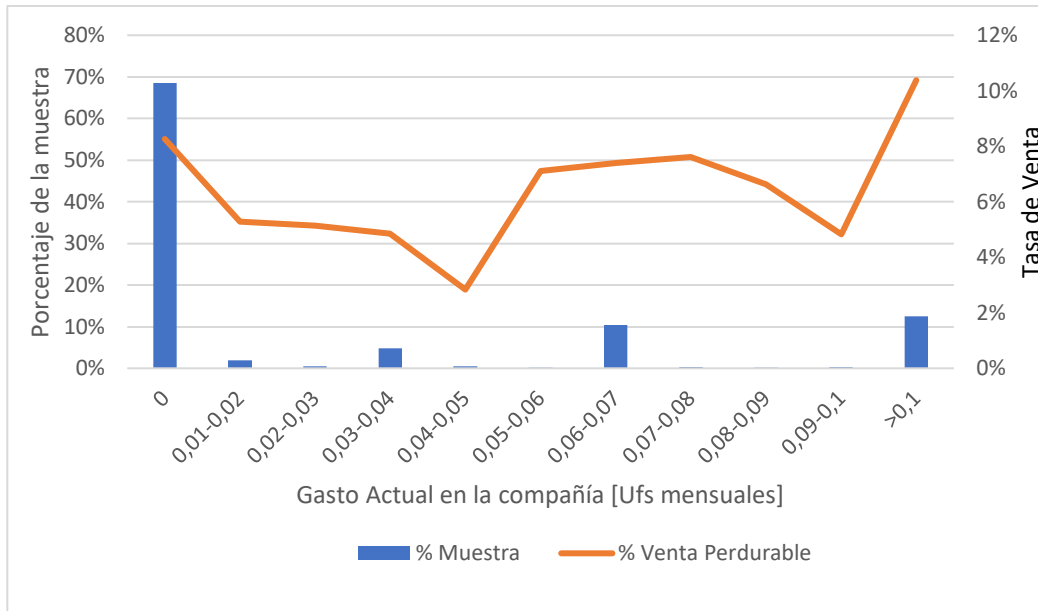


Ilustración 24: Análisis variable Monetary actual

### 8.2.3 Conclusiones del análisis exploratorio

De la Tabla 9, se observa que con respecto a la variable edad, existe un mínimo de 18 años, esto debido a que los clientes para poder abrir una tarjeta en el holding deben tener la mayoría de edad. Con respecto al máximo, dado que se ofrecen seguros de vida, estos no pueden ser ofrecidos a clientes mayores de edad.

Tabla 9: Análisis variables numéricas

Variable	Media	Desviación Estándar	Máximo	Mínimo	1° Cuartil	3° Cuartil
EDAD	32,98	10,67	59	18	25	41
N° HIJOS	0,90	1,20	14	0	0	2
N° VEHÍCULOS	0,51	0,94	70	0	0	1
CUPO	440.047	468.282	15.000.000	10.000	150.000	510.000
LEFT	6,72	19,20	205	0	0	5
RIGHT	2,33	7,30	194	0	0	1
FREQUENCY	1,74	1,03	20	0	0	2
MONETARY	0.016	0.16	10	0	0	0

Sobre los números de hijos, se observa que el promedio es 0,9 hijos por persona, y que el máximo es 14 hijos, muy lejano al promedio, lo que da señales de problemas de outliers en los datos. Este problema se ve reforzado en las variables N° HIJOS,

N° VEHÍCULOS, CUPO, LEFT, RIGHT, FREQUENCY, Dado esto en la etapa de pre procesamiento de los datos se debe tener una etapa de limpieza.

Además, se observa que, si bien existen muchas variables interesantes de los clientes, la gran mayoría de estas no son utilizables en el presente trabajo, debido a que se estimara la propensión de compra de clientes nuevos, y la información que se tiene de estos es limitada. Pese a esto, en el análisis exploratorio se encontraron insights interesantes con respecto al comportamiento de los clientes. Por ejemplo, la mayoría de los clientes son jóvenes menores a 35 años solteros, siendo además este grupo quienes tienen mayor tasa de ventas perdurables. Este resultado es interesante, dado que se están estudiando los productos de seguros de vida, y se podría asumir erróneamente, que los clientes mayores estarían más interesados.

Otra conclusión importante del análisis exploratorio, es que los clientes con cupos medios bajas tienen mayor tasa de venta perdurable, mientras que los clientes con mayor cupo tienen tasas más bajas, por lo que pese a que son clientes con mayor poder de compra parecen no estar tan interesados en los productos.

### **8.3 Preparación de los Datos**

Se comenzó realizando una limpieza de los datos nulos o mal ingresados, cabe destacar que no se encontró muchos problemas en estos términos, de 192.583 datos que se tenían inicialmente quedaron 189.064, en donde el principal problema eran Ruts repetidos.

Posterior a esto, se realizó la eliminación de datos outliers en las variables numéricas. Se utilizó la metodología de los diagramas de caja y el rango inter cuartil, para la detección de valores atípicos. De esta manera, la sabana final con la que se trabajará cuenta con 180 mil datos aproximadamente.

Se creó la variable Prima/Cupo con el fin de determinar cómo afecta la capacidad de pago del cliente a cada producto, a la probabilidad de compra perdurable. Y se categorizaron algunas variables para evaluar si los modelos obtenían un mejor ajuste.

La Tabla 10 resume las variables que serán utilizadas en la etapa de minería de datos.

Tabla 10: Variables que se utilizaran en la etapa de minería de datos

<b>Variable</b>	<b>Tipo</b>	<b>Categorías</b>
Venta Perdurable	Binaria dependiente	1 = Venta perdurable
		0 = No venta
Cupo	Numérico/Categórico	-
Edad	Numérico/Categórico	-
Left	Numérico	-
Right	Numérico	-
Frecuency	Numérico	-
Monetary Actual	Numérico	-
N de Hijos	Numérico	-
N de Vehículos	Numérico	-
Estado Civil	Categórico	Soltero
		Casado
Zona	Categórico	Norte
		Sur
		Centro
		Oriente
		Poniente
Género	Categórico	Masculino
		Femenino
Monto Renta	Categórico	Bajo (Bajo Categoría 6)
		Alto (Categoría 6 hacia arriba)
Actividad	Categórico	Administrativo
		Empl. Tecnico
		Profesional Univ. (Empleado)
		Obreros y Auxiliares
		Trabajador Independiente
		Profesional Univ. (Independiente)
		Otros
Bonificación	Binaria	1
		0
Descuento de farmacia	Binaria	1
		0
Muerte Natural	Binaria	1
		0

Por último, se hizo un análisis de correlación de las variables numéricas, de tal manera de determinar variables que puedan causar problemas de colinealidad en el modelo. En Anexos B, se encuentra la matriz correlación de algunas de las variables del modelo.

Destaca la alta correlación entre las variables Bonificación y ITP por accidentes, en donde se muestran linealmente dependientes, por lo que para efectos de los modelos se trabajará solo con la variable Bonificación, dado que esta variable está presente en el producto nuevo. Junto con lo anterior, se ve en general alta correlación entre las variables de atributos de los productos, esto señala la poca varianza que existe entre el mix de productos evaluados (Todos son seguros de vida). Además, se tiene que el precio es linealmente dependiente de las otras variables de los atributos, por lo que por sí sola la variable no puede entrar en el modelo

Con respecto a las otras variables, se observa que la edad y el Número de hijos, Left y Right, y Frequency y Right, se encuentran correlacionadas entre sí, pese a esto la correlación no es lo suficientemente alta como para encontrar problemas de multicolinealidad en las variables.

## **8.4 Minería de datos.**

Dentro de las preguntas más relevantes que se hace el marketing es ¿Cómo responderá el cliente a una promoción o producto? Durante este capítulo, se realizarán 3 metodologías de minería de datos con el fin de estimar la propensión de compra de los clientes.

### **8.4.1 Árbol de decisiones**

Para la estimación de los clientes más propensos a las compras perdurables, se calibraron arboles de decisión. Se dividió la muestra en dos categorías, se utilizó el 70% de los datos para entrenar los modelos y el 30 % restante como grupo de entrenamiento del modelo, con el fin medir evaluar los modelos.

Dado que existe una baja proporción de clientes que son ventas perdurables (solo el 7% de los datos.) se realiza un over-sampling de esta variable con el fin de conseguir un mejor ajuste [16]. Se utiliza la metodología de random over-sampling, es importante señalar que esta metodología solo fue aplicada al grupo de entrenamiento, mientras que el de testeo no se modificó.

Se utilizaron 2 metodologías distintas de árboles de decisión: C.H.A.I.D y C.A.R.T, en las cuales se fueron ingresando todas las variables descritas en Tabla 10.

Posteriormente, se evaluaron los resultados según las métricas señaladas en la sección, con el fin de elegir el mejor modelo.

En Anexos C, se muestran las curvas de crecimiento y las matrices confusión de los 3 árboles evaluados en el grupo de testeo, es importante señalar que para las matrices confusiones se utilizó un corte de 0,5, debido a que se está trabajando con clases balanceadas.

Al compararlos el árbol con mejor desempeño es el de corte tipo C.H.A.I.D. Las métricas de desempeño de los distintos árboles se muestran a continuación resumidos en la siguiente tabla.

Tabla 11: Resultados de los árboles de decisiones.

Metodología	% Acumulado 1° decil	Sensibilidad	Especificidad	Exactitud
C.H.A.I.D	% 19	0,60	0,58	0,58
C.A.R.T	% 16	0,63	0,55	0,56

Se observa que el modelo con mejor desempeño con respecto al % acumulado en el primer decil es el árbol tipo C.H.A.I.D, lo que señala que este modelo ordena de mejor manera las propensiones de los clientes, además se observa que tanto en la especificidad (detección de verdaderos negativos) como en la Exactitud (detección de verdaderos negativos y verdaderos positivos) se tienen mejor desempeño con este tipo de árbol. El único indicador en donde no destaca el árbol tipo C.H.A.I.D es en la sensibilidad, en donde se encuentra 3 puntos porcentuales bajo el resto. Se considera en porcentaje acumulado en el 1° decil como el indicador con más peso a la hora de seleccionar un modelo, es por eso que se selecciona el árbol tipo C.H.A.I.D.

El esquema del árbol se puede ver en Anexos C, en donde se destaca que no se realizan cortes respecto a los atributos de los productos, lo que señala que para el modelo estos atributos no aportan información relevante en la propensión de compra perdurable, esto da indicios a que modelo

#### 8.4.2 Modelo Logit al ramo vida.

Como se mencionó anteriormente, es interesante estudiar el comportamiento de los clientes a la compra perdurable de seguros de vida. Se realiza por lo tanto un modelo genérico sin diferenciar por producto. Al igual que con los modelos de

árboles decisiones, se dividió la muestra en dos categorías, el 70% de ella se utilizó para entrenar el modelo, mientras que el 30% restante fue utilizado de testeo, de tal manera de medir el rendimiento de los modelos.

Dado que existe una baja proporción de clientes que son ventas perdurables se volvió a utilizar la metodología del random over-sampling para entrenar el modelo, mientras que el grupo de testeo se dejó igual. Al ejecutar el modelo Logit se obtiene un R2 de negerleke 0.05 lo cual es bastante bajo. Esto puede ser debido a que como se está trabajando con clientes nuevos, se tiene muy poca información sobre ellos por lo que existen otros factores que no se están considerando que podrían afectar la probabilidad de compra perdurable. La Tabla 12 muestra los coeficientes obtenidos.

