



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

**SEGMENTACIÓN DE CLIENTES DE UNA EMPRESA MULTISERVICIO PARA LA  
RENTABILIZACIÓN DE SU CARTERA**

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL**

**IGNACIO ANDRÉS PORRAS SILVA**

**PROFESOR GUÍA:  
ALEJANDRA PUENTE CHANDÍA**

**MIEMBROS DE LA COMISIÓN:  
MAXIMO BOSCH PASSALACQUA  
MANUEL REYES JARA**

**SANTIAGO DE CHILE  
NOVIEMBRE 2011**

RESUMEN DE LA MEMORIA PARA OPTAR AL  
TITULO DE INGENIERO CIVIL INDUSTRIAL  
POR: IGNACIO PORRAS SILVA  
FECHA: 11/07/2011  
PROF.GUIA: SRTA. ALEJANDRA PUENTE

## **SEGMENTACIÓN DE CLIENTES DE UNA EMPRESA MULTISERVICIO PARA LA RENTABILIZACIÓN DE SU CARTERA**

Hoy en día la industria aseguradora es muy competitiva, es por esto, que para mantenerse en los primeros lugares los esfuerzos de marketing deben ser máximos y eficientes. Para esto será primordial entender y entregar los productos que desean los consumidores.

Actualmente la empresa posee una baja efectividad en las campañas de cross y up selling, es por esto que se busca encontrar una clasificación de los clientes de la empresa, que permita diseñar acciones de marketing focalizado con el objetivo de aumentar la tenencia de productos por cliente y aumentar la efectividad de las campañas.

El objetivo de esta memoria es realizar una metodología de segmentación que aporte información relevante capaz de asistir la oferta de productos, junto a esto encontrar un perfil de los individuos a abordar, que aumente el nivel de eficacia de las acciones comerciales.

Se aplicaron modelos predictivos para analizar las características que gatillan la compra de los diferentes productos. Los métodos utilizados para este análisis fueron: Logit, Análisis Discriminante y Árbol de Clasificación C.5. Luego se aplicó una segmentación con el modelo no supervisado k-means, a partir de las variables de propensión y tenencia de productos. Con el fin de caracterizar de la mejor manera posible cada uno de los perfiles y segmentos encontrados, se realiza un estudio de mercado levantando información socio demográfica y de estilos de vida de los clientes.

A partir de lo anterior, se encontraron grupos para la oferta de los tres productos analizados y las coberturas que pueden contratar. Además se obtuvieron 8 segmentos que reflejan necesidades y características particulares, que permiten el desarrollo de acciones comerciales diferenciadas, desde el producto a ofrecer hasta la forma de hacerlo.

Se realizan recomendaciones para futuras acciones de marketing, un ejemplo es que grupos se debe intentar desarrollar y cuáles sólo se debe tratar de fidelizar, ya que han alcanzado gran parte de su potencial dentro de la empresa.

Cómo trabajos futuros se recomienda replicar la metodología con mayor información demográfica, con el fin de aumentar los índices de desempeño de los modelos predictivos. Además sería interesante poder cuantificar el aumento de la efectividad debido a la aplicación de esta metodología, a través de una campaña real.

## ÍNDICE

1	INTRODUCCIÓN.....	1
1.1	DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN .....	3
1.2	OBJETIVOS .....	4
1.3	METODOLOGÍA.....	5
1.3.1	Estudio de bibliografía y modelo de negocio.....	5
1.3.2	Estructuración y análisis de base de datos .....	6
1.3.3	Diseño de estudio de mercado y muestreo .....	6
1.3.4	Segmentación preliminar .....	7
1.3.5	Desarrollo de modelo de propensión de compra.....	7
1.3.6	Segmentación por tenencia y propensión de productos.....	7
1.3.7	Recomendación de acciones comerciales .....	7
1.4	ALCANCES .....	8
1.5	RESULTADOS ESPERADOS .....	8
2	MARCO CONCEPTUAL.....	8
2.1	SEGMENTACIÓN.....	9
2.1.1	Métodos de partición .....	10
2.2	MODELOS DE PROPENSIÓN DE COMPRA.....	13
2.2.1	Análisis discriminante .....	13
2.2.2	Regresión logística .....	14
2.2.3	Arboles de decisión .....	15
2.2.4	Índices de comparación de desempeño.....	18
2.3	MARKETING FOCALIZADO.....	19
2.4	MUESTREO .....	19
2.4.1	Muestreo Aleatorio Simple.....	20
2.4.2	Muestreo Aleatorio Sistemático .....	20
2.4.3	Muestreo Aleatorio Estratificado .....	20
2.4.4	Muestreo aleatorio por conglomerados.....	20
2.5	ESTUDIO DE MERCADO.....	21
3	DESARROLLO DE LA METODOLOGÍA.....	22
3.1	ESTUDIO DE BIBLIOGRAFÍA Y MODELO DE NEGOCIO .....	22

3.2	ESTRUCTURACIÓN Y ANÁLISIS DE LA BASE DE DATOS .....	22
3.3	REDUCCIÓN DE DIMENSIONES .....	31
3.4	DISEÑO DE LA ENCUESTA Y MUESTREO .....	32
3.5	RESULTADO DE LA ENCUESTA .....	33
3.6	SEGMENTACIÓN PRELIMINAR .....	39
3.6.1	C.H.A.I.D .....	39
3.6.2	QUEST .....	42
3.6.3	K- medias .....	44
3.6.4	Conglomerado en dos fases .....	46
3.7	DESCRIPCIÓN DE CONGLOMERADOS.....	48
3.8	APLICACIÓN DE MODELOS DE PROPENSIÓN DE COMPRA.....	56
3.8.1	Productos .....	58
3.8.2	Coberturas.....	64
3.9	CREACIÓN DE VARIABLES OFERTA.....	71
3.9.1	Perfil por oferta .....	71
3.9.2	Perfil por coberturas .....	76
3.10	CLASIFICACIÓN POR PROPENSIÓN Y TENENCIA DE PRODUCTOS .....	79
3.11	RECOMENDACIONES DE ACCIONES COMERCIALES.....	89
4	CONCLUSIONES.....	92
4.1	PROBLEMA DE NEGOCIO .....	92
4.2	METODOLOGÍA.....	93
4.3	TRABAJOS FUTUROS.....	93
5	BIBLIOGRAFÍA Y FUENTES DE INFORMACIÓN .....	94
6	ANEXOS.....	96
6.1	ANEXO A- Sectores de residencia de clientes.....	96
6.2	ANEXO B - Encuesta.....	97
6.3	ANEXO C – Análisis de arboles.....	99
6.4	ANEXO D – Resultados Encuesta .....	100
6.5	ANEXO E- Arboles de clasificación .....	106
6.5.1	Flexible protección.....	106
6.5.2	APV Flexible.....	107
6.5.3	APV- Protección .....	108

6.6	ANEXO F – Anova tenencia de productos .....	109
6.7	ANEXO G – Análisis de varianza de segmentos.....	110
6.8	ANEXO H – Análisis de segmentación .....	112

## Índice de figuras

Figura 1: Productos de Consorcio .....	1
Figura 2: Crecimiento de la industria de seguros .....	2
Figura 3: Participación de la empresa en el mercado de seguros .....	2
Figura 4: Metodología .....	5
Figura 5: Función Logística .....	15
Figura 6: Ejemplo de árbol de decisión .....	16
Figura 7: Formas de levantar información .....	21
Figura 8: Distribución edad.....	24
Figura 9: Distribución geográfica de clientes.....	25
Figura 10: Distribución de valor del cliente.....	25
Figura 11: Participación de productos .....	26
Figura 12: Participación de coberturas .....	27
Figura 13: Número de coberturas.....	27
Figura 14: Estado civil .....	34
Figura 15: Nivel de educación .....	34
Figura 16: Actividades de preferencia .....	35
Figura 17: Frecuencia de la actividad.....	35
Figura 18: Medio favorito de información .....	36
Figura 19: Tipo de compra por internet .....	37
Figura 20: Red social preferida .....	37
Figura 21: Frecuencia de uso de la red social.....	38
Figura 22: Tamaño grupo familiar .....	38
Figura 23: Regla del codo K-Medias .....	44
Figura 24: Regla del codo Conglomerado en dos fases.....	47
Figura 25: Coberturas por conglomerado.....	49
Figura 26: Nivel de educación de cada segmento.....	50
Figura 27: Valor del cliente de cada grupo .....	51
Figura 28: Participación de productos por grupo .....	52
Figura 29: Seguros en competencia por grupo .....	53
Figura 30: Cláusulas por cada grupo .....	53
Figura 31: Tamaño familiar de cada grupo.....	54
Figura 32: Capital de coberturas por grupo .....	55
Figura 33: estado civil por grupo .....	55
Figura 34: Tipos de productos.....	57
Figura 35: Utilización de datos .....	58
Figura 36: Curva de ganancia Flexible-Protección.....	59
Figura 37: Curva de ganancia APV-Flexible .....	61
Figura 38: Curva de ganancia APV-Protección .....	63
Figura 39: Curva de ganancia Cobertura de accidentes .....	65
Figura 40: Curva de ganancia de Cobertura de Invalidez .....	66
Figura 41: Curva de ganancia cobertura de salud.....	67
Figura 42: Curva de ganancia cobertura de Gastos médicos .....	68
Figura 43: Curva de ganancia cobertura oncológica .....	69
Figura 44: Actividad preferida por oferta de producto .....	73

Figura 45: Red social preferida por cada grupo .....	74
Figura 46: Nivel educacional por oferta de producto .....	75
Figura 47: Estado Civil por oferta de producto .....	76
Figura 48: Regla del codo 2º K-medias .....	80
Figura 49: N° de casos por grupo.....	81
Figura 50: Actividad preferida por grupo .....	82
Figura 51: Estado civil por grupo.....	83
Figura 52: Pagos por internet por grupo.....	84
Figura 53: Pagos por internet por grupo.....	85
Figura 54: Red Social preferida por grupo .....	88
Figura 55: Mapa de la división zonal de la ciudad de Santiago.....	96
Figura 56: Estado Civil de la muestra.....	100
Figura 57: Nivel educacional de la muestra .....	100
Figura 58: Actividades de preferencia .....	101
Figura 59: Frecuencia de actividades de preferencia.....	101
Figura 60 Medio favorito de información .....	101
Figura 61: Tipo de compra por internet .....	102
Figura 62: Red Social Preferida .....	102
Figura 63: Frecuencia de uso de la red social.....	103
Figura 64: Tamaño del grupo familiar.....	103
Figura 65: Principales competencias.....	104
Figura 66: APV en la competencia .....	105
Figura 67: Seguros de protección en la competencia .....	105
Figura 68: Seguros en competencia.....	105
Figura 69: Árbol flexible protección .....	106
Figura 70: Árbol APV-Flexible .....	107
Figura 71: Árbol APV-Protección.....	108

## Índice de Tablas

Tabla 1: Tipos de segmentación .....	10
Tabla 2: Ejemplo de tabla de Confusión.....	18
Tabla 3: Variables de la Base de Datos .....	23
Tabla 4: Tabla resumen de limpieza de base de datos .....	24
Tabla 5: Caracterización de clientes por productos.....	28
Tabla 6: Lift entre los diferentes productos.....	29
Tabla 7: Caracterización de clientes bi-producto.....	30
Tabla 8: Caracterización de clientes por cobertura .....	30
Tabla 9: Resultado análisis factorial.....	31
Tabla 10: Matriz de componentes rotados .....	32
Tabla 11: Variable productos.....	39
Tabla 12: Resultado CHAID .....	40
Tabla 13 : Capacidad de predicción para CHAID.....	41
Tabla 14: Lambda de Wilks para CHAID.....	41
Tabla 15: Resultados QUEST .....	42
Tabla 16: Capacidad de predicción QUEST.....	43
Tabla 17: Lambda de Wilks para QUEST .....	43
Tabla 18: Resultado K-Medias .....	45
Tabla 19: Capacidad de predicción K-Medias .....	45
Tabla 20: Lambda de Wilks K-Medias .....	46
Tabla 21 : Resultado de Conglomerado en dos fases.....	47
Tabla 22: Capacidad de predicción Conglomerado en dos fases .....	48
Tabla 23: Lambda de Wilks de Conglomerado en dos fases .....	48
Tabla 24: Calidad de modelos predictivos para cruce Flexible-Protección .....	59
Tabla 25: Desempeño de modelos para cruce APV-Flexible .....	61
Tabla 26: Análisis Discriminante .....	62
Tabla 27: Desempeño de modelos para Cruce APV-Protección.....	63
Tabla 28: Desempeño de modelo para cobertura de accidentes .....	65
Tabla 29: Desempeño modelo para cobertura de invalidez .....	66
Tabla 30: Desempeño de modelo para cobertura de salud.....	67
Tabla 31: Desempeño de modelo para cobertura en Gastos médicos.....	69
Tabla 32: Desempeño de modelos para cobertura oncológica .....	70
Tabla 33: Función Discriminante para cobertura oncológica.....	70
Tabla 34: Perfil por oferta .....	71
Tabla 35: Lift entre los diferentes perfiles por oferta .....	72
Tabla 36: Perfil por oferta de coberturas .....	76
Tabla 37: Lift con el promedio de los clientes por cobertura .....	77
Tabla 38: Preguntas de la encuesta.....	98
Tabla 39: Aplicación de Chaid.....	99
Tabla 40: Aplicación de Quest.....	99
Tabla 41: Anova por posesión de productos .....	109
Tabla 42: ANOVA segmentos .....	110
Tabla 43: Comparación de medias para los segmentos .....	111
Tabla 44: Datos de Segmentos 1 .....	112



Tabla 45: Datos de segmentos 2.....	113
Tabla 46: Datos de segmentos 3.....	113
Tabla 47: Test de significancia bilateral .....	114
Tabla 48: Error en encuesta para segmentos .....	114

# 1 INTRODUCCIÓN

El presente trabajo se realiza para Consorcio, empresa multiservicio del área financiera y seguros, que posee alrededor de 12 variados productos, que acompañan a los clientes en distintas etapas de su vida.

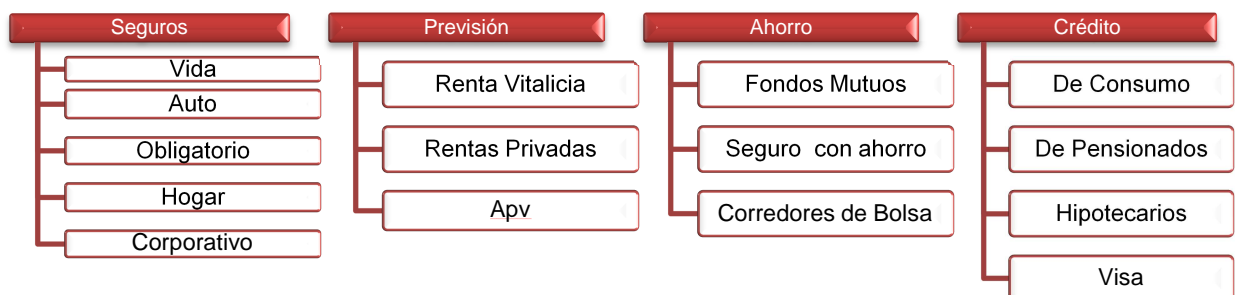
Dentro de su oferta de productos se puede mencionar cuatro líneas diferentes. La primera familia es de seguros, donde se ofrecen seguros de: vida, auto, hogar, SOAP y Corporativos. Esta línea de negocios, consiste básicamente en que el cliente paga una prima, que le da derecho a exigir que la compañía se haga cargo, en caso de algún siniestro. Este tipo de servicios nace de la aversión al riesgo de cada uno de los individuos, los cuales prefieren pagar un monto mensual que les brinda la seguridad, que ante un accidente serán atendidos sin la necesidad de pagar, o pagando sólo un deducible.

Esta empresa también ofrece servicios de previsión, en donde se encuentran: rentas vitalicias, rentas privadas y APV. Todos estos servicios están dedicados a la administración de los fondos que cada uno de los clientes han ahorrado durante su periodo de actividad laboral. En donde éstos, en el momento de la jubilación determinan la forma en que prefieren utilizar el dinero ahorrado, donde una de estas es la renta vitalicia, asegurándose de no quedarse sin dinero en años posteriores por no haber invertido de buena manera sus ahorros.

La tercera familia de productos es la de ahorro con acceso a productos como: fondos mutuos, seguro con ahorro y la corredora de bolsa. Estos productos buscan satisfacer la necesidad de aquellas personas que han logrado tener ciertos ahorros y buscan generar utilidades a partir de éstos. El producto que elijan dependerá de los riesgos que estén dispuestos a correr.

Por último se encuentran los créditos, como: créditos de consumo, créditos a pensionado, crédito hipotecario y la nueva Visa Consorcio. En la figura 1 se muestran los productos que se ofrecen y la línea de negocio a la que pertenecen.

Figura 1: Productos de Consorcio

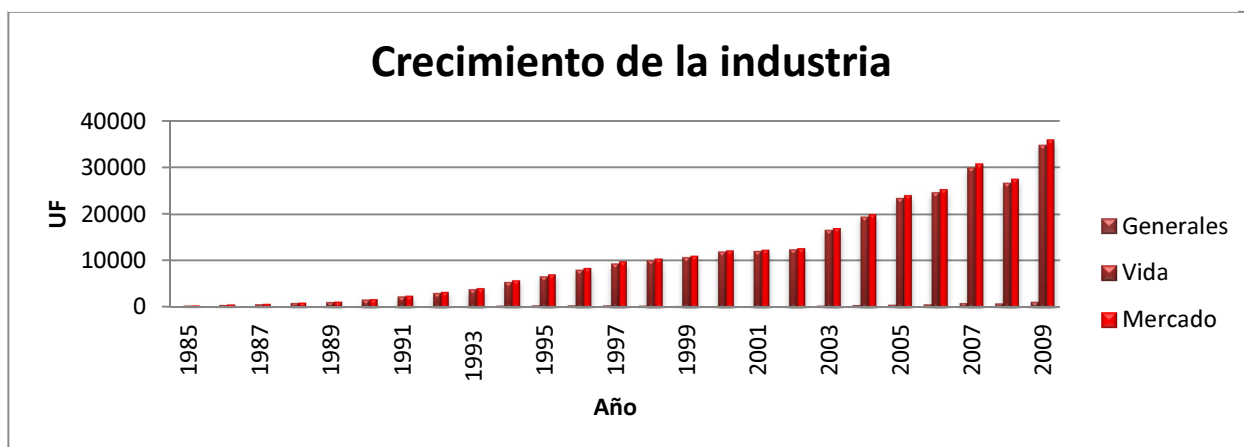


Fuente: "Elaboración propia"

La competencia que enfrenta esta organización es múltiple, ya que posee productos tanto de bancos como de las aseguradoras y en ambas industrias existe una amplia oferta, esto hace difícil encontrar un competidor principal. Estas industrias han tenido un crecimiento sostenido en los últimos años, tal como muestra la figura 2, donde se puede observar el crecimiento en la inversión de seguros de vida y seguros generales desde el año 1985<sup>1</sup>.

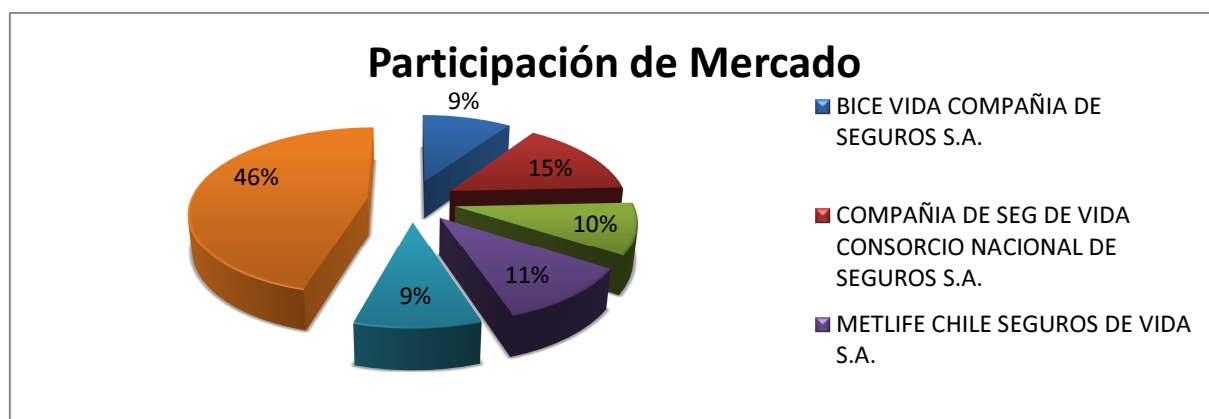
Ante tal crecimiento la competencia aumenta, por lo que los esfuerzos por ser más eficientes son mayores. Dentro del sector asegurador Consorcio es la empresa de mayor tamaño, donde el valor de sus activos es equivalente al 14% de la industria. Sus principales competidores dentro de esta industria son Metlife<sup>2</sup> con un 10,6%, Vice Vida y Principal con un poco menos del 10%, tal como puede observarse en la figura 3.

Figura 2: Crecimiento de la industria de seguros



Fuente: "Superintendencia de valores y seguros"

Figura 3: Participación de la empresa en el mercado de seguros



Fuente: "Superintendencia de valores y seguros"

<sup>1</sup> Fuente: Superintendencia de valores y seguros

<sup>2</sup> Fuente: Superintendencia de valores y seguros

El principal negocio de esta compañía son los seguros de vida, donde se ha posicionado como el más importante agente del mercado, lo que la ha transformado en el conglomerado financiero no bancario más grande del país, con capitales 100% chilenos. Consorcio administra fondos por más de US\$ 5.000 millones<sup>3</sup> y más de 2.000 empleados a lo largo del país.

Su oferta de seguros de vida es de tres tipos: seguros de protección, ofrecen una cobertura por cierto monto mientras el contrato este vigente, es decir, mientras se pague la póliza mensual. El segundo son los seguros flexibles, que son seguros con ahorro, al igual que los anteriores, ofrecen una cobertura mientras el contrato esté al día. Estos se diferencian en que las pólizas son más caras pero ofrecen la posibilidad de hacer un retiro de lo invertido en caso de no haber sido utilizado en algún siniestro. Por último está el ahorro previsional voluntario, que además de ofrecer una forma de ahorro con beneficios tributarios está asociado a una cobertura de fallecimiento.

Por su gran variedad de clientes y la poca información que se maneja de éstos en esta industria, ha sido muy difícil poder realizar un estudio que entregue datos significantes que los caractericen.

El presente trabajo busca entregar una clasificación de los clientes que permita el desarrollo de una estrategia de marketing focalizado, que entregará información de que producto se le debe entregar y de qué manera hacerlo. A partir de los resultados de este trabajo, la empresa espera poder diseñar sus acciones comerciales de manera diferenciada para cada segmento y que genere campañas más eficientes y efectivas.

## **1.1 DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN**

La competencia en esta industria es muy fuerte, por lo que el cliente es más exigente y obliga a esforzarse día a día a entregar un mejor servicio; por lo tanto es perentorio conocerlos, identificar a los más rentables y mejorar el servicio, realizando promociones más efectivas, logrando clientes satisfechos y leales a la empresa y generando una relación de largo plazo con ellos.

La gran diversidad de clientes y productos, hace primordial que se realicen acciones de marketing diferenciado, para esto es fundamental incurrir en estudios que entreguen información más profunda de cada individuo. A partir de esta información se podrían tomar acciones oportunas en la entrega de servicios financieros en el momento preciso en que más lo necesitan. Con esto se realizarían campañas de fidelización y retención, de manera eficiente y efectiva, al enfocar los recursos exactamente a aquellos que se quiere llegar, y entregándoles lo que quieren. Junto a esto, se podrían hacer campañas de up y cross selling para tener una cartera más rentable.

---

<sup>3</sup> Fuente: *Consorcio.cl*

Hoy en día, Consorcio no posee una segmentación de su cartera de clientes que le permita realizar acciones como las antes mencionadas. Se realizan acciones comerciales enfocadas en ciertos grupos como los clientes de mayor valor, que ellos conocen debido a su modelo de valor del cliente, en otras ocasiones se realizan campañas para un producto o para clientes de cierta edad, sin tener un conocimiento completo y global de estos. En este momento su capacidad de cross selling es muy baja, muy cercana a 1.5 productos por clientes y las campañas de up selling que se han realizado han tenido una baja eficiencia, cercana al 3% de efectividad.

Para rentabilizar la cartera es necesario mejorar estos índices, por lo que se desarrollaran acciones tácticas diferenciadas para cada segmento, identificando los diferentes tipos de comportamiento por conglomerado. Es por esto que este trabajo realizará una segmentación de los clientes de la empresa, en donde en primer lugar se utilicen las bases de datos de la organización, para luego complementar con una encuesta a una muestra de cada uno de los segmentos encontrados, que permita revelar los intereses, necesidades y estilos de vida de éstos, que luego se puedan extrapolar al resto de la población.

El hecho de poder realizar una caracterización por estilos de vida podrá agregar valor a la hora de la realización de promociones, ya que este tipo de análisis es mucho más completo y preciso. Respecto de los gustos y necesidades de los clientes que los estudios tradicionales que sólo usan datos socio demográfico o transaccional para caracterizarlos.

Con esto se desarrollarán acciones comerciales únicas para cada uno de los tipos de clientes identificados, de los cuales ya se conocerán sus necesidades, intereses y estilo de vida, por lo que las estrategias serán realizadas en base a un conocimiento global de éstos, lo cual se traducirá en índices de cross y up selling mucho mayores, además de alargar su ciclo de vida. Lograr esto significaría un aumento considerable de la rentabilidad de la cartera de la empresa.

## **1.2 OBJETIVOS**

### **o Objetivo General**

- Diseñar una metodología para segmentar y realizar recomendaciones de marketing, con el fin de aumentar el número de productos por cliente que se consumen dentro de la empresa.

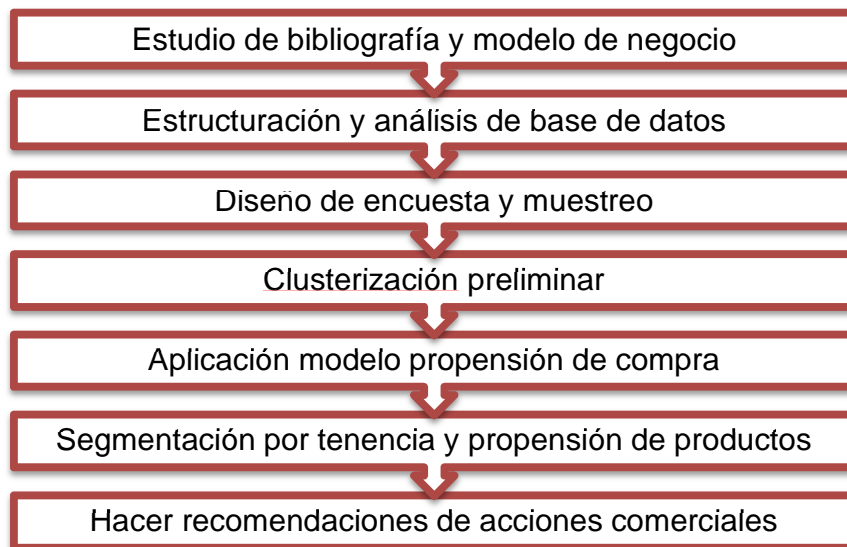
### **o Objetivos Específicos**

- Identificar y caracterizar a los clientes propensos a comprar los diferentes productos

- Segmentar a los clientes en base a tenencia y propensión compra de productos
- Encontrar características que permitan diseñar protocolos de atención diferenciada para cada segmento.

### 1.3 METODOLOGÍA

Figura 4: Metodología



Fuente: "Elaboración Propia"

#### 1.3.1 Estudio de bibliografía y modelo de negocio

El objetivo de esta etapa es poder comprender los métodos que se aplicaran durante este trabajo, para así poder definir cómo se realizará cada etapa del proyecto una vez que se hayan analizado las mejores opciones.

Para poder desarrollar una buena investigación es fundamental conocer el contexto en donde ésta se sitúa, ya que una de las complejidades del proyecto a realizar es la naturaleza de la industria a la cual pertenece. Por otra parte es necesario conocer los productos que ésta ofrece, para saber las necesidades que estos suplen y así definir las variables más precisas que permitan la caracterización de cada uno de los conglomerados que se pretende encontrar en etapas posteriores.

Ante la necesidad de usar software para la aplicación de cada uno de los métodos, se hace fundamental el conocimiento y manejo previo de éstos. Para lo

cual es necesario estudiar, practicar y familiarizarse con cada uno de los softwares que se ocuparán durante esta investigación.

Para poder recabar mayor información sobre los estilos de vida, será necesario aplicar una encuesta a una muestra del total de los clientes de la empresa. Para esto es importante elegir la herramienta adecuada, por lo que hay que estudiar las opciones que existen y ver cual se adecua de mejor manera al estudio que se realizará.

### **1.3.2 Estructuración y análisis de base de datos**

Antes de realizar la segmentación es necesario realizar este paso, ya que la calidad de los resultados que entregará el modelo de segmentación dependerá en gran parte de las variables que se le introduzcan. Si se le entrega una cantidad de variables muy alta complicará los análisis posteriores, mientras que si se le entregan muy pocas, los resultados serán muy básicos. Es por esto que se estudiará de manera cuantitativa y en base a juicio de expertos, que variables incluir. Para luego realizar un análisis de factores que nos permita reducir el número de variables al mostrar cuales son las que están más correlacionadas y sólo tomar la más importante de cada grupo.

Con el fin de realizar mejores análisis posteriores, es indispensable entender acabadamente cada una de las variables y realizar estadísticos descriptivos que ayuden a su entendimiento. Además, es importante que se limpien los datos que ensuciaran el estudio, como valores fuera de rango y valores faltantes.

### **1.3.3 Diseño de estudio de mercado y muestreo**

Para describir a los grupos de clientes en base a los estilos de vida, se deben definir los atributos que caracterizaran a cada uno de éstos. Para esto será necesario el estudio de la metodología Vals y de otras metodologías que cumplen esta función.

En esta etapa se diseñará una encuesta a partir de los atributos antes mencionados, que pretenden describir a cada uno de los clientes, con el fin de encontrar estilos de vida.

Para poder aplicar una encuesta será necesario encontrar una muestra representativa. Se realizará un muestreo aleatorio simple que permita lo antes mencionado.

#### **1.3.4 Segmentación preliminar**

Luego de que se hayan definido las variables de estudio, se desarrollará la primera segmentación a partir de la información existente en las bases de datos de la empresa. En esta etapa se aplicarán dos métodos de segmentación que permitan comparar los resultados y así poder elegir el más conveniente para los objetivos de la empresa.

Una vez realizada la segmentación, será necesario analizar y caracterizar cada uno de los segmentos. A partir de toda la información levantada durante el estudio, se pretende caracterizar cada uno de los segmentos encontrados, utilizando atributos que permitirán describir estilos de vida.

#### **1.3.5 Desarrollo de modelo de propensión de compra**

En esta etapa se aplicarán modelos predictivos para analizar las características de compra de los diferentes productos de la empresa, con el fin de poder asistir la oferta de productos a los clientes actuales de la organización.

#### **1.3.6 Segmentación por tenencia y propensión de productos**

A partir de las variables de propensión de compra generadas, se realizará una clasificación de los clientes de la empresa. Con esto se busca encontrar las diferentes necesidades que la empresa está resolviendo, con el fin de poder diseñar las campañas para cada uno de estos conjuntos. Además se espera encontrar diferencias de estilos de vida en cada una de estas necesidades encontradas, que nos darán información relevante para el diseño de campañas de marketing

#### **1.3.7 Recomendación de acciones comerciales**

Una vez segmentados los clientes, se podrá realizar una recomendación a la organización de cómo deberá enfrentar a cada uno de los segmentos. Estas decisiones estarán apoyadas en los conocimientos de qué producto y de la forma de ofrecerlo, de acuerdo a la personalidad de cada uno de estos grupos.