Tabla 12: Coeficientes modelo Logit

Variable	Categorías	Coeficiente	Error Estándar	P valor
Intercepto	-	-0,64	0,03	0
Left	-	0,01	0	0
Right	-	0,02	0	0,04
Frecuency	-	0,38	0,01	0
N de Hijos	-	0,02	0,01	0
N de Vehículos	-	-0,06	0,01	0
Estado Civil	Soltero	-0,02	0,01	0,08
	Casado	-	-	-
Zona	Norte	0,2	0,01	0
	Sur	0,16	0,01	0
	Centro	-	-	-
	Oriente	-0,15	0,04	0
	Poniente	-0,01	0,01	0,49
Género	Masculino	0,31	0,01	0
	Femenino	-	-	-
Monto Renta	Bajo	-0,17	0,02	0
	Alto	-	-	-
Actividad	Administrativo	-	-	-
	Empl. Tecnico	-0,03	0,01	0,03
	Profesional Univ. (Empleado)	-0,3	0,02	0
	Obreros y Auxiliares	-	-	-

Variable	Categorías	Coefficiente	Error Estándar	P valor
	Trabajador Independiente	0,05	0,02	0,01
	Profesional Univ. (Independiente)	-0,4	0,04	0
	Otros	0,18	0,01	0
Rango Edad	18-25	-	-	-
	26-35	-0,36	0,01	0
	36-45	-0,36	0,02	0
	46-55	-0,29	0,02	0
	>56	-0,13	0,02	0
Rango Cupo	<100.000	-	-	-
	100.001-200.000	0,32	0,01	0
	200.001-300.000	0,33	0,02	0
	300.001-500.000	0,05	0,02	0
	500.001-750.000	-0,17	0,02	0
	750.001-1.000.000	-0,21	0,02	0
	1.000.001-1.250.000	-0,1	0,03	0
	1.250.001-1.500.000	-0,34	0,03	0
	1.500.001-2.000.000	-0,44	0,15	0
>2.000.001	-0,55	0,22	0,01	

#### 8.4.2.1 Análisis de los coeficientes:

- Cupo: El modelo señala que en general los clientes que tienen cupos bajos, son más propensos a la compra perdurable de los productos. Esto puede deberse al segmento de clientes que apunta a la compañía, que en general es masivo y no de elite. Se eligió esta variable sobre Precio/Cupo, debido a que estaban correlacionadas y entregaba mayor información.
- Rango Edad: Con respecto a los coeficientes relacionados a la edad, se tienen valores negativos en todos los rangos, por lo que todas estas categorías con respecto a la categoría base (edad entre 18 y 25 años) tienen una propensión de compra perdurable más baja. Lo que confirma lo observado en el análisis exploratorio.
- Género: Se observa que los clientes cuyo género es masculino son más propensos a la compra perdurable de los productos.

- Estado Civil: Contrario a lo que se podía pensar, se observa un coeficiente negativo en el estado civil casado. Por lo tanto, aquellos clientes solteros son más propensos a la compra perdurable de los seguros estudiados.
- Actividad: Se observa que los clientes que están en la categoría de otras actividades son los más propensos a comprar seguros de vida. Por otro lado, los profesionales universitarios son aquellos menos propensos.
- Monto Renta: Se observa que quienes tienen Monto Renta en la categoría baja, son menos propensos a la compra perdurable de los productos. Puede deberse a que al tener una renta más limitada no están dispuestos a gastar mensualidades en seguros.
- N de Hijos: Al igual que en el análisis exploratorio se observa que a mayor cantidad de hijos, los clientes son más propensos a la compra de seguros de vida. Esto se puede dar debido a que los clientes al tener más hijos son más conscientes de la importancia de asegurar su vida de tal manera que ante cualquier problema sus hijos tengan ayuda económica.
- N de Vehículos: La variable N de Vehículos no resultó significativa con un intervalo del 95% de confianza, por lo que no se puede concluir nada al respecto.
- Left: Se observa que el coeficiente de esta variable es negativa, por lo que entre más lejana haya sido la primera compra del cliente en el negocio, más baja la probabilidad de compra del cliente. Su coeficiente es bastante bajo, por lo que el efecto de la variable con respecto a la probabilidad de compra no es mucho, pese a ser significativa.
- Right: Con respecto a la variable Right, este tiene un coeficiente positivo, por lo que entre más tiempo ha pasado desde la última compra mayor será la probabilidad de compra. El coeficiente es bastante bajo, por lo que el efecto de la variable con respecto a la probabilidad de compra no es mucho, pese a ser significativa.
- Frecuency: Se observa que entre más seguros tenga el cliente con la compañía, más propenso es a la compra perdurable de algún producto ofrecido. Esto hace sentido, dado que, si un cliente tiene varios seguros con la compañía, significa por un lado que le interesa tener seguros, y por otro lado podría demostrar que confía en la compañía.
- Zona: El modelo entrega que los clientes de las Zonas Norte, Sur y Poniente de Santiago, son más propensos a la compra de seguros de vida. Mientras que la zona oriente de Santiago son los menos propensos a comprar.



### 8.4.2.2 Resultados

Se utilizaron las mismas métricas que para el árbol de decisiones sobre los datos del grupo de testeo. Respecto a la matriz de confusión, Tabla 13, del total de las compras reales, el modelo fue capaz de predecir el 60%, además fue capaz de predecir el 64% de los resultados que no terminaron en compra. Por último, catalogo correctamente el 63% de los datos.

Tabla 13: Matriz confusión modelo Logit para el ramo vida con corte de 0,5

		Real	
		No Compra	Compra
Predicho	No Compra	29.230	1.637
	Compra	18.841	2.563

Respecto a la curva de ganancia acumulada, se observa de la Ilustración 25, que con el 10% de los registros, se pueden capturar el 21% de las ventas. Este indicador es el más relevante debido a que al momento de tener que asignar, es importante poder ordenar de buena manera las propensiones de los productos.

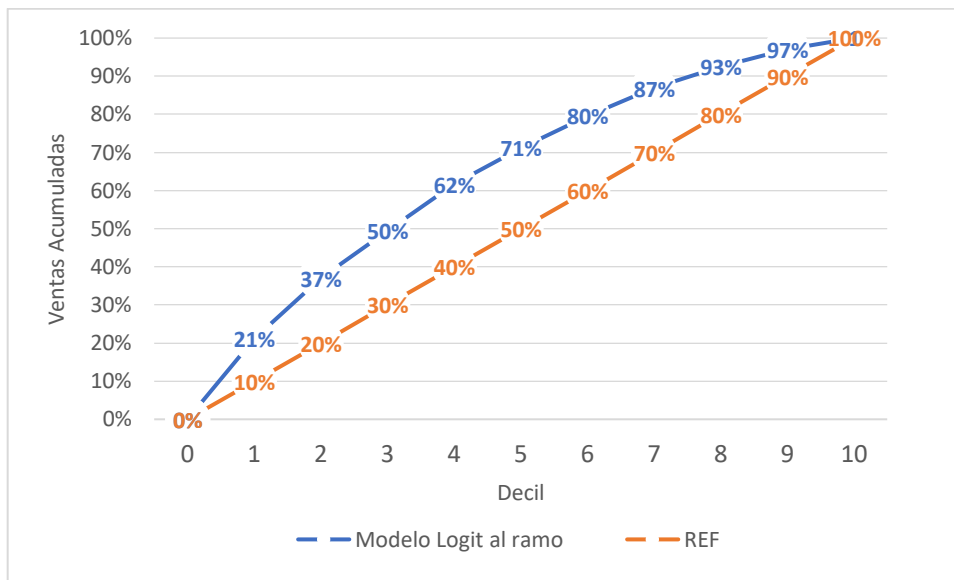


Ilustración 25 Curva acumulada de ganancia del modelo Logit para ramo vida

### **8.4.3 Modelo Logit con interacción entre variables demográficas y atributos de los productos.**

Para el cálculo de las propensiones por productos, se realizó un modelo Logit que incorporara variables de interacción entre las variables de los clientes y los atributos de los clientes, de esta manera se pretende evaluar la propensión de los clientes a los distintos atributos y como afectan los atributos a las distintas variables. [17]. Se dividió la muestra en dos grupos, en donde el 70% se utilizó para entrenar el modelo, mientras que el 30% restante se utilizó como grupo de testeo. Se volvió a utilizar la metodología del random over sampling para igualar las clases.

Al agregar las variables interaccionadas, se generan 36 nuevas variables (la interacción entre 12 variables demográficas y 3 variables de atributos). Se fueron agregando las variables una a una de tal manera de observar cuales aportaban al ajuste del modelo. Los resultados del modelo resultante se pueden ver en Anexos D, en donde destaca que la variable Prima/Cupo no fue seleccionada, esto debido a que la variable Rango cupo aportaba mayor información y al estar correlacionadas se eligió esta última.

#### **8.4.3.1 Análisis de los coeficientes:**

Debido a que este modelo solo difiere del anterior en la incorporación de los atributos de los productos, se analizarán solo estas variables, puesto que las otras tienen en general la misma interpretación de antes, con la salvedad que se deben considerar que no se le está ofreciendo ningún atributo:

- **Bonificación:** Se observa que ofrecer el atributo de bonificación aumenta la probabilidad de compra de los clientes por sí sola. Dado los coeficientes, es recomendado ofrecer esta cobertura a clientes que tengan varios hijos y que vivan en la zona oriente de Santiago, debido a que se observan coeficientes positivos. Por otro lado, según los coeficientes, no es recomendable ofrecer esta cobertura a clientes hombres, que ya tengan varios seguros con la compañía y que vivan en el sector poniente de Santiago.
- **Descuento de Farmacia:** Según los coeficientes del modelo, se observa que el atributo de descuento de farmacia por sí solo aumenta la probabilidad de compra de los clientes. Dado los resultados del modelo, se recomienda ofrecer esta cobertura a los clientes de la zona centro, casados y con autos. Por otro lado, no se recomienda ofrecer esta cobertura a clientes solteros, que pertenezcan a la zona oriente o poniente de Santiago.

- Muerte Natural: Esta variable tiene un coeficiente negativo, por lo que por sí sola disminuye la probabilidad de compra de los clientes. Se recomienda ofrecer esta cobertura a clientes con monto renta bajo que vivan en la zona norte del país, o poniente de Santiago. A su vez, no se recomienda ofrecerle este producto a clientes solteros que tengan hijos.

#### 8.4.4 Resultados

Para realizar la matriz de confusión es relevante señalar el nivel de corte, dado que se balancearon las clases mediante el método de random over-sampling. Se trabajará con un modelo de corte de 0,5. Las matrices confusión fueron construidas a partir de los datos de testeo y se pueden observar en la Tabla 14.

Tabla 14: Matriz Confusión modelo Logit con interacciones corte 0,5

		Real	
		No Compra	Compra
Predicho	No Compra	28.240	1.554
	Compra	19.831	2.646

De esta manera se obtiene una sensibilidad del 63%, con respecto a la exactitud se observa una ratio similar, tan solo el 59% de los registros fueron catalogados correctamente. Sobre la especificidad se observa resultados un poco mejores, en donde el 58% de los registros catalogados como no compra, efectivamente no compraron.

Con respecto a la curva de ganancia acumulada, también fue construida a partir de los datos de testeo que se observa en la Ilustración 26. Con el 10% de los registros se pueden obtener el 20% de las ventas totales. Tal como se mencionó anteriormente, la curva de ganancia se considera el indicador con mayor relevancia a la hora de comprar los modelos.

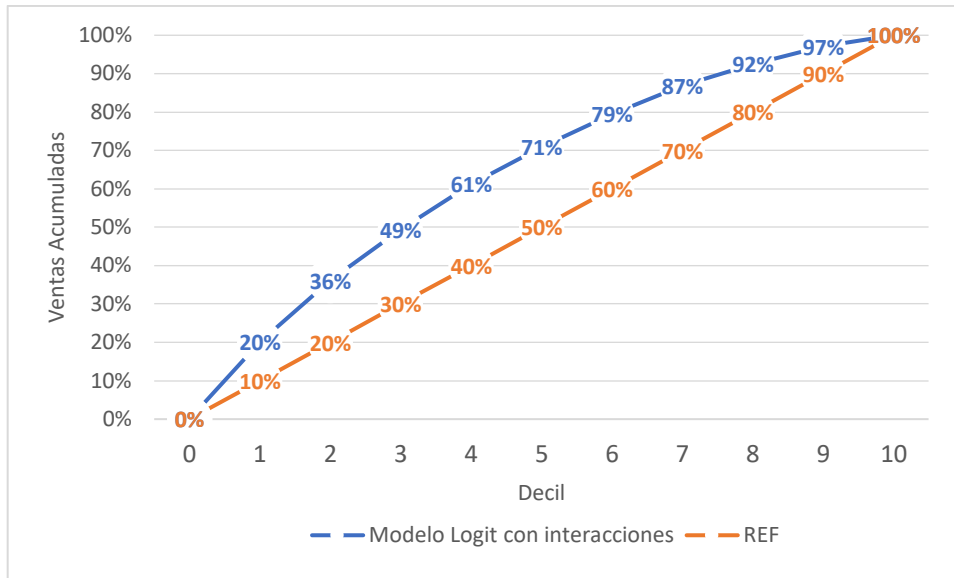


Ilustración 26 Curva de ganancia acumulada para el modelo Logit con interacciones

#### 8.4.5 Elección del mejor Modelo

En Tabla 15 se resumen las ratios estudiados, para la elección final del modelo. Se observa el modelo Logit enfocado al a la categoría vida tiene valores más altos, por lo que se tiene un mejor ajuste con este modelo. Esto señala que al diferenciar los modelos por producto no se está ganando en ajuste de los modelos, por lo que los clientes son propensos a comprar seguros de vida independiente de los atributos ofrecidos por los productos. Dado el problema de negocio que se está evaluando requiere poder discernir entre que productos ofrecer, se optara por elegir el modelo de Logit con interacciones, aunque signifique perder un poco de ajuste del modelo. De esta manera el modelo seleccionado da mayor flexibilidad a la hora de evaluar distintos productos, incluso productos que con los que no fue entrenado. Es importante señalar, que los valores obtenidos en las métricas de comparación, no son muy altos en ambos modelos, esto se debe a que se está trabajando con clientes nuevos, por lo que la información que se tiene de ellos es limitada y pueden existir otros factores que influyan en la probabilidad de compra que el modelo no está considerando.

Tabla 15: Resultados Métricas de comparación

<b>Técnica de minería de Datos</b>	<b>% 1° decil acumulado</b>	<b>Sensibilidad</b>	<b>Especificidad</b>	<b>Exactitud</b>
Modelo Logit a la categoría Vida	21%	61%	61%	61%
Modelo Logit con variables atributos.	20%	63%	59%	59%
Árbol C.H.A.I.D	19%	60%	58%	58%

#### 8.4.6 Validación del modelo seleccionado

Los datos utilizados para la validación fueron de los meses de Junio y Julio del 2018. Durante estos dos meses se estuvo vendiendo los productos VCB, VA5, AP y APA, mientras que el producto VA4 ya no se estaba vendiendo. La Tabla 16 muestra el número de registros y ventas perdurables por producto.

Tabla 16: Datos utilizados para la validación del modelo final

<b>Productos</b>	<b>AP</b>	<b>APA</b>	<b>VA5</b>	<b>VCB</b>
<b>Registros</b>	7.753	19.309	947	5.480
<b>Ventas Perdurables</b>	743	2.096	56	461

De esta manera se validará cómo se comporta el modelo con respecto a los productos con datos fuera de la muestra que se utilizó como entrenamiento y testeo. Además, se podrá evaluar como el modelo estima la probabilidad de compra de los clientes a los que se les ofrece el producto nuevo APA, dado que durante estos dos meses fue la primera vez que se ofreció. Las curvas de crecimiento y las matrices de confusión para cada producto se encuentran en Anexos D.

Tabla 17: Resultados de la validación

<b>Producto</b>	<b>% Acumulado Primer decil</b>	<b>Sensibilidad</b>	<b>Especificidad</b>	<b>Exactitud</b>
AP	21%	67%	58%	59%
VCB	20%	64%	64%	64%
VA5	33%	91%	78%	81%
APA	19%	65%	56%	57%

De la Tabla 17, se extrae que el modelo se comporta de manera similar al testeo en los distintos productos, lo que señala que es un modelo robusto. Se observa que en

los productos AP y VCB se obtiene un 21% y un 20% de ventas acumuladas en el primer decil, lo que está dentro de lo esperado según lo visto en el testeó. Además, se observa que el producto que tienen los mejores indicadores es el VA5, pero esto puede deberse a que se tienen muy pocos registros (947 datos y 56 de estos son ventas perdurables), por lo que los números no son muy representativos. Con respecto al producto APA, el cual es el producto nuevo que se lanzó en la compañía, se observa que el modelo se comportó de manera similar al grupo de testeó, por lo tanto, pese a que no se entrenó el modelo con datos de este producto, se logró obtener ratios esperados, dado lo estudiado en el grupo de testeó.

Por último, se observó la correlación de las propensiones obtenidas por productos. En la Tabla 18 se observa que las propensiones por productos tienen una alta correlación, y que por lo tanto aquellos clientes que son propensos a comprar un producto en general son altos a comprar los otros productos también.

Tabla 18: Correlación de las propensiones por productos

Productos	AP	APA	VA5	VCB
AP	1,0	0,8	0,9	0,9
APA	0,8	1,0	0,8	0,8
VA5	0,9	0,8	1,0	0,9
VCB	0,9	0,8	0,9	1,0

## 8.5 Modelo de Asignación

En este capítulo se detallan modelo de asignación propuesto. Para ello se resuelve un problema de optimización, considerando como inputs las propensiones calculadas en el capítulo anterior. Además, se propone un modelo de asignación más completo a la compañía, que considera más variables y que por lo tanto podría tener un mayor impacto.

### 8.5.1.1 Aplicación al caso real.

Con el fin de tomar la decisión final de que producto ofrecer a cada cliente, se desarrolló un modelo de optimización. Para la elaboración del modelo se conversó con el canal con el fin de levantar las restricciones del canal, además de consultar cual era el objetivo que se buscaba al realizar una asignación, de esta manera se determinó la función objetivo y las restricciones. Se toma como problema base el modelo de la multi mochila mencionado en el marco conceptual, en donde las

capacidades de las mochilas son la capacidad del canal de ofrecer cada producto. Los valores utilizados para modelar la optimización son los siguientes:

$p_{ij}$ : son las propensiones de compra dado que el cliente contesta. Estos valores fueron calculados según el modelo Logit con interacciones desarrollado anteriormente.

$CAPC_j$  = Con respecto a la capacidad, el Canal decide la capacidad de cada producto según el número de ejecutivos con los que dispone. Es importante señalar que el canal no considera este valor como una capacidad máxima, si no que exige que se cumpla en N de registros señalado de manera exacta, es por esto, que el valor puede ir cambiando mes a mes.

El canal solo ofrece un producto a cada cliente, esto debido a que los productos son muy similares. Además, todos los clientes deben ser asignados a algún producto y llamados.

Es importante recalcar que la probabilidad de compra fue calculada suponiendo que el cliente siempre contesta, por lo que es una probabilidad condicional. De esta manera, la probabilidad de compra se puede observar en la formula 12.

$$\begin{aligned}
 PC &= \text{Probabilida de compra} \\
 &= P(\text{Compra}|\text{Contesta}) * \text{Probabilidad Contesta} \\
 &\quad + P(\text{Compra}|\text{No contesta}) * \text{Probabilidad No Contesta}
 \end{aligned}
 \tag{12}$$

Donde  $P(\text{Compra}|\text{No contesta})$  es 0. De esta manera la probabilidad de compra se puede calcular como se ve en la formula 13.

$$\begin{aligned}
 PC &= \text{Probabilida de compra} \\
 &= P(\text{Compra}|\text{Contesta}) * \text{Probabilidad Contesta}
 \end{aligned}
 \tag{13}$$

El canal busca maximizar el número de ventas. La variable decisión del problema será  $X_{ij}$ , en donde tomará el valor de 1 si al cliente i se le asigna el producto j. Con esto la formulación del problema será: dado lo anterior, el problema puede ser escrito tal y como se ve en la formula 14.

Función Objetivo:

$$Max: \sum_{i=0}^C \sum_{j=0}^P p_{ij} * X_{ij} * PC \quad (14)$$

Sujeto a:

$$\sum_{i=0}^C X_{ij} = CAPC_j$$
$$\sum_{j=0}^P X_{ij} = 1 \quad \forall i \in Clientes$$

- La función objetivo busca maximizar las ventas esperadas del canal, según la probabilidad de compra del cliente al producto que se ofreció, ajustado por la probabilidad que conteste cuando sea llamado.
- La primera restricción es la restricción de capacidad que entrega Call-center según su número de ejecutivos para cada producto
- Por último, la segunda restricción limita el número de productos que se le puede ofrecer a cada cliente, en donde por decisión del canal, se debe ofrecer a todos los clientes un producto.

### 8.5.1.2 Simulación de caso.

Se realizó una simulación del modelo de optimización con el fin de evaluar el potencial incremento que podría generar. Para ello se utilizaron los datos de lo sucedido en junio del 2018, y se trabajaron 4 productos: VA5, AP, APA y VCB, se calcularon las propensiones de compra para cada producto para luego realizar la asignación mediante el modelo de optimización y comparar las ventas esperadas obtenidas, con las ventas esperadas según la asignación realizada por los supervisores del canal.

Para el mes de junio Call-center determino que su capacidad para trabajar cada producto, el cual se detalla en la Tabla 19



Tabla 19: Capacidad máxima por producto

Productos	CAPC <sub>j</sub>
AP	7.750
APA	19.310
VA5	950
VCB	5.480

El modelo propuesto fue ejecutado en Python. En Anexos E, se tiene un ejemplo de cómo muestra el output de la variable decisión. De la Tabla 20, se puede observar como el modelo propuesto, el número de ventas esperadas no aumenta en todos los productos, pero en valor total se ve un aumento de 290 ventas esperadas en el modelo de asignación óptima, lo que significa un aumento de un 8,7% en las ventas aumenta las ventas esperadas.

Tabla 20: Resultados de la simulación del modelo aplicado al caso real

Productos	Registros	Ventas Esperadas según Asignación Canal	Ventas Esperadas según Asignación Propuesta	% Var Esperada
AP	7.753	758	647	- 14,6 %
APA	19.309	1.984	2.231	+ 12,4 %
VA5	947	76	121	+ 59,2%
VCB	5.480	483	591	+ 22,3 %
	<u>TOTAL</u>	<u>3.301</u>	<u>3.591</u>	<u>+ 8,7%</u>

### 8.5.1.3 Robustez de la metodología.

Debido a que el modelo de optimización toma como input los resultados de los modelos de propensión, se evalúa que tanto difieren los resultados de la optimización si se cambian los parámetros obtenidos por los modelos de propensión. Para ello se simuló el caso anterior utilizando el modelo escogido y otro modelo con peor desempeño y se realizó una “matriz de confusión” para contrarrestar la asignación. De esta manera se observó que el 70% de los registros fueron asignados al mismo producto, pese a utilizar un modelo con peor desempeño. Se puede concluir que la metodología es robusta, debido a que solo

un 30% de los registros es asignado a otro producto, y considerando los resultados vistos en la Tabla 18, aquellos cambios no deberían tener grandes efectos en las ventas esperadas del canal.