## **1.4 ALCANCES**

Para la realización del trabajo, sólo se tomará en cuenta información de las bases de datos de la empresa y la levantada en la encuesta. Es por esto que no influye la competencia u otras variables exógenas que no se encuentran en los datos antes descritos. Sólo será analizada la información de seguros de vida, ya que es la base más completa.

Dentro del estudio, será utilizado como variable el modelo de valoración al cliente que la empresa posee, por lo que no se creará un modelo nuevo para este trabajo.

El objetivo es agrupar el total de clientes en conglomerados más pequeños que puedan describirse, aportando gran información para diseñar estrategias de marketing diferenciado. No se contempla el diseño de estas acciones en el marco de este estudio.

## **1.5 RESULTADOS ESPERADOS**

Se espera obtener una base de datos con los clientes separados en diferentes conglomerados, los cuales deben ser medibles, identificables y de un tamaño suficiente como para que sea rentable hacer una estrategia de marketing dirigida exclusivamente a ellos.

Estos grupos deberán entregar información relevante para poder entender las principales necesidades e intereses de cada uno. Principalmente saber cuál es el producto, los canales y la forma más adecuada para cada perfil.

Lo anterior para facilitar la creación de acciones de marketing focalizado que a final de cuentas se traduzcan en una cartera de clientes más rentable para la empresa.

## **2 MARCO CONCEPTUAL**

En este capítulo se explica brevemente parte importante del conocimiento utilizado para el desarrollo y entendimiento de los diferentes capítulos de este trabajo, destacando los conceptos de segmentación y propensión de compra.

## 2.1 SEGMENTACIÓN

La segmentación es una herramienta que busca agrupar datos en diferentes conglomerados, en donde cada uno de los datos pertenecientes a un grupo son similares entre sí y diferentes a los de los demás clúster. Se puede definir cómo “El proceso de identificación y caracterización, mediante técnicas estadísticas, de subgrupos de consumidores dentro de un mismo mercado que presentan distintas necesidades, así como la selección de aquellos que podrían ser objeto de una oferta comercial diferente adaptada a los mismos, con el fin de incrementar la eficacia y rentabilidad de la actuación comercial de la empresa”<sup>4</sup>.

Esta herramienta nos permite analizar el mercado de una manera menos agregada, bajo el supuesto de que los clientes no son iguales, por lo que tratarlos a todos de la misma manera es un error. Es por esto que en marketing se ha utilizado la segmentación de mercado para detectar oportunidades de negocio, segmentos más rentables y desarrollar estrategias diferenciadas para cada uno de ellos.

Una buena segmentación debe tener grupos de tamaño significativo, accesibles e identificables. Lo que implica que los grupos sean lo suficientemente grandes y fácilmente ubicables para poder apoyar una estrategia en cada uno de estos y que sea rentable.

Existen diferentes tipos de segmentación, las cuales dependerán de las diferentes variables a utilizar, que serán elegidas a partir de los objetivos que busca la organización. Estos se pueden apreciar en la Tabla 1.

---

<sup>4</sup> Talaya, Agueda. “Principios de marketing”

Tabla 1: Tipos de segmentación

Tipos de segmentación	Variables utilizadas
Demográfica	Sexo
	Edad
	Raza
	Residencia
Socio-Económica	Niveles de ingreso
	Nivel de educación
	Profesión
	Clase social
Uso	Cantidad de Uso
	Tipo de Uso
	Oportunidad de Uso
	Lealtad a la marca
Estilos de vida	Engloban diferentes criterios, los cuales buscan describir segmentos a partir de la forma en que enfrenta la vida cada cliente.

Fuente: Santiago Nuñez.2010.

Para lograr lo anterior se ocupan diferentes métodos matemáticos, a continuación se explicarán algunos de los que serán utilizados en este trabajo.

### 2.1.1 Métodos de partición

Para la utilización de estos métodos es necesario entregar el número de conglomerados y variables a utilizar, antes de comenzar la segmentación. A partir de esta información se comienza un problema de optimización, en el cual se busca minimizar la distancia entre los datos individuales y los posibles centros de grupos, los que se modificaran constantemente hasta llegar a un óptimo. Dentro de estos métodos podemos encontrar dos familias.

- **Partición rígida**

Cómo en todas las variantes de segmentación, estos buscan separar la población existente en k grupos diferentes, a partir de ciertas variables involucradas. En los métodos de partición rígida cada individuo posee una pertenencia a un solo grupo.

- **Partición difusa**

A diferencia del modelo anterior, los métodos de partición con lógica difusa no entregan un valor de pertenencia a un solo grupo, sino que entregan la probabilidad de que el elemento i forme parte del conglomerado k ( $\mu_{ik}$ ) para todos los i,k.

### 2.1.1.1 K- medias

Pertenece a los modelos de partición rígida que se puede encontrar en la mayoría de los paquetes estadísticos del mercado, llevándolo a ser uno de los métodos de segmentación más populares.

Este método se basa en la asignación de cada uno de los elementos al grupo del centroide más cercano. Para definir esta cercanía utiliza la distancia euclidiana, es por esto que necesita que cada una de las variables sea de razón. El centroide de un conglomerado es el promedio en todas las dimensiones de los elementos pertenecientes a ese grupo como se aprecia en la ecuación 1, donde  $N_k$  es el número de observaciones de Y en el clúster k.

$$v_k = \frac{\sum_{i=1}^{N_k} y_i}{N_k} \quad \text{Ecuación 1}$$

Recibe una matriz de mxn, es decir, con m observaciones y n variables, las que buscará reagrupar en k grupos distintos. Esto último lo hará encontrando el mínimo de la ecuación 2, donde se muestra la suma de los cuadrados de las distancias entre cada punto  $y_i$  y el centro del cada cluster  $v_k$ .

$$\sum_{k=1}^K \sum_{i=1}^m \|y_i - v_k\|^2 \quad \text{Ecuación 2}$$

La heurística utilizada en este modelo se puede describir de la siguiente manera:

1. Definir los centroides iniciales y definir el número de clúster a calcular.
2. Asignación de cada elemento al conjunto cuyo centroide esté más cerca.
3. Cálculo de los nuevos centroides.
4. Se repiten los pasos 2 y 3 hasta que ninguno de los elementos cambie de conglomerado.

Además destaca por su fácil aplicación, pero su solución muchas veces es sólo un óptimo local, lo cual depende mucho de sus centroides iniciales. Otra desventaja de este método es que no calcula el número óptimo de clúster y además sólo utiliza variables de carácter numérico.

### 2.1.1.2 C- medias

Este es un modelo de segmentación basado en la lógica difusa, es decir, no asigna a los elementos a un único conglomerado, sino que asigna una probabilidad de pertenencia a éste.

En este caso el método busca encontrar el óptimo a la función objetivo que se puede observar en la ecuación 3, que a su vez busca encontrar el grado de pertenencia de cada objeto de la matriz de datos Y a cada uno de los clusters.

$$J(Y, U, V) = \sum_{k=1}^K \sum_{i=1}^m (\mu_{ik})^\delta d_{ik}^2 \quad \text{Ecuación 3}$$

$$d_{ik}^2 = |x_i - v_k|^2 \quad \text{Ecuación 4}$$

$$\mu_{ik} \in [0, 1], \quad 1 \leq i \leq m, \quad 1 \leq k \leq K \quad \text{Ecuación 5}$$

$$\sum_{k=1}^K \mu_{ik} = 1, \quad 1 \leq i \leq m \quad \text{Ecuación 6}$$

$$0 < \sum_{i=1}^m \mu_{ik} < m, \quad 1 \leq k \leq K \quad \text{Ecuación 7}$$

En donde las ecuaciones 5, 6 y 7 son las restricciones a las cuales el problema de optimización está sujeto.  $\delta$  es un índice de difusividad el cual va de  $[1, \infty)$ , por lo que si se toman valores de  $\delta$  muy cercanos a 1, el resultado será una partición rígida y mientras más alejado de uno el resultado será más difuso. El valor recomendado y más usado en la literatura es 2,5.<sup>5</sup>

### 2.1.1.3 Conglomerado en dos fases

Este método posee una gran eficiencia en bases de datos muy grandes, una de sus principales ventajas es que es capaz de clasificar utilizando variables categóricas como continuas, además define automáticamente el número ideal de clúster.

Primero realiza un modelo C.F. tree, que consiste en un árbol con un número determinado de niveles (SPSS usa 3), donde cada nodo puede tener hasta 8 hojas que se consideran como sub clúster. Esta primera segmentación se lleva a cabo con las variables categóricas, luego con índices como B.I.C.<sup>6</sup> o A.I.C.<sup>7</sup> se determina el número ideal de conglomerados.

---

<sup>5</sup> Teppola et al., 1999

<sup>6</sup> B.I.C.: Criterio de información Bayesiano

<sup>7</sup> A.I.C.: Criterio de información de Akaike

Para la segunda etapa se ocupa un modelo de clúster jerárquico aglomerativo, el cual, toma como entrada el resultado del árbol CF y comienza a juntar los sub conglomerados en relación a las distancias calculadas entre ellos a partir de las variables continuas, hasta que llega al número ideal de clúster antes determinado.

## 2.2 MODELOS DE PROPENSIÓN DE COMPRA

Los modelos de propensión de compra son métodos predictivos aplicados a la compra o no compra de un producto o servicio determinado. Estos tienen como objetivo predecir cuan probable es la compra de un producto  $i$  por la persona  $k$ , para todos los  $i, k$ . Esto lo hacen a partir de las diferentes características de los clientes, las cuales vienen a ser las variables predictivas.

Existe una amplia gama de métodos predictivos utilizados en estos casos, a continuación se mencionan los más populares:

- Análisis Discriminante
- Regresión Logística
- Clasificadores Bayesianos
- Árboles de decisión
- Redes Neuronales

En la siguiente sección se describirán los métodos necesarios para la realización de este trabajo.

### 2.2.1 Análisis discriminante<sup>8</sup>

Este modelo se puede definir según Abaira y Perez (1996, pág.341) como “El conjunto de técnicas estadísticas que permiten estudiar el grado en que diferentes poblaciones, establecidas a priori, difieren entre sí”. Por lo tanto, es una técnica de clasificación, que se realiza mediante funciones lineales y cuadráticas. Este algoritmo puede ser utilizado con fines tanto predictivos como explicativos.

El método discriminante más popular es el de Fisher, el cual posee la siguiente función discriminante,

$$Z_{ij} = U_1 X_{i1j} + \dots + U_n X_{inj}. \quad \text{Ecuación 8}$$

Dónde:

$i=1, \dots, k$  (grupos)

---

<sup>8</sup> [V] Lévy, Jean. Varela, Jesús (2003)

$j=1,\dots,m$  (observaciones)  
 $k=1,\dots,n$  (variables)

El análisis discriminante bajo los siguientes supuestos según Sánchez (1984).

- Necesidad de que haya al menos dos grupos.
- Para cada grupo se necesitan dos o más casos.
- Se puede utilizar cualquier número de variables discriminantes.
- La variable que define los grupos ha de ser nominal, mientras que las variables discriminantes tienen que ser de intervalo.
- Ninguna variable discriminante puede ser combinación lineal de otras variables discriminantes.
- El número máximo de funciones discriminantes que se pueden calcular podrá ser igual al número de variables discriminantes, siempre y cuando su número no sea mayor que el número de grupos menos uno.
- Las matrices de varianzas-covarianzas de cada grupo han de ser aproximadamente iguales.
- Las variables discriminantes han de comportarse según una distribución normal multivariante.

Muchos de estos supuestos pueden relajarse sin afectar en gran manera la calidad del análisis, pero sí condicionan los resultados que podrían obtenerse.

El criterio utilizado para la determinación del eje discriminante  $Z$  consiste en maximizar  $\lambda$ .

$$\lambda = \frac{\text{Variación entre grupos}}{\text{Variación intragrupos}} \quad \text{Ecuación 9}$$

### 2.2.2 Regresión logística<sup>9</sup>

Es un modelo no lineal, en donde la variable a estudiar no es cuantitativa, sino que refleja el resultado de la elección de un individuo, o la pertenencia a una categoría. La función logística se encuentra acotada entre cero y uno, es por esto que sus resultados se interpretan como la probabilidad del suceso en estudio.

Este modelo parte por la definición de la variable observada, la cual tendrá un valor de cero si el suceso no ocurre y 1 si pasa lo contrario. A partir de las variables explicativas del fenómeno de compra, se puede calcular la esperanza de ocurrencia del suceso, como se puede apreciar en la ecuación número 10.

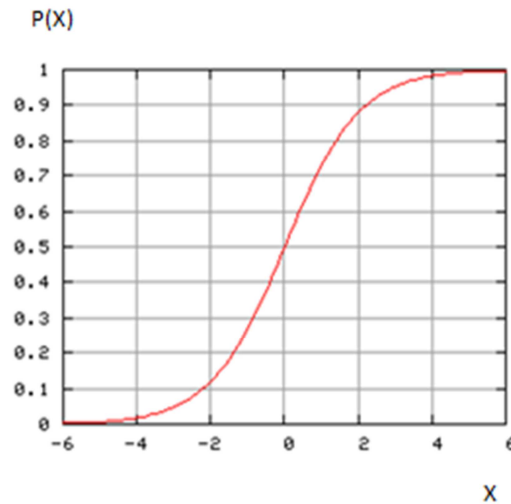
---

<sup>9</sup> [V] Lévy, Jean. Varela, Jesús (2003)

$$E[Y_i] = P_i = F(\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}) = \frac{e^{(\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}}{1 + e^{(\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}} \quad \text{Ecuación 10}$$

Donde F es la función logística

Figura 5: Función Logística



Fuente: "Elaboración propia"

### 2.2.3 Árboles de decisión

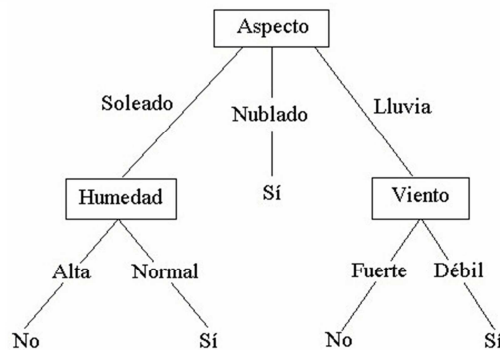
Un árbol de decisión es un grafo formado por un nodo padre que contiene a la población, a partir de divisiones de ésta, se comienzan a generar nodos hijos, que representan segmentos del total del universo en estudio. Estos métodos son utilizados para explicar el comportamiento de una variable a través de una serie de variables independientes.

Los diferentes algoritmos utilizados como árboles de decisión se diferencian principalmente por cómo definen cada una de las particiones realizadas en el árbol.

En la figura 6 se puede apreciar cómo se comporta una variable en relación a otras, en este caso se podrían clasificar los días por las variables conocidas, o también se podría hacer una predicción de si ocurrirá o no el evento estudiado dependiendo de las variables independientes.



Figura 6: Ejemplo de árbol de decisión



Fuente: "Elaboración propia"

Dentro de esta clasificación se encuentran dos tipos:

**Árboles de regresión:** Estos árboles permiten pronosticar un valor continuo, se tiene una variable supervisada conteniendo estos valores. Hay dos criterios para determinar la ramificación del árbol

- Minimizar el error: minimiza el error general del árbol.
- Minimizar la varianza: este criterio busca minimizar la variabilidad de los registros que asignará a cada rama.

$$Gini(v) = \sum_{i \neq j} p_v \sum_{i \neq j} p(i/v) * p(j/v) \quad \text{Ecuación 11}$$

*v*: nodo

*p*(*v*): Probabilidad de estar en el nodo *v*

*p*(*i/v*): Probabilidad de pertenecer a la clase *i* dado que estoy en el nodo *v*.

*p*(*j/v*): Probabilidad de pertenecer a la clase *j* dado que estoy en el nodo *v*.

**Árboles de clasificación:** Estos árboles permiten clasificar registros. La variable supervisada está compuesta por valores nominales que indican la pertenencia de cada registro a cada clase. Se destacan los siguientes tipos de árbol de regresión:

A continuación se profundizará en algunos de los algoritmos ocupados en este trabajo<sup>10</sup>.

<sup>10</sup> [IX] MENA, SEBASTIÁN. 2008

### 2.2.3.1 C.H.A.I.D

Este algoritmo es un árbol de clasificación, el cual se ha utilizado bastante en las empresas, con diferentes fines como clasificar conjuntos de datos y tratar de predecir sucesos de diferentes tipos.

No es necesario que las ramificaciones sean binarias, cómo es exigido por otros métodos, éstas se definen a partir del estadístico  $\chi^2$  (Se utilizará la partición con mayor valor en este estadístico). Luego de la realización de esta separación se generan nodos hijos, los que se dividirán de igual manera en base al mayor  $\chi^2$  y así sucesivamente hasta que cumpla con alguno de los criterios de parada. Este método es capaz de utilizar variables nominales como variable dependiente.

Alguna de las ventajas que posee este modelo son:

- Construcción de árboles no binarios
- Clasificación de nodos mutuamente excluyentes
- Creación de reglas de decisión valiosas para el negocio

### 2.2.3.2 C.5

Este algoritmo proviene de la familia de los TDIDT (top down induction decision tree), donde destaca el Id3 que es el predecesor de este árbol, diseñado por Quinlan en la década de 1980. Este método utilizaba un indicador de ganancia para la definición de las particiones, el cual consiste en la diferencia de entropía que se produce al realizar la división, se observa en la ecuación 12.

$$H(D) = - \sum_{i=1}^m p_i \text{Log}_2(p_i) \quad \text{Ecuación 11}$$

$$\text{Gain} = H(D) - \sum_{j=1}^v \frac{|D_j|}{|D|} * H(D_j) \quad \text{Ecuación 12}$$

$H(D)$  = Entropía en el nodo D

$|D|$  = N° de individuos en nodo D padre

$|D_j|$  = N° de individuos en nodo  $D_j$  nodo hijo de D

El Id3 fue mejorado a partir del algoritmo C4.5, que consiste en un pequeño cambio en el criterio de definición de las particiones realizadas, ya que éste utiliza el ratio de ganancia, el cual corrige una tendencia de su predecesor a preferir las particiones con mayor número de categorías:

$$Gain\ ratio = \frac{H(D) - \sum_{j=1}^v \frac{ID_j I}{IDI} * H(D_j)}{- \sum_{i=j}^v \frac{ID_j I}{IDI} * \log_2 \left( \frac{ID_j I}{IDI} \right)}$$

Ecuación 13

El C.5 es la versión comercial mejorada del anterior, sus ventajas son<sup>11</sup>:

- Mayor velocidad en la creación de modelos en comparación al C.45.
- Más eficiencia en la utilización de memoria en comparación al C.45.
- Se pueden utilizar árboles de decisión muchos más pequeños y obteniendo resultados similares en comparación a otros TDIDT.
- Creación de *boosting* para generar mayor precisión.
- Creación de ponderadores de peso para la nivelar cada tipo de error.
- Mejor tolerancia al ruido y datos nulos.

## 2.2.4 Índices de comparación de desempeño

Para poder conocer el desempeño de los métodos predictivos es necesario observar ciertos índices que se basan en la matriz de contingencia que a su vez capta los valores pronosticados correctamente como negativos y positivos. Además muestra los positivos indicados como negativos y los negativos como positivos. A partir de esta información se generan todos los índices.

Tabla 2: Ejemplo de tabla de Confusión

Tabla de confusión				
		Valor observado	total	
		0	1	
Valor pronosticado	0	A	B	A+B
	1	C	D	C+D
total		A+C	B+D	total

Fuente Elaboración Propia

Los índices son:

- Calidad Global : Busca explicar la cantidad predicha correctamente, versus el total de la población en estudio

<sup>11</sup>Parraguez,2010 VII

$$CG = \frac{A+D}{Total}$$

- Precisión : Explica cuáles de los valores positivos, realmente fueron identificados como positivos por el modelo.

$$CG = \frac{D}{B+D}$$

- Sensibilidad : Muestra que parte de los valores pronosticados como positivos realmente lo son.

$$CG = \frac{D}{C+D}$$

- Especificidad : Análogo al anterior pero para el caso de los pronosticados negativos.

$$CG = \frac{A}{A+B}$$

- Exactitud : Este último es una ponderación de los dos anteriores.

$$CG = \frac{Especificidad+Sensibilidad}{2}$$

## 2.3 MARKETING FOCALIZADO

Es la rama del marketing que es diseñada a partir de un objetivo mucho más pequeño, ya que éste no apunta a todo el mercado, sino que a algunos segmentos específicos que fueron elegidos de antemano. Debido a que al diseñar una campaña para un grupo específico, en base a sus intereses y necesidades, será mucho más efectiva que una que está diseñada para todo el mercado. Es por esto que el marketing focalizado permite aumentar tanto la eficacia como la eficiencia de las campañas. Con esto las organizaciones son capaces de aumentar la rentabilidad su empresa, con inversiones de menor monto.

## 2.4 MUESTREO<sup>12</sup>

En ciertas ocasiones cuando se requiere realizar un estudio y no es posible abarcar todo el universo, es necesario realizar un muestreo. Este método permite tomar un

---

<sup>12</sup> Reyes, M. 2008

pequeño pedazo del total del universo que permita hacer inferencias hacia la totalidad de los datos. La muestra debe ser representativa de todos los datos, es por esto que el muestreo es una herramienta científica, que se debe realizar con mucho cuidado ya que se pueden cometer errores importantes, debido a que al extrapolar los datos a toda la población aumentarían todos los errores que se produzcan por una mala muestra.

Existen varios tipos de muestreo con diferentes grados de dificultad y calidad. A continuación se describen los más conocidos:

#### **2.4.1 Muestreo Aleatorio Simple**

Consta de elegir a  $N$  individuos de un universo de 50.000, de manera aleatoria sin mayor cantidad de restricciones, que sea un método simple y fácil de aplicar, pero no tiene mucha utilidad práctica cuando la población enfrentada es muy grande.

#### **2.4.2 Muestreo Aleatorio Sistemático**

Para poder realizar este proceso es necesario ordenar a toda la muestra asignándole un número, luego se toma un número aleatorio, el cual es una semilla para recorrer la muestra completa. El tamaño del paso depende de la cantidad de individuos que se decidió tomar para el estudio. Este muestreo da mejores resultados que el anterior y es un poco más complejo.

#### **2.4.3 Muestreo Aleatorio Estratificado**

Este método trata de separar la población en estratos y luego toma una cantidad de individuos determinada de cada uno para realizarle un muestreo aleatorio simple. La muestra podrá tomarse en proporción al tamaño de cada estrato. Este método asegura que cada grupo quede representado de manera significativa en un posterior estudio.

#### **2.4.4 Muestreo aleatorio por conglomerados**

Este último método toma conglomerados que se encuentran en la realidad y estudia la totalidad de los elementos pertenecientes a ellos, por ejemplo si se necesita saber la opinión de la educación en una zona geográfica, se tomará como muestra los habitantes de esa zona y no a todo el país.

## 2.5 ESTUDIO DE MERCADO<sup>13</sup>

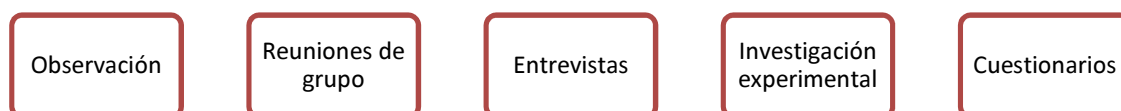
La investigación de mercado es fundamental en cualquier negocio, ya que busca conocer a los clientes actuales y a los potenciales. Esto es esencial para el desarrollo de toda la estrategia de marketing de la empresa, desde el desarrollo de los productos, hasta la definición de las promociones, la plaza y el precio.

Aquel que sea capaz de conocer mejor a sus clientes, podrá adquirir una ventaja competitiva considerable a los demás participantes de la industria. Es por esto que la investigación de mercados es el proceso de recolección y análisis de información, para ayudar la toma de decisiones de la empresa.

Este proceso consta de los siguientes pasos:

- **Definición del problema:** El entendimiento del problema es necesario para determinar los datos que se necesitan del mercado, el valor de esta información quedará reflejado por la eficacia de las decisiones tomadas. Los objetivos e hipótesis se elaboran en relación a la naturaleza del problema, que puede ser descriptivo, dónde se caracteriza la situación respecto de la frecuencia de aparición del fenómeno, o causal, que busca un nivel de explicación completo acerca de las causas de determinados fenómenos. Siempre que el problema sea bien definido, el costo de recoger la información será menor, “Un problema bien definido es un problema bien resuelto”.
- **Definición de los objetivos de la investigación:** Deben ser claros para definir con precisión el alcance que tendrá el trabajo y como se realizará. La clara definición de estos, se verá reflejada en los resultados, por lo tanto es necesario que estos objetivos justifiquen llevar a cabo el proyecto.
- **Desarrollo del plan de investigación:** Es el cómo se llevará a cabo el proyecto, definiendo costos, personal y tiempo. Lo importante de esta etapa es que ayudara a los ejecutivos a decidir si se justifica invertir en este conocimiento. Los datos pueden ser recogidos de diferentes maneras:

Figura 7: Formas de levantar información



Fuente: "Elaboración propia"

---

<sup>13</sup> Bernal, M. 2005

- **Levantamiento de la información:** Hay dos tipos de datos: primarios, que se recogen directamente de la fuente a un costo mayor y secundarios, que son constituidos por fuentes internas.
- **Análisis de la información:** Se debe realizar un análisis completo de toda la información levantada, tanto de datos primarios como secundarios. A partir de éstos se sacarán conclusiones, luego de tabularlos y calcular estadísticos descriptivos, además de la aplicación de métodos estadísticos más avanzados con la esperanza de encontrar información adicional.
- **Presentación de los resultados:** Es necesario exponerlos de la manera correcta, para que los conocimientos y la información encontrada no se pierda. Los resultados deben ser claros y precisos con implicancias directas a decisiones estratégicas del negocio.
- **Cronograma de ejecución y control:** Es necesario definir tiempos para cada una de las etapas a realizar y así cumplir con los plazos totales del trabajo. Esta se transforma en una herramienta de control durante la realización del trabajo.

### 3 DESARROLLO DE LA METODOLOGÍA

#### 3.1 ESTUDIO DE BIBLIOGRAFÍA Y MODELO DE NEGOCIO

Como punto de partida, se investiga en papers y libros sobre segmentación, tanto del punto de vista matemático como del comercial. Dentro de los temas a estudiar se encuentra todo lo relacionado con estudios de mercado, ya que este conocimiento es necesario para poder levantar información relevante y precisa y no cometer errores que luego bajen la calidad de los análisis o limiten la realización de otros.

Además estudia la realidad de la industria aseguradora, con el fin de conocer el lenguaje y particularidades de esta.

#### 3.2 ESTRUCTURACIÓN Y ANÁLISIS DE LA BASE DE DATOS

La empresa en este momento posee alrededor de 250 mil clientes, los que poseen uno o más productos. Los datos utilizados en este trabajo pertenecen a los clientes de seguros vida individual, la cual consta de 58.907 clientes. Éstos poseen por lo menos alguno de los cuatro productos de esta cartera, como: A.P.V., seguros de protección y seguros flexibles.

Como herramienta de apoyo informático se utilizó Access, para poder estructurar una base de datos única que concentre la información necesaria para el trabajo. La base está compuesta por las siguientes tablas:

- Pólizas existentes en la empresa
- Cruce de los clientes con el registro nacional de automóviles
- Coberturas respectivas a cada póliza
- Datos sociodemográficos y transaccionales de los rut asociados a ellas.

A partir de lo anterior, se realizó un análisis de las variables que constituían esta base de datos, para analizar su importancia dentro del estudio. Se tomó en cuenta si son relevantes para el negocio o si serán útiles estadísticamente para la segmentación. Finalmente se utilizaron las siguientes variables.

Tabla 3: Variables de la Base de Datos

Variables		
Categóricas		Continuas
Sexo	Cobertura Salud	Valor Cliente
Sector	A.P.V	Edad años
Cláusulas	Protección	Capital en invalidez
Contacto	Flexibles	Capital en Gastos Médicos
Cobertura Oncológica	productos	Nº personas
Cobertura Accidentes		Pago anual
Cobertura GM		Capital accidentes
Cobertura Invalidez		Capital fallecimiento

Fuente: "Elaboración propia".

Luego se desarrolló un estudio de los registros de los clientes, donde se eliminaron aquellos que no tenían los datos completos. Además de esto se identificarán los datos fuera de rango para cada una de las variables, que podrían bajar la calidad del estudio.

A continuación en la Tabla 4 se pueden observar las magnitudes de la limpieza realizada a los datos que se utilizarán. Además en anexo se encuentra un detalle de los criterios utilizados



Tabla 4: Tabla resumen de limpieza de base de datos

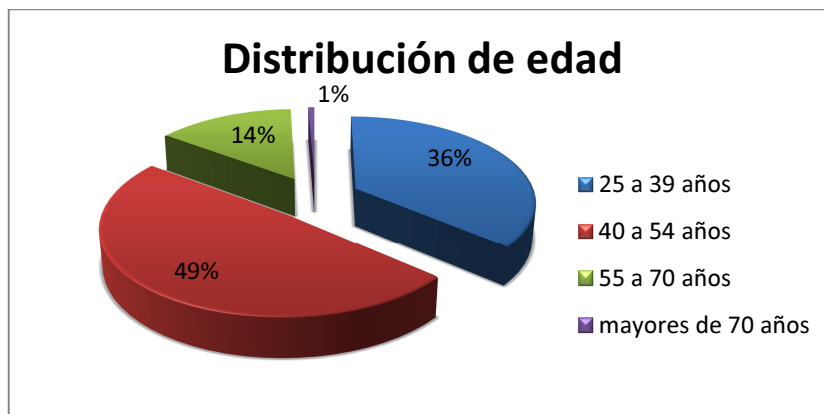
Tabla de resumen de limpieza de la base de datos			
Datos iniciales	Datos incompletos	%	Datos finales
58.097	7.345	12,64%	
	Datos fuera de rango		49,147
	656	1,13%	

Fuente: "Elaboración propia".

La base de datos en la cual se trabajará está compuesta por un 61% de hombres con una edad promedio de 44 años.

Además se puede observar en la figura 8, que la mitad de los clientes se encuentran en el tramo de 40 a 54 años y que existe una baja población en el tramo de mayores de 70 años, es más, sólo un 15% de los clientes son mayores de 55 años.

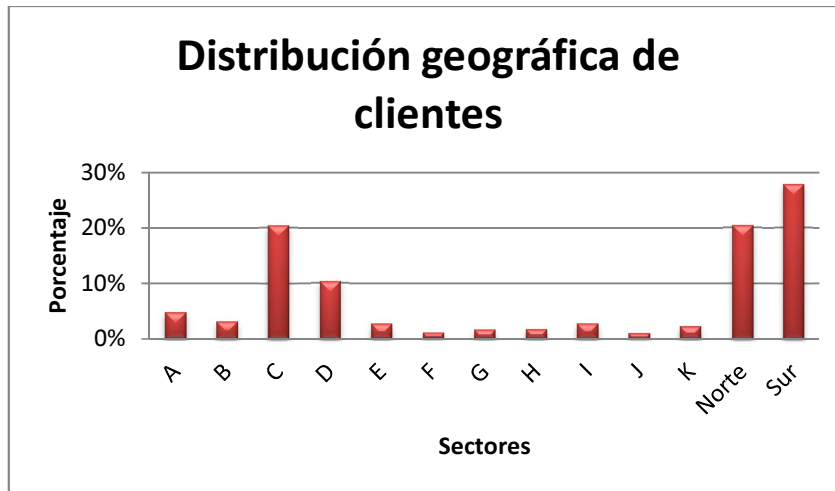
Figura 8: Distribución edad



Fuente "Elaboración propia"

La mayoría de los clientes son de Santiago (alrededor del 60%), es por esto que para el análisis se agrupó la zona norte y la zona sur del país. Además Santiago se separó de acuerdo a las áreas delimitadas por el Transantiago, como se puede apreciar en el Anexo A. También se puede observar una gran población en el sector C el cual consta de las comunas de Providencia, Las Condes, Vitacura y Lo Barnechea.