Tabla 21: Análisis de robustez del modelo de asignación

		Modelo Seleccionado en 8.4.5			
		Productos	AP	APA	VA5
Modelo con peor desempeño	AP	3.872	2.482	438	961
	APA	1.810	15.905	349	1.245
	VA5	524	220	139	64
	VCB	1.547	702	21	3.210

### 8.5.2 Modelo de asignación propuesto

Si bien el modelo de optimización anterior logra aumentar las ventas esperadas del canal, se propone generar un modelo más completo que busque maximizar los ingresos esperados del canal. Para ello se relajarán las restricciones del modelo anterior, de tal manera de que el modelo escoja cuántos clientes asignar a cada producto y cuántos clientes es óptimo llamar. De esta manera las variables para el nuevo problema de optimización son:

$p_{ij}$  = Probabilidad de compra del cliente  $i$  al producto  $j$ , dado que el cliente contesta. Este valor es estimado en base a los modelos propensión de la etapa anterior.

$PC$  = Probabilidad de que el cliente conteste. Se asumirá igual para todos y el valor será la tasa de conecta del canal. (70%)

$I_j$  = Ingreso obtenido si el cliente  $i$  compra el producto  $j$ .

$C_j$  = Costo variable asociado a ofrecer el producto  $j$  al cliente  $i$ .

$CAPC_j$  = Capacidad del canal para ofrecer el producto  $j$ . Al igual que el modelo anterior, la capacidad es determinada por el canal.

$MIN_j$  = Número mínimo de clientes a los que se le debe ofrecer el producto. Este valor es relevante debido a que la compañía se compromete con los proveedores ofrecer sus productos, por lo que la corredora debe asegurar que todos los productos sean ofrecidos.

De esta manera el problema puede ser escrito como se muestra en la formula 15.

Función Objetivo:

$$Max: \sum_{i=0}^C \sum_{j=0}^P p_{ij} * X_{ij} * I_{ij} * PC - X_{ij} * C_{ij} \quad (15)$$

Sujeto a:

$$\sum_i^C X_{ij} \leq CAPC_j \quad \forall j \in \text{Productos}$$

$$\sum_i^C X_{ij} \geq MIN_j \quad \forall j \in \text{Productos}$$

$$\sum_j^P X_{ij} \leq 1 \quad \forall i \in \text{Clientes}$$

- La función objetivo busca maximizar la ganancia generada por el canal, eligiendo que producto ofrecer a los clientes tomando en consideración la propensión de compra.
- La primera restricción limita la cantidad de clientes que se le puede ofrecer cada producto según la capacidad que se tienen.
- La segunda restricción permite cumplir la cuota mínima acordada con los proveedores, de tal manera que no haya productos a los que no se le asigne ningún cliente.
- Por último, la tercera restricción limita el número de productos que se le puede ofrecer a cada cliente.

### 8.5.2.1 Simulación del modelo propuesto

Al igual que el modelo aplicado al caso real, se realiza una simulación del modelo propuesto. Se determinaron los ingresos como el precio de venta de los productos, con respecto a los costos variables, según conversaciones con el canal, se asumirán estos costos como el 5% de los ingresos recibidos por una venta. Por otro lado, el número mínimo de clientes a los que se les debe ofrecer los productos serán el 30% de la capacidad máxima. La Tabla 22 resume todos los valores que se utilizaran en la simulación.

Tabla 22: Resumen valores simulación caso de optimización

<b>Producto j</b>	<b><math>CAPC_j</math></b>	<b><math>MIN_j</math></b>	<b><math>I_j</math></b>	<b><math>C_j</math></b>
AP	7.750	2.325	\$ 10.125	\$ 500
APA	19.310	5.792	\$ 11.700	\$ 585
VA5	950	284	\$ 6.750	\$ 340
VCB	5.480	1.644	\$ 12.375	\$ 620

La Tabla 23 resume los resultados de la simulación. Se observa que la asignación del canal es la que tiene peor desempeño tanto en ventas esperadas como en ingresos totales. Por otro lado, se tiene que en la asignación óptima base es la que tiene mayor cantidad de ventas esperadas, pero 2 MM de pesos menos que la asignación óptima del caso simulado, esto es debido a que se llama a todos los clientes para conseguir esa cantidad de ventas esperadas, en cambio en el modelo óptimo final, se llama solamente al 83% de los clientes. De esta manera se concluye que hoy el canal al no tener en cuenta todos los valores asociados al trabajo que realiza, toma decisiones de manera sesgada, por ejemplo, llamar a todos los clientes. Dado los alcances del trabajo, y el problema de optimización resuelto, se obtiene que no es óptimo llamar a un 17% de la base y que por lo tanto se tiene capacidad ociosa en el canal, este resultado por si solo no es concluyente, puesto que el canal trabaja con otras bases y fuentes, obteniéndose por lo tanto un óptimo local al resolver el problema de optimización. Para realizar una recomendación mas completa, se debería realizar un problema de optimización que considere todas las fuentes y bases con las que trabaja en canal.

Tabla 23: Resultados simulación del modelo de asignación propuesto

<b>Método de asignación</b>	<b>Registros llamados</b>	<b>Ventas Esperadas</b>	<b>Ingresos Totales Esperados</b>	<b>Aporte Esperado</b>
Asignación Call	33.489	3.379	\$ 19 MM	-
Modelo aplicado al caso real	33.489	3.591	\$ 21 MM	+11%
Modelo de asignación propuesto	28.062	3.279	\$ 23 MM	+21%

## 8.6 Diseño de Productos

Dado el contexto en el que se desarrollada la memoria es importante para la compañía el estar constantemente ofreciendo productos nuevos a sus clientes. La formulación de los modelos de propensiones presentados en este trabajo, permite evaluar el potencial impacto de nuevas configuraciones de productos con el fin de recomendar la generación de nuevos productos.

### 8.6.1.1 Potenciales Productos.

Los potenciales productos a ofrecer en el canal son determinados por la mix de los atributos estudiados de los productos. De esta manera existen 4 potenciales productos nuevos que se podrían diseñar. Es importante destacar, que uno de estos, tiene la misma configuración que el producto VA4 con la que fue entrenado el modelo. Por lo que se evaluará si conviene volver a poner en el mercado dicho producto, o uno de los nuevos.

Tabla 24: Potenciales nuevos productos según mix de atributos

<b>Producto</b>	<b>Muerte Natural</b>	<b>Bonificación</b>	<b>Descuento en Farmacia</b>
VA4	0	0	0
Nuevo Producto 1	1	0	0
Nuevo Producto 2	1	0	1
Nuevo Producto 3	1	1	0

### 8.6.1.2 Simulación

Para evaluar el potencial impacto de cada configuración de productos, se simuló la situación en donde el canal Call-center desea que el 10% de las bases del mes de Junio sean asignadas a un producto nuevo. Para determinar el producto se calcularán las propensiones de todos los potenciales productos de la Tabla 24. Luego se simulará 4 veces el problema mencionado en la sección 8.5.1.2 cambiando las restricciones de capacidad por las que se muestra en Tabla 25.

Tabla 25: Capacidad del canal para el caso de simulación de diseño de productos

Productos	CAPC <sub>j</sub>
AP	6.975
APA	17.379
VA5	855
VCB	4.932
Nuevo Producto	3.348

### 8.6.1.3 Resultados:

La tabla resume los resultados de las simulaciones, donde se observa que en todos los casos las ventas esperadas del canal aumentarían en comparación a los resultados vistos en la simulación del caso real (Ver Tabla 20), por lo que al aumentar el número de productos que ofrece el canal, aumenta su diversificación de productos aumentando así sus ventas esperadas.

Respecto al producto con mayor venta esperada, resulto ser el Producto número 1, el cual ofrece el atributo muerte natural. Lo cual puede deberse a que es el producto más alejado los productos ofrecido hoy en día en el canal respecto a sus atributos, pues de los productos estudiados, tan solo uno ofrece cobertura de muerte natural. De esta manera, si el canal considerara ofrecer un nuevo producto, se recomienda que tenga tan solo muerte natural dentro de sus coberturas.

Tabla 26: Resultados de la simulación del diseño de nuevos productos.

<b>Producto Incluido</b>	<b>Aumento de Ventas Esperadas</b>
Nuevo Producto 1	+ 1,1%
Nuevo Producto 2	+ 0,14%
Nuevo Producto 3	+ 0,66%
VA4	+ 1,0%

## 8.7 Diseño Experimental

Se busca evaluar el impacto de la asignación propuesta en las ventas perdurables del canal Call-center, de tal manera de determinar si existe un incremento estadísticamente significativo.

Se propone generar los siguientes grupos:

- Grupo de Control: Se realiza la asignación utilizando la metodología actual (opinión experta de los supervisores del canal).
- Grupo de Tratamiento: Se realiza la asignación utilizando la metodología propuesta.

La hipótesis a validar en este experimento es: “la metodología propuesta aumenta las tasas de ventas en comparación con la metodología actual”. Para validar esto, se utiliza el test de proporciones. Detallado en el marco conceptual (Formula 11) Donde:

- $x_1$  = Casos favorables en grupo de tratamiento.
- $n_1$  = Casos totales en el grupo de tratamiento.
- $x_2$  = Casos favorables en grupo de control.
- $n_2$  = Casos totales en el grupo de control.
- $p_1$  = Tasa de venta perdurable del grupo de tratamiento.
- $p_2$  = Tasa de venta perdurable del grupo de control.

De esta manera dado el estadístico z, se obtendrá si existe diferencia significativa entre los grupos.

Para la generación de los grupos es importante utilizar una muestra significativa, utilizando la formula mostrada en el marco conceptual (Ecuación 10) y los siguientes parámetros

- $X^2$  es igual a 1,96 para obtener un nivel de confianza del 95%.
- $P$  es igual a 0,5.
- $d$  es igual a 5%.



Se obtiene que el tamaño de muestra necesaria es de 384 clientes. Es importante mencionar que el experimento se realiza en un Call-center, por lo que existe una probabilidad de que el cliente no conteste. Dado esto se recomienda que la muestra considere la tasa histórica de contactabilidad del 70%. De esta manera, para obtener 384 clientes que contesten se deben llamar aproximadamente 550. Para que los resultados no estén sesgados, ambos grupos deben ser conformados de la siguiente manera:

- Ambos grupos se les debe ofrecer todos los productos disponibles.
- Ambos grupos deben tener una similar distribución de probabilidades de propensión de compra de los distintos productos.

La experimentación debe tener una duración de un mes, debido a que el canal tiene un mes para recorrer las bases entregadas por el área de inteligencia de negocio. Es importante señalar que los ejecutivos no deben tener noción de la experimentación, para evitar que fracase la experimentación. Además, se debe tener en consideración que se tienen ejecutivos expertos por productos, por lo que un ejecutivo solo puede vender un producto.

La experimentación no fue realizada debido a la situación actual en el que se encuentra el canal, por lo que no se pueden asegurar la conformación de un grupo de control del tamaño necesario para obtener resultados estadísticamente significativos.

## 9 Conclusiones y trabajos futuros

El objetivo principal de este trabajo es desarrollar una metodología que aumenta las ventas perdurables del canal Call-center de una Corredora de seguros, mediante una asignación de productos óptima. Se trabaja con los clientes catalogados aperturas, debido a que son los más importantes para el canal representado el 70% de sus ventas.