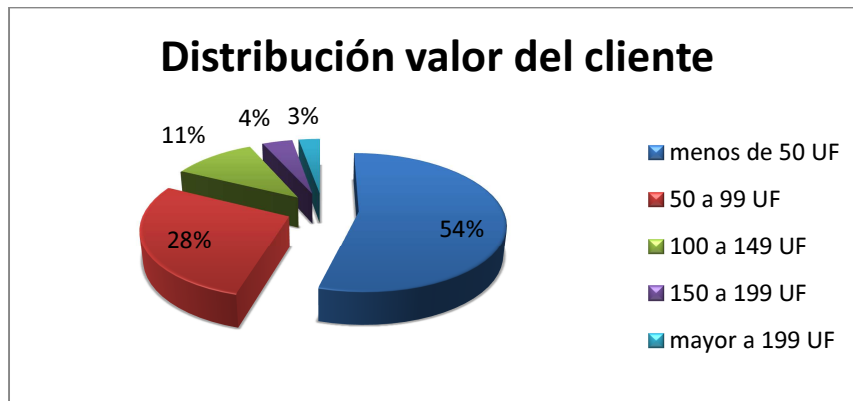
Figura 9: Distribución geográfica de clientes



Fuente "Elaboración Propia"

La organización le ha asignado a cada uno de sus clientes un valor, el cual tiene un promedio de 61 UF. En la figura 10 se puede apreciar la distribución del valor de los clientes de la compañía, en dónde la mayor parte no supera las 50 UF, además sólo un 3% supera las 200 UF y el máximo valor alcanza las 1.200 UF.

Figura 10: Distribución de valor del cliente



Fuente: "Elaboración propia"

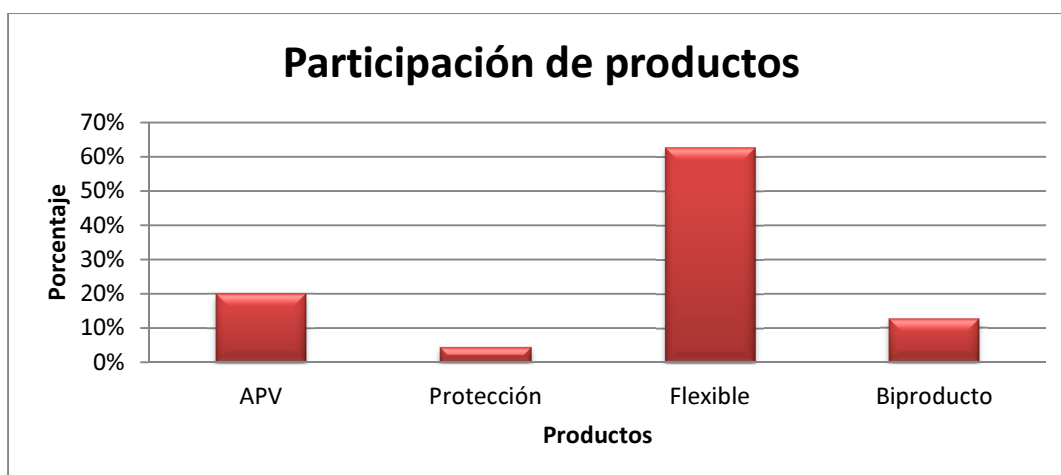
A continuación la figura 11 muestra la participación de los diferentes productos dentro de la cartera. Es importante notar la gran presencia de los seguros flexibles, los cuales durante los últimos años han aumentado notablemente su participación, causando la desaparición de algunos productos antiguos.

Los seguros flexibles, se caracterizan por tener una componente de ahorro, ya que pese a ser un poco más caras, las primas en comparación con los seguros de

protección, éstos permiten realizar retiros del capital invertido una vez que el cliente lo necesite; esta característica ha sido la que ha conquistado a muchos clientes que han dejado sus seguros de protección y han optado por los flexibles.

Se observa el porcentaje de personas del total de la población que ha contratado cada uno de los productos. El 20% de la base de datos sólo tiene contratado APV, alrededor del 5% tienen seguros de protección, mientras que un 63% tienen seguros flexibles. Es importante notar que alrededor de un 12% poseen más de un producto, y es aquí donde se presenta una oportunidad de negocio al hacer esfuerzos por que este grupo aumente su participación.

Figura 11: Participación de productos

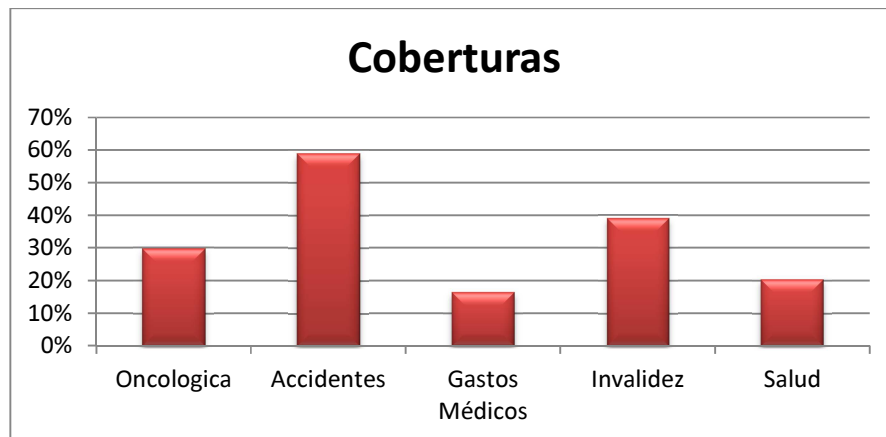


Fuente: "Elaboración propia"

Cada uno de los seguros antes descritos poseen diferentes coberturas, las que pueden complementar los productos que cada una de las personas, una forma de poder rentabilizar a los clientes sería ofrecerles más coberturas.

A continuación en la figura 12 se puede ver la participación que poseen cada una de las coberturas, dónde la más popular es la por accidentes, con casi un 60% de la población total. Con un 39% la segunda cobertura en participación es la de invalidez, con un 30% se encuentra la cobertura oncológica y cerca del 20 % posee las coberturas en gastos médicos y salud.

Figura 12: Participación de coberturas



Fuente: "Elaboración propia"

También es importante notar que el 60% de los clientes posee al menos 2 o 3 coberturas y sólo un 14% posee sólo la cobertura de fallecimiento. Otro aspecto a considerar es que existe sólo un 20% con más de 3 coberturas, de 6 posibles, lo que significa que los demás clientes tienen un alto potencial de crecimiento en este sentido. Lo anterior se puede observar en la figura 13, donde se muestra el detalle del número de coberturas que poseen los clientes.

Figura 13: Número de coberturas



Fuente: "Elaboración propia"

Como primer input para el análisis de los clientes es importante poder entender que diferencias existen entre ellos respecto del producto que poseen. De manera intuitiva se espera poder encontrar perfiles de clientes diferentes para cada producto.

Es importante notar diferencias entre aquellos clientes mono producto, que son la mayor parte de la base y los clientes multi producto, ya que el objetivo es lograr traspasar a los clientes que se encuentran en el primer grupo al segundo.

En la Tabla 5 y 6 se muestra la diferencia de los clientes según sus productos, en la primera se muestran los valores y en la segunda se observa la variación porcentual con respecto al promedio. Los clientes analizados en esta ocasión sólo poseen el producto señalado.

Tabla 5: Caracterización de clientes por productos

Caracterización de clientes por producto				
	APV	Protección	Flexible	
	Media	Media	Media	Población
Valor del Cliente	46,47	30,40	65,71	62
Sexo (Masculino)	66%	75%	58%	61%
Cláusulas	1%	17%	10%	9%
Contacto	43%	35%	46%	45%
Nº de coberturas	1,5	2,7	2,9	2,64
Tipo de Agente	60%	64%	34%	44%
Pago anual	36	24	27	33
Edad	43,07	42,64	44,41	43,9
Cobertura Oncológica	1%	18%	38%	30%
Cobertura Accidentes	1%	53%	75%	59%
Cobertura en Gastos Médicos	0%	31%	19%	16%
Cobertura en Invalidez	44%	71%	30%	39%
Cobertura Salud	0%	0%	28%	20%

Fuente: "Elaboración propia"

Los clientes de A.P.V. poseen menos coberturas que los que poseen otros productos, esto se puede explicar ya que el cliente está interesado en el ahorro y los beneficios tributarios que éste ofrece, más que por las coberturas; todos estos clientes se ven obligados a contratar una cobertura de fallecimiento, pero un 44% ha contratado cobertura de invalidez como cobertura adicional. Algunos además han contratado coberturas de accidentes y oncológica.

No se aprecia diferencia en las edades promedio de los distintos productos, especialmente en los clientes de APV que poseen una edad muy cercana al promedio.

El valor del cliente promedio por cada producto es bastante diferente, dónde aquellos que han contratado un seguro flexible son más valiosos que uno de APV, con una diferencia de un 30% respecto al promedio. Con los clientes de los seguros de protección existe una diferencia de más del 50%.

Un 75% de los clientes de protección son hombres, lo cual es un 7% sobre el promedio de la base de datos completa, además son los principales consumidores de las

coberturas en Gastos Médicos e Invalidez. En este segmento se puede encontrar un porcentaje mayor de individuos con cláusulas. Estos son clientes más atrapados por la compañía debido a sus condiciones físicas.

Por último los clientes de los seguros flexibles son los principales contratantes de las coberturas de accidentes, oncológicas y salud. Con esto se puede determinar que necesidades tratan de satisfacer con cada uno de los diferentes productos.

Este grupo es el de mayor edad promedio, pese a que la diferencia de edad es pequeña, es estadísticamente significativa.

Tabla 6: Lift entre los diferentes productos

	Lift		
	APV	Protección	Flexible
Valor del Cliente	-25%	-51%	7%
Sexo	8%	22%	-6%
Cláusulas	-90%	84%	8%
Contacto	-4%	-24%	1%
Nº de coberturas	-45%	4%	10%
Tipo de Agente	36%	47%	-23%
Pago anual	6%	-29%	-20%
Edad	-2%	-3%	1%
Cobertura Oncológica	-98%	-39%	28%
Cobertura Accidentes	-99%	-10%	28%
Cobertura en Gastos Médicos	-99%	93%	14%
Cobertura en Invalidez	15%	83%	-22%
Cobertura Salud	-100%	-99%	36%

Fuente: "Elaboración propia"

A continuación se realiza un análisis de los clientes que poseen dos productos, con el fin de entender que características los diferencian de aquellos que sólo poseen uno, para así poder diseñar estrategias en base a este conocimiento.

Cómo es de esperarse, los clientes multi-producto valen más para la empresa y esto se ve reflejado en los números, lo que confirma el interés de aumentar el tamaño de cada uno de estos conglomerados. Además se puede notar tendencia en que los hombres son más propensos a contratar más de un producto. Una de las características importantes a destacar, es que poseen en general un mayor número de cláusulas, lo que refleja una necesidad mayor por los productos que el resto de los clientes.

El grupo de clientes más valioso son los que poseen un APV y un seguro flexible. Este grupo de clientes se caracteriza por no tener demasiadas coberturas, lo cual se puede explicar por la característica del APV.

Tabla 7: Caracterización de clientes bi-producto

	APV- Flexible	APV- Protección	Protección- Flexible	Población
	Media	Media	Media	Media
Valor del Cliente	<b>80,33</b>	67,18	74,26	62
Sexo (Masculino)	63%	<b>79%</b>	76%	61%
Cláusulas	14%	18%	<b>27%</b>	9%
Contacto	52%	50%	50%	45%
Nº de coberturas	3,2	2,9	<b>3,6</b>	2,6
Tipo de Agente	65%	<b>73%</b>	57%	44%
Pago anual	72	<b>78</b>	64	33
Edad	43,23	42,89	<b>44,79</b>	43,9
Cobertura Oncológica	41%	19%	<b>45%</b>	30%
Cobertura Accidentes	74%	59%	<b>81%</b>	59%
Cobertura en Gastos Médicos	24%	27%	<b>35%</b>	16%
Cobertura en Invalidez	58%	<b>79%</b>	71%	39%
Cobertura Salud	25%	7%	<b>30%</b>	20%

Fuente: "Elaboración propia"

De la misma forma que se caracterizó a los clientes por su tenencia de productos, se realiza por la tenencia de coberturas. Es necesario conocer las razones que diferencian a los distintos clientes en base a necesidades de seguros.

Tabla 8: Caracterización de clientes por cobertura

Caracterización de clientes por cobertura						
	Oncológica	Accidentes	Gastos Médicos	Invalidez	Salud	Población
	Media	Media	Media	Media	Media	
Valor del Cliente	82,82	73,19	60,19	55,04	97,57	72
Sexo (Masculino)	56%	63%	60%	62%	60%	61%
Cláusulas	12%	11%	33%	14%	14%	14%
Contacto	49%	46%	45%	43%	51%	47%
Nº de coberturas	3,69	3,27	3,9	3,17	3,7	3,44
Nº personas	5,3	4,83	5,58	4,96	5,65	5,12
Tipo de agente	38%	37%	60%	63%	28%	44%
Pago anual	40,39	35,73	37,74	37,54	41,58	37,9
Edad	46,3	44,6	41,6	40,6	46,5	43,9

Fuente: "Elaboración propia"

En la tabla número 8 se pueden observar las principales diferencias entre las coberturas, donde destaca la cobertura de salud como la que posee mayor valor para la compañía. Una de las características que se puede apreciar de ésta, es que es un grupo de mayor edad, con 3 años más promedio.

La distribución de sexo para cada una de las coberturas es bastante homogénea, sólo en la cobertura oncológica se puede observar una mayor cantidad de mujeres.

### 3.3 REDUCCIÓN DE DIMENSIONES

Ante la gran cantidad de variables que se posee y la dependencia que existe entre ellas, es necesaria la aplicación de un análisis factorial, para resumir la información que se tiene en un menor número de variables y además que éstas sean independientes.

Se aplicó un análisis de componentes principales a las variables continuas de la base de datos. Con esto se obtuvieron 3 constructos que representan el 73% de la varianza. Y un índice de K.M.O. de 0,784 el cual indica que la calidad de este análisis es adecuado.

Tabla 9: Resultado análisis factorial

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		0,784
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	161888,586
	gl	28
	Sig.	0,000

Fuente: "Elaboración propia"

Para facilitar la interpretación de los constructos, se realizó una rotación de los ejes utilizando el método Varimax, con el cual se obtuvo los resultados que se muestra en la Tabla 12.



Tabla 10: Matriz de componentes rotados

Matriz de componentes rotados			
	Componente		
	Capital	Valor	Nº coberturas
Valor Cliente	0,119	0,838	0,271
Edad	-0,12	0,759	-0,033
Pago anual	0,529	0,588	-0,109
Capital Accidentes	0,818	0,105	0,266
Capital Fallecimiento	0,923	0,155	0,059
Capital en gastos médicos	0,638	-0,168	0,282
Capital en invalidez	0,856	-0,08	0,072
Nº personas	0,805	0,215	0,181
Suma coberturas	0,266	0,119	0,916

Fuente: "Elaboración propia"

Se puede concluir que el constructo número 1 representa los capitales de las diferentes coberturas; el segundo representa el valor de cada uno de los clientes, la edad y monto de pago anual; y el tercero representa la cantidad de coberturas que poseen.

### 3.4 DISEÑO DE LA ENCUESTA Y MUESTREO

Parte del trabajo consiste en levantar información que complemente los datos transaccionales. Para esto se realizó el diseño y aplicación de encuesta.

El foco de la encuesta está en recopilar información que permita caracterizar los estilos de vida de sus clientes, para así poder entenderlos, adaptarse a ellos y que aumente la información de los clientes para que se puedan clasificar adecuadamente.

En primera instancia, en la encuesta se realizan preguntas para determinar en qué etapa del ciclo de vida se encuentran, donde se puede observar preguntas como el estado civil y tamaño del grupo familiar. Esto junto con la edad nos da buenos indicios de qué tipo de necesidades posee y cuáles pueden ser cubiertas por la organización.

A continuación se realizan preguntas para determinar su nivel de ingresos, como el nivel de educación y la ocupación. Detectar esto es muy importante a la hora de caracterizar el potencial de desarrollo dentro de la empresa.

Uno de los aspectos que se busca identificar, son los medios de comunicación a los que son más sensibles. Dado que las tecnologías y las redes sociales se han masificado tanto, se pretende observar cómo ha influido esto en los clientes de la

compañía. Además conocer los canales adecuados, es importante a la hora de diseñar estrategias comerciales para diferentes grupos.

Para caracterizar los estilos de vida y complementar la información anterior se preguntará por la preferencia de actividades recreativas para determinar gustos y personalidades de los clientes. Esto permitirá ofrecer incentivos más adecuados al tener un pequeño indicio del perfil psicográfico de los individuos.

Por último se llevan a cabo 2 preguntas por otros productos de la compañía, para poder identificar si es que no tiene el producto o lo tiene en la competencia. A través de toda esta información se espera poder llevar a cabo una descripción precisa del cliente.

Se realizará un muestreo aleatorio simple con el fin de obtener una muestra significativa del universo de clientes que sea representativa del resto de la población.

Para esto se utilizará la fórmula:

$$\varepsilon = z_{1-\alpha/2} \sqrt{p(1-p)} \sqrt{\frac{1}{n} - \frac{1}{N}} \quad \text{Ecuación 14}$$

Dónde:  $\varepsilon$  = error de la muestra

$z_{1-\alpha/2} = 1,96 \approx 2$  = Nivel de confianza al 95%

N=tamaño de la población

n=Tamaño de la muestra

P=probabilidad de un evento en una distribución binomial.

Haciendo un supuesto de máxima varianza ( $p = 0,5$ ) y N muy grande, en este caso cincuenta mil, la formula queda de la siguiente manera.

$$\varepsilon = \frac{1}{\sqrt{n}} \quad \text{Ecuación 15}$$

Por lo tanto para un error del 3%, al 95% de confianza se necesita una muestra de 1.000 clientes para un universo de 50.000.

Este procedimiento se realiza con la herramienta de planillas Excel, haciendo uso de la variable aleatorio y ordenar.

### 3.5 RESULTADO DE LA ENCUESTA

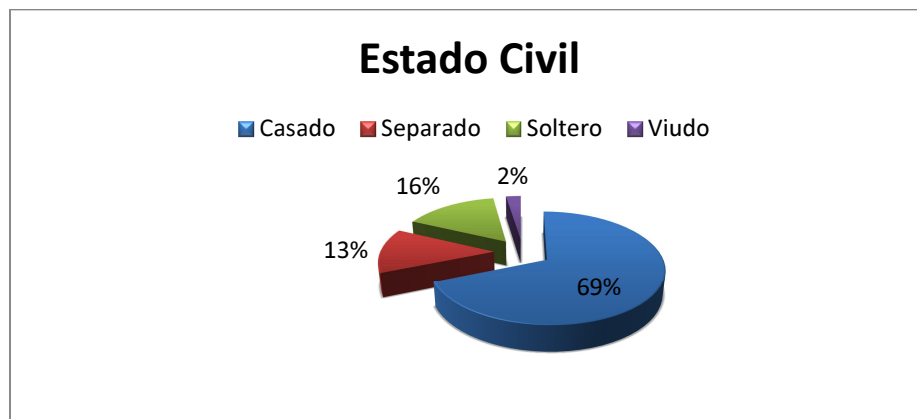
La encuesta fue realizada desde el Lunes 9 de Mayo hasta el Viernes 24 de Junio, Durante este periodo se intentó contactar a 959 clientes de los cuales sólo 246 contestaron, esto corresponde a una tasa de respuesta del 26%.El error muestral para

cada uno de los conjuntos analizados, será el explicado en la sección anterior, donde 246 clientes corresponde a un error  $\epsilon$  del 6%.

- Resultados de la encuesta

Se puede apreciar que la base de datos está compuesta por un 70% de personas casadas, lo cual nos muestra que los productos a entregar por la empresa son en gran parte enfocados en la familia, con solo un 16% de personas solteras.

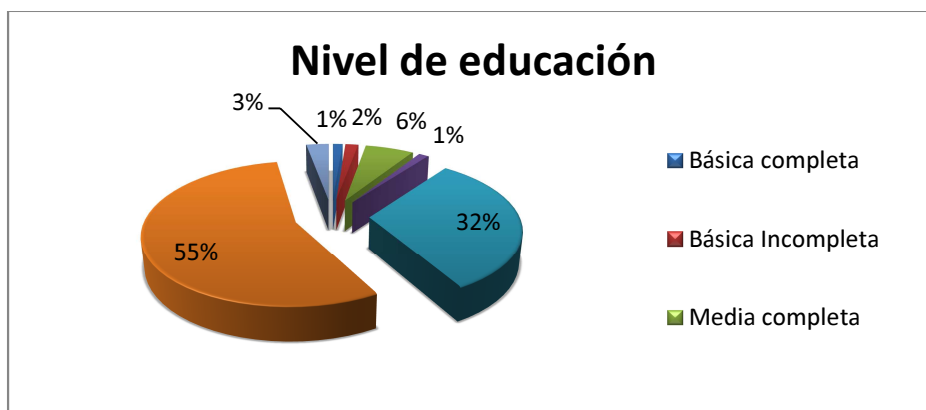
Figura 14: Estado civil



Fuente: "Elaboración propia"

A partir de la Figura 15 se puede apreciar que la mayoría de los clientes son profesionales, un 32% posee un título de postgrado y sólo un 12% no posee título profesional, lo cual comprueba la hipótesis anterior de la alta capacidad económica de estos.

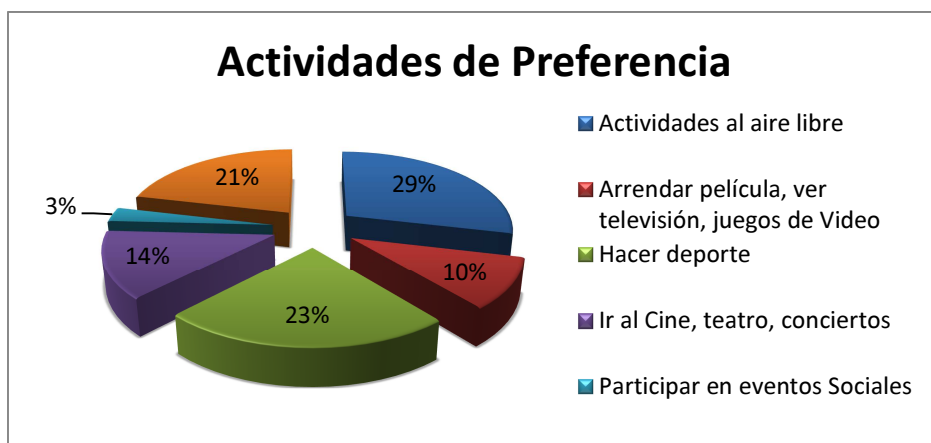
Figura 15: Nivel de educación



Fuente: "Elaboración propia"

A continuación se aprecia la preferencia de las distintas actividades, donde todas tuvieron cierta adhesión a excepción de los eventos sociales que solo un 3% prefirió. La actividad preferida por los clientes fueron las realizadas al aire libre, seguidas por hacer deporte con un 23% y actividades nocturnas como salir a comer, ir a un pub o bailar con un 21%. Cabe destacar que un 14% a preferido actividades como el cine, teatro o conciertos, y un 10% destaca entre sus actividades preferidas aquellas que se realizan en el hogar como arrendar películas, ver televisión y jugar videojuegos.

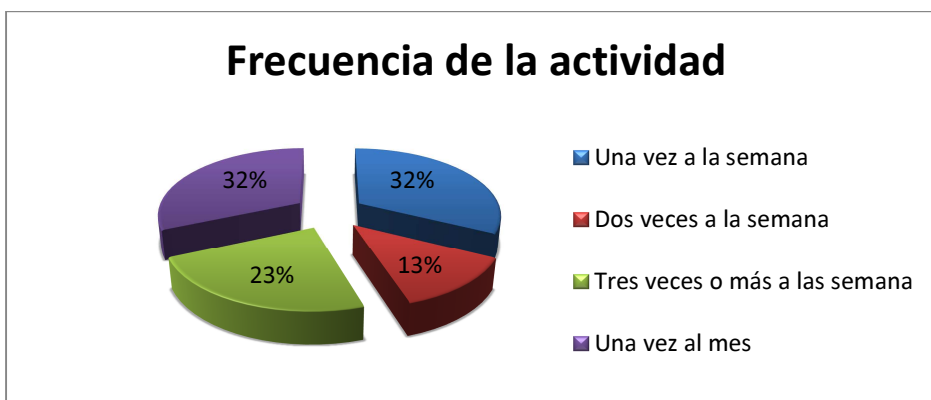
Figura 16: Actividades de preferencia



Fuente: "Elaboración propia"

Respecto de la frecuencia con que realizan las actividades antes mencionadas, se puede ver que un 32% las realiza una vez al mes o menos, lo que es una frecuencia bastante baja. En segundo lugar como los menos activos está un 32% que realizan esta actividad una vez a la semana y en tercer lugar los que realizan la actividad dos veces a la semana con un 13%. Por último se encuentra el grupo más activo que realiza la actividad tres o más veces a la semana con un 23% de la población.

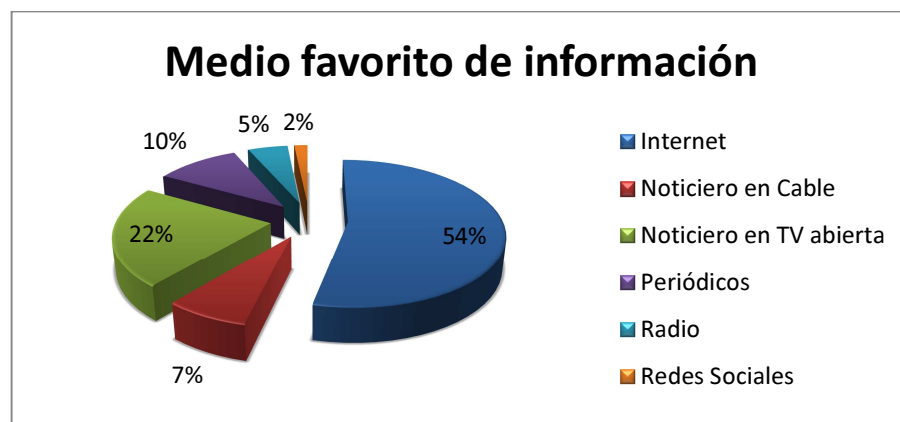
Figura 17: Frecuencia de la actividad



Fuente: "Elaboración propia"

Al preguntar por los medios de comunicación se puede observar que las redes sociales y la radio son poco populares, ya que entre ambos sólo logran un 7%, también con un 7% se encuentran los noticieros especializados que se pueden sintonizar en el cable. Con un 10% se encuentran los periódicos y los noticieros en tv abierta con un 22% de las preferencias. El medio favorito es el internet con un 54%, lo ocupan como su medio principal para informarse, lo cual entrega gran importancia a los esfuerzos de la organización por mejorar el servicio a través de este medio.

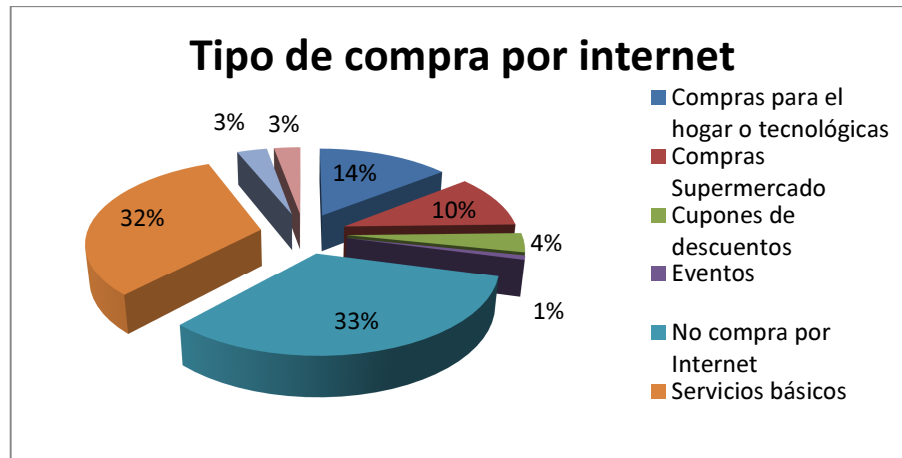
Figura 18: Medio favorito de información



Fuente: "Elaboración propia"

Es importante notar que pese al gran uso que tienen los clientes del internet, existe un 33 % que no hace compras de ningún tipo a través de este útil medio de compra. De los que lo ocupan la gran mayoría lo utiliza para el pago de los servicios básicos con un 32%. Un 14 % lo utiliza para compras del hogar y sólo un 10 % para las compras de los supermercados, el resto de las opciones de compra por internet fueron seleccionadas sólo marginalmente.

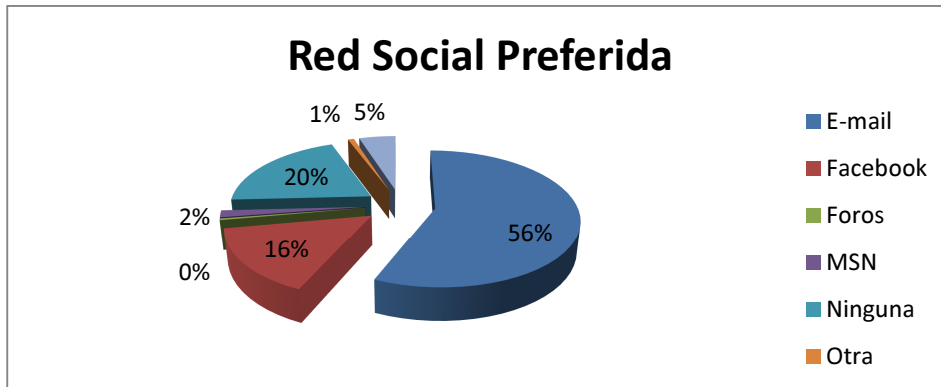
Figura 19: Tipo de compra por internet



Fuente: "Elaboración propia"

Dentro del uso de las redes sociales se puede destacar dos principalmente, Facebook y el mail, con un 16% y 56% respectivamente. Existe un 20 % que no utiliza ninguna red social.

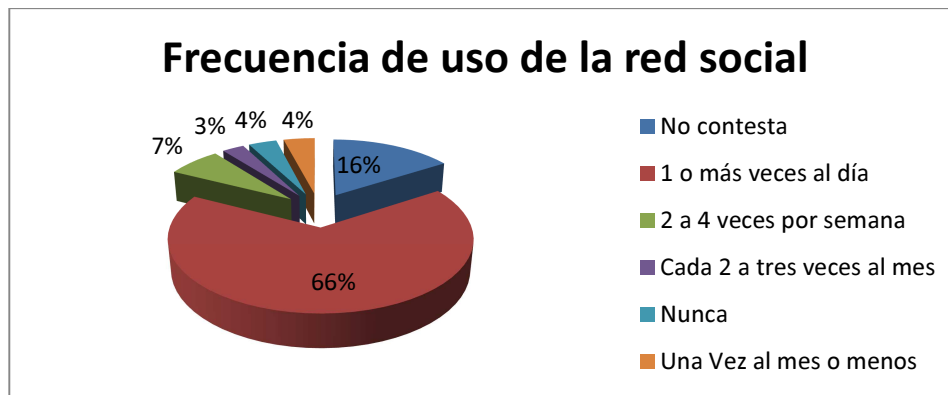
Figura 20: Red social preferida



Fuente: "Elaboración propia"

De los que utilizan las redes sociales, se puede observar que la frecuencia de uso es alta, ya que el 66% la utiliza por lo menos una vez al día.

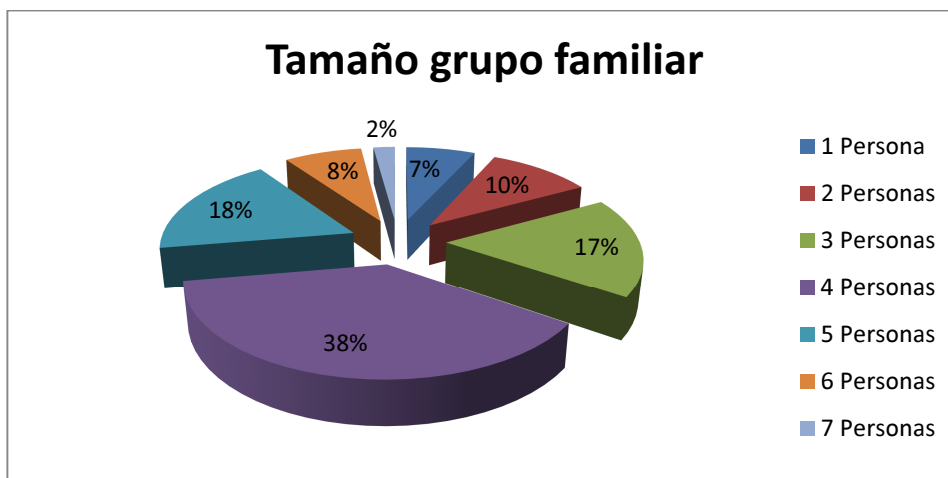
Figura 21: Frecuencia de uso de la red social



Fuente: "Elaboración propia"

Se puede observar en la siguiente Figura que el 38% de la muestra posee un grupo familiar de 4 personas, el 73% son familias de entre 3 a 5 personas y sólo 17% tiene un grupo familiar de menos de 3 personas.

Figura 22: Tamaño grupo familiar



Fuente: "Elaboración propia"

Dentro de los productos fuera de la empresa, se puede observar presencia importante de algunas entidades como, Banco de Chile, BCI, Cuprum, ING, Santander y Security. La mayoría de estas instituciones son bancos, los cuales probablemente ofrecen APV. Dentro de las aseguradoras destaca Ing.

Un 30% posee un APV fuera de la empresa, 14% un seguro de protección y un 3% flexibles fuera de la empresa, además un 18% posee un seguro de algún tipo fuera de la compañía.

### 3.6 SEGMENTACIÓN PRELIMINAR

Esta etapa tiene una finalidad descriptiva, en donde se podrá apreciar los distintos tipos de clientes que se encuentran en la base de datos y caracterizarlos apropiadamente. Esto será fundamental para la comprensión de los clientes a los que se enfrenta la compañía.

En esta fase se realizarán diferentes métodos de segmentación, y se pondrán a prueba métodos supervisados y no supervisados. Dentro de los no supervisados se ha utilizado k-medias y el conglomerado en dos fases. Por su parte, en los supervisados se utilizarán los siguientes árboles de clasificación: Chaid y Quest.

El fin es comparar los diferentes algoritmos, elegir la clasificación de mayor calidad estadística y que al mismo tiempo resuelva el problema de negocio que se busca solucionar.

Para validar cada una de las siguientes segmentaciones se ha utilizado el método de análisis discriminante, con el cual se busca determinar si existe diferencia o no entre los diferentes grupos, encontrar si por cada variable las medias son estadísticamente diferentes y por último el porcentaje de clasificación comparado con el grupo inicial.

A continuación se muestra los resultados de cada uno de los métodos antes mencionados.

#### 3.6.1 C.H.A.I.D

El primer método supervisado a aplicar será éste, en donde la variable a predecir que se utiliza es la variable producto. Ésta es una variable categórica que representa la posesión de los diferentes productos que tiene cada cliente. Los valores que puede tomar esta variable se pueden apreciar en la Tabla 13.

Tabla 11: Variable productos

Variable producto	
Valor	Producto
1	Sólo APV
2	Sólo Protección
3	Sólo Flexible
4	APV-Protección
5	APV-Flexible
6	Protección Flexible
7	Todos los productos

Fuente: "Elaboración propia"



Esta variable fue definida a partir de las necesidades del negocio, el que busca poder caracterizar a los poseedores de cada producto para determinar a los clientes potenciales de cada uno.

Para el desarrollo de este árbol se introdujeron todas las variables, ya que este determinará cuáles son las más importantes por sí solo. El resultado obtenido fue un árbol de tres niveles con 8 nodos finales, representando los segmentos encontrados. Las variables utilizadas fueron los tres constructos derivados del análisis factorial (capital, valor, número de coberturas), además de cobertura de accidentes, cobertura invalidez y tipo de agente.

En la Tabla 14, se puede observar el número de casos por cada segmento. Es importante resaltar que el segmento de mayor población es el conglomerado 1 que representa el 26% de la población, mientras que el menor consta del 3% de ésta. En este aspecto se logran clúster de tamaños adecuados desde el punto de vista del negocio.

Tabla 12: Resultado CHAID

Número de casos en cada conglomerado		%	
Conglomerado	1	12666	26%
	2	4861	10%
	3	5443	11%
	4	6046	12%
	5	5099	10%
	6	11784	24%
	7	1725	4%
	8	1523	3%
Válidos	49.147	100%	
Perdidos	-	0%	

Fuente: "Elaboración propia"

Se validó la segmentación utilizando análisis discriminante. En la Tabla 15 se aprecia la capacidad de reclasificación, muy cercana al 70%, lo cual es bajo, ya que existe un 30% mal clasificado.

Tabla 13 : Capacidad de predicción para CHAID

	Clúster	Grupo de pertenencia pronosticado								Total
		1	2	3	4	5	6	7	8	
%	1	87	-	2	0	0	-	7	4	100,0
	2	-	76	0	0	-	-	6	17	100,0
	3	27	14	48	1	1	1	2	5	100,0
	4	10	15	-	55	0	2	1	17	100,0
	5	11	5	1	-	51	30	2	1	100,0
	6	0	1	3	4	16	74	1	0	100,0
	7	14	6	0	-	0	0	61	18	100,0
	8	2	3	2	10	0	2	17	64	100,0
Se han clasificado 69,3% de los datos agrupados originales										

Fuente: "Elaboración propia"

Es importante revisar los Lambdas de Wilks de las distintas funciones discriminantes para ver si están realmente separados los diferentes segmentos y poder decidir si esta clasificación es válida estadísticamente.

En la Tabla 16 se aprecia que los p-valor de todas las funciones discriminantes es igual a cero, por lo que se determina que existen diferencias estadísticamente significativas entre los diferentes conglomerados. Pero su calidad no de buena calidad debido al porcentaje de reclasificación.

Tabla 14: Lambda de Wilks para CHAID

Lambda de Wilks				
Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 7	0,087	119810	77	0,000
2 a la 7	0,245	69104	60	0,000
3 a la 7	0,493	34712	45	0,000
4 a la 7	0,674	19399	32	0,000
5 a la 7	0,845	8250	21	0,000
6 a la 7	0,960	2019	12	0,000
7	0,983	857	5	0,000

Fuente: "Elaboración propia"

En consecuencia, con este árbol de clasificación se han obtenido conjuntos de clientes de tamaños apropiados para futuros análisis desde el punto de vista del negocio, pero un 30% de los datos han sido clasificados de manera diferente por el análisis discriminante, lo que baja la calidad del análisis.

### 3.6.2 QUEST

Se utilizará la variable producto como dependiente y el resto serán introducidas para que el algoritmo tenga la libertad de elegir cuál de estas son las que disgregan de mejor manera la población.

En este caso el algoritmo seleccionó las siguientes variables para segmentar: El constructo número uno, cobertura de accidentes, cobertura salud y cobertura oncológica, lo cual lo diferencia de las dos clasificaciones anteriores, que habían escogido todos los constructos del análisis factorial antes realizados.

Producto de éste se consiguió 8 conglomerados, que se pueden apreciar en la Tabla 17 que muestra el número de casos en cada uno de los grupos. En esta clasificación se puede ver que el clúster 1 posee casi la mitad de la población, es más, entre dos de sus 8 grupos se puede observar que se encuentra el 80 % de la población.

Desde el punto de vista del negocio, se puede notar que existen por lo menos dos grupos importantes respecto a su tamaño, en los que podrían realizarse acciones comerciales. Los demás son muy pequeños y a menos que sean muy importantes para la empresa no serán significativos para futuros análisis.

Tabla 15: Resultados QUEST

Número de casos en cada conglomerado		%	
Conglomerado	1	23717	48,3%
	2	2766	5,6%
	3	887	1,8%
	4	1646	3,3%
	5	16587	33,7%
	6	786	1,6%
	7	2437	5,0%
	8	321	0,7%
Válidos	49.147	100%	
Perdidos	-	0%	

Fuente: "Elaboración propia"

De acuerdo a lo realizado a cada uno de los métodos antes aplicados, se efectuará un análisis discriminante que permita validar los resultados encontrados. El análisis aplicado en esta ocasión separó la muestra a través de 7 funciones discriminantes, las cuales predijeron de igual manera el 83% de los casos previamente clasificados con este árbol. Este resultado es de mejor calidad que los encontrados en los arboles anteriores y da indicios de que estos segmentos son realmente diferentes.

Tabla 16: Capacidad de predicción QUEST

Terminal Node Identifier	Grupo de pertenencia pronosticado								Total
	1	2	3	4	5	6	7	8	
% 1	81	4	-	-	3	0	11	0	100
2	11	70	-	0	4	2	3	10	100
3	-	19	65	2	-	3	0	11	100
4	0	21	8	46	1	10	-	15	100
5	3	2	-	-	93	2	-	-	100
6	0	12	1	8	11	67	-	-	100
7	3	1	-	-	-	0	90	5	100
8	1	6	4	10	0	5	7	67	100

Se han clasificado 83,1% de los datos agrupados originales

Fuente: "Elaboración propia"

Al concluir el análisis estadístico de este algoritmo, revisaremos los lambda de Wilks para cada una de las funciones discriminantes. En la Tabla 19 se distingue que para todas ellas salvo para la última son significativas, lo que revela la buena calidad del análisis.

Tabla 17: Lambda de Wilks para QUEST

Lambda de Wilks					
Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.	
1 a la 7	0,09	118214	77	0,000	
2 a la 7	0,29	60235	60	0,000	
3 a la 7	0,65	21117	45	0,000	
4 a la 7	0,85	8001	32	0,000	
5 a la 7	0,94	2853	21	0,000	
6 a la 7	1,00	246	12	0,000	
7	1,00	10	5	0,075	

Fuente: "Elaboración propia"

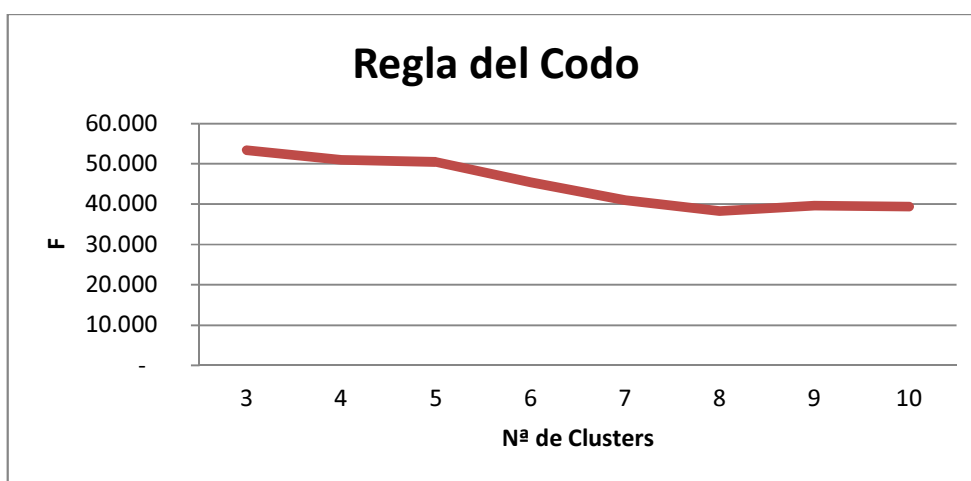
En síntesis, éste último fue el que mejor se comportó estadísticamente hablando, dentro de los métodos supervisados, pero se pierde calidad en el tamaño de cada uno de los conglomerados.

### 3.6.3 K- medias

Este método a diferencia de los demás sólo acepta variables numéricas, por lo tanto se utilizarán sólo los constructos antes encontrados a través del análisis de componentes principales.

Para determinar el número óptimo de conglomerados se utilizó la regla del codo a partir de la distancia a los centros para cada número de conglomerados, como se puede apreciar en la Figura 23, donde se observa la inflexión cuando el número de clúster es igual a 8. En consecuencia se realizó una segmentación con 8 conglomerados.

Figura 23: Regla del codo K-Medias



Fuente: "Elaboración propia"

Una vez elaborada la segmentación, se procede a analizar el número de casos en cada uno de los grupos encontrados. En la Tabla 20 se puede observar como el conglomerado más pequeño posee apenas 1 registro; además el conglomerado de mayor envergadura es el número 5 que representa el 53% de la población. Por otro lado tres de los 8 conglomerados, suman el 93% del universo. Estos resultados son menos atractivos desde el punto de vista del negocio debido a lo pequeño de 5 de los ocho grupos, ya que para estos puede que no valga la pena desarrollar una estrategia comercial. Por otra parte la mitad de la población se encuentra en un solo grupo, lo cual tampoco es lo óptimo, ya que se busca localizar conglomerados con tamaños no tan diferentes como lo ocurrido en este caso.

Estos resultados son de muy baja calidad en comparación a los antes encontrados con los métodos supervisados. Los tamaños no recomiendan la utilización de este algoritmo para la clasificación.

Tabla 18: Resultado K-Medias

Número de casos en cada conglomerado		%	
Conglomerado	1	74	1,5%
	2	776	1,6%
	3	1.833	3,7%
	4	11.412	23,2%
	5	26.260	53,4%
	6	1	0,0%
	7	201	0,4%
	8	7.918	16,1%
Válidos		49.147	100,0%
Perdidos		-	0,0%

Fuente: "Elaboración propia"

Por esto se analizará desde el punto matemático la calidad de esta segmentación. De igual manera que el caso anterior se utilizará un análisis discriminante para saber cuan efectivo fue el método desarrollado.

Tabla 19: Capacidad de predicción K-Medias

Resultados de la clasificación <sup>b,c</sup>											
	Número inicial de casos	Grupo de pertenencia pronosticado								Total	
		1	2	3	4	5	6	7	8		
%	1	90,5	,0	2,4	,0	,0	,0	,0	,0	7,1	100,0
	2	,3	86,9	2,1	10,1	,0	,0	,8	,0	,0	100,0
	3	1,9	,3	93,7	,6	1,5	,0	,4	1,6	100,0	
	4	,1	,2	,9	87,5	4,0	,0	,0	7,3	100,0	
	5	,0	,0	,8	2,0	95,0	,0	,0	2,2	100,0	
	6	,0	,0	,0	,0	,0	100,0	,0	,0	100,0	
	7	1,0	,0	1,5	,0	,0	,0	97,5	,0	100,0	
	8	,9	,0	1,2	14,0	2,0	,0	,0	81,9	100,0	
Se han clasificado 90,9% de los datos agrupados originales											

Fuente: "Elaboración propia"

En la Tabla 21 se puede observar el producto de la reclasificación a partir del análisis discriminante. Donde se obtuvo una tasa del 90.9% de eficacia, lo que indica que estadísticamente es un modelo correcto y que se comporta de buena manera.

Por otra parte es necesario analizar si los grupos son realmente diferentes. A partir del estadístico Lambda de Wilks se refleja que parte de la varianza total es representada

por la varianza intragrupo. Como se puede apreciar en la Tabla 22 todas las funciones discriminantes (a excepción de la número 7) muestran que los grupos son diferentes entre sí, esto debido a que cada uno de los p-valor son igual cero, lo que hace rechazar la hipótesis nula de que cada uno de los grupos son iguales. El hecho de que una de las funciones discriminantes no clasifique de la mejor manera, puede ser a causa de que uno de los segmentos posee solo un registro, lo cual no es significativo para aplicar una diferencia estadística con un grupo.

Tabla 20: Lambda de Wilks K-Medias

Lambda de Wilks				
Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 7	,026	179.167	84	,000
2 a la 7	,095	115.748	66	,000
3 a la 7	,260	66.273	50	,000
4 a la 7	,677	19.169	36	,000
5 a la 7	,873	6.679	24	,000
6 a la 7	,947	2.696	14	,000
7	1,000	3	6	,756

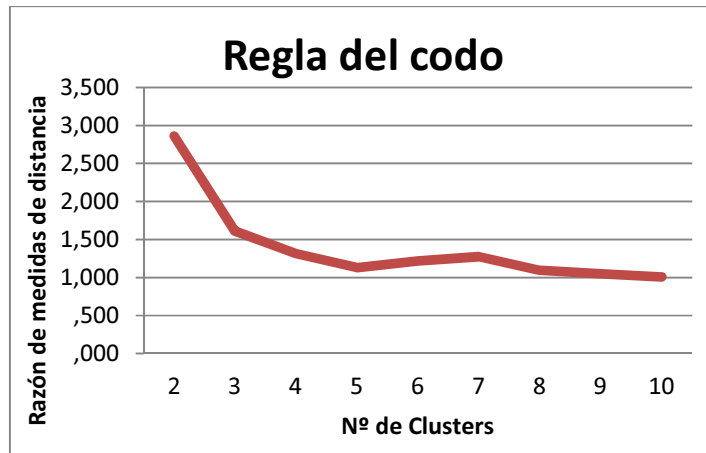
Fuente: "Elaboración propia"

En conclusión, desde el punto de vista del negocio esta segmentación no es adecuada debido al tamaño de los clúster encontrados, cabe destacar que estadísticamente el método ha resultado y ha encontrado 8 conglomerados significativamente diferentes.

### 3.6.4 Conglomerado en dos fases

A partir del criterio BIC el algoritmo decidió que el número óptimo de grupos es dos, como este número no es útil desde el punto de vista del negocio se determinó buscar el número ideal de forma manual a partir del mismo criterio. En la figura 24, se puede observar el cambio de pendiente producido en el número 5, es por esto que se resolvió realizar este método con 5 clúster.

Figura 24: Regla del codo Conglomerado en dos fases



Fuente: "Elaboración propia"

La segmentación realizada permite observar a través de la Tabla 21, que los tamaños de los grupos no varían demasiado, donde el grupo con menor cantidad de elementos es el número 4 con un 8% de la población, mientras que el mayor es el 3 con un 32%. Esta distribución es bastante buena para realizar acciones comerciales en el futuro si es que estos grupos son realmente diferentes.

Tabla 21 : Resultado de Conglomerado en dos fases

Distribución de conglomerados			
		N	% del total
Conglomerado	1	8590	17%
	2	9591	20%
	3	15593	32%
	4	4165	8%
	5	11208	23%
	Combinados	49147	100%
Total		49147	

Fuente: "Elaboración propia"

En la Tabla 22 se puede apreciar la calidad de la clasificación, en donde se logró clasificar a través del análisis discriminante el 98,3% de los datos agrupados originalmente mediante el conglomerado en dos fases.



Tabla 22: Capacidad de predicción Conglomerado en dos fases

Resultados de la clasificación <sup>a</sup>							
	Número de conglomerados en dos fases	Grupo de pertenencia pronosticado					Total
		1	2	3	4	5	
%	1	99,9	,0	,1	,0	,0	100,0
	2	,0	100,0	,0	,0	,0	100,0
	3	,0	,0	100,0	,0	,0	100,0
	4	4,2	9,3	2,7	79,5	4,3	100,0
	5	,0	,0	,0	,3	99,7	100,0
Clasificados correctamente el 98,3% de los casos agrupados originales							

Fuente: "Elaboración propia"

Se puede apreciar en la Tabla 23 el contraste de las funciones discriminantes, para todas ellas el p-valor es 0, lo que significa que se rechaza la hipótesis nula de que los segmentos son iguales, por lo que se concluye que efectivamente cada uno de los grupos posee diferencias estadísticamente significativas.

Tabla 23: Lambda de Wilks de Conglomerado en dos fases

Lambda de Wilks				
Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 4	,001	326363,020	44	,000
2 a la 4	,030	171643,224	30	,000
3 a la 4	,124	102668,527	18	,000
4	,404	44502,543	8	,000

Fuente: "Elaboración propia"

### 3.7 DESCRIPCIÓN DE CONGLOMERADOS

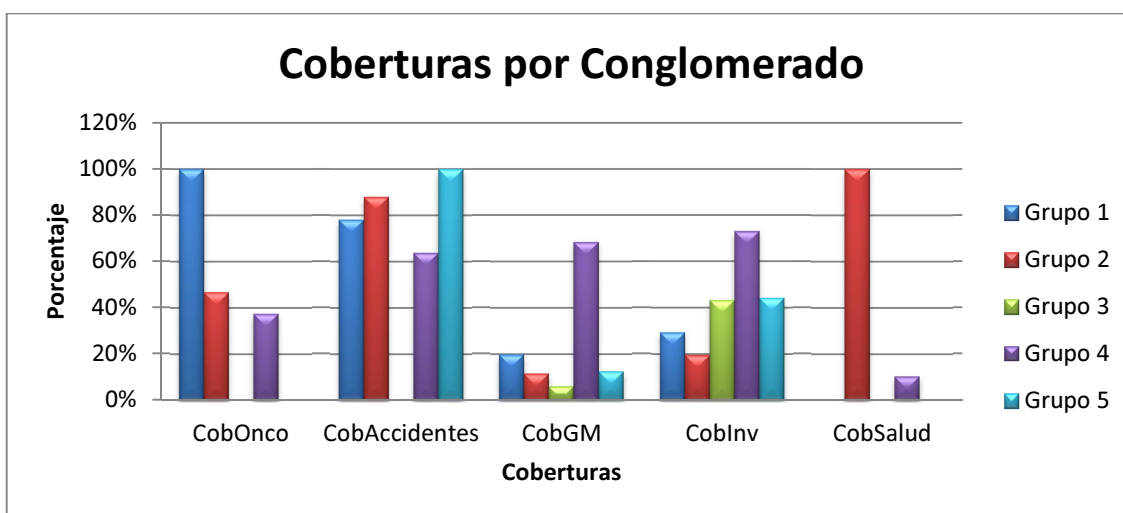
A continuación se presenta una caracterización de cada uno de los conglomerados encontrados. De las 5 segmentaciones antes explicadas se seleccionará sólo la mejor de éstas para su análisis de segmentos, se tomará la de mayor calidad estadística y de tamaños más útiles para el negocio.

Se eligió el conglomerado en dos fases, ya que su predicción fue ampliamente la de mayor calidad, tanto estadísticamente como por el tamaño de sus conglomerados.

### Clúster 1 (preocupados por el cáncer):

La principal característica de este conglomerado es su preocupación por el cáncer, ya que todos poseen cobertura por este mal, siendo que sólo un 30 % de la población posee esta cobertura. Es un grupo valioso para la empresa ya que su valor asignado promedio es de 70 UF (superior a las 61 UF) que es el promedio de la población. La cantidad de cláusulas es muy baja y en promedio poseen más de 3 coberturas, lo que es bastante para la realidad de esta población.

Figura 25: Coberturas por conglomerado



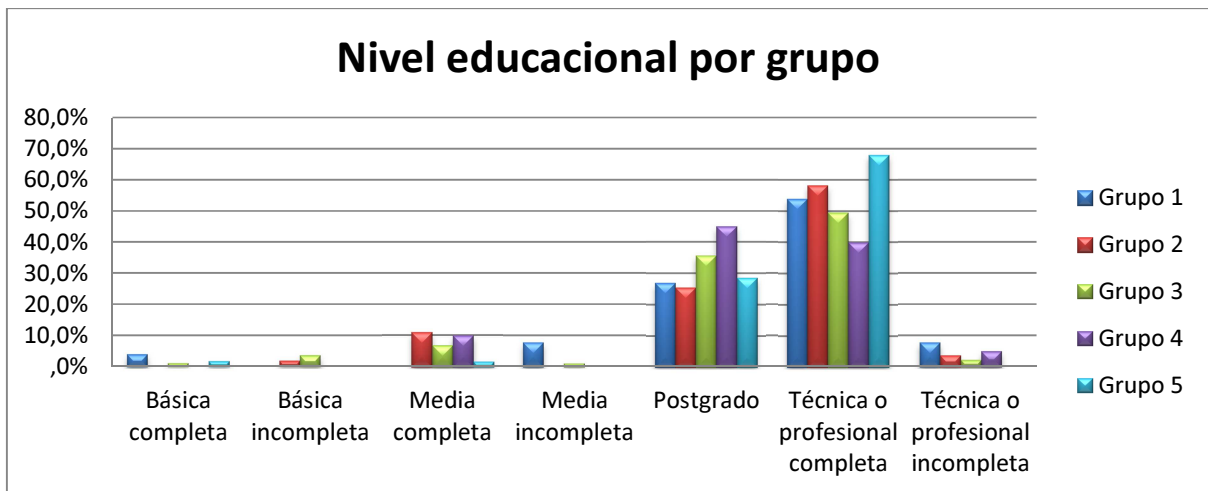
Fuente: "Elaboración propia"

### Descripción a partir de datos encuesta $\varepsilon = 20\%$

Es un grupo compuesto por personas que poseen un grupo familiar reducido, destacando que un 10% define el tamaño de su grupo familiar como un individuo y prácticamente no participan en los grupos familiares de más de cinco integrantes.

Se puede observar que es el con menor nivel de estudios, ya que predomina en todos los niveles menores al profesional y es uno de los grupos con menor porcentaje de postgrado. Esto se puede traducir en que es el grupo con menor poder adquisitivo de los analizados en esta sección.

Figura 26: Nivel de educación de cada segmento



Fuente: "Elaboración propia"

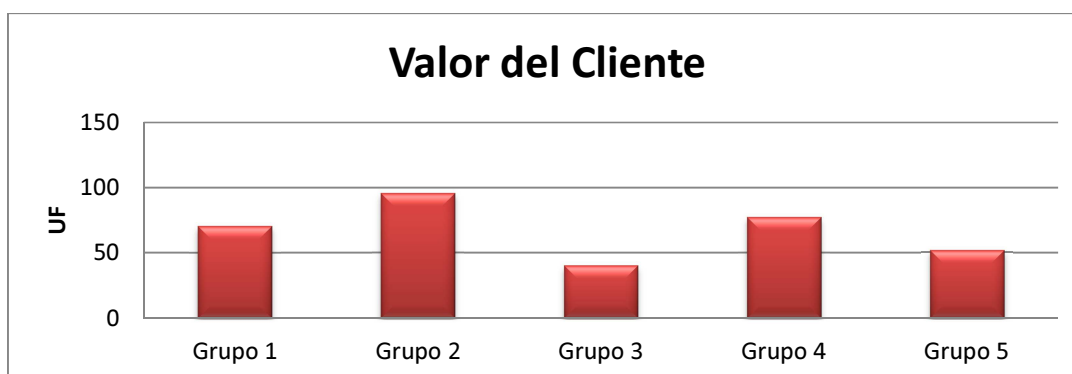
Dentro de sus actividades preferidas se encuentran las realizadas al aire libre y el deporte, donde la mayoría las realiza dos veces a la semana.

Entre los medios favoritos para informarse se pueden mencionar los noticieros de la TV abierta y la radio; un gran porcentaje de estos no utiliza ninguna red social.

Es importante mencionar que entre los grupos éste es el que tiene más clientes que no posee APV en la competencia, con un 88,5%, y donde solo un 7,7% si los posee.

**Clúster 2 (Gold):** La característica principal de este segmento es su alto valor, que se muestra en la Figura 27, ya que el promedio de éste es de 95 UF, superior al resto de los conglomerados. Además se caracteriza por ser prácticamente el único cliente de la cobertura de salud, lo que puede explicar el alto valor del segmento. Por otra parte, un 40% de este segmento posee cobertura en oncología. Su relación con la compañía es principalmente a través de un agente de mantención, ya que el 75% tiene asignado uno de estos agentes. Este segmento es el que posee la mayor edad promedio, superando los 46 años.

Figura 27: Valor del cliente de cada grupo



Fuente: "Elaboración propia"

### Descripción a partir de datos encuesta $\varepsilon = 13\%$

Posee un grupo familiar más numeroso que los otros grupos, donde destaca que un 13 % tiene 6 integrantes en su familia. Es el con menor porcentaje de clientes con postgrado, pero la cantidad de profesionales es alta. Se puede decir que es de alto poder adquisitivo debido a que casi un 85 % por lo menos son profesionales.

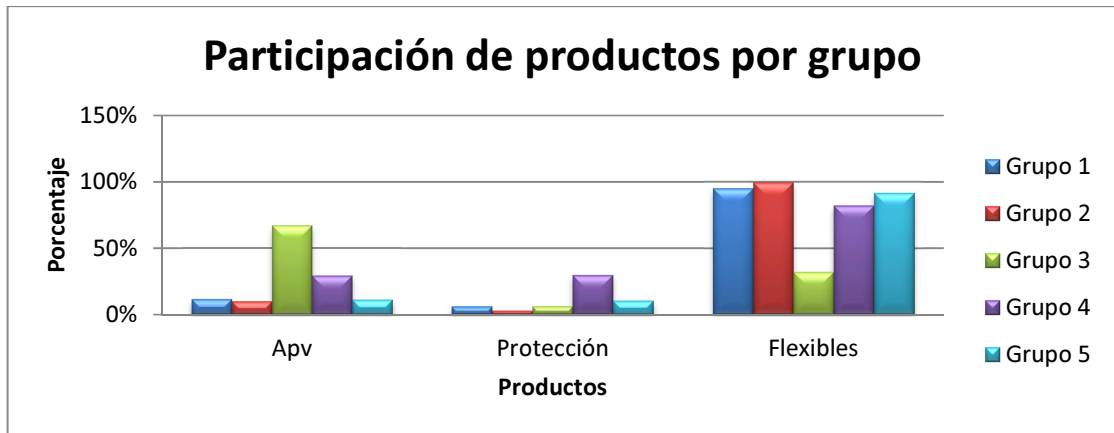
Este grupo es el que más sale a Pubs, a comer o bailar, también disfruta bastante con actividades como ir al cine, teatro o a un concierto. Se puede concluir que sus gustos son por organizar panoramas fuera de casa por lo general pagados. La mayor parte de estos no realizan las actividades muy frecuentemente ya que la mayoría lo hace una vez al mes.

Casi un 60% utiliza el internet como medio de información, lo que los transforma en uno de los principales grupos para este medio. Son menos asiduos a la televisión que los demás grupos. El 61% de este grupo ocupan el email como su red social preferida, superior a los otros grupos.

Es importante percatarse que un 30% de este grupo posee un APV en la competencia, siendo el grupo donde más se acentúa este problema. Lo mismo ocurre con los seguros de protección que un 20% de este grupo posee uno de estos fuera de la empresa.

**Clúster 3 (Ahorradores):** Este grupo posee productos en proporción inversa al resto de la cartera, ya que poseen un alto nivel de APV y bajo de Flexibles, esos clientes se encuentran en esta base debido a que su producto principal posee una cobertura por fallecimiento asociada a estos. Son el grupo más joven con una edad promedio menor a los 42 años. Estos clientes no poseen mucho valor, pero debido a su edad y al nivel de ingresos se puede decir que poseen un gran potencial, para llevarlos a comprar nuevos productos dentro de la empresa.

Figura 28: Participación de productos por grupo



Fuente: "Elaboración propia"

### Descripción a partir de datos encuesta $\varepsilon = 11\%$

Este grupo destaca por tener un porcentaje de solteros mayor, con casi un 20%. Poseen familias más pequeñas, donde la mayor parte son de 2 a 3 integrantes. Su nivel educacional es bastante alto con un 35% de postgrado lo que nos indica que posee un nivel de ingresos superior al de los grupos anteriores.