Para lograr el objetivo principal, se propone desarrollar una asignación mediante la resolución de un problema de optimización, que busque maximizar las ventas perdurables del canal, y utilizando como input las propensiones de los clientes a comprar.

El trabajo se abordó en base a la metodología CRISP-DM, el cual pone énfasis en el proceso de entendimiento del negocio, lo que permite una mejor comprensión de los procesos actuales de la empresa. Además, es relevante que el proceso sea iterativo, debido a que al avanzar en la metodología pueden encontrarse nuevas oportunidades que pueden ser interesantes incluir.

Se trabajó en conjunto con el área de productos de la compañía, con el fin de levantar atributos relevantes de los productos ofrecidos al segmento de clientes estudiados. Se determinó que las coberturas de los seguros de vida eran atributos interesantes a incorporar en el estudio. De esta manera los atributos de los productos son: Bonificación, Muerte Natural, Descuento en Farmacia e ITP por accidentes. Todas estas son consideradas como variables binarias.

En la etapa del análisis exploratorio de los datos, se observa, la gran mayoría de las variables con la que dispone la empresa, no pueden ser utilizadas en el segmento de clientes que se estaba estudiando. Dado esto la información disponible de los clientes catalogados como aperturas era bastante limitado.

El primer objetivo específico del trabajo corresponde a la estimación de la probabilidad de compra por productos de los clientes. Se evaluaron las metodologías C.A.R.T y C.H.A.I.D de árboles de decisión, en donde se obtuvo mejores resultados en el árbol tipo C.H.A.I.D. Bajo el criterio de la curva de ganancia acumulada, se logró capturar un 19% de las ventas en el primer decil según los datos de testeo. Es importante señalar que ninguno de los dos árboles consideró relevante las variables de los atributos de los productos. Posterior a esto, se realizó un modelo Logit genérico para todos los productos y un modelo Logit que incorpora los atributos de los productos. El modelo Logit genérico, logró capturar el 21% de

las ventas en el primer decil, mientras que el modelo que incluye los atributos de los productos, sólo captura un 20%. De esta manera se observa un trade off entre la flexibilidad del modelo y el ajuste de este. Dado el contexto en el cual se encuentra la empresa, en donde los productos ofrecidos van variando en el tiempo, se escoge el modelo que incorpora los atributos de los productos.

El modelo escogido fue validado con datos fuera de la muestra, en donde se observó que tenía un desempeño similar a lo estudiado en el grupo de testeo, demostrando la robustez de este. Además, se validó cómo funcionaba el modelo prediciendo las compras de un producto nuevo que se empezó a vender en el canal. Los resultados fueron similares a los resultados de los productos con los que fue entrenado, obteniendo un 19% de las ventas capturadas en el primer decil. De esta manera se corrobora que el modelo propuesto tiene la flexibilidad necesaria para evaluar las propensiones de productos próximos a ofrecer por la compañía.

El segundo objetivo específico es el desarrollar una metodología de asignación de productos a ofrecer a los clientes, se propuso resolver un modelo de optimización el cual fue construido en conjunto con el canal Call-center y que tomaba como input las propensiones calculadas en la etapa de minería de datos. Se realizó una simulación en donde se observó que el canal podría aumentar en un 8% las ventas esperadas mediante la asignación propuesta. Pese a que se obtuvieron buenos resultados, el modelo construido en conjunto con Call-center es muy restringido y no necesariamente se están considerando todas las variables relevantes de la asignación. Dado lo anterior, se propuso otro modelo de optimización que maximiza la utilidad del canal. Se simuló este caso y se observó que la nueva asignación óptima decide llamar tan sólo al 80% de la base y obtiene una utilidad esperada de 4 Millones. Este resultado hace pensar que el hecho que el canal quiera llamar a todos los clientes de la base no es lo óptimo, y por lo tanto deberían disminuir su capacidad máxima para así no tener que llamar a todos los clientes y evitar los tiempos ociosos.

Por último, se generó una simulación del modelo de optimización propuesto, de tal manera de determinar que producto adicional podría ofrecer el canal. Se concluyó que el mejor producto a ofrecer, en conjunto a los ya ofrecidos, debe tener como cobertura la muerte natural, esto debido a que aumenta las ventas perdurables esperadas del canal en 1%.

Como trabajos futuros se propone:

- Utilizar la metodología en otro segmento de clientes, de tal manera de tener mayor información en los clientes y aumentar el ajuste de los modelos.

Además, se tendrá mayor variabilidad de los productos ofrecidos, por lo que pese a que sean segmento que hoy en día no son tan relevantes para el canal, se tiene un espacio de mejora mayor.

- Se propone realizar un estudio sobre otras variables interesantes para incorporar en el modelo de optimización. Por ejemplo, estimar la probabilidad que responda el llamado el cliente mediante un modelo o una segmentación. También se propone utilizar el valor esperado del cliente o la permanencia esperada como variables en la asignación.
- Se propone, realizar un problema de optimización del canal Call-center completo, considerando todas las fuentes con las que este trabaja y la dotación del canal, de tal manera de encontrar un óptimo global al realizar la asignación.
- Por último, se recomienda utilizar la metodología para campañas multi canales, de tal manera de determinar no solo el mejor producto a ofrecer a los clientes, sino que también por cual canal hacer la campaña.

## 10 Bibliografía

- [1] Comisión para el mercado Finciero, «Comisión para el mercado financiero,» [En línea]: <http://www.svs.cl/educa/600/w3-propertyvalue-1160.html>. [Último acceso: 28 Mayo 2018].
- [2] Comisión para el mercado Financiero, «Informe Corredoras de Seguros Grandes Tiendas,» 2017.
- [3] M. K. J. P. Jiawei Han, Data Mining Concepts and Techniques, Third Edition, 2012.
- [4] M. A. ALVARADO NARANJO, «Generación de un modelo de propensión de compra en la industria de telecomunicaciones,» Memoria para optar al título de ingeniero civil industria, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial , Santiago, Universidad de Chile, 2012.
- [5] P. A. PEREZ VILLANUEVA, «Modelo de predicción de fuga de clientes de la telefonía móvil post pago,» Memoria para optar al título de ingeniero civil industria, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial , Santiago, Universidad de Chile, 2016.
- [6] B. A. CALDERON TESTA, «Determinación de la propensión al aumento de consumo con tarjeta de credito de clientes de una institución financiera,» Memoria para optar al título de ingeniero civil industria, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial , Santiago, Universidad de Chile, 2016.
- [7] F. I. GARCIA BACCHIEGA, «Diseño de un modelo predictivo para el aumento de pólizas principales de una compañía de seguros,» Memoria para optar al título de ingeniero civil industria, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial , 2017.
- [8] P. I. ROMERO VÁSQUEZ, «Estimación del riesgo asociado a potenciales clientes jóvenes en el negocio de seguros de automóviles,» Memoria para optar al título de ingeniero civil industria, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial , Santiago de Chile, 2015.
- [9] P. T. Silvano Martello, KNAPSACK PROBLEMS Algorithms and Computer Implementations, 1990.

- [10] P. U. P. D. Kellerer H., «The Multiple-Choice Knapsack Problem,» de Knapsack Problems, Berlin, Heidelberg, 2004, pp. 317-347.
- [11] W. M. DANIEL, «Modelos de propensión integrados para la optimización de campañas de marketing,» Tesis para optar al grado de magister en gestión de operaciones, Santiago, Universidad de Chile, 2013.
- [12] M.-D. COHEN, «Exploiting response model-optimizing cross-sell and up-sell opportunities in banking,» Information System, nº 29, pp. 327-341, 2004.
- [13] R. L. A. F. T. N. S DELANOTE, «Optimization of the annual planning of targeted offers in direct marketing,» Journal of the Operational Research Society, nº 64, pp. 1770-1779, 2013.
- [14] M. A. ARAYA ABADIE, «Estimación de la efectividad promocional en un supermercado mayorista,» Memoria para optar al título de ingeniero civil industria, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial , Santiago, Universidad de Chile, 2011.
- [15] C. J. K. R. K. T. S. C. W. R. CHAPMAN P, CRISP-DM 1.0 step-by-step data mining guide, The modeling agency.
- [16] S. F. SVEN F. CRONE, «Instance sampling in credit scoring: An empirical study of sample size and balancing,» International Journal of Forecasting, nº 28, pp. 224-238, 2012.
- [17] R. J. FRIEDRICH, «In Defense of Multiplicative Terms in Multiple Regression Equations,» American Journal of Political Science, vol. 26, nº 4, pp. 797-833, 1982.

# 11 Anexos

## 11.1 Anexos A

Tabla 27: Fuentes de información relevantes

Casar sf	TMP_BI_CMI_CC_FINAL	TMK_BI_RFM
RUT	TIPO_BASE	RUT
EDAD	PRODUCTO	PERIODO
GÉNERO	RUT	L
ESTADO_CIVIL	FECHA	R
N_HIJOS	HORA	FREQUENCY
MONTO_RENTA	SUPERVISOR	MONETARY
N_VEHICULOS	BASE	ACTUAL
COMUNA	ESTADO	
REGION	RES_CONTACTO	
DIR_GSE	RES_TIPO	
CLI_C_12M	VENTA	
CLI_F_12M	EJECUTIVO	
CLI_T_12M	RUT_AGENTE	
CLI_VIA_12M		
CLI_SEG_12M		
RECENCY_C		
GASTO_C_12M		
RECENCY_F		
GASTO_F_12M		
VISITA_F_12M		
GASTO_FC_12M		
RECENCY_S		
GASTO_S_12M		
VISITA_S_12M		
TIPO_TARJETA		
ESTADO_CUENTA		
CUPO		
PUNTOS_PROX_VENCER		
SEGMENTO_VALOR_C		
SEGMENTO_USO_C		
SEGMENTO_VALOR_F		
SEGMENTO_VALOR_S		
SEGMENTO_VALOR_T		
SEGMENTO_VALOR_VIA		
GASTO_EXTRANJERO_12M		
TRX_EXTRANJERO_12M		
RECENCY_EXTRANJERO		

GASTO_RUBRO_AUTOMOTRIZ_1 2M
PRINC_AUTOMOTRIZ
PRINC_COMBUSTIBLE
PRINC_COMUNIC_OT
PRINC_COMUNIC_OU
PRINC_COMUNICACIONES
PRINC_EDUCACION
PRINC_ENTRETENCION
PRINC_FARMACIAS
PRINC_MEJ_HOGAR_OT
PRINC_MEJ_HOGAR_OU
PRINC_MEJ_HOGAR
PRINC_RECAUDACION
PRINC_RESTAURANT
PRINC_SALUD
PRINC_SEGUROS_OT
PRINC_SEGUROS_OU
PRINC_SEGUROS
PRINC_SUPERMERCADOS_OT
PRINC_SUPERMERCADOS_OU
PRINC_SUPERMERCADOS
PRINC_TIENDAS_DPTO_OT
PRINC_TIENDAS_DPTO_OU
PRINC_TIENDAS_DPTO
PRINC_TRANSPORTE
PRINC_VIAJES_OT
PRINC_VIAJES_OU
PRINC_VIAJES
PRINC_VIVIENDA
PRINC_C

Tabla 28: Variables que aplican al problema

<b>Variable</b>	<b>Aplica al problema</b>	<b>Razón</b>
RUT	si	-
EDAD	si	-
GÉNERO	si	-
ESTADO_CIVIL	si	-
N_HIJOS	si	-
MONTO_RENTA	si	-
N_VEHICULOS	si	-
COMUNA	si	-
REGION	si	-