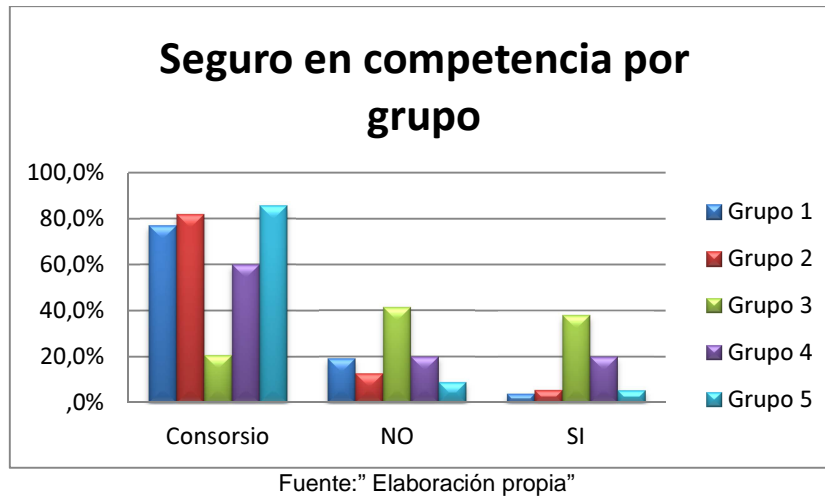
Al igual que el segmento Gold a este grupo le gusta bastante salir a un Pub, a comer o a bailar, pero se diferencia en que prefiere las actividades al aire libre, en vez de los cines, teatros o conciertos. Al mismo tiempo tienen una intensidad de uso mayor que otros, ya que realizan esta actividad al menos una vez a la semana.

Este conglomerado ocupa el internet de manera promedio, al igual que los noticieros de televisión abierta, se destacan por informarse en mayor cantidad que los otros segmentos a través de los noticieros que proporciona la señal de cable.

Sobre el 20% de este segmento no utiliza ninguna red social, lo que es más que la mayoría de los otros grupos, el uso del email es de los más reducidos en la muestra y solo en el uso de Facebook son cercanos al promedio.

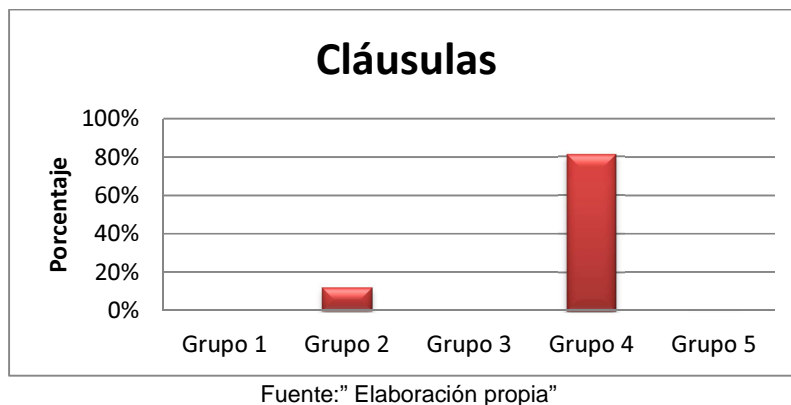
Este grupo posee poca cantidad de seguros en la empresa, pero se puede observar que estos clientes si poseen seguros fuera de la organización como se aprecia en la Figura 29.

Figura 29: Seguros en competencia por grupo



**Clúster 4 (Prisioneros):** La figura 30 muestra que el 80% de este grupo posee cláusulas, lo que los lleva ser más dependientes de la compañía. Tienen capitales en gastos médicos mucho más alto que los demás y en general poseen mayores capitales en todas las coberturas. Además su pago anual es significativamente mayor, cercano al doble del que lo sigue, pese a esta característica no es el grupo de mayor valor, esto debido a que las cláusulas indican un mayor riesgo de estas personas.

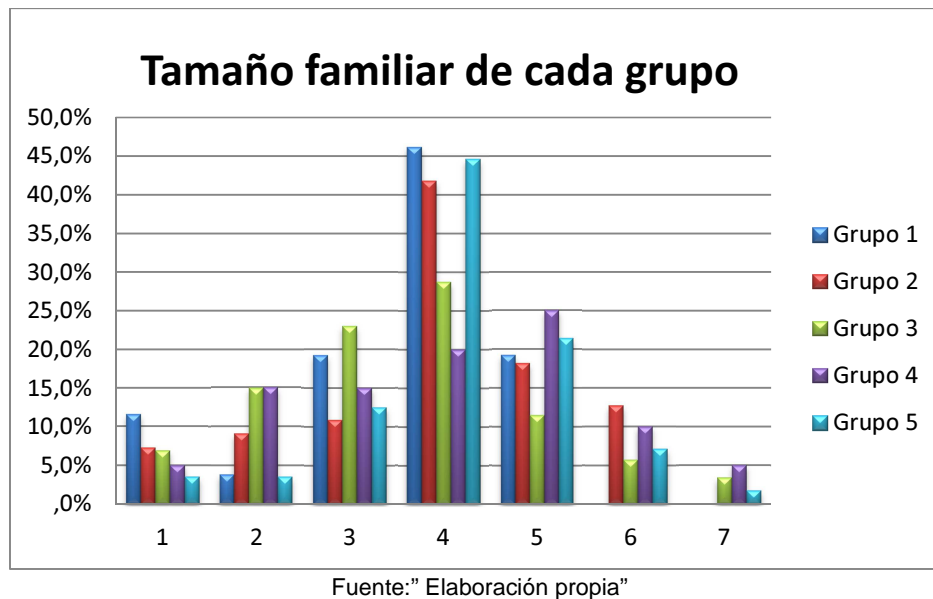
Figura 30: Cláusulas por cada grupo



#### Descripción a partir de datos encuesta $\varepsilon = 22\%$

Con un 20% es el grupo con el mayor índice de personas separadas, además un 5% son viudos, lo cual también es bastante alto. Su grupo familiar es de los más numerosos de la base, donde destaca su superioridad en familias de 5, 6 y 7 personas como se aprecia en la figura 31.

Figura 31: Tamaño familiar de cada grupo



Este grupo es el con mayor nivel educacional, donde un 45% posee un postgrado y casi no tiene participación en solo estudios primarios.

Prefieren realizar deporte en su tiempo libre, además de esto disfrutan de las reuniones sociales.

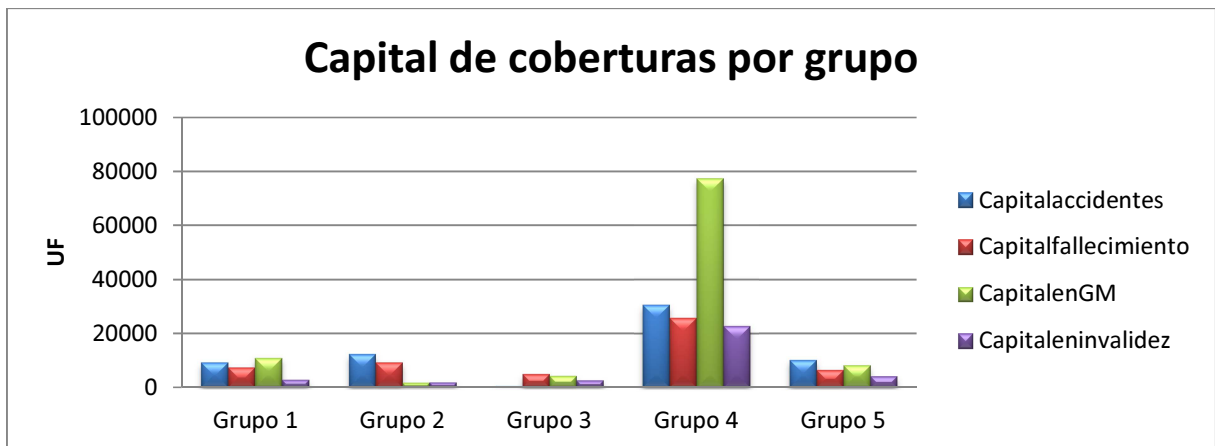
Tiene un bajo consumo de internet como medio de información en comparación a los demás grupos, pero ocupan bastante los noticieros de la televisión abierta, además son el grupo que más ocupa las redes sociales con este fin, pese a que solo es un 5 %.

Respecto de las redes sociales se puede señalar que la mayoría ocupa al menos una de estas, ya que sólo un 15% no posee. Es el grupo que más utiliza twitter con casi un 10% pero también son fuertes usuarios del email.

Pese a que no poseen muchos flexibles, tampoco los han contratado fuera de la empresa, por lo que se puede concluir que sus necesidades de seguro ya están cubiertas con los productos ya contratados.

**Clúster 5 (Preocupados por accidentes):** Cómo lo dice su nombre, este grupo se caracteriza porque todos poseen coberturas por accidentes, además son un grupo joven con bastante potencial, ya que el capital de sus coberturas no es muy alto.

Figura 32: Capital de coberturas por grupo

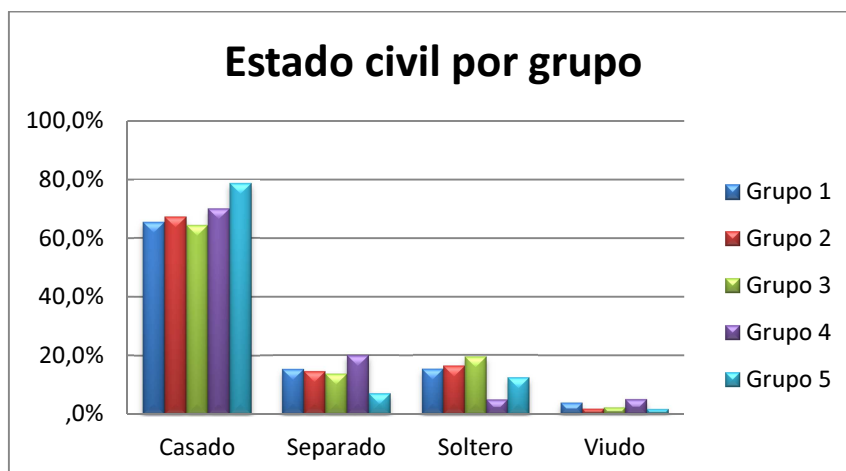


Fuente: "Elaboración propia"

**Descripción a partir de datos encuesta  $\varepsilon = 13\%$**

Casi el 80% de este grupo se encuentra casado y poseen un grupo familiar bastante grande como se puede ver en la figura 33.

Figura 33: estado civil por grupo



Fuente: "Elaboración propia"

Pese a que el nivel de postgrado en este grupo es menor que en los demás, es primordial darse cuenta de que esto se ve compensado por el alto porcentaje de profesionales, y al analizar la suma de profesionales y postgrado es la mayor de todos los grupos llegando al 95%.

Dentro de sus actividades preferidas se puede apreciar que son el grupo que más disfruta de las realizadas en el hogar con un 15%. No poseen una clara preferencia por alguna en especial.



Es uno de los grupos que más ocupan internet como medio para informarse con un 60%, también un 15% utiliza los periódicos, siendo el que más los utiliza. Este es el segmento que mayor uso le da a Facebook con un 18%. Son los que más ocupan las redes sociales, donde más del 70% de la población las utilizan 1 o más veces al día.

Pese a no poseer ni APV ni Protección en la compañía, se puede observar que tampoco los han contratado en otra empresa. Por otra parte los seguros flexibles parecen ser los que realmente necesitan, ya que sí los han contratado en otras empresas a pesar que la gran mayoría los poseen en Consorcio.

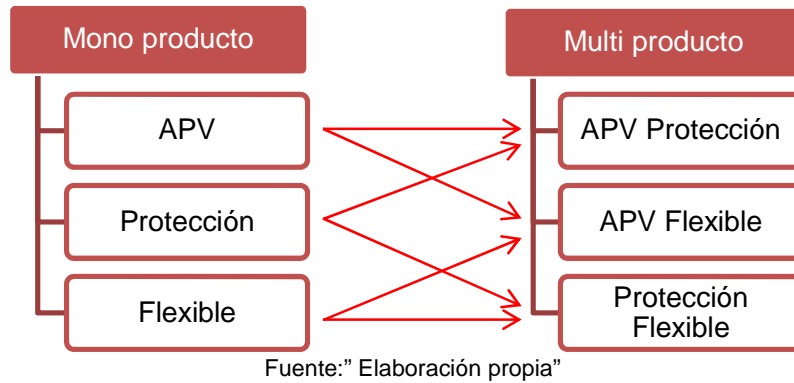
En esta sección se ha mostrado el resultado de una de las segmentaciones, la que muestra claras diferencias de comportamiento de compra, ya que las principales variables que se poseen en esta base de datos son correspondientes a las características de los productos que cada individuo posee. Además fue caracterizado por los datos levantados en la encuesta que han entregado información relevante de las personas que se encuentran detrás de los productos.

### **3.8 APLICACIÓN DE MODELOS DE PROPENSIÓN DE COMPRA**

Esta sección tiene como objetivo entender las características que diferencian a los clientes mono productos, con aquellos que son multi producto; con el fin de encontrar clientes con un sólo producto parecidos a los que tengan varios, para poder ofrecerle aquel que complete su necesidad adicional.

Por ejemplo, se busca encontrar parte de los clientes pertenecientes a APV que se parezcan a los clientes que poseen APV y Flexible, para así ofrecerles un seguro Flexible. De esta manera se traspasaría una parte de los clientes mono producto a los multi producto haciendo a la cartera más rentable. La idea es repetir este procedimiento para cada uno de los productos seleccionados tal cómo se observa en la figura 34.

Figura 34: Tipos de productos



Se utilizarán modelos predictivos para analizar las características de los clientes que permiten reconocer a los que les podría interesar tener dos productos. Los métodos ocupados en este trabajo son:

- Regresión Logística
- Análisis Discriminante
- Árbol de clasificación C5

Las variables a predecir en este caso, serán las variables que indican la posesión de algunos de los multiproductos, se pueden observar en la figura 34.

Las variables descriptivas de los clientes, en este caso son las predictivas para los modelos a aplicar. Las que se muestran a continuación:

- Sexo
- Residencia
- Cláusulas
- Contacto
- N° de Autos
- N° de Autos lujosos
- Tipo de Agente
- Edad

Es ambicioso tratar de predecir con estas variables de manera precisa si son o no propensos a la compra de un producto, pero sí puede aportar información importante para el análisis de qué producto ofrecer, a qué persona y también definir cuál ayudaría en futuros estudios en que se posea más información de estos.

Para la aplicación de cada uno de los modelos antes explicados, será necesario sacar una muestra balanceada, lo que quiere decir que debe existir un 50% que posea el producto analizado y el resto que no lo tenga. Esto se realiza para evitar que el

algoritmo no prefiera definir que nadie tiene el producto en caso de que en la base hubiese un bajo porcentaje de posesión de éste.

Además dentro de la muestra sacada se utilizará un 90% para el entrenamiento y el 10% restante será utilizado para validación de los métodos ocupados y descartar el sobre entrenamiento.

Figura 35: Utilización de datos



Fuente: "Elaboración propia"

A continuación se mostrarán los resultados entregados por los métodos predictivos, en cada uno de los casos correspondientes.

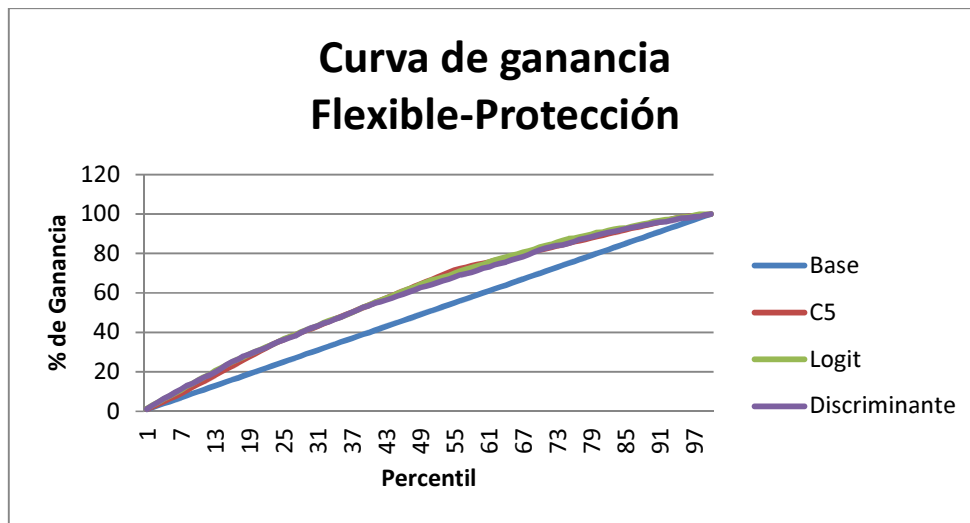
### 3.8.1 Productos

#### 3.8.1.1 Resultado a clientes de Flexible Protección

El primero de los casos a analizar será el de los clientes que han contratado seguros de Protección y Flexible. Esta combinación de productos nos muestra que son personas que poseen la necesidad de un seguro.

Al comparar los diferentes métodos ocupados, se puede apreciar en la curva de ganancia que su desempeño es bastante parecido, ya que el área bajo las curvas correspondientes a los diferentes métodos es bastante similar y no es muy grande, pese a esto se ve una leve predominancia del árbol de clasificación C5. Lo que se puede observar en la Figura 36.

Figura 36: Curva de ganancia Flexible-Protección



Fuente: "Elaboración propia"

Observando los índices de calidad de cada uno de los diferentes métodos se puede apreciar que al igual que las curvas de ganancia, el desempeño de los diferentes modelos es bastante similar. Los de mejor desempeño global fueron el A.D. y el C5, donde el primero destaca por su precisión y el segundo por su sensibilidad.

Tabla 24: Calidad de modelos predictivos para cruce Flexible-Protección

Modelo predictivo Cruce Flexible y Protección					
	Global	Precisión	Sensibilidad	Especificidad	Exactitud
Logit	68,3%	61,8%	70,8%	66,4%	68,6%
C5	69,6%	71,6%	68,2%	71,0%	69,6%
A.D	69,3%	61,8%	74,1%	65,8%	70,0%

Fuente: "Elaboración propia"

Tomando en cuenta lo anterior, es importante darse cuenta que los arboles de clasificación entregan reglas que son muy importantes desde el punto del negocio, por lo que ante desempeños predictivos muy parecidos se pondrá por delante este tipo de heurísticas debido a su aporte de conocimiento adicional.

Es por esto que en este caso se utilizó el árbol de clasificación C5, del cual a partir de las diferentes reglas de clasificación se destaca lo siguiente:

- Una de las variables más importantes a la hora de saber si un cliente contratará este cruce de productos es la tenencia de cláusulas, ya que aquellos que las poseen son altamente más propensos a esta contratación. Esto es fácil de explicar ya que personas más enfermas tratarán de estar más cubiertas que las demás, aspecto que está incluido en este cruce.

- Otro aspecto importante son las variables relacionadas con los autos, las cuáles pueden ser interpretadas cómo un reflejo del nivel socioeconómico, aquellos que poseen más autos son más propensos a comprar estos cruces, lo cual es traducido a que a mayor nivel socioeconómico es más probable lograr un cruce. La posesión de motos, se puede tomar cómo un sustituto más económico de los autos por lo tanto es entendible pensar que está relacionado con un nivel socioeconómico más bajo y por ende menor probabilidad de contratación de estos productos.
- Otro factor que marca una diferencia es si es hombre o mujer, ya que las mujeres son muy poco propensas a la contratación de este cruce.
- Los menores de 32 años también son un grupo de bajo interés para este cruce ya que probablemente no poseen muchas enfermedades.
- Respecto a la residencia, se puede decir que en el sur y la zona oriente son los más propensos a la contratación de este producto, además la zona sur oriente es también un candidato a estos productos.

Las conclusiones anteriores pueden ser complementadas observado en el Anexo E, donde se muestra el detalle de las reglas de decisión definidas por el árbol C5 para este caso.

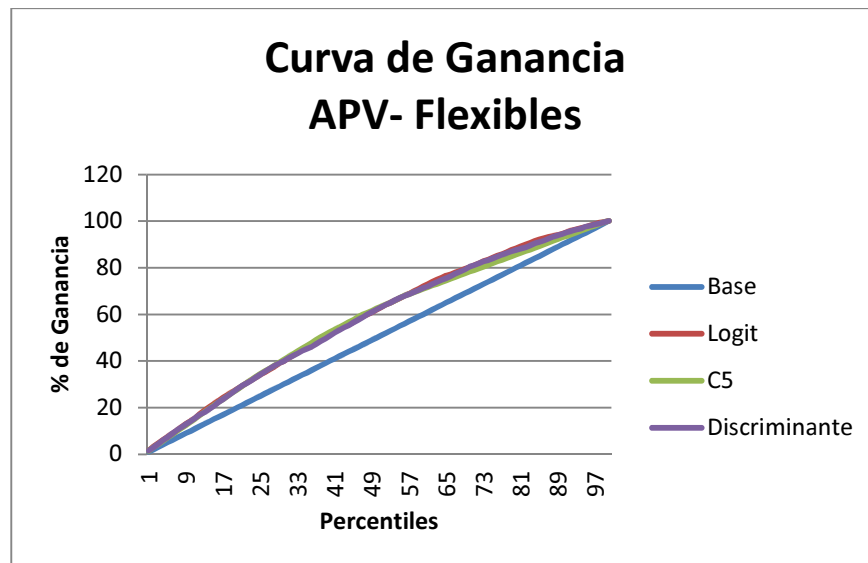
### **3.8.1.2 Resultado de Clientes APV Flexibles**

Por las características propias de los productos, los clientes que los poseen son aquellos muy preocupados por el futuro y los ahorros necesarios para tener un buen pasar en éste. Además los clientes de APV se conocen por poseer niveles de ingreso mayores, con lo cual se espera que en aquellos que poseen éste cruce se acentúe esta situación.

Igual que en el caso anterior se aplicaron los diferentes modelos predictivos para poder entender los factores que desencadenan la compra de este producto.

A partir de las curvas de ganancia correspondientes a cada uno de los modelos utilizados para esta variable, se puede apreciar que el área bajo las distintas curvas es bastante similar, pese a que se podría decir que el análisis discriminante se comporta de una manera un tanto superior a los demás.

Figura 37: Curva de ganancia APV-Flexible



Fuente: "Elaboración propia"

De igual forma que en las curvas de ganancia se puede observar una ligera superioridad del análisis discriminante. La calidad de los modelos no es muy alta, pero éste método es mejor en todos los aspectos analizados.

Tabla 25: Desempeño de modelos para cruce APV-Flexible

Modelo predictivo Cruce Flexible y APV					
	Global	Precisión	Sensibilidad	Especificidad	Exactitud
Logit	64%	63%	58%	69%	63%
C5	65%	63%	60%	70%	65%
A.D	67%	69%	61%	73%	67%

Fuente: "Elaboración propia"

En la tabla 25 se pueden apreciar los coeficientes de la función discriminante. A partir de esto se concluye:

- Importancia del tipo de agente, donde la mayoría de los que posee este cruce tienen contacto con la organización a través de un agente de compañía.
- Se caracterizan por ser un grupo un tanto más joven que los otros dos cruces. Aunque la diferencia es pequeña
- Poseen un número elevado de cláusulas, característica común en los clientes con cruce de producto.

- Son de un nivel socio económico alto, esto se deduce del alto número de autos de lujo que se pueden observar en estos clientes.

Tabla 26: Análisis Discriminante

Análisis Discriminante	
Agente de compañía	,800
Fuma	-,324
Cláusulas	,324
Contacto	,258
Edad	-,208
Autos nuevos	,189
Autos	,047
Sexo	,042
Autos de Lujo	,018
Motos	-,007

Fuente: "Elaboración propia"

Complementando lo anterior con la información entregada por el árbol de clasificación:

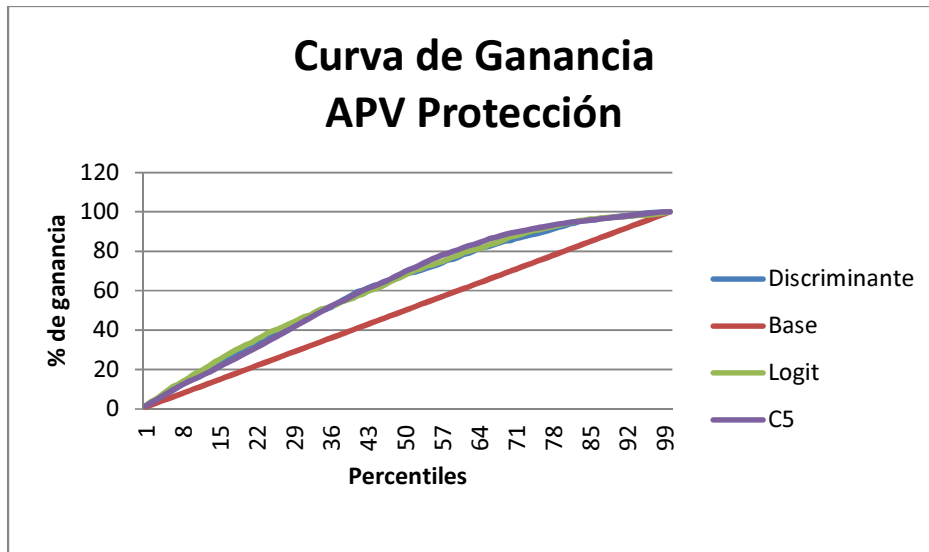
- Al igual que el producto anterior, los hombres tienen mayor probabilidad de contratar este producto.
- Por otra parte es importante notar que los mayores de 32 pero menores de cuarenta son los más propensos a la contratación de este producto.

### 3.8.1.3 Resultado de los clientes APV y Protección

Este cruce de clientes posee la característica de que resuelven cada una de sus necesidades con productos especializados. Por una parte cubren su necesidad de ahorro con un APV haciendo uso de sus beneficios tributarios, mientras que para contratar un seguro, llevan uno que no posee capacidad de ahorro pero con primas más baratas para cada una de las diferentes coberturas.

En relación a la calidad de cada uno de los modelos, se puede observar que el árbol C5 posee un área bajo la curva mayor con respecto a los diferentes algoritmos.

Figura 38: Curva de ganancia APV-Protección



Fuente: "Elaboración propia"

Lo anterior se verifica al observar los diferentes índices de desempeño, en donde sólo en relación a la sensibilidad este método baja su calidad. De todas maneras se comporta mejor y como antes se explicó, este método entrega variables de decisión muy importantes desde el punto del negocio, por lo que se preferirá éste.

Tabla 27: Desempeño de modelos para Cruce APV-Protección

Modelo predictivo Cruce APV y Protección					
	Global	Precisión	Sensibilidad	Especificidad	Exactitud
Logit	66%	58%	78%	57%	67%
C5	69%	73%	72%	64%	68%
A.D	65%	61%	76%	56%	66%

Fuente: "Elaboración propia"

De las reglas de decisión definidas por el árbol de clasificación C5 destacan las siguientes conclusiones:

- En este caso se puede apreciar que aquellos que poseen agente de compañía son mucho más cercanos a los que han contratado este servicio en ocasiones anteriores.
- Al igual que en los productos anteriores se puede apreciar que los hombres son mucho más propensos a la compra de estos dos productos juntos.



- Otra de las variables importantes es si poseen o no cláusulas, ya que de igual manera que los demás cruces, aquellos que poseen más cláusulas son más cercanos a la compra de este producto.
- Las variables de la cantidad de autos nuevamente sirven para determinar la propensión de compra, por lo que los clientes de mayores ingresos están más cercanos a la compra de estos productos.
- Además los sectores C,D,A,E y H son más propensos a la compra.
- Los clientes que han contratado este producto se caracterizan por tener mayor contacto con la compañía, lo cual es un buen indicador para saber si compran o no este producto.

#### **3.8.1.4 Multi productos**

Los modelos de propensión de compra aplicados en esta sección han buscado diferencias entre un tipo de cliente mono producto y otro multi producto. Al observar estas características se puede apreciar que coinciden muchas veces para diferentes tipos de multi productos. Ante esto se puede concluir que las diferencias encontradas por los modelos son por el número de productos más que por tipo de estos.

Características que se pueden adjudicar a los clientes multi productos:

- Mayor cantidad de cláusulas
- Hombres son más propensos a poseer más productos
- Clientes de un mayor nivel socioeconómico, son más adecuados para la oferta de más de un producto.

A partir de lo anterior, se pueden sacar varias conclusiones acerca de las diferencias que existen entre los clientes que poseen sólo un producto y aquellos que poseen varios.

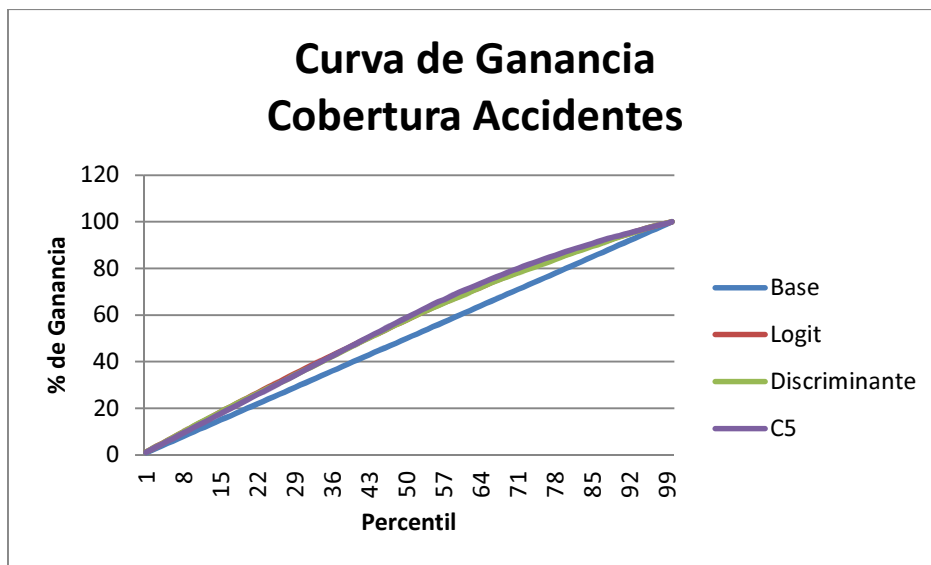
#### **3.8.2 Coberturas**

Así cómo se realizó un estudio de los clientes multi producto para poder aumentar el cross selling, se realizará un estudio de los clientes de cada una de las coberturas para poder aumentar el up-selling existente y el promedio de coberturas que poseen todos los clientes.

### 3.8.2.1 Resultado clientes cobertura de accidentes

A continuación se analizará a los clientes que poseen cobertura de accidentes. Cómo se puede observar en la curva de ganancia, los diferentes modelos no funcionan de la mejor manera y todos tienen un desempeño muy similar.

Figura 39: Curva de ganancia Cobertura de accidentes



Fuente: "Elaboración propia"

Los diferentes índices ratifican lo antes dicho, ya que el Logit y el C5 funcionan bastante parecidos, por lo que se decidió por el árbol de clasificación.

Tabla 28: Desempeño de modelo para cobertura de accidentes

Modelo predictivo Cobertura Accidentes					
	Global	Precisión	Sensibilidad	Especificidad	Exactitud
Logit	66%	84%	67%	61%	64%
C5	66%	84%	68%	61%	64%
A.D	64%	69%	70%	55%	63%

Fuente: "Elaboración propia"

A partir de las reglas que este árbol ha entregado se sacan las siguientes conclusiones:

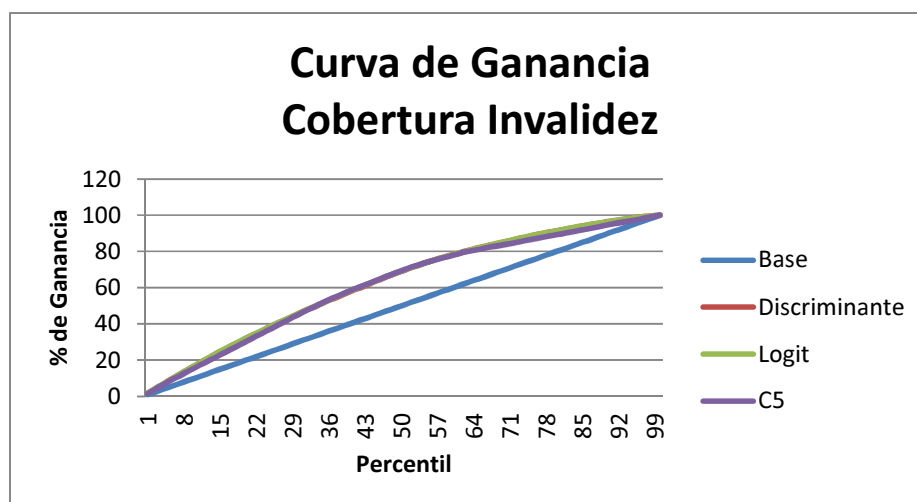
- Las variables más importantes para definir la compra o no de esta cobertura está muy relacionada a la edad de los individuos, ya que los de mayor edad son más propensos a la contratación de esta cobertura.

- Se puede ver que los jóvenes que viven en los sectores E, F e I son propensos a la compra de los diferentes productos. Mientras que los que provienen del sur pero que pertenecen a un sector más acomodado son propensos a la compra de esta cobertura.

### 3.8.2.2 Resultado clientes de cobertura de Invalidez

A partir de las curvas de ganancia para cada uno de los modelos aplicados se puede apreciar que su desempeño en este caso es mejor que para otros productos analizados. Además es importante notar que el modelo Logit y C5 funcionan un poco mejor que el Análisis discriminante.

Figura 40: Curva de ganancia de Cobertura de Invalidez



Fuente: "Elaboración propia"

Pese a que los índices muestran un desempeño muy parejo se aprecia una pequeña diferencia de calidad del árbol de clasificación utilizado. Es por esto que se utilizará el modelo Logit para obtener los resultados.

Tabla 29: Desempeño modelo para cobertura de invalidez

Modelo predictivo Cobertura Invalidez					
	Global	Precisión	Sensibilidad	Especificidad	Exactitud
Logit	69%	63%	68%	69%	69%
C5	68%	69%	66%	71%	68%
A.D	69%	66%	68%	71%	69%

Fuente: "Elaboración propia"

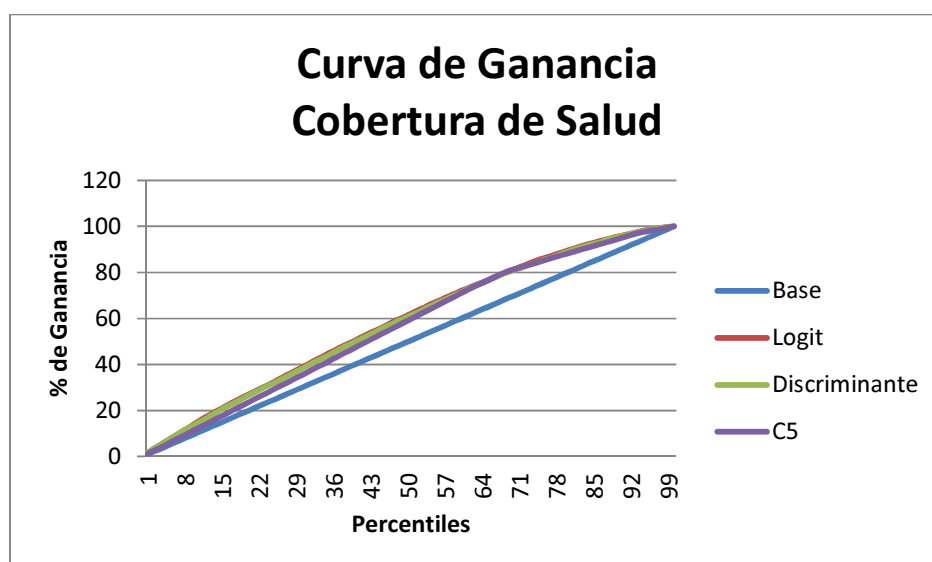
A partir del resultado de los modelos se pueden sacar algunas conclusiones:

- Se puede apreciar que los clientes de menor edad son más cercanos a esta cobertura.
- Los hombres compran más esta cobertura.
- Los sectores K y F son los predominantes de esta cobertura.

### 3.8.2.3 Resultado de clientes de Cobertura de Salud

Al analizar las curvas de ganancia se puede ver una pequeña predominancia de la calidad del análisis discriminante.

Figura 41: Curva de ganancia cobertura de salud



Fuente: "Elaboración propia"

A pesar de que se confirma con el índice de calidad global que el análisis discriminante es el de mejor desempeño, es importante tomar en cuenta la calidad del árbol C5 ya que se destaca por una alta precisión, pese a esto se prefiere este último debido a que la diferencia no es muy grande y la información adicional que ofrece este método explica la decisión tomada.

Tabla 30: Desempeño de modelo para cobertura de salud

Modelo predictivo Cobertura Salud					
	Global	Precisión	Sensibilidad	Especificidad	Exactitud
Logit	60,7%	65,2%	59,9%	61,6%	60,8%
C5	60,6%	78,0%	58,0%	66,2%	62,1%
A.D	62,2%	66,7%	61,3%	63,2%	62,2%

Fuente: "Elaboración propia"

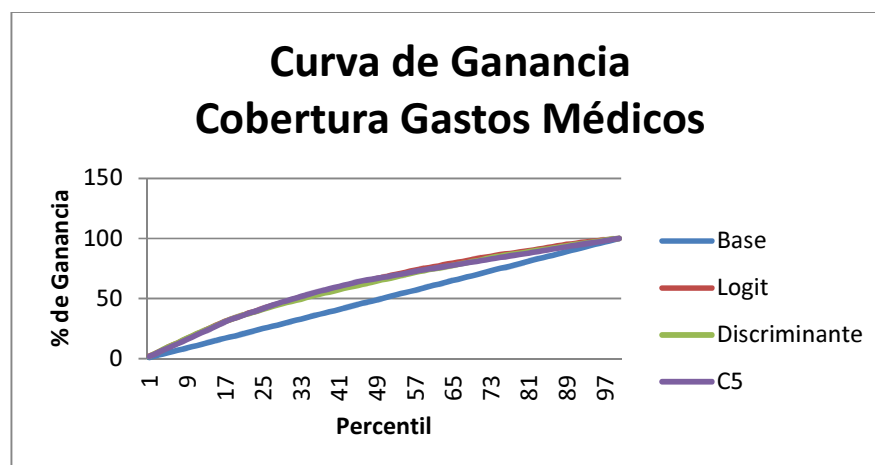
Además como se explicó anteriormente el árbol de clasificación entrega reglas muy importantes para el negocio. Es por esto que se decide su utilización para el análisis de los clientes propensos a esta cobertura. A continuación se aprecian las conclusiones más importantes encontradas en este análisis:

- Los clientes menores de 31 años son muy poco propensos a la contratación de esta cobertura.
- Por otra parte los que poseen agente de compañía y tienen más de 53 años son bastante propensos a la compra de este producto.

### 3.8.2.4 Resultados de clientes de cobertura en Gastos Médicos

Las curvas de ganancia demuestran una mejoría en la calidad de los resultados para esta variable. Aunque se ve una pequeña ventaja en la calidad del árbol de clasificación utilizado.

Figura 42: Curva de ganancia cobertura de Gastos médicos



Fuente: "Elaboración propia"

Lo anterior se ve reflejado en los índices, pero el desempeño es prácticamente el mismo para los diferentes modelos. Por las razones antes explicadas se preferirá el Árbol C5 por sobre los demás.

Tabla 31: Desempeño de modelo para cobertura en Gastos médicos

Modelo predictivo Cobertura Gastos Médicos					
	Global	Precisión	Sensibilidad	Especificidad	Exactitud
Logit	68,6%	60,4%	72,2%	66,0%	69,1%
C5	68,9%	58,2%	74,0%	65,7%	69,8%
A.D	68,4%	55,1%	75,7%	64,1%	69,9%

Fuente: "Elaboración propia"

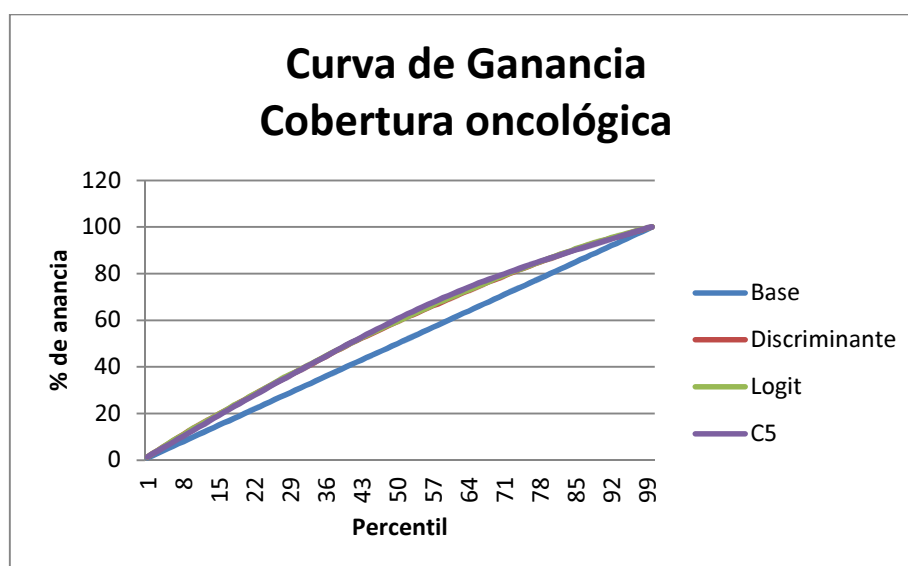
De las reglas generadas por este árbol destacan las siguientes:

- Los clientes de mayor edad y que poseen una mayor cantidad de autos, es decir, son parte de un nivel socioeconómico más elevado, son bastante más propensos a la contratación de este servicio.
- Los mayores que pertenezcan a los sectores B, E, F, G, J, Sur y Norte son bastante propensos a la contratación de este servicio.
- Una de las variables más importantes es el número de cláusulas, las que reflejan una necesidad directa de este tipo de seguros. Aquellos que poseen cláusulas son mucho más propensos a las diferentes coberturas.

### 3.8.2.5 Resultado de clientes de Cobertura oncológica

Lamentablemente se aprecia en la Figura 43 que la calidad del resultado de los diferentes modelos para el análisis de esta variable está lejos del óptimo, ya que las áreas bajo las diferentes curvas de ganancia son bastante bajas.

Figura 43: Curva de ganancia cobertura oncológica



Fuente: "Elaboración propia"

Lo anterior se confirma al analizar los índices de desempeño de los diferentes métodos ocupados, donde todos son bastante bajos. El análisis discriminante posee una pequeña ventaja por sobre los demás algoritmos, pese a que el árbol destaca por su precisión.

Tabla 32: Desempeño de modelos para cobertura oncológica

Modelo predictivo Cobertura Oncológica					
	Global	Precisión	Sensibilidad	Especificidad	Exactitud
Logit	61%	63%	60%	62%	61%
C5	60%	64%	59%	61%	60%
A.D	62%	60%	63%	62%	62%

Fuente: "Elaboración propia"

Esta vez se prefirió utilizar el análisis discriminante debido a su mejor calidad de predicción respecto a los demás métodos.

La función discriminante encontrada se puede observar en la tabla 33.

Tabla 33: Función Discriminante para cobertura oncológica

Función discriminante	
Sexo	-,432
Fuma	,253
Cláusulas	,342
Contacto	,150
Autos	,133
Autos de Lujo	,028
Motos	,036
Autos nuevos	-,058
Agente de Compañía	-,279
Edad	,736

Fuente: "Elaboración propia"

A partir de la función discriminante se pueden sacar algunas conclusiones respecto a los clientes que contratarían este servicio:

- Los clientes de mayor edad son mucho más cercanos a la compra de este producto
- Además se puede observar que las mujeres tienen mucha más cercanía a este producto
- Son atendidos por agentes de mantención.

### 3.9 CREACIÓN DE VARIABLES OFERTA

Luego de la aplicación de los modelos de propensión de compra, se crean las variables ofertas. Estas variables tienen como objetivo identificar a los clientes que son predichos como potenciales para alguno de los cruces, para poder ofrecerles el producto complementario. Por ejemplo, la variable oferta APV, posee unos en los clientes que fueron predichos como APV Flexible o APV Protección y que no tienen ninguno de estos dos productos ni tampoco poseen APV, por lo que son clientes o de protección o de flexible, que son potenciales a pasar a ser multi producto a través de la contratación de APV.

Lo mismo se realizó para las coberturas, se crearon las variables propensión, las cuales toman a los clientes predichos por el modelo en cuestión, pero que realmente no poseen esas coberturas.

Para cada una de las diferentes ofertas potenciales se realizó un perfil de estos clientes, para conocerlos de manera más profunda.

#### 3.9.1 Perfil por oferta

Se busca perfilar a los clientes potenciales de los diferentes productos, para mejorar el diseño de la estrategia comercial para cada uno de ellos. En la tabla 34 se puede observar algunas características propias de cada una de estas ofertas.

Tabla 34: Perfil por oferta

Perfil por oferta				
	APV	Protección	Flexible	Sin Oferta
	Media	Media	Media	Media
Sexo	29%	90%	42%	38%
Cláusulas	8%	12%	7%	1%
Contacto	35%	47%	36%	44%
Autos de lujo	7%	10%	7%	6%
Tipo de agente	94%	30%	83%	18%
Pago anual	32,2	44,2	40,4	32,3
Valor del cliente	51,1	77,6	45,0	60,2
Edad	41,9	44,4	41,9	44,6

Fuente: "Elaboración propia"

En la tabla 35 se puede observar la diferencia porcentual entre las medias de cada una de las ofertas con el promedio de la población, que será utilizado para la comprensión y caracterización de estas ofertas.



Tabla 35: Lift entre los diferentes perfiles por oferta

Lift			
	1	2	3
	Media	Media	Media
Sexo	-52%	46%	-32%
Cláusulas	-10%	25%	-24%
Contacto	-22%	3%	-21%
Autos de lujo	-7%	34%	-9%
Tipo de agente	112%	-32%	88%
Pago anual	-6%	29%	18%
Valor del cliente	-17%	26%	-27%
Edad	-5%	1%	-5%

Fuente: "Elaboración propia"

### 3.9.1.1 Oferta de APV

Los clientes clasificados en este grupo se caracterizan por poseer agente de compañía, lo que podrá ser utilizado para el diseño de la campaña, ya que el principal canal de acceso a los clientes es a través de su agente. Además es un grupo levemente más joven que el promedio. Está principalmente compuesto por mujeres, lo cual nos entrega información muy relevante para el enfoque de la campaña. Este segmento prácticamente no posee cláusulas por lo que está menos aferrado a los productos de la compañía y el esfuerzo por retenerlos debe ser mayor.

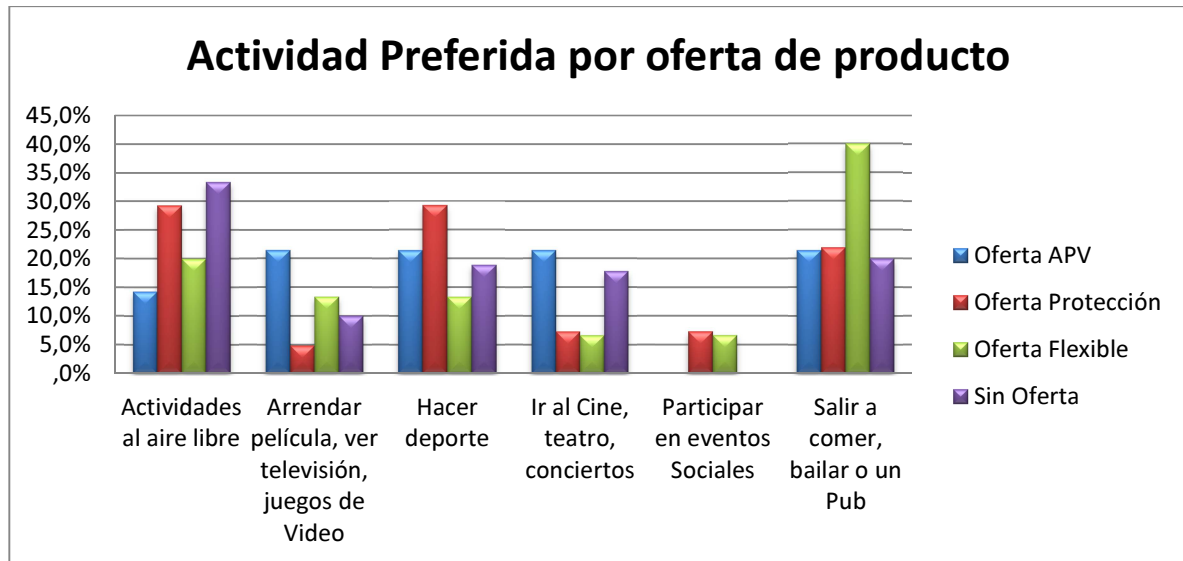
#### Descripción a partir de datos encuesta $\varepsilon = 13\%$

Los clientes pertenecientes a esta clasificación son en un 78% personas casadas, con el menor porcentaje de separados. Es el grupo con menor cantidad de postgrados, pero si un alto nivel de profesionales.

En relación a su actividad preferida se puede observar que es el grupo al que más le gustan las actividades hogareñas como arrendar películas, ver televisión o jugar video juegos; además destacan por su fascinación hacia el cine, el teatro y los eventos en vivo. El 60% de estos realiza estas actividades una vez a la semana.

Un 60% utiliza internet como medio de información, lo que los posiciona en el primer lugar de uso de esta herramienta, además es el grupo que más ocupa twitter.

Figura 44: Actividad preferida por oferta de producto



Fuente: "Elaboración propia"

### 3.9.1.2 Oferta de Protección

Esta oferta se caracteriza por la predominancia de los hombres con un 90 % de participación. Es el grupo que ha mantenido mayor contacto con la compañía y por lo general poseen agentes de mantención. Es el con mayores ingresos, reflejado en el número de autos de lujo que poseen. Su edad promedio es apenas mayor que la de los otros grupos, con 44 años. Este grupo es el de mayor valor.

#### Descripción a partir de datos encuesta $\varepsilon = 9\%$

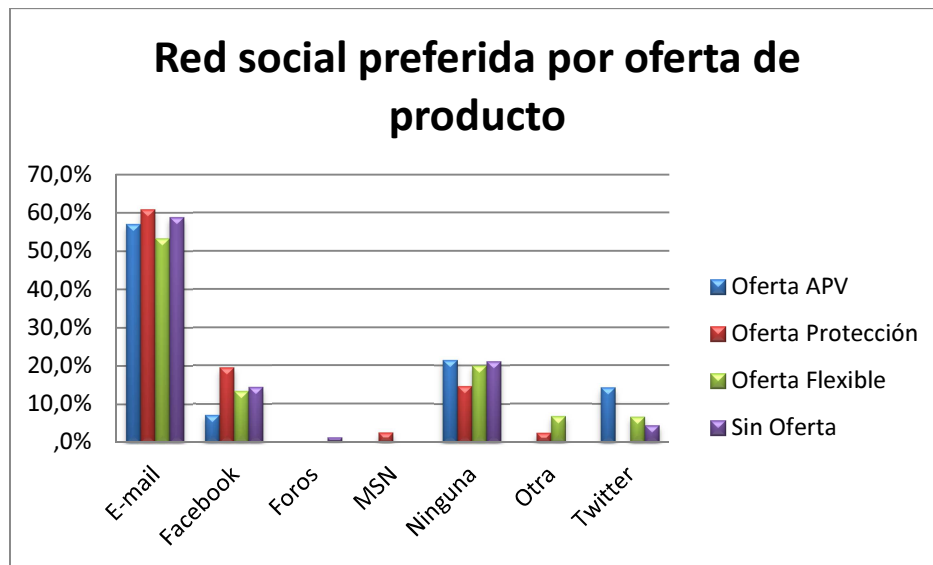
Este grupo posee un alto nivel de personas casadas, por sobre el 70%. El tamaño de su grupo familiar es menor al del resto de los grupos.

En cuanto a sus actividades recreativas cabe destacar que es el grupo más deportista con casi un 30% y otro porcentaje importante prefiere las actividades al aire libre. Estas actividades las realizan tres veces o más a la semana, lo que los hace uno de los grupos más activos.

Son los que más utilizan los periódicos para informarse, en desmedro de los noticieros de televisión abierta, pese a esto su principal medio de información como en todos los otros grupos es el internet.

Para casi un 60% de este grupo su red social favorita es el email, además de esto casi el 90% utiliza por lo menos una red social.

Figura 45: Red social preferida por cada grupo



Fuente: "Elaboración propia"

### 3.9.1.3 Oferta de Flexible

Este grupo se caracteriza por tener mayor participación de mujeres, con un 56%. La cantidad de cláusulas que poseen es bastante baja y la mayoría posee agente de compañía. No poseen un gran valor, pero esto se puede solucionar con una oferta de productos complementarios.

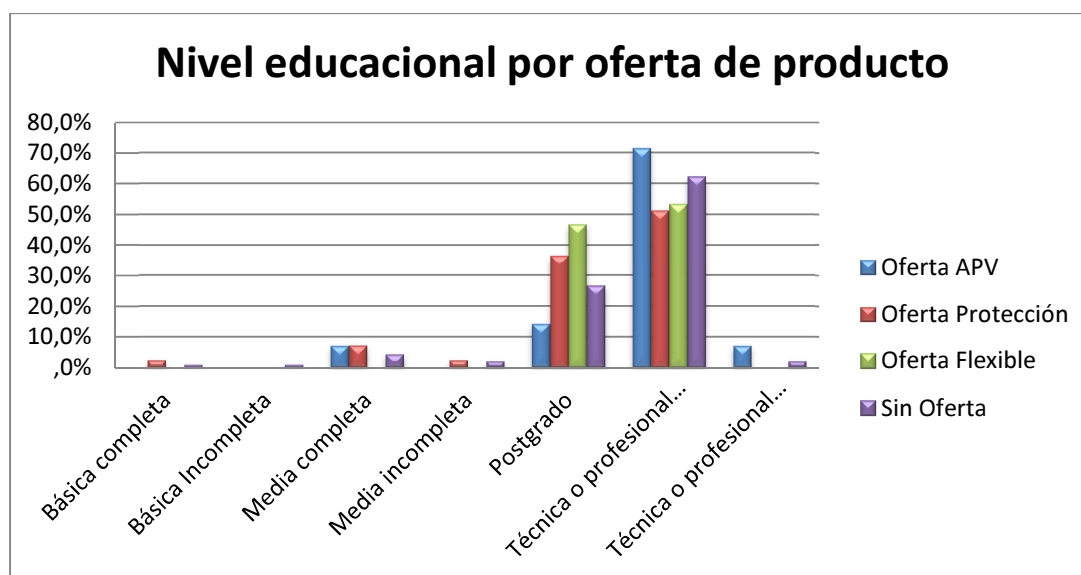
#### Descripción a partir de datos encuesta $\varepsilon = 13\%$

Es el grupo con mayor nivel educacional, ya que más de un 40% posee postgrado y más de un 50% posee por lo menos algún título profesional o técnico.

Respecto a las actividades recreativas este conglomerado posee un gusto bien definido, ya que cercano al 40% prefieren actividades como salir a un pub, a comer o bailar.

Es el que menos utiliza el internet como medio de información, con menos de un 40%. Por otra parte es el que más sintoniza los noticieros en televisión por cable o abierta. Además es el grupo que menos utiliza el email como su red social preferida con sólo un 60%.

Figura 46: Nivel educacional por oferta de producto



Fuente: "Elaboración propia"

### 3.9.1.4 Sin Oferta

Existe un grupo de clientes a los cuales no se determinó como propenso a la compra de algún producto en especial, ya que los modelos predictivos no lo determinaron así o estos ya poseían el producto.

Es un grupo con mayor cantidad de mujeres que poseen agente de mantención, además poseen un valor del cliente bastante cercano al promedio.

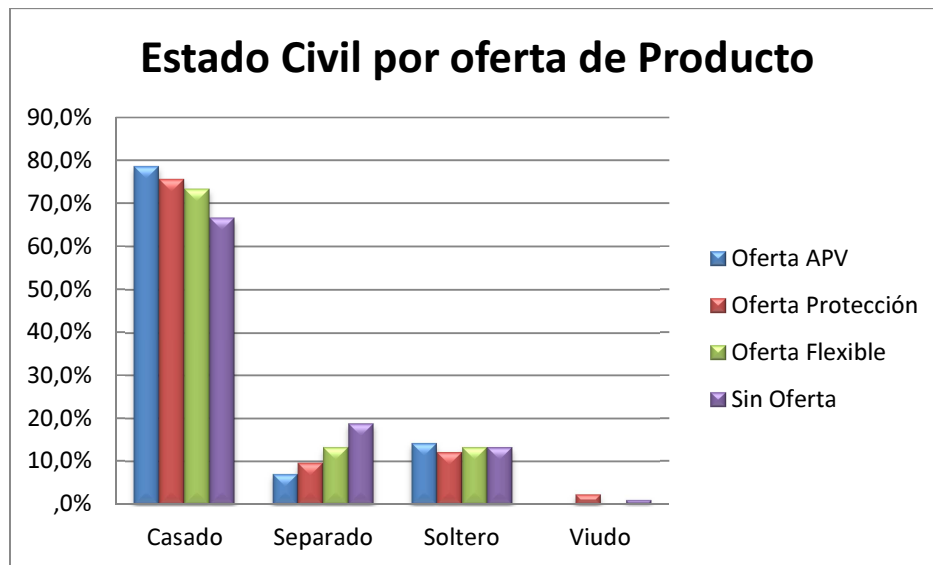
#### Descripción a partir de datos encuesta $\varepsilon = 8\%$

Este grupo tiene un menor nivel de casados, cercano al 65%, mientras que el nivel de separados es el más alto llegando al 20%.

Al analizar las actividades escogidas se puede ver que prefieren las realizadas al aire libre o ir al cine, teatro o algún evento en vivo. En relación a esto no es un grupo muy activo ya que más de un 30% las realiza una vez al mes o menos.

Son los que más utilizan los noticieros de la televisión abierta para informarse con más de un 20% y su nivel de uso de internet es cercano al 50%. Tienen una alta utilización del email como red social favorita, aunque la ocupan en menor grado ya que un 60% la ocupa una o más veces al día.

Figura 47: Estado Civil por oferta de producto



Fuente: "Elaboración propia"

También se realizó lo anterior para la oferta de cada una de las coberturas.

### 3.9.2 Perfil por coberturas

Tabla 36: Perfil por oferta de coberturas

	PGM	Pinv	POncO	PSalud	Paccidentes
	1	1	1	1	1
	Media	Media	Media	Media	Media
Sexo	64%	62%	47%	63%	62%
Cláusulas	23%	16%	15%	8%	11%
Contacto	49%	40%	54%	50%	46%
Autos de lujo	10%	7%	10%	8%	7%
Tipo de agente	69%	84%	29%	13%	37%
Pago anual	30,7	26,6	35,6	34,7	34,1
Valor del cliente	53,4	47,8	64,1	61,2	48,3
Edad	39,7	39,3	49,1	47,2	45,2

Fuente: "Elaboración propia"

Tabla 37: Lift con el promedio de los clientes por cobertura

	Lift				
	PGM	Pinv	POnc	PSalud	Paccidentes
	1	1	1	1	1
	Media	Media	Media	Media	Media
Sexo	4%	1%	<b>-23%</b>	3%	2%
Cláusulas	<b>146%</b>	72%	58%	<b>-13%</b>	18%
Contacto	8%	<b>-13%</b>	<b>19%</b>	10%	1%
Autos de lujo	<b>22%</b>	<b>-6%</b>	<b>24%</b>	7%	<b>-10%</b>
Tipo de agente	<b>57%</b>	<b>90%</b>	-35%	<b>-71%</b>	-15%
Pago anual	-10%	<b>-22%</b>	4%	1%	0%
Valor del cliente	-13%	<b>-22%</b>	4%	-1%	<b>-22%</b>
Edad	<b>-10%</b>	<b>-11%</b>	<b>12%</b>	7%	3%

Fuente: "Elaboración propia"

### 3.9.2.1 Oferta Cobertura Gastos Médicos

- Edad promedio 39 años
- 64% son hombres
- Valor del cliente 53,4 UF
- Alta cantidad de cláusulas 23%

#### Descripción a partir de datos encuesta $\varepsilon = 16\%$

Este es el grupo con mayor nivel de solteros, llegando casi a un 30%; además el tamaño de su grupo familiar es menor al de los demás grupos.

En cuanto a su nivel educacional se puede apreciar que más de un 60% posee un título profesional o técnico, pero solo un poco más de un 20% posee postgrado.

Prefieren en general las actividades al aire libre, las cuales son realizadas con una alta frecuencia por este conglomerado.

Son los principales usuarios de internet a la hora de informarse, donde su red social favorita es el email y un 80% utiliza su red social por lo menos una vez al día.

### 3.9.2.2 Oferta Cobertura Invalidez

- Edad promedio 39 años
- 62% son hombres
- Valor del cliente 47,8 UF

- Cláusulas 16%

#### **Descripción a partir de datos encuesta $\varepsilon = 16\%$**

Este es el grupo con menor porcentaje de casados, con solo un 50%, el nivel de separados es el más alto alcanzando más del 10%.

Su nivel educacional es bastante alto ya que un 30% posee algún tipo de postgrado y más de un 60% posee al menos algún título de pregrado.

Es el segmento más deportista ya que más de un 30% prefiere esta actividad, además de esto son el grupo al que más les gusta disfrutar de una obra de teatro, una película en el cine o asistir a un concierto.

Tiene un alto uso de internet a la hora de informarse y para pagar servicios básicos, son los más fanáticos del Facebook y se preocupan bastante de revisar frecuentemente sus redes sociales.

#### **3.9.2.3 Oferta Cobertura Oncológica**

- 47% hombres
- Mayor porcentaje de autos de lujo
- 71% posee agente de mantención
- Edad promedio 49 años

#### **Descripción a partir de datos encuesta $\varepsilon = 11\%$**

Es el grupo donde hay mayor porcentaje de casados, por sobre el 70%. Su nivel educacional es más bajo ya que sólo un 20% posee postgrado y menos de un 60% pregrado.

Su actividad preferida es salir a bailar, comer o a un pub y son los que más disfrutan de actividades en la casa como ver televisión o jugar videojuegos. Tienen una baja frecuencia respecto de estas actividades, ya que la mayoría las realizan una vez al mes o menos.

A la hora de informarse solo un 40% prefiere el internet y son los que más sintonizan los noticieros de televisión abierta.

Un 20% de estos no posee ninguna red social, junto a esto es importante recalcar que este segmento es el que menos compra por internet.

#### **3.9.2.4 Oferta Cobertura Salud**

- Menor cantidad de cláusulas

- 87% posee agente de mantención
- Edad promedio 47 años

#### **Descripción a partir de datos encuesta $\varepsilon = 10\%$**

Posee un alto nivel de casados. Su nivel educacional es bajo al igual que el grupo anterior.

Su preferencia por actividades recreacionales es bastante heterogénea por lo que no se puede definir una como la más importante.

Poseen un bajo uso de internet, sólo un poco mayor que el grupo anterior.