<b>Variable</b>	<b>Aplica al problema</b>	<b>Razón</b>
DIR_GSE	si	-
CLI_C_12M	no	Clientes nuevos sin información
CLI_F_12M	no	Clientes nuevos sin información
CLI_S_12M	no	Clientes nuevos sin información
CLI_T_12M	no	Clientes nuevos sin información
CLI_VIA_12M	no	Clientes nuevos sin información
CLI_SEG_12M	no	Clientes nuevos sin información
RECENCY_C	no	Clientes nuevos sin información
GASTO_C_12M	no	Clientes nuevos sin información
RECENCY_F	no	Clientes nuevos sin información
GASTO_F_12M	no	Clientes nuevos sin información
VISITA_F_12M	no	Clientes nuevos sin información
GASTO_FC_12M	no	Clientes nuevos sin información
RECENCY_S	no	Clientes nuevos sin información
GASTO_S_12M	no	Clientes nuevos sin información
VISITA_S_12M	no	Clientes nuevos sin información
TIPO_TARJETA	no	Clientes nuevos comienzan con tarjeta normal
ESTADO_CUENTA	no	Clientes nuevos sin información
CUPO	si	-
PUNTOS_PROX_VENCER	no	Clientes nuevos sin información
SEGMENTO_VALOR_C	no	Clientes nuevos sin información
SEGMENTO_USO_C	no	Clientes nuevos sin información
SEGMENTO_VALOR_F	no	Clientes nuevos sin información

<b>Variable</b>	<b>Aplica al problema</b>	<b>Razón</b>
SEGMENTO_VALOR_S	no	Clientes nuevos sin información
SEGMENTO_VALOR_T	no	Clientes nuevos sin información
SEGMENTO_VALOR_VIA	no	Clientes nuevos sin información
GASTO_EXTRANJERO_12M	no	Clientes nuevos sin información
TRX_EXTRANJERO_12M	no	Clientes nuevos sin información
REGENCY_EXTRANJERO	no	Clientes nuevos sin información
GASTO_RUBRO_AUTOMOTRIZ_12M	no	Clientes nuevos sin información
PRINC_AUTOMOTRIZ	no	Clientes nuevos sin información
PRINC_COMBUSTIBLE	no	Clientes nuevos sin información
PRINC_COMUNIC_OT	no	Clientes nuevos sin información
PRINC_COMUNIC_OU	no	Clientes nuevos sin información
PRINC_COMUNICACIONES	no	Clientes nuevos sin información
PRINC_EDUCACION	no	Clientes nuevos sin información
PRINC_ENTRETENCION	no	Clientes nuevos sin información
PRINC_FARMACIAS	no	Clientes nuevos sin información
PRINC_MEJ_HOGAR_OT	no	Clientes nuevos sin información
PRINC_MEJ_HOGAR_OU	no	Clientes nuevos sin información
PRINC_MEJ_HOGAR	no	Clientes nuevos sin información
PRINC_RECAUDACION	no	Clientes nuevos sin información
PRINC_RESTAURANT	no	Clientes nuevos sin información
PRINC_SALUD	no	Clientes nuevos sin información
PRINC_SEGUROS_OT	no	Clientes nuevos sin información

<b>Variable</b>	<b>Aplica al problema</b>	<b>Razón</b>
PRINC_SEGUROS_OU	no	Clientes nuevos sin información
PRINC_SEGUROS	no	Clientes nuevos sin información
PRINC_SUPERMERCADOS_OT	no	Clientes nuevos sin información
PRINC_SUPERMERCADOS_OU	no	Clientes nuevos sin información
PRINC_SUPERMERCADOS	no	Clientes nuevos sin información
PRINC_TIENDAS_DPTO_OT	no	Clientes nuevos sin información
PRINC_TIENDAS_DPTO_OU	no	Clientes nuevos sin información
PRINC_TIENDAS_DPTO	no	Clientes nuevos sin información
PRINC_TRANSPORTE	no	Clientes nuevos sin información
PRINC_VIAJES_OT	no	Clientes nuevos sin información
PRINC_VIAJES_OU	no	Clientes nuevos sin información
PRINC_VIAJES	no	Clientes nuevos sin información
PRINC_VIVIENDA	no	Clientes nuevos sin información
PRINC_C	no	Clientes nuevos sin información

## 11.2 Anexos B

Tabla 29: Matriz de correlación de las variables

	Edad	Left	Right	Frecuency	N Hijos	N Vehículos	CUPO	Monetary Actual	Bonificación	Descuento Farmacia	Muerte Natural	ITP Accidentes
Edad	1	0,12	0,05	-0,1	0,56	0,19	0,11	0,04	0,1	0,11	-0,04	-0,1
Left	0,12	1	0,38	-0,13	0,08	0,14	0,05	0,09	0,03	0,03	-0,01	-0,03
Right	0,05	0,38	1	-0,45	0,01	0,04	0,05	0,01	0,01	0,02	0,01	-0,01
Frecuency	-0,1	-0,13	-0,45	1	0	-0,04	-0,07	0,11	-0,01	-0,03	0,01	0,01
N Hijos	0,56	0,08	0,01	0	1	0,15	0,04	0,02	0,04	0,06	-0,03	-0,04
N Vehículos	0,19	0,14	0,04	-0,04	0,15	1	0,2	0,10	0,03	0,03	-0,06	-0,03
CUPO	0,11	0,05	0,05	-0,07	0,04	0,2	1	0,09	0,01	-0,02	0	-0,01
Monetary Actual	0,04	0,09	0,01	0,11	0,02	0,10	0,09	1	0,01	0,01	-0,01	0,02
Bonificación	0,1	0,03	0,01	-0,01	0,04	0,03	0,01	0,01	1	0,47	0,31	-1
Descuento Farmacia	0,11	0,03	0,02	-0,03	0,06	0,03	-0,02	0,01	0,47	1	0,14	-0,47
Muerte Natural	-0,04	-0,01	0,01	0,01	-0,03	-0,06	0	-0,01	0,31	0,14	1	-0,31
ITP Accidentes	-0,1	-0,03	-0,01	0,01	-0,04	-0,03	-0,01	0,02	-1	-0,47	-0,31	1

### 11.3 Anexos C

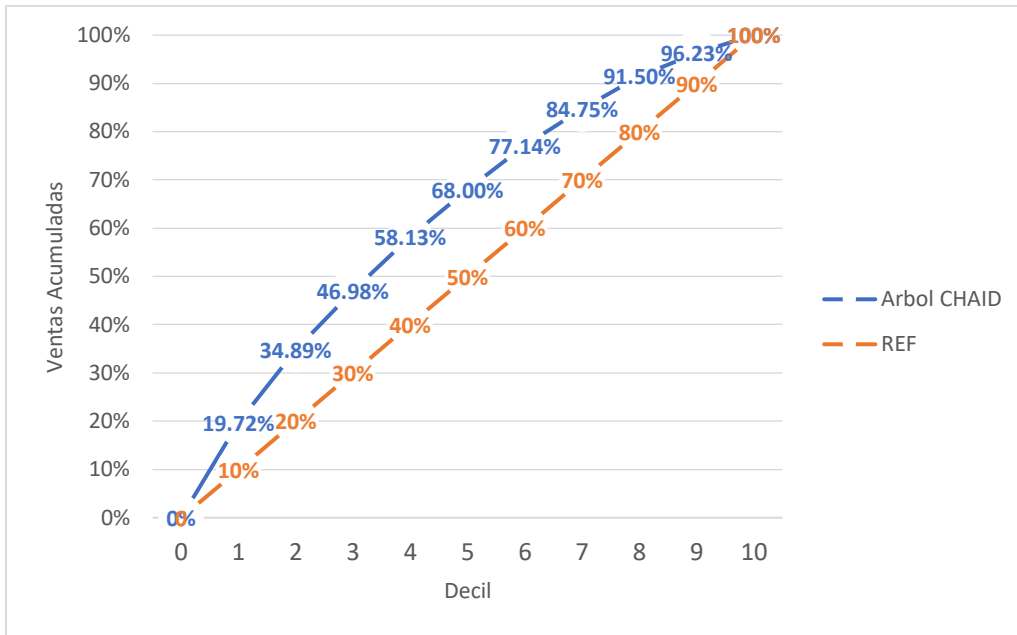


Ilustración 27: Curva de ganancia acumulada del Árbol tipo C.H.A.I.D

Tabla 30: Matriz confusión del árbol tipo C.H.A.I.D con corte 0,5

		Real	
		No Compra	Compra
Predicho	No Compra	28.279	19.792
	Compra	1.664	2.536

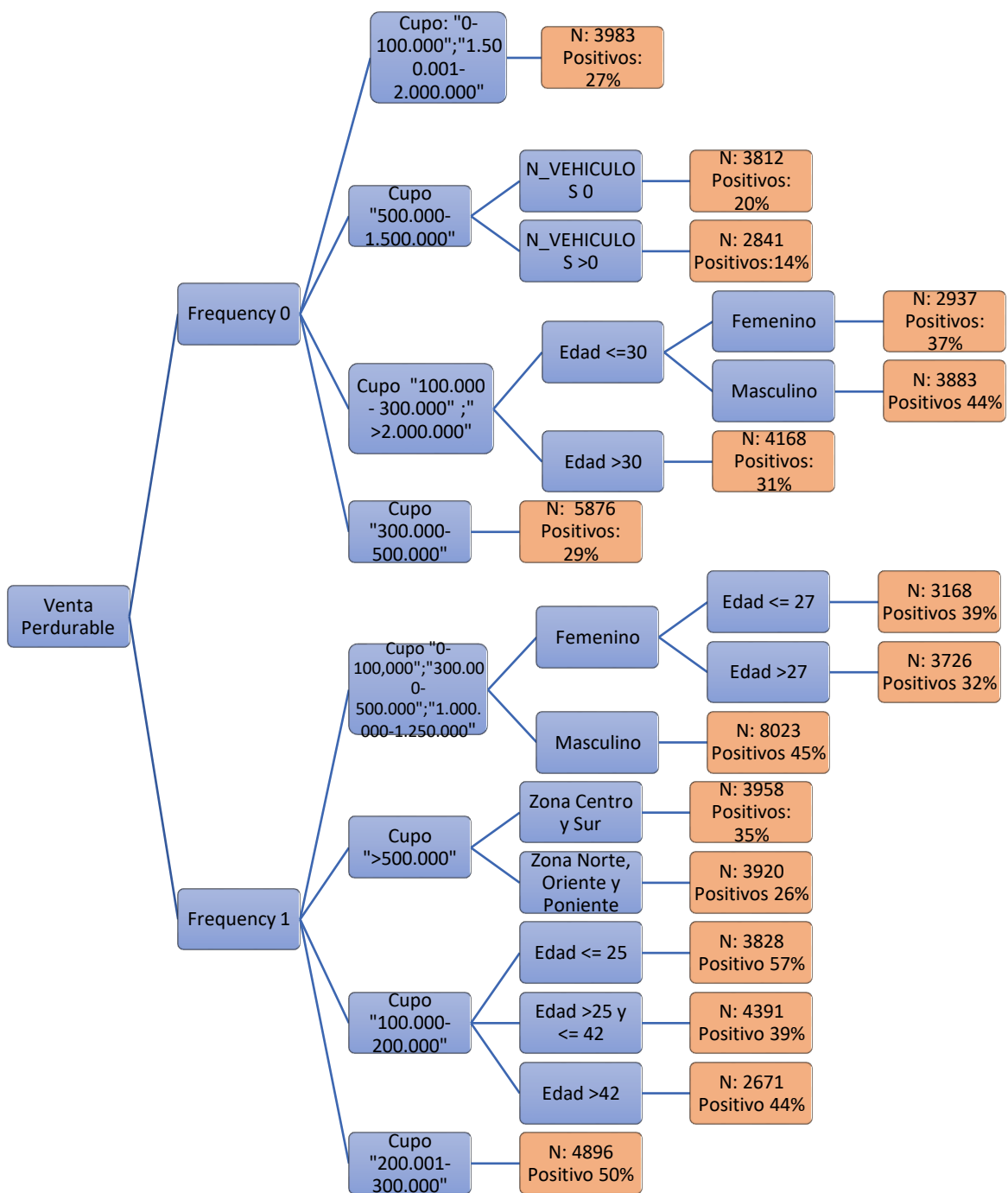


Ilustración 28: Diagrama del árbol de decisión tipo C.H.A.I.D nodo 1 y 2

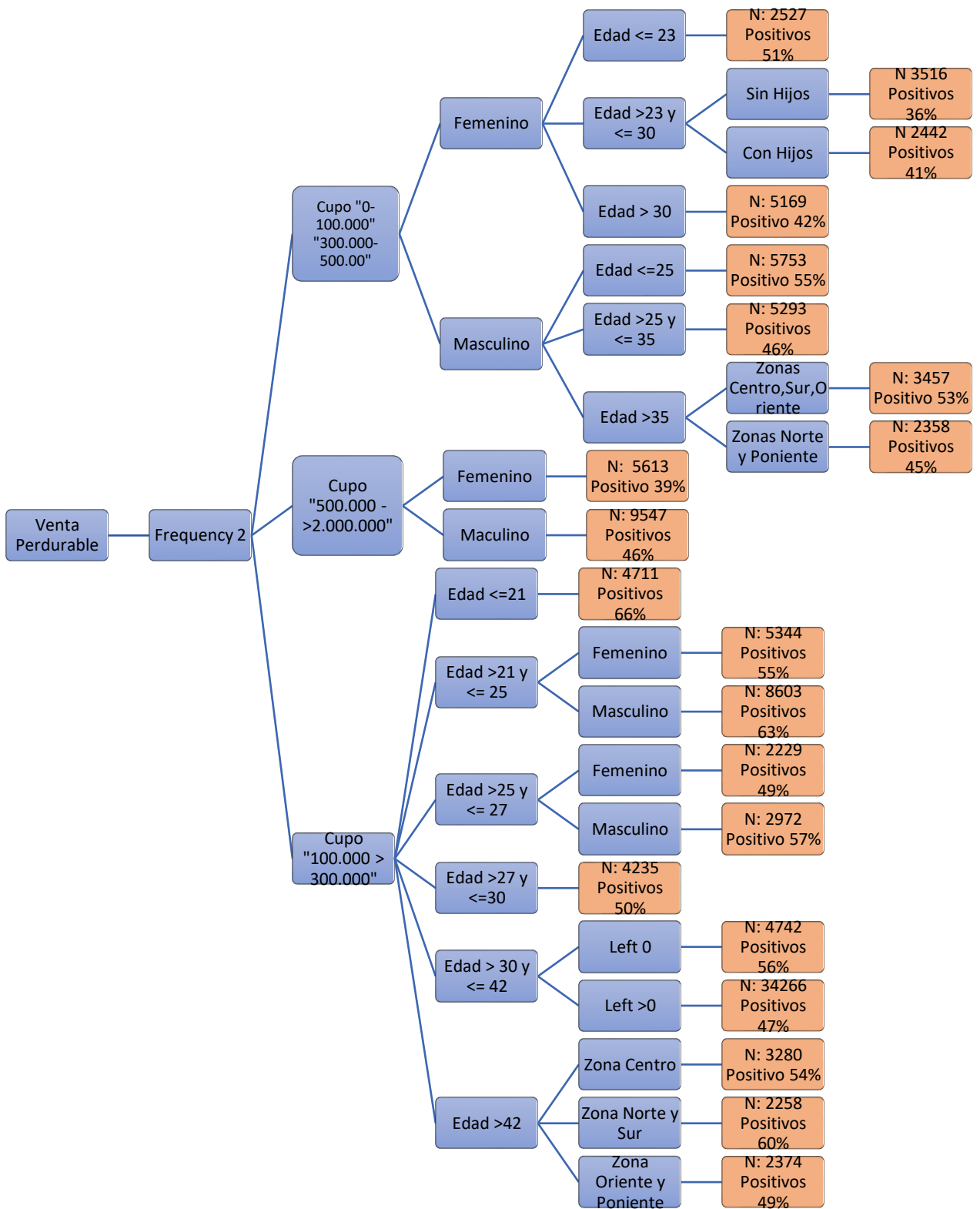


Ilustración 29: Diagrama del árbol de decisión tipo C.H.A.I.D nodo 3.

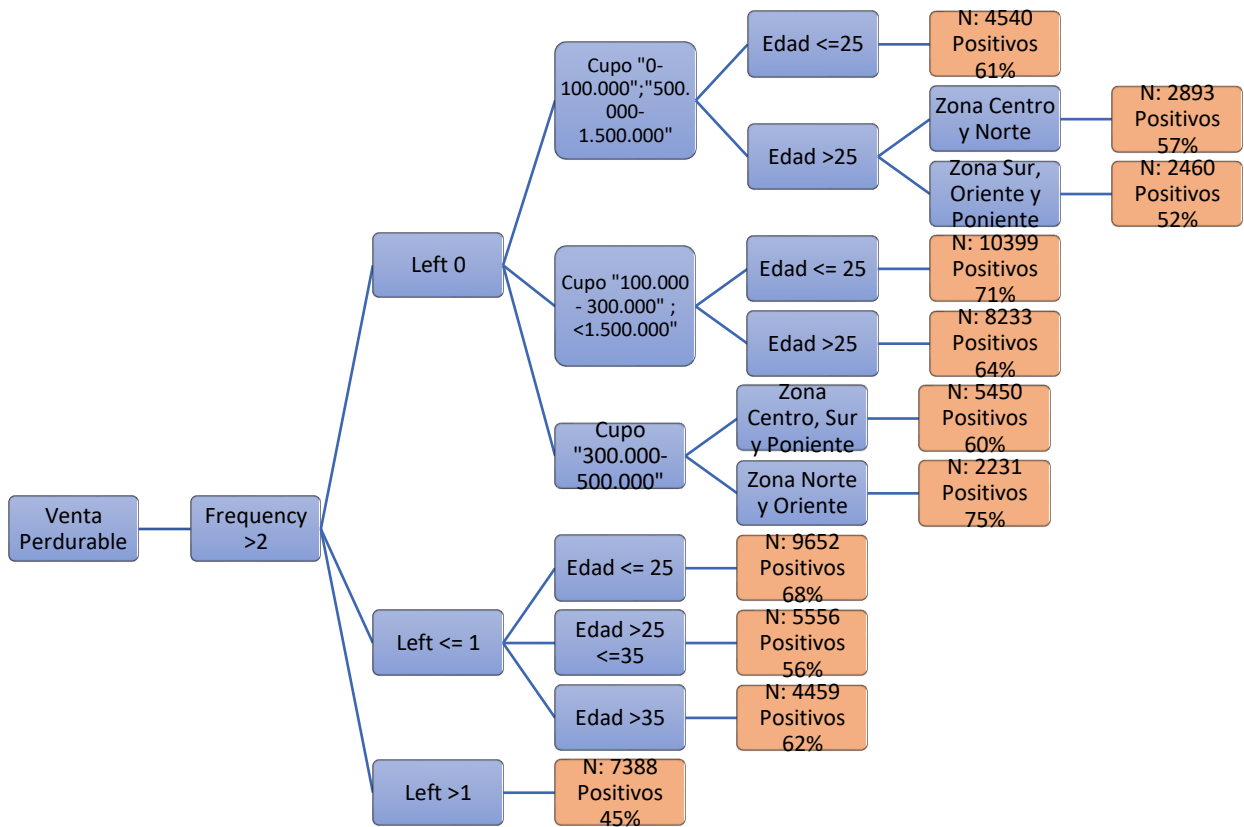


Ilustración 30: Diagrama del árbol de decisión tipo C.H.A.I.D nodo 4.



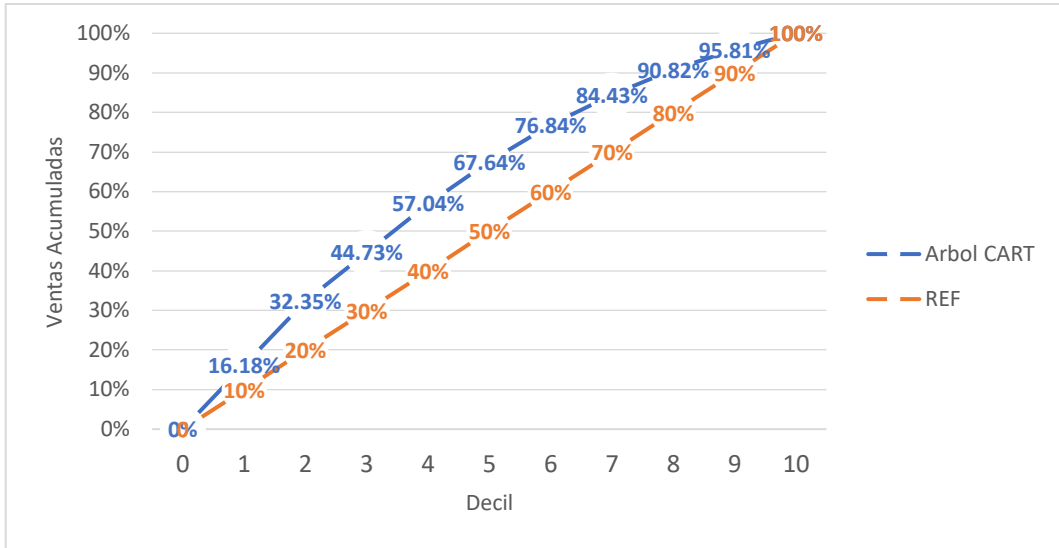


Ilustración 31: Curva de ganancia acumulada del árbol tipo C.A.R.T

Tabla 31: Matriz confusión del árbol tipo C.H.A.I.D con corte 0,5

		Real	
		No Compra	Compra
Predicho	No Compra	27.268	20.803
	Compra	1.553	2.647

## 11.4 Anexos D

Tabla 32: Coeficientes del modelo Logit con atributos de los productos

Variable	Categorías	Coeficiente	Error Estándar	P valor
Intercepto	-	- 0,99	0,31	0,00
Left	-	- 0,01	0,00	0,00
Frecuency	-	0,39	0,02	0,00
N de Hijos	-	- 0,04	0,02	0,03
N de Vehículos	-	- 0,40	0,08	0,00
Estado Civil	Soltero	0,65	0,29	0,03
	Casado	-	-	-
Zona	Norte	0,62	0,12	0,00
	Sur	0,22	0,13	0,10
	Centro			
	Oriente	0,48	0,30	0,10
	Poniente	0,07	0,09	0,46
Género	Masculino	0,48	0,30	0,10
	Femenino	-	-	-
Monto Renta	Bajo	- 0,31	0,03	0,00
	Alto	-	-	-
Actividad	Administrativo	-	-	-
	Empl. Tecnico	- 0,03	0,01	0,03
	Profesional Univ. (Empleado)	- 0,30	0,02	0,00
	Obreros y Auxiliares	-	-	-
	Trabajador Independiente	0,04	0,02	0,02
	Profesional Univ. (Independiente)	- 0,40	0,04	0,00
	Otros	0,18	0,01	0,00
Bonificación	1	0,15	0,05	0,01
	0	-	-	-
Descuento Farmaci	1	0,37	0,30	0,22
	0	-	-	-