#### **3.9.2.5 Oferta Cobertura Accidentes**

- Valor del cliente 48 UF
- Edad 45 años
- 63% Agente de mantención

#### **Descripción a partir de datos encuesta $\varepsilon = 12\%$**

De igual manera que los últimos dos grupos, cercano al 70% de los integrantes de este conglomerado son casados. La mayoría prefiere las actividades al aire libre y las realizan una vez a la semana.

A la hora de informarse es el que menos ocupa internet con menos de un 40%, lo reemplazan a través de los noticiarios de televisión abierta y los periódicos. Es uno de los grupos que menos utiliza las redes sociales.

### **3.10 CLASIFICACIÓN POR PROPENSIÓN Y TENENCIA DE PRODUCTOS**

En este capítulo se buscará clasificar a los clientes por los productos que poseen y por los que están más cercanos a contratar. La finalidad es encontrar perfiles de clientes a partir de sus necesidades de seguros, así al haber realizado una buena caracterización de ellos, se les podrá ofrecer el producto adecuado de la manera precisa. Esta clasificación combinará la información proporcionada por los modelos de propensión de compra y entregará perfiles de clientes a partir de sus necesidades, ya sea de coberturas como de productos

Para esto se ocupará el método no supervisado de K-medias, el cual a través de las siguientes variables buscará clasificar a los clientes.

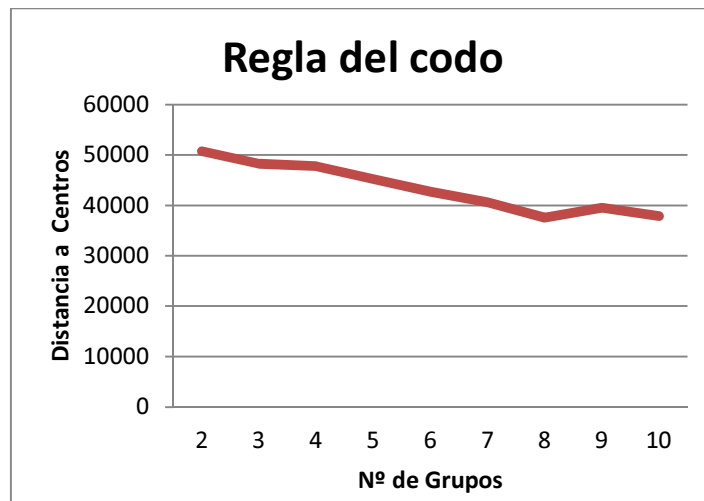


VARIABLES DE SEGMENTACIÓN:

- Oferta APV
- Oferta Flexible
- Oferta Protección
- Oferta Cobertura invalidez
- Oferta Cobertura Salud
- Oferta Cobertura Gastos Médicos
- Oferta Cobertura Oncológica
- Oferta Accidentes

Para definir el número óptimo de conglomerados se utilizó la regla del codo, a partir de las distancias de cada uno de los clientes al centro de cada uno de los grupos, el gráfico se puede apreciar en la figura 48. A partir de ahí se definió que el número ideal de conglomerados es 8.

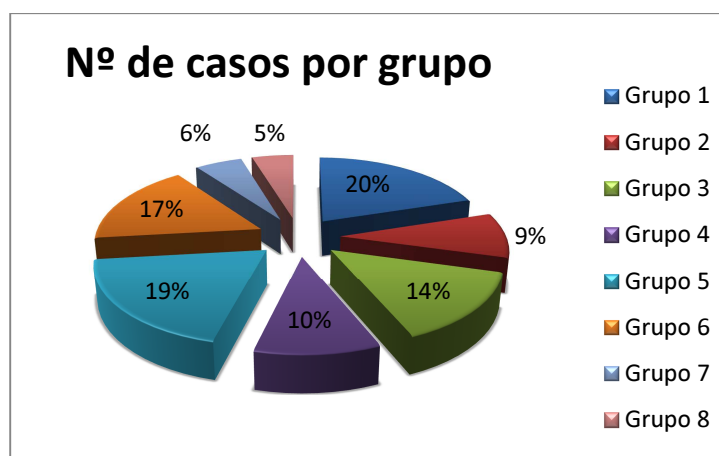
Figura 48: Regla del codo 2º K-medias



Fuente: "Elaboración propia"

Se obtuvo una segmentación con 8 clusters, con grupos de tamaños bastante parecidos, los cuales se pueden apreciar en la Figura 49, donde el más grande posee el 20 % de la población, mientras que el menor posee un 5% de ésta.

Figura 49: N° de casos por grupo



Fuente: "Elaboración propia"

### Grupo 1

- 47% hombres
- 48 años
- Valor del cliente 52 Uf
- 87% posee agente de mantención
- Sólo un 5% posee cobertura oncológica
- Sólo un 1% es propenso a cobertura de invalidez
- 75% propenso a cobertura oncológica
- 96% propenso a cobertura Salud
- 67% propensos a cobertura accidentes
- 40% pose sólo APV

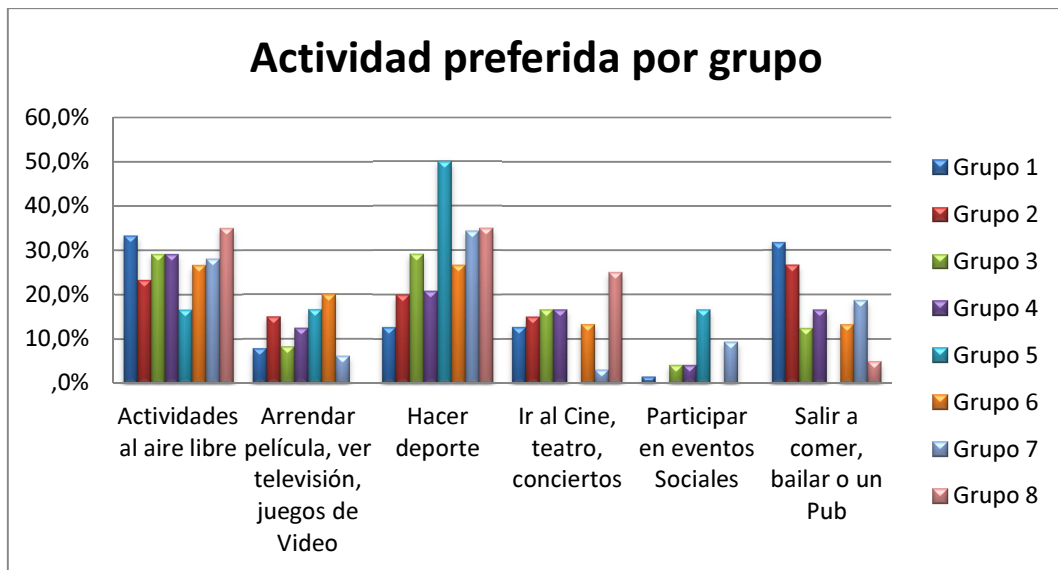
### Descripción a partir de datos encuesta $\varepsilon = 13\%$

El 76% de este grupo se encuentra casado, lo que lo transforma en uno de los grupos con mayor cantidad en esta condición. Su grupo familiar es un poco mayor al promedio.

El nivel educacional es bastante alto pero no sobresale por encima de los demás grupos.

Dentro de sus actividades preferidas se encuentran las actividades al aire libre, salir a un Pub, a comer o a bailar, donde son los principales realizadores de estas actividades en la base. Realiza la actividad una vez al mes y el resto solo una vez a la semana.

Figura 50: Actividad preferida por grupo



Fuente: "Elaboración propia"

Este conglomerado es el que menos utiliza el internet para informarse, el medio utilizado por éstos es el noticiero en la televisión abierta. El bajo uso de internet se ve reflejado en que es de los grupos en los que más de un 20% no ocupa ningún tipo de red social y la utilización del email es cercana al 50%.

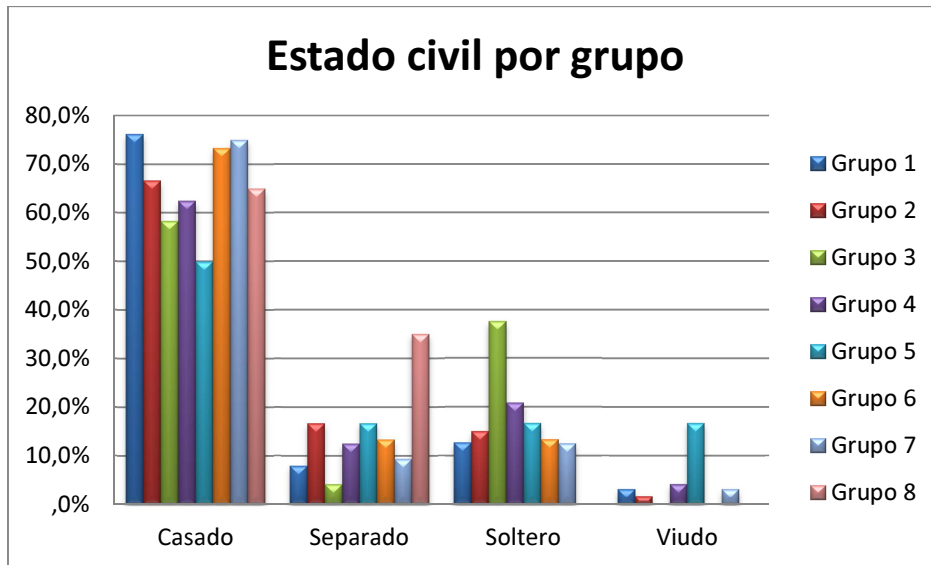
## Grupo 2

- 44% hombres
- 42 años
- Valor del cliente 62 Uf
- 63% posee agente de compañía
- 0,92 autos promedio
- Indiferentes a coberturas de Salud, Accidentes y gastos médicos adicionales
- 39% posee cobertura de salud
- 63% posee cobertura accidentes

## Descripción a partir de datos encuesta $\varepsilon = 13\%$

Un 16% de este grupo son personas separadas, lo cual es bastante alto, además el tamaño de su grupo familiar es menor que el promedio. El nivel educacional de este grupo es más bajo, se puede apreciar que es el grupo con menor cantidad de postgrado y el con mayor número de personas con solo enseñanza media completa. Esto se podría ver reflejado en niveles de ingreso no tan altos como en otros conglomerados.

Figura 51: Estado civil por grupo



Fuente: "Elaboración propia"

De igual manera que el clúster anterior, este grupo disfruta de salir a comer, bailar o asistir a algún pub, pero no gustan de actividades al aire libre, cabe destacar que disfrutan de actividades hogareñas como arrendar una película, ver televisión, los juegos de video, o salir al teatro, conciertos o el cine. Este tipo de actividades las realizan por lo general 1 vez a la semana.

Solo un 50% utiliza el internet como medio para informarse, tienen participaciones sobre el promedio en la utilización de periódicos o de la radio con este objetivo. Su uso de redes sociales es bastante promedio. De todas maneras es uno de los grupos que utiliza este medio para pagar servicios básicos.

### Grupo 3

- 62% hombres
- 37 años
- Valor del cliente 53 Uf
- 72% posee agente de compañía
- Todos son propensos a contratar cobertura en Gastos médicos
- 36% son propensos a compra de cobertura de invalidez

### Descripción a partir de datos encuesta $\varepsilon = 20\%$

Sólo el 58% de este grupo es casado, lo que es muy bajo en comparación al resto de los clientes, esta diferencia se explica porque es el grupo más joven que se encuentra

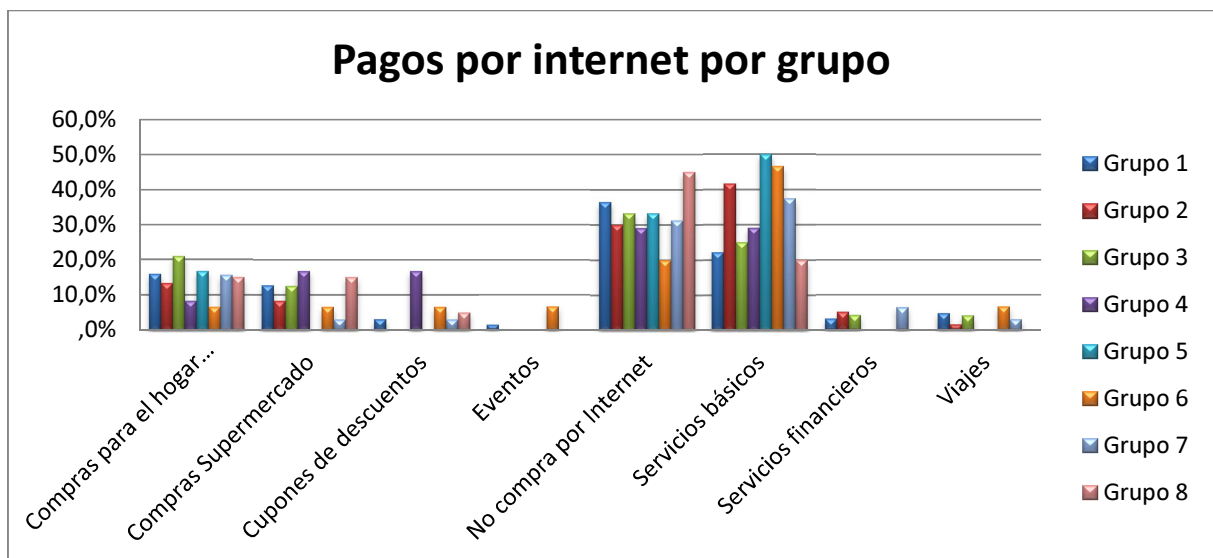
en otra etapa del ciclo de vida, donde un 38% son solteros. Esto no se ve del todo reflejado en el tamaño del grupo familiar que apenas es bajo el promedio.

Este grupo destaca por tener el mayor número de profesionales, pero no muchos postgrado, a pesar de ésto es un grupo de un alto nivel socioeconómico.

Dentro de sus actividades preferidas se pueden ver las realizadas al aire libre y los deportes, es un grupo bastante activo en relación a estas actividades ya que más de un 40% realiza esta actividad 3 o más veces en la semana.

Este grupo utiliza mayormente internet para informarse, con casi un 80%. Además es importante destacar que su red social favorita es el email con más de un 60 %, pero su nivel de uso es mayor al de los demás grupos ya que más del 80% lo utiliza varias veces durante el día. Este conglomerado no sólo usa el internet para mantenerse informado sino que también realiza compras para el hogar en un 20% como se aprecia en la Figura 52.

Figura 52: Pagos por internet por grupo



Fuente: "Elaboración propia"

#### Grupo 4

- 83% hombres
- 43 años
- Valor del cliente 70 Uf
- 65% posee agente de Compañía
- Poseen bastantes coberturas en accidentes y Gastos médicos
- Sólo poseen flexibles
- No están interesados en un nuevo producto

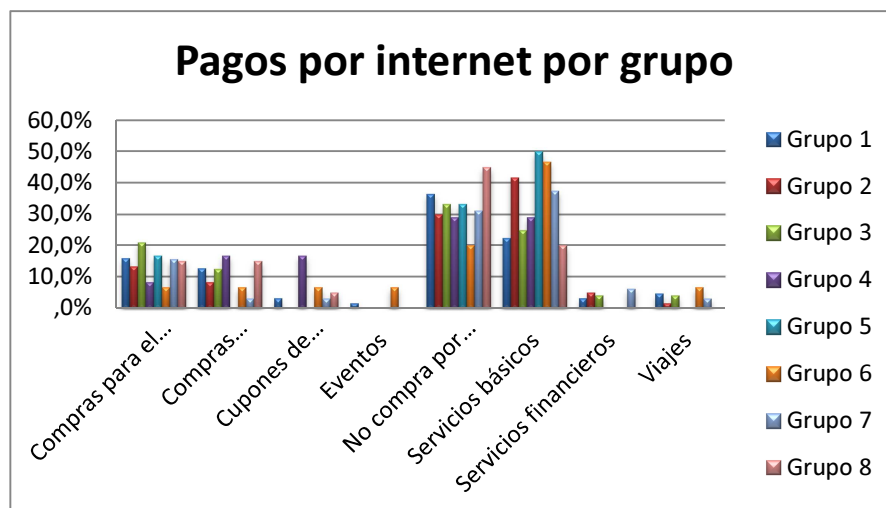
## Descripción a partir de datos encuesta $\varepsilon = 20\%$

Un 62,5% de este grupo son casados, lo que es bastante bajo. El tamaño del grupo familiar es un poco más bajo que el promedio y el nivel educacional es alto pero no más que el promedio.

En general disfrutan de las actividades al aire libre, pero también disfrutan de actividades como ir al cine, ver una obra de teatro o asistir a algún concierto de sus artistas favoritos. Estas actividades no las realizan muy a menudo ya que más de un 30% sólo las realiza una vez al mes o menos.

Tienen un alto uso del internet como medio para informarse, un 10% lo hace a través de los noticieros que ofrece la señal de cable, siendo estos uno de los principales clientes de estos canales. Este conglomerado se diferencia del resto por ser el de menor utilización del email, como red social preferida con sólo un 45%, siendo uno de los grupos en que existe gran parte que no utiliza ninguna red social. Pese a esto es el principal comprador de cupones de descuento y compras de supermercado mediante esta herramienta.

Figura 53: Pagos por internet por grupo



Fuente: "Elaboración propia"

### Grupo 5

- 77% hombres
- 49 años
- Valor del cliente 62 Uf

- 62% posee agente de mantención
- Mayor contacto 58%
- 1,35 autos promedio por persona
- Todos poseen Flexibles
- No les interesa otro producto
- Propensos a cobertura de Salud (90%) y oncológica (72%)

#### **Descripción a partir de datos encuesta $\varepsilon = 41\%$**

Dado que los datos recolectados en la encuesta para la descripción de este grupo generan un error demasiado alto como para poder extrapolar los datos de la muestra al resto del universo, estos no serán utilizados en el estudio.

#### **Grupo 6**

- 66% hombres
- 42 años
- Valor del cliente 49 Uf el más bajo
- 87% posee agente de compañía
- Menor contacto, sólo el 38%
- 50% posee sólo APV
- 9% propenso a contratar APV
- 7% propenso a contratar Protección
- 9% propenso a contratar Flexible
- Todos propensos a contratar cobertura de accidentes
- Indiferentes a cobertura de salud y oncológica
- Propensos a invalidez (46%)

#### **Descripción a partir de datos encuesta $\varepsilon = 26\%$**

El error para este grupo también es bastante alto, por lo que las características enunciadas a continuación sólo serán tomadas como tendencias de cómo es la realidad en este segmento.

Gran parte de los que pertenecen a este grupo son casados, poseen el mayor tamaño de grupo familiar y niveles de estudios bastante semejantes al resto de la población.

Sus actividades preferidas están dominadas por las realizadas en el hogar, como ver televisión, arrendar una película o disfrutar de los videos juegos, También disfrutaban del teatro, cine y actividades al aire libre. Por lo general realizan estas actividades una vez a la semana.

Junto con el primer grupo son los más asiduos a ver los noticieros que se exhiben en la televisión abierta. Un alto porcentaje usa el internet como principal medio de información.

En cuanto al uso de redes sociales se puede observar en Anexo D, que este segmento es bastante promedio tanto en el uso como en la frecuencia de estos servicios de internet. Ocupan bastante este medio para el pago de servicios básicos, pero también lo ocupan para comprar tickets de eventos o viajes.

### **Grupo 7**

- 91% hombres
- 44 años
- Valor del cliente 79 Uf, los más valiosos
- 74% posee agente de mantención
- Propensos a seguros de Protección
- 30% posee APV Flexible
- No destaca ninguna cobertura

### **Descripción a partir de datos encuesta $\varepsilon = 18\%$**

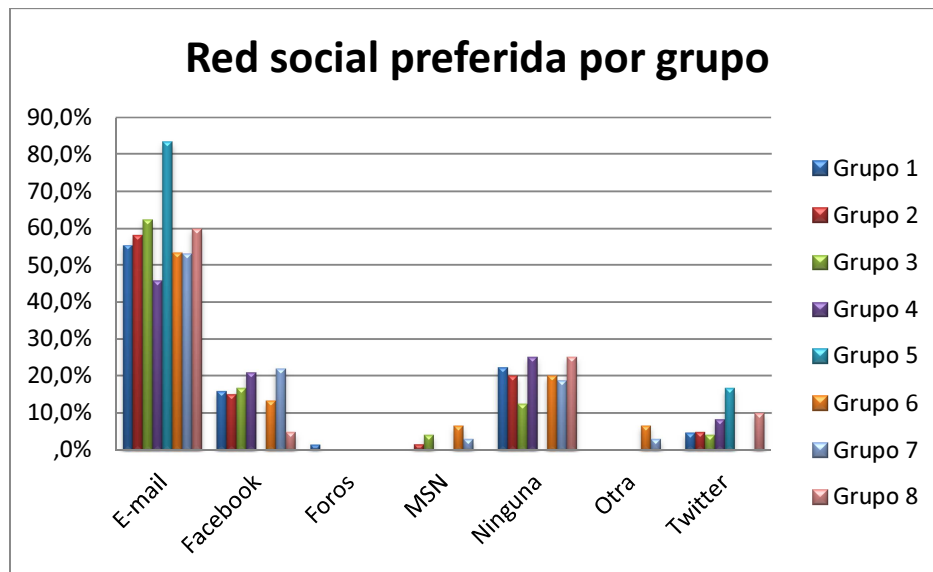
Más del 70% de los integrantes de este grupo son personas casadas, el tamaño del grupo familiar promedio de este grupo no se diferencia del resto, con menos de 4 integrantes, se puede apreciar que el número de personas con postgrado es bajo en comparación a los demás conglomerados.

Cómo actividades recreativas prefieren las que se puedan realizar al aire libre, incluyendo los deportes. En cuanto a la frecuencia de estas actividades es bastante alta, donde más de un 30% las realiza 3 o más veces a la semana.

Este grupo se caracteriza por ser el mayor lector de periódicos, pese a que la utilización del internet es su principal medio de información. A pesar de que utilizan bastante el email, son los principales usuarios de Facebook con un 20%.



Figura 54: Red Social preferida por grupo



Fuente: "Elaboración propia"

## Grupo 8

- 50% hombres
- 46 años
- Valor del cliente 67 Uf
- 92% posee agente de mantención
- Son propensos a cobertura de salud
- Indiferente al resto de las coberturas
- Poseen cobertura de accidentes (92%) y oncológica (66%)

## Descripción a partir de datos encuesta $\varepsilon = 22\%$

El error para este grupo es alto por lo que sólo se explicarán las características en que más se diferencie del resto de la base.

Se puede apreciar que un 30% de los pertenecientes a este clúster son personas separadas, pese a esto el porcentaje de personas casadas es superior al 60%. El tamaño del grupo familiar de este segmento es superior a los demás, muy cercano a 4 personas por familia. En cuanto el nivel educacional se puede ver que es el grupo con mayor número de carreras incompletas.

Dentro de sus preferencias a la hora de recrearse, se puede ver que son los que más le gustan de actividades al aire libre, ir al cine o teatro y además a una gran parte les

gusta hacer deporte. No son muy frecuentes en estas actividades, ya que la mayoría las realiza una vez al mes en promedio.

Es un grupo que ocupa bastante internet para informarse, pero a pesar de esto existe un alto porcentaje que no ocupa ninguna red social. Los que poseen redes sociales las ocupan frecuentemente. Es el grupo que menos compra por internet, diferenciándose de otros grupos por bastante ya que un 45% no realiza ningún tipo de compra.

### **3.11 RECOMENDACIONES DE ACCIONES COMERCIALES**

En esta sección se realizarán sugerencias de acciones comerciales para los diferentes perfiles de clientes encontrados y para los clientes pertenecientes a los diferentes clúster de la clasificación realizada en el apartado anterior.

#### **Grupo 1**

Este grupo consta de un 20% de la población, con un valor del cliente promedio de 52 UF. Es un grupo propenso a coberturas oncológicas y de salud, por lo que se espera realizar campañas que generen la contratación de estas coberturas, lo que aumentaría su valor en la compañía. El 87% de estos clientes poseen un agente de mantención, por lo que se deberá entrenar a este tipo de agentes para la venta de estas dos coberturas, de la manera más adecuada a esta descripción.

Con 48 años promedio es un grupo bastante mayor, por lo que se sugiere un enfoque menos juvenil, más serio y enfocado a los beneficios. Además este grupo está balanceado entre hombres y mujeres por lo que no se debe realizar distinciones de sexo en el enfoque de la campaña.

No se debería ofrecer coberturas de invalidez ya que según el resultado de los modelos aplicados, la probabilidad de que éstos la contraten es bajísima.

Este grupo no es un gran usuario del internet, por lo que no debería usarse como medio de difusión en este segmento. Como incentivo para este grupo se pueden utilizar descuentos para restaurantes y pub debido a que disfrutan de estos lugares.

#### **Grupo 2**

Es un grupo joven, su valor del cliente es de 62 UF, prácticamente lo mismo que el promedio. Es un grupo que se ha desarrollado bastante dentro de la compañía, ya que poseen gran cantidad de clientes con más de un producto y que además poseen coberturas de Salud y Accidentes, por lo que no son especialmente propensos a alguno de los productos.

Se recomienda realizar acciones de fidelización más que de desarrollo de estos clientes. Es un grupo Joven y se espera poder mantenerlos la mayor cantidad de años posibles y que empiecen a contratar nuevos productos a medida que estos avancen en su ciclo de vida.

Como campaña de fidelización se podrían ofrecer pequeños descuentos en compras para el hogar o para actividades como restaurantes, pubs y discoteck.

Debido al pequeño tamaño familiar y el gran porcentaje de personas separadas, las acciones no podrán ser enfocadas en la familia.

### **Grupo 3**

Es el grupo más joven con 37 años promedio, tienen un valor para la empresa de solo 53 UF lo que lo hace ser un grupo con alto potencial. Son propensos a coberturas en gastos médicos y cobertura de invalidez, por lo tanto ésta es una oportunidad de negocio que se deberá aprovechar.

Con el fin de acceder a este segmento se recomiendan incentivos tecnológicos, debido a que es un grupo más joven. Las ofertas deberán ser enfocadas a partir de los agentes de compañía, ya que estos son los que generan en mayor parte la relación con la compañía.

Para poder mejorar el desarrollo de este grupo será importante estar atento a su avance dentro del ciclo de vida, para ofrecerle los productos más adecuados en cada momento.

El desarrollo de esta oferta dependerá fuertemente de internet, por lo que la construcción adecuada de página web será primordial. También deberá ser acompañada por una campaña de mailing.

Este segmento es el que va más atrasado en el ciclo de vida de todos los clientes, por lo que se podría concluir que las personas con estas características podrían comenzar a contratar los productos de la empresa, es por esto que se recomienda hacer publicidad masiva en eventos deportivos en estratos sociales altos, para tratar de buscar la adquisición de nuevos clientes.

### **Grupo 4**

Este segmento representa el 10% de la población, es un grupo muy valioso para la compañía, con 70 UF promedio de valor del cliente. El 83 % son hombres, clientes de seguros flexibles y poseen alta participación en coberturas de accidentes y gastos

médicos. No están interesados en productos adicionales. Por esto se recomienda realizar acciones de fidelización, hacer esfuerzos adicionales por mejorar el servicio hacia ellos y ofrecer incentivos sin la necesidad de que contraten algo de por medio. Debe ser enfocado el servicio para hombres a través de los agentes de compañía.

Como incentivo se pueden ofrecer descuentos para el cine, obras de teatro o conciertos.

### **Grupo 5**

Es el grupo de mayor edad, con 49 años promedio y un valor del cliente de 62 UF. Este grupo no está interesado en nuevos productos, pero si en complementar su producto con coberturas de salud y oncológica.

Se le recomienda la oferta exclusiva de estas dos coberturas, de manera adecuada para hombres alrededor de 50 años ABC1.

### **Grupo 6**

Este es un grupo un tanto más joven con 42 años promedio, donde el 66% son hombres. Es el clúster de menor valor para la compañía pero con un alto potencial de desarrollo, ya que existen clientes de este segmento con intereses por cada uno de los productos. Además todos poseen la necesidad de cobertura por accidentes; por otra parte no están interesados en coberturas de salud y Oncológica.

Se recomienda la oferta de las coberturas antes mencionadas y cómo complemento algún producto con el enfoque en la cobertura. Las campañas televisivas deberían enfocarse en este segmento, ya que es el que más sintoniza los noticiarios.

### **Grupo 7**

Se recomienda realizar una campaña inspirada en los hombres de familia. Se deben realizar acciones de fidelización, por su alto valor para la compañía y por qué no son propensos a la compra de un nuevo producto.

Son personas sanas que realizan bastante deporte y disfrutan las actividades al aire libre. Para poder acceder a este grupo, Facebook es una buena opción. Ya que es de los de mayor utilización de esta red social.

## **Grupo 8**

Es un grupo bastante valioso para la compañía, con 67 UF. Es equilibrado entre hombres y mujeres. Además este conglomerado tiene potencial de desarrollo ya que son propensos a comprar coberturas de salud.

La manera más fácil de acceder a este grupo es a través de los agentes de mantención, por lo que es perentorio el entrenamiento de estos funcionarios para la venta de esta cobertura.

## **4 CONCLUSIONES**

### **4.1 PROBLEMA DE NEGOCIO**

El objetivo central de esta memoria fue diseñar una metodología de segmentación que permitiese apoyar la toma de decisiones de la empresa, que aumentaran la eficiencia y eficacia de las acciones comerciales que se realizan cada año a los clientes de la compañía. Se buscó encontrar las variables adecuadas que permitan la diferenciación de los clientes y definir un perfil de los clientes propensos a la compra de algún producto.

Este trabajo comienza con una parte de carácter descriptivo, que a través de los datos de la empresa y los levantados en el estudio de mercado se pretende conocer de la manera más completa posible las diferencias que existen entre los clientes de cada uno de los productos, donde se encuentran los tres productos en estudio más las 5 coberturas. Esto permitirá desarrollar protocolos de atención diferenciado de acuerdo a los diferentes productos.

En segunda instancia, se aplicaron modelos de propensión de compra y métodos de segmentación, todo dentro del ámbito del marketing cuantitativo. A partir de estos métodos, se han encontrado grupos de clientes que poseen una alta probabilidad de comprar alguno de los diferentes productos. Esto a partir de variables socio demográficas que la empresa ha levantado durante los años. De esta manera es posible realizar una atención diferenciada para los clientes de cada uno de los productos, que mejore la satisfacción de cada uno de ellos respecto de la organización.

Como resultado final se encuentra una clasificación de los clientes a partir de las variables de propensión de compra y de los productos que poseen. Se encontraron 8 segmentos diferentes, donde 6 de estos tienen propensión a comprar un producto adicional, por lo que se podrían desarrollar. Mientras que los otros dos no están interesados en más productos pero son muy valiosos para la empresa, por lo que será primordial la mejora del servicio con el fin de fidelizarlos y retenerlos el mayor tiempo posible. Cada uno de estos segmentos ha sido descrito con las características de los productos contratados, variables sociodemográficas y una introducción a sus estilos de

vida, lo que permitirá diferenciar la atención y las ofertas a partir de un conocimiento amplio del cliente objetivo.

## **4.2 METODOLOGÍA**

En relación a la metodología que se ha desarrollado en este trabajo de título, es importante destacar el análisis exploratorio, tanto de la base completa, como al ser dividida por la tenencia de productos. Este análisis previo ha permitido obtener conclusiones valiosas de los resultados de la segmentación final.

Los modelos de propensión de compra aplicados no lograron niveles de calidad muy altos, con un índice de desempeño global promedio cercano al 70%, esto se puede explicar por la baja cantidad de información que se posee de los clientes, lo que no permite tener niveles de predicción muy altos.

Dentro de los métodos utilizados, el de mejor desempeño fue el árbol de clasificación C.5., pese a esto, es importante recalcar que la calidad de los otros métodos en varios casos fue similar y se prefirió este modelo por la información adicional que se obtiene al crear reglas de clasificación que entregan claves desde el punto de vista del negocio.