<b>Variable</b>	<b>Categorías</b>	<b>Coefficiente</b>	<b>Error Estándar</b>	<b>P valor</b>
Muerte Natural	1	- 0,28	0,05	0,00
	0	-	-	-
Rango Edad	18-25	-	-	-
	26-35	- 0,35	0,01	0,00
	36-45	- 0,36	0,02	0,00
	46-55	- 0,29	0,02	0,00
	>56	- 0,12	0,02	0,00
Rango Cupo	<100.000	-	-	-
	100.001-200.000	0,31	0,01	0,00
	200.001-300.000	0,32	0,02	0,00
	300.001-500.000	0,05	0,02	0,00
	500.001-750.000	- 0,18	0,02	0,00
	750.001-1.000.000	- 0,22	0,02	0,00
	1.000.001-1.250.000	- 0,10	0,03	0,00
	1.250.001-1.500.000	- 0,35	0,03	0,00
	1.500.001-2.000.000	- 0,46	0,15	0,00
	>2.000.001	- 0,52	0,22	0,02
Interacción de variable	Masculino * Bonificación	- 0,17	0,04	0,00
	Monto Renta Bajo * Muerte Natural	0,26	0,04	0,00
	Left*Muerte Natural	0,004	0,00	0,00
	Frequency*Bonificación	- 0,04	0,02	0,08
	Frequency*Muerte Natural	0,04	0,01	0,00
	N de Hijos*Bonificación	0,08	0,02	0,00
	N de hijos * muerte Natural	- 0,02	0,01	0,06
	N de vehículos * descuento Farmacia	0,34	0,08	0,00
Zona Norte * Bonificación	0,08	0,07	0,28	

Variable	Categorías	Coeficiente	Error Estándar	P valor
	Zona Sur * Bonificación	0,01	0,06	0,82
	Zona Oriente * Bonificación	0,40	0,23	0,08
	Zona Poniente * Bonificación	-	0,05	0,11
	Zona Norte * Desc Farmacia	-	0,14	0,00
	Zona Sur * Desc Farmacia	0,04	0,15	0,79
	Zona Oriente * Desc Farmacia	0,99	0,37	0,01
	Zona Poniente * Desc Farmacia	-	0,10	0,96
	Zona Norte * Muerte Natural	0,09	0,03	0,00
	Zona Sur * Muerte Natural	0,07	0,03	0,01
	Zona Oriente * Muerte Natural	0,07	0,08	0,37
	Zona Poniente * Muerte Natural	0,01	0,02	0,72
	Estado civil Soltero * Desc Farmacia	0,64	0,30	0,03
	Estado civil Soltero * Muerte Natural	0,09	0,03	0,00

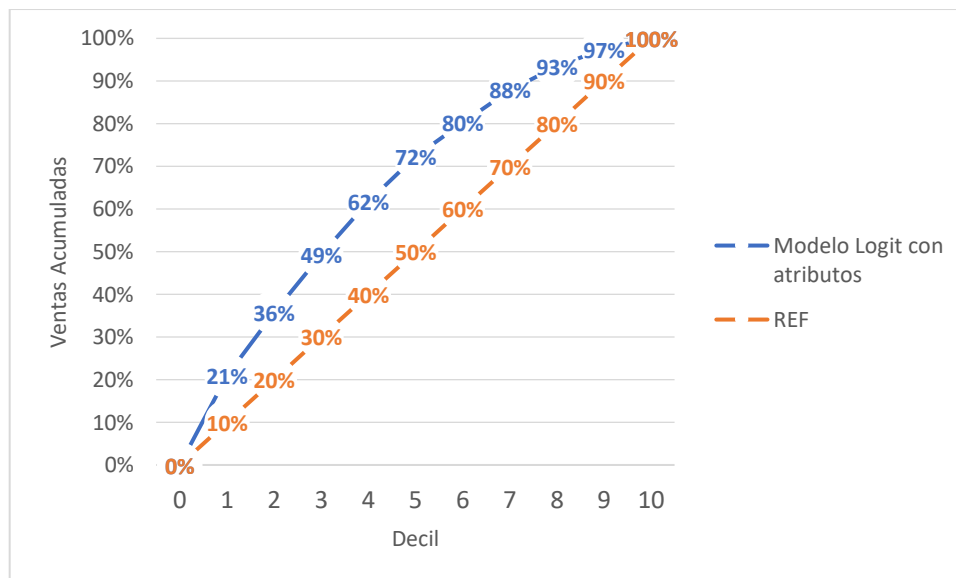


Ilustración 32 Curva de ganancia acumulada en grupo de validación para el producto AP

Tabla 33: Matriz confusión en grupo de validación para el producto AP

		Real	
		No Compra	Compra
Predicho	No Compra	27.268	20.803
	Compra	1.553	2.647

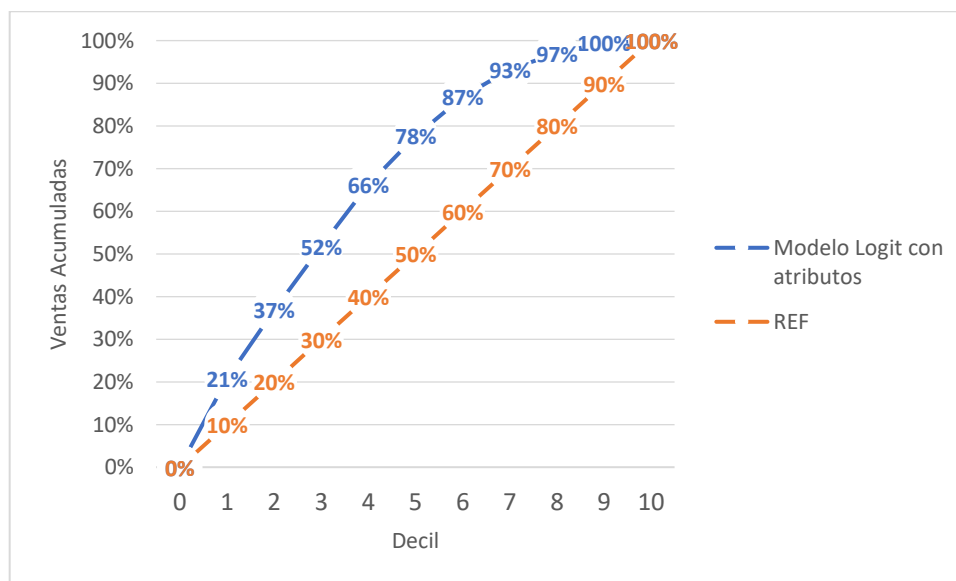


Ilustración 33: Curva de ganancia acumulada en grupo de validación para el producto VCB

Tabla 34: Matriz confusión en grupo de validación para el producto VCB

		Real	
		No Compra	Compra
Predicho	No Compra	3208	165
	Compra	1811	296

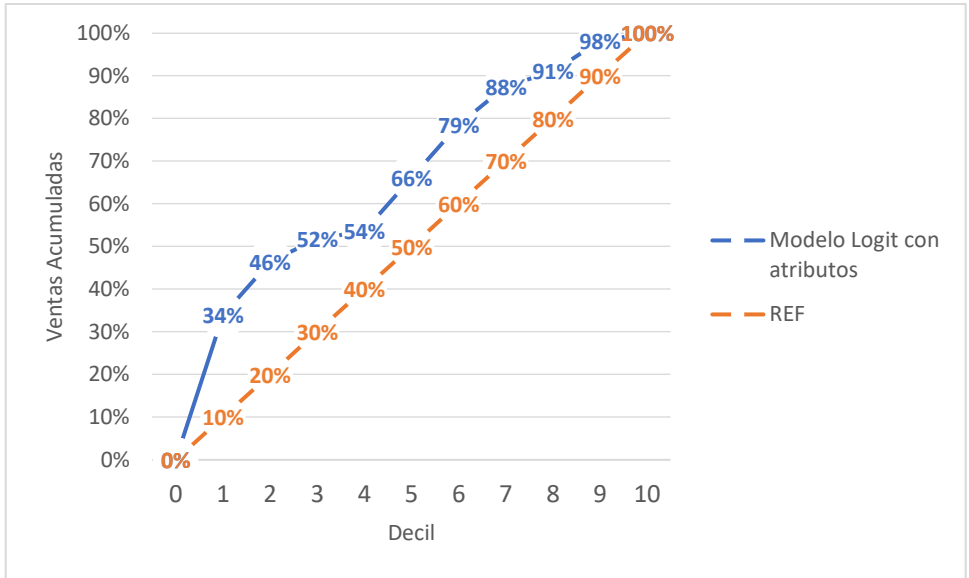


Ilustración 34: Curva de ganancia acumulada en grupo de validación para el producto VA5

Tabla 35: Matriz confusión en grupo de validación para el producto VA5

		Real	
		No Compra	Compra
Predicho	No Compra	694	29
	Compra	197	296

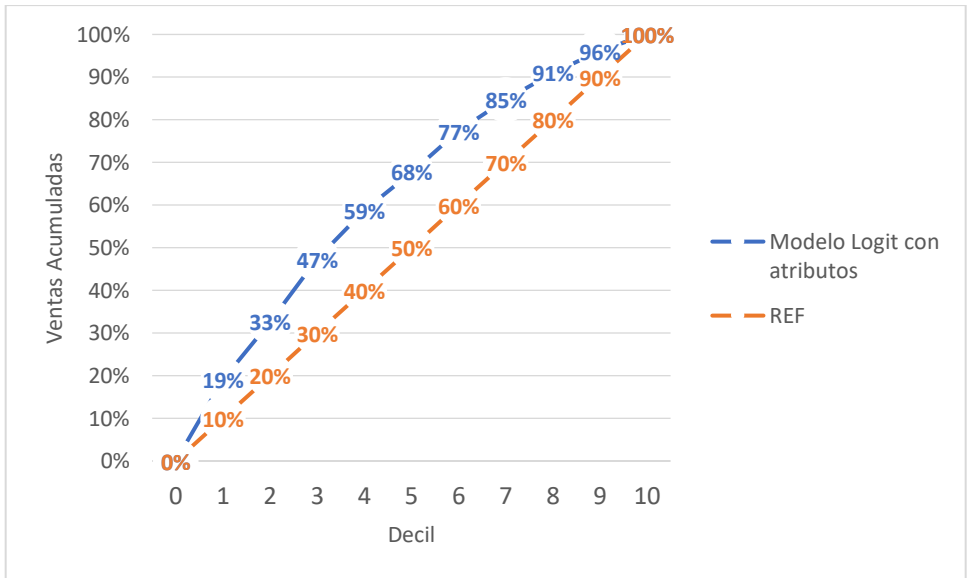


Ilustración 35: Curva de ganancia acumulada en grupo de validación para el producto APA

Tabla 36: Matriz confusión en grupo de validación para el producto APA

		Real	
		No Compra	Compra
Predicho	No Compra	9673	735
	Compra	7540	1361

## 11.5 Anexos E

Tabla 37: Ejemplo del output modelo de asignación

Rut	AP	APA	VA5	VCB
1	0	1	0	0
2	0	1	0	0
3	0	1	0	0
4	0	0	0	1
5	0	1	0	0
6	0	0	0	1
7	0	1	0	0
8	1	0	0	0
9	0	1	0	0
10	0	0	1	0