La segmentación realizada fue a partir de la tenencia y propensión de productos. Esto ha permitido clasificar a los clientes por sus necesidades, lo que permitirá realizar estrategias a cada uno de los segmentos con una oferta en particular, en la cual deberían tener una mayor probabilidad de compra, lo que aumentará la efectividad de las ofertas tanto de up selling como de cross selling.

## **4.3 TRABAJOS FUTUROS**

Luego del desarrollo de la metodología actual y en base a los resultados de ésta, se puede recomendar el diseño de investigaciones futuras, que mejoren o complementen lo ya realizado.

- Uno de los puntos débiles de los resultados actuales tiene referencia con el desempeño de los modelos predictivos, lo cual está muy correlacionado con la falta de información de los clientes como personas. Ante esto se propone replicar la metodología una vez que la organización haya levantado más información.
- Sería interesante poder replicar la metodología para todos los clientes de la empresa, abarcando los diferentes productos que ellos ofrecen. Para esto necesitarán hacer mayores esfuerzos por poblar las bases de datos existentes.
- A partir de lo anterior y con el fin de cuantificar el aporte de una segmentación de este tipo, se propone realizar una campaña comercial, que permita analizar la mejora real en la efectividad como en la eficiencia.

## 5 BIBLIOGRAFÍA Y FUENTES DE INFORMACIÓN

[I] AAKER, D.A. y DAY, G.S. 1989. Investigación de Mercados. 3ª ed. México, Mc Graw Hill.

[II] ALBARRAN, LOZANO. 2001. El perfil del consumidor en el sector asegurador español. Universidad complutense de Madrid. Madrid, España.

[III] PRINCIPIOS DE MARKETING. 2006. Por Talaya Agueda "et al". 2ª ed. Madrid, Esic.

[IV] BERNAL, MARILYN. 2005. Procedimiento para desarrollar estudios de mercados en organizaciones.

[V] CONSORCIO [en línea] [www.consorcio.cl](http://www.consorcio.cl) [consulta: 05 Octubre 2010]

[VI] GARRIDO, MARCEL. 2009. Estimación del valor de clientes de fidelización de una tienda de retail. Santiago, Chile.

[VII] LÉVY, JEAN. VARELA, JÉSUS. 2003. Análisis multivariable para las ciencias sociales. 1ª ed. Prentice Hall.

[VIII] MALLEA, V. 2004. Aplicación de los modelos de clase latente a la segmentación: Caso Economax. Memoria para otra al título de ingeniero civil industrial. Santiago, Universidad de Chile.

[IX] MENA, SEBASTIÁN. 2008. Clasificación de shoppers de una cadena de supermercados. Santiago, Chile.

[X] NÚÑEZ, S. 2010. Segmentación de clientes de una cadena de supermercados en base a estilos de vida. Memoria para optar al título de ingeniero civil industrial. Santiago, Universidad de Chile.

[XI] PARRAGUEZ, M. 2010. Generación de ofertas focalizadas para clientes de baja vinculación en una institución bancaria. Memoria para otra al título de ingeniero civil industrial. Santiago, Universidad de Chile.

[XI] REYES, MANUEL. 2009. Apuntes del Curso "Métodos estadísticos para economía y gestión". Universidad de Chile. Santiago, Chile.

[XII] SUPERINTENDENCIA DE VALORES Y SEGUROS [en línea] [www.svs.cl](http://www.svs.cl) [consulta: 20 Noviembre 2010]

[XIII] TEPPOLA, PEKKA. MINKKINEN, PENTTI. 1999. Possibilistic and fuzzy C-means clustering for process monitoring in an activated sludge waste-water treatment plant. Journal of chemometrics. Volumen 13: 3-4,445-459.

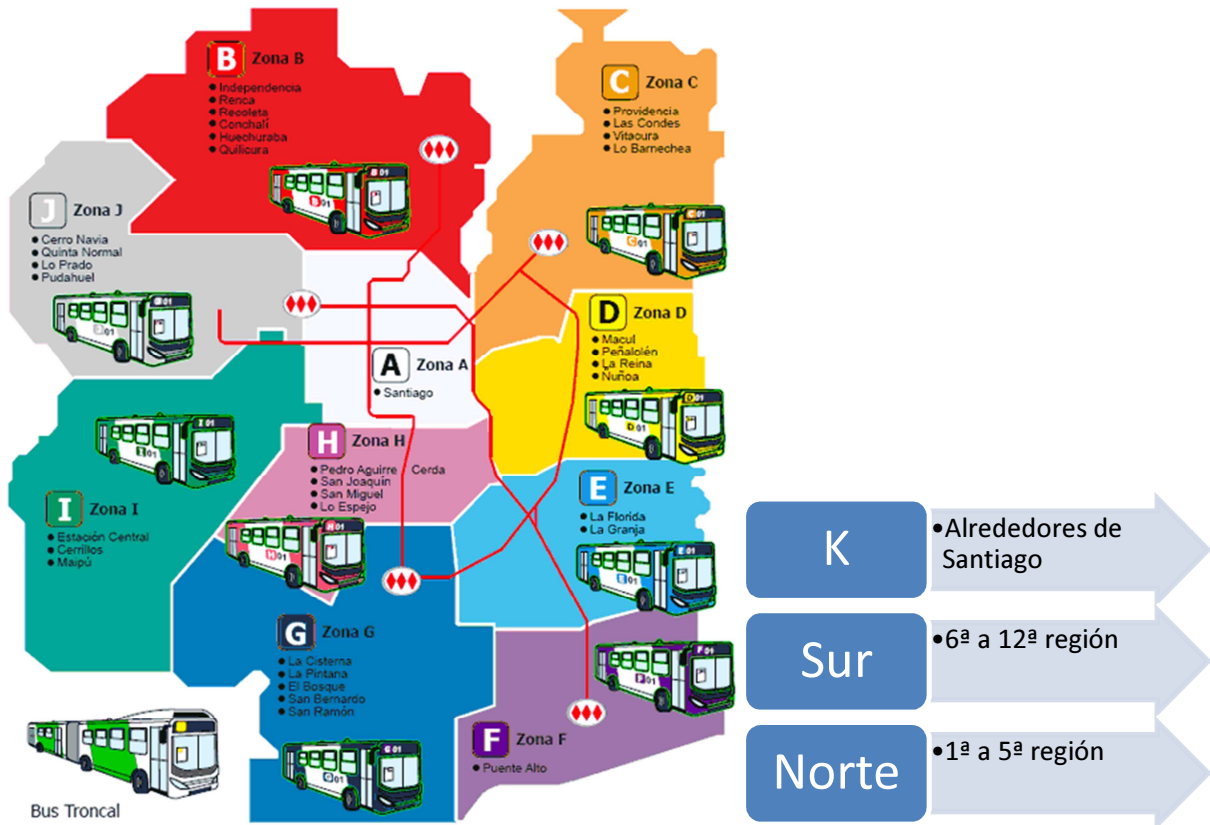
[XIV] VIDELA, MARIA INES. 2011. Metodología para diseñar acciones de retención de clientes no contractuales en una empresa de retail. Memoria para otra al título de ingeniero civil industrial. Santiago, Universidad de Chile.



## 6 ANEXOS

### 6.1 ANEXO A- Sectores de residencia de clientes

Figura 55: Mapa de la división zonal de la ciudad de Santiago



Fuente: "Elaboración propia"

## 6.2 ANEXO B - Encuesta

Cuestionario de estudio de la cartera de clientes de Consorcio

Introducción:

Buenos días (tardes), mi nombre es XX, ejecutivo de Consorcio,  
Con el Objetivo de brindar a Ud\_ Un mejor servicio, Nos interesa responder las siguientes preguntas:

1. ¿Es usted?
  - a) Soltero
  - b) Casado
  - c) Separado
  - d) Viudo
  
2. ¿Cuál es su nivel de educación?
  - a) Básica incompleta
  - b) Básica completa
  - c) Media incompleta
  - d) Media completa
  - e) Técnica o profesional incompleta
  - f) Técnica o profesional completa
  - g) Postgrado
  
3. ¿Cuál es su actividad laboral?
  
4. ¿Cuál de las siguientes actividades usted prefiere?
  - a) Salir a comer, bailar o a un Pub
  - b) Arrendar película, ver televisión, juegos de video
  - c) Participar en eventos sociales
  - d) Actividades al aire libre
  - e) Ir al cine, teatro, conciertos
  - f) Hacer deporte
  
5. ¿Con que frecuencia realiza esta actividad?
  
6. ¿Cuál es el medio favorito para informarse?
  - a) Periódicos
  - b) Internet
  - c) Noticiero en Cable
  - d) Noticiero en TV abierta
  - e) Radio
  - f) Redes sociales
  
7. ¿Qué compra o paga por internet usualmente?
  - a) Servicios básicos

- b) Compras supermercado
- c) Compras para el hogar o tecnológicas
- d) Servicios financieros
- e) Viajes
- f) Cupones de descuento
- g) Eventos
- h) No compra por internet

8. ¿Qué red social es la que más ocupa?

- a) Twitter
- b) Facebook
- c) E-mail
- d) Foros
- e) MSN
- f) Otra
- g) Ninguna

9. ¿Cada cuánto la ocupa?

- a) 1 o más veces al día
- b) 2 a 4 veces por semana
- c) 2 o tres veces por mes
- d) Una vez al mes o menos
- e) Nunca

10. ¿Cuál es el número de integrantes de su grupo familiar?

11. y 12 Se preguntará según la siguiente tabla

Tabla 38: Preguntas de la encuesta

Si tiene	Pregunta
<b>Apv o Protección</b>	¿Usted posee algún seguro Flexible?
<b>Flexible o Protección</b>	¿Usted posee algún APV?
<b>Flexible o Apv</b>	¿Usted posee algún seguro de Protección?

Fuente: "Elaboración propia"

Protocolo de despedida de la empresa.

### 6.3 ANEXO C – Análisis de arboles

Tabla 39: Aplicación de Chaid

Resumen del modelo		
Especificaciones	Método de crecimiento	CHAID
	Variable dependiente	productos
	Variables independientes	Sector, Sexod, Clausulas, Contacto, ACompañia, CobOnco, CobAccidentes, CobGM, CobInv, CobSalud, REGR factor score 1 for analysis 2, REGR factor score 2 for analysis 2, REGR factor score 3 for analysis 2
	Validación	Dividir muestra
	Máxima profundidad de árbol	3
	Mínimo de casos en un nodo filial	100
	Mínimo de casos en un nodo parental	50
Resultados	Variables independientes incluidas	CobAccidentes, REGR factor score 1 for analysis 2, ACompañia, CobInv, REGR factor score 3 for analysis 2
	Número de nodos	15
	Número de nodos terminales	8
	Profundidad	3

Fuente: "Elaboración propia"

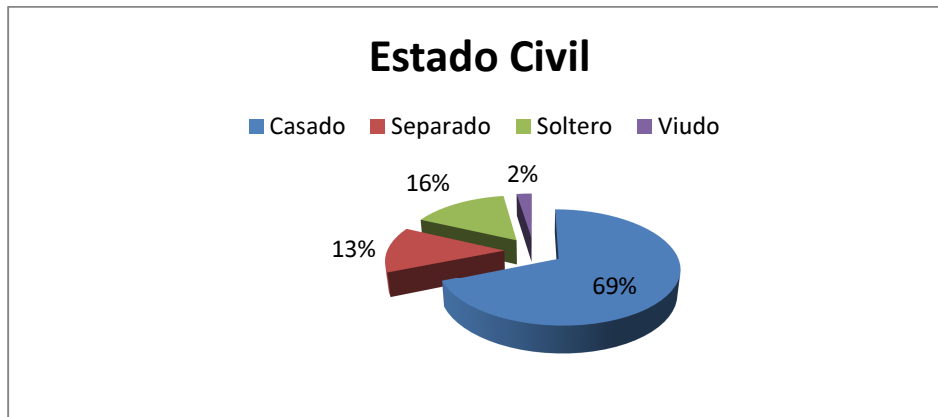
Tabla 40: Aplicación de Quest

Resumen del modelo		
Especificaciones	Método de crecimiento	CRT
	Variable dependiente	productos
	Variables independientes	Sector, Sexod, Clausulas, Contacto, ACompañia, CobOnco, CobAccidentes, CobGM, CobInv, CobSalud, REGR factor score 1 for analysis 2, REGR factor score 2 for analysis 2, REGR factor score 3 for analysis 2
	Validación	Dividir muestra
	Máxima profundidad de árbol	3
	Mínimo de casos nodo filial	100
	Mínimo de casos nodo parental	50
Resultados	Variables independientes incluidas	REGR factor score 3 for analysis 2, CobAccidentes, REGR factor score 2 for analysis 2, REGR factor score 1 for analysis 2, CobSalud, CobGM, CobInv, Sector, CobOnco, Clausulas
	Número de nodos	15
	Número de nodos terminales	8
	Profundidad	3

Fuente: "Elaboración propia"

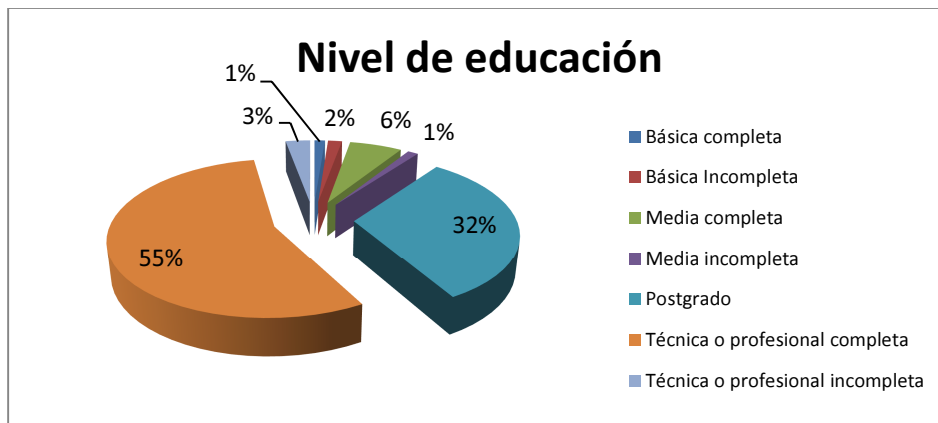
## 6.4 ANEXO D – Resultados Encuesta

Figura 56: Estado Civil de la muestra



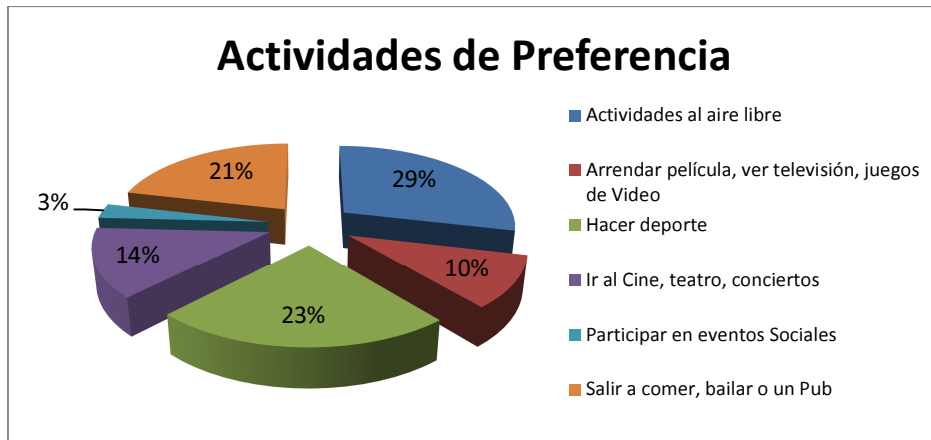
Fuente: "Elaboración propia"

Figura 57: Nivel educacional de la muestra



Fuente: "Elaboración propia"

Figura 58: Actividades de preferencia



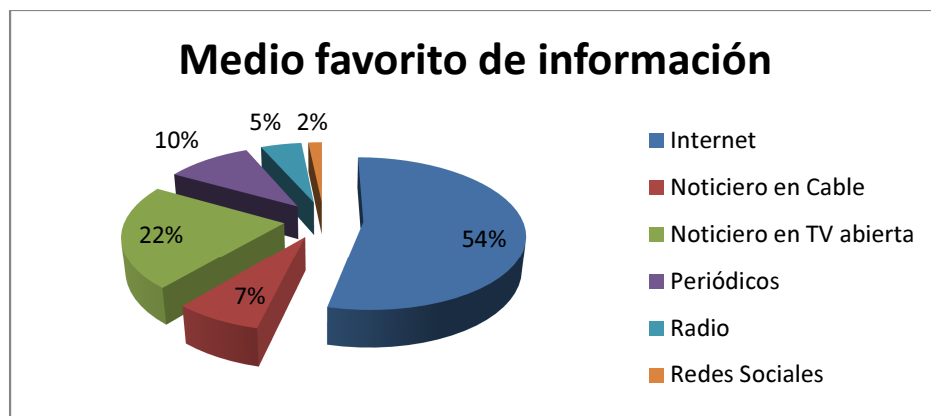
Fuente: "Elaboración propia"

Figura 59: Frecuencia de actividades de preferencia



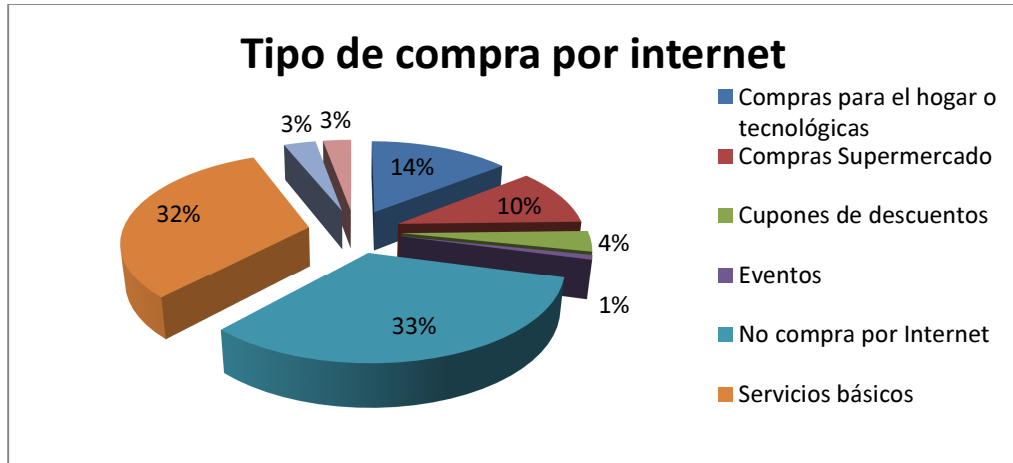
Fuente: "Elaboración propia"

Figura 60 Medio favorito de información



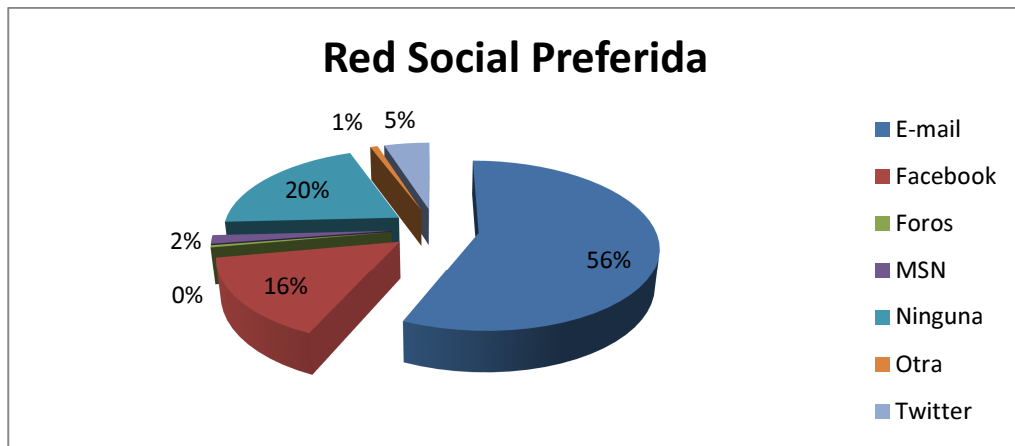
Fuente: "Elaboración propia"

Figura 61: Tipo de compra por internet



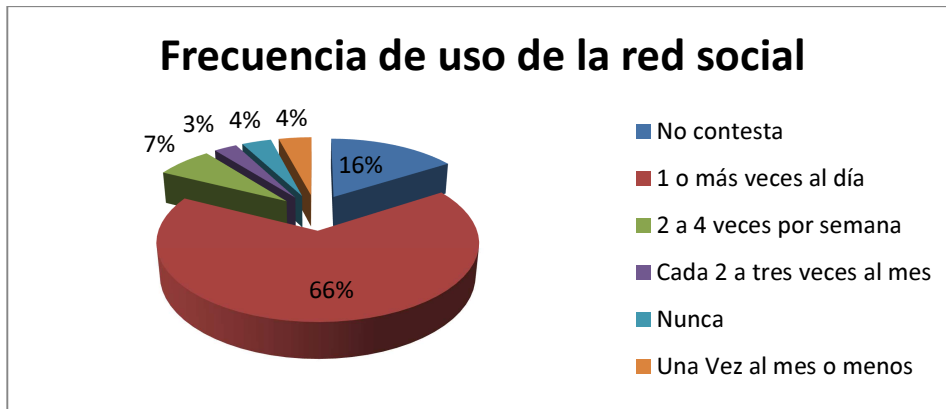
Fuente: "Elaboración propia"

Figura 62: Red Social Preferida



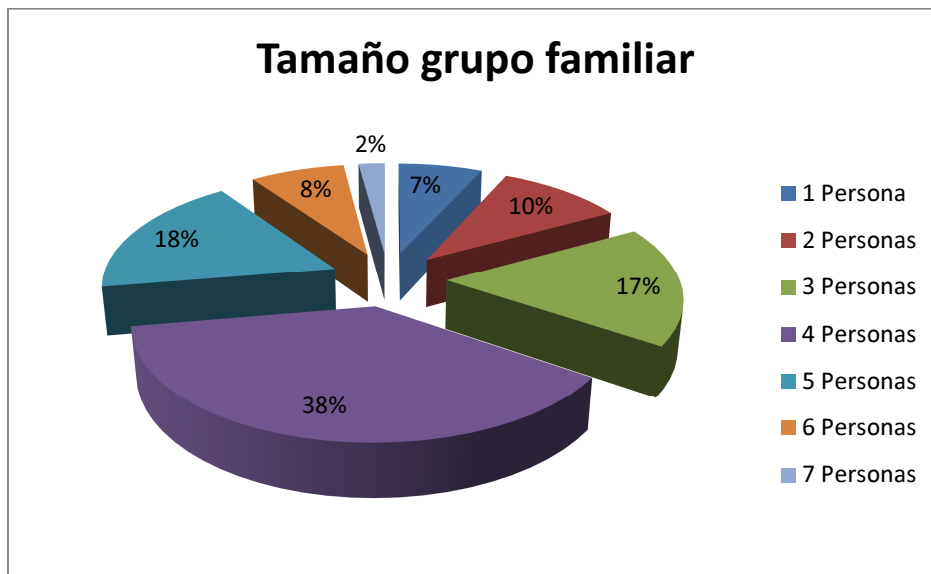
Fuente: "Elaboración propia"

Figura 63: Frecuencia de uso de la red social



Fuente: "Elaboración propia"

Figura 64: Tamaño del grupo familiar



Fuente: "Elaboración propia"

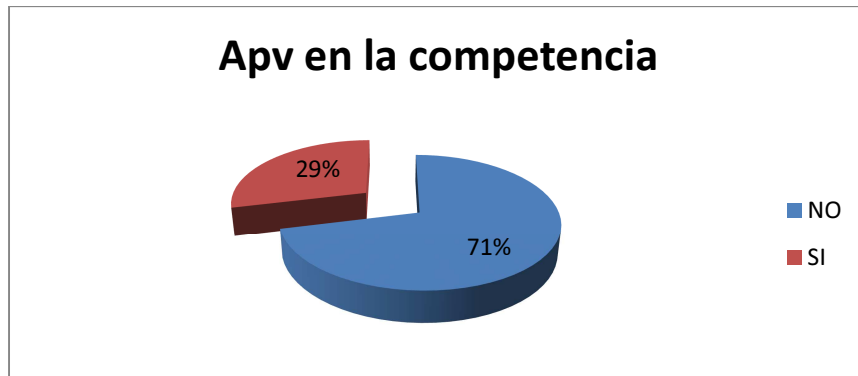


Figura 65: Principales competencias

	Frecuencia	Porcentaje	Porcentaje acumulado
Afp Habitat	4	1,6	1,6
AFP Provida	1	,4	2,0
Afp, Vida corp	1	,4	2,5
Banco de chile	6	2,5	4,9
Banco Estado	1	,4	5,3
Bancos Scotiabank y Desarrollo	1	,4	5,7
BCI	5	2,0	7,8
bicevida	1	,4	8,2
Celfin Capital	2	,8	9,0
Chilena Consolidada	7	2,9	11,9
Concha y toro	1	,4	12,3
Corp banca	1	,4	12,7
Cruz del Sur	1	,4	13,1
Cuprum	10	4,1	17,2
en la Armada de Chile	1	,4	17,6
escandia	1	,4	18,0
Euamerica	1	,4	18,4
ING	10	4,1	22,5
interamericana	3	1,2	23,8
Interamericana	1	,4	24,2
Larrain vial	1	,4	24,6
Metlife	3	1,2	25,8
mutual de seguridad	2	,8	26,6
NA	160	65,6	92,2
Paris	3	1,2	93,4
Penta	2	,8	94,3
PRINCIPAL	1	,4	94,7
Santander	7	2,9	97,5
SECURITY	6	2,5	100,0
Total	244	100,0	

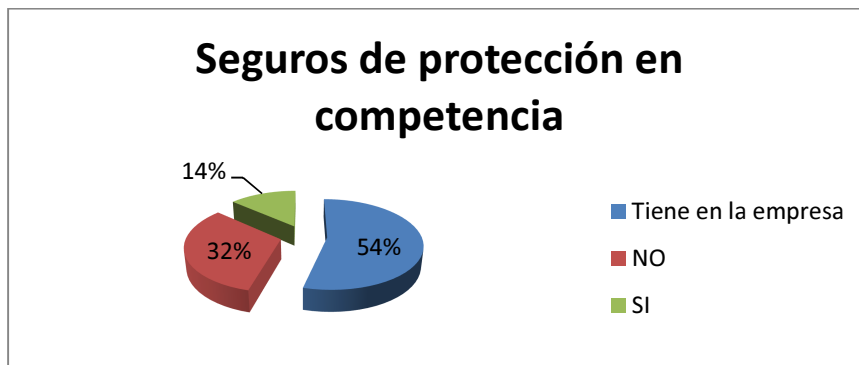
Fuente: "Elaboración propia"

Figura 66: APV en la competencia



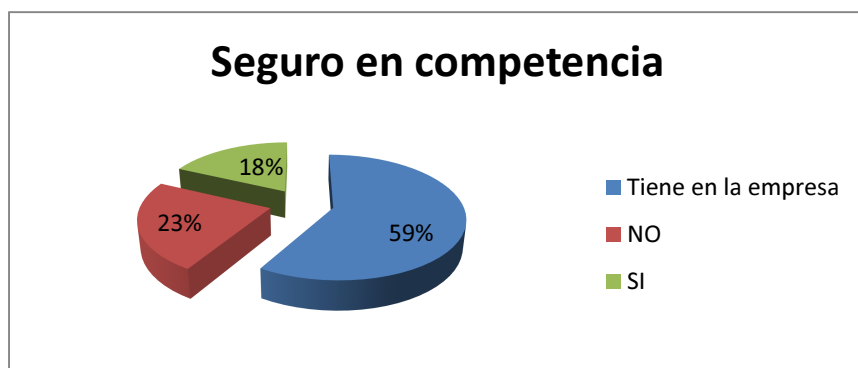
Fuente: "Elaboración propia"

Figura 67: Seguros de protección en la competencia



Fuente: "Elaboración propia"

Figura 68: Seguros en competencia

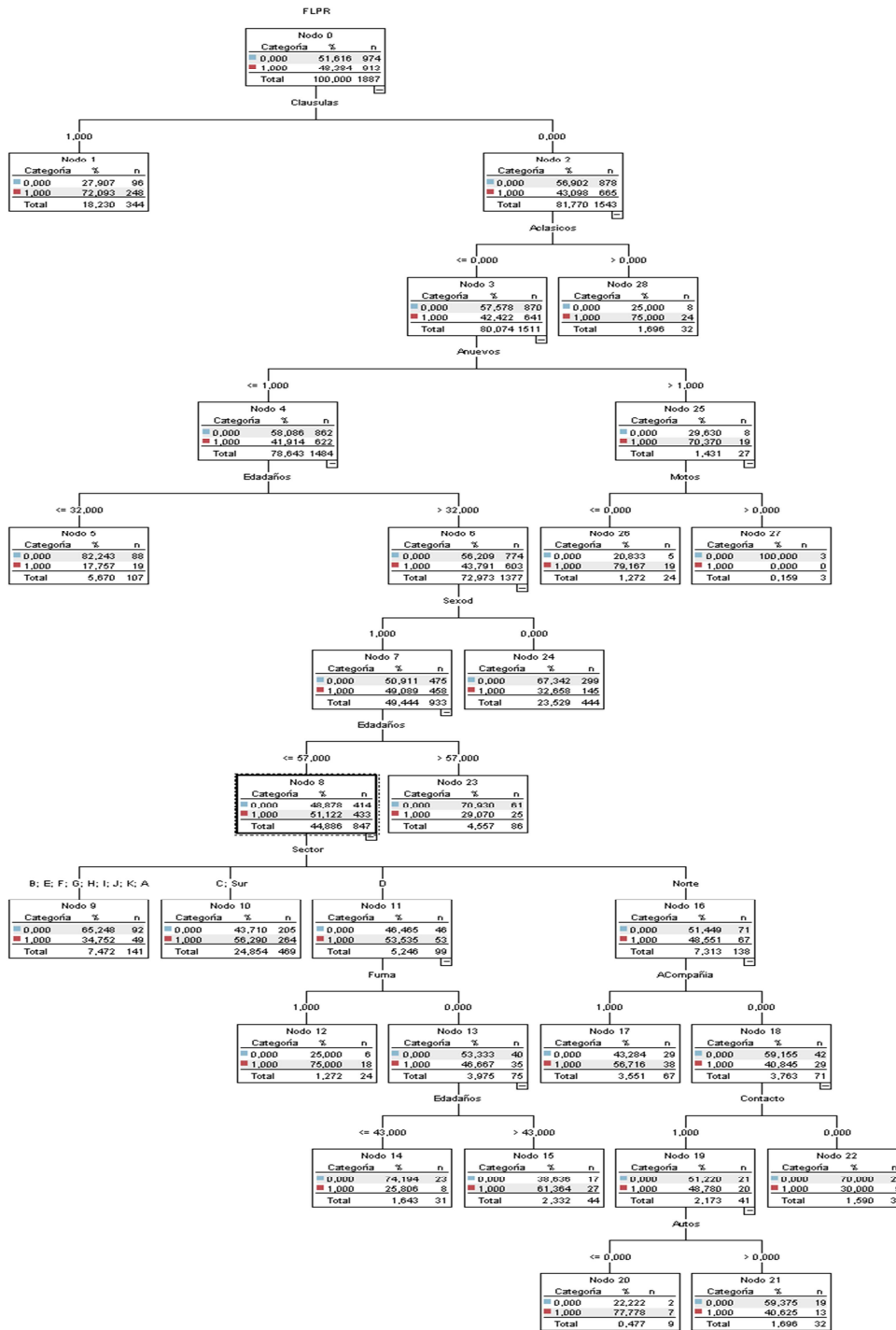


Fuente: "Elaboración propia"

## 6.5 ANEXO E- Árboles de clasificación

### 6.5.1 Flexible protección

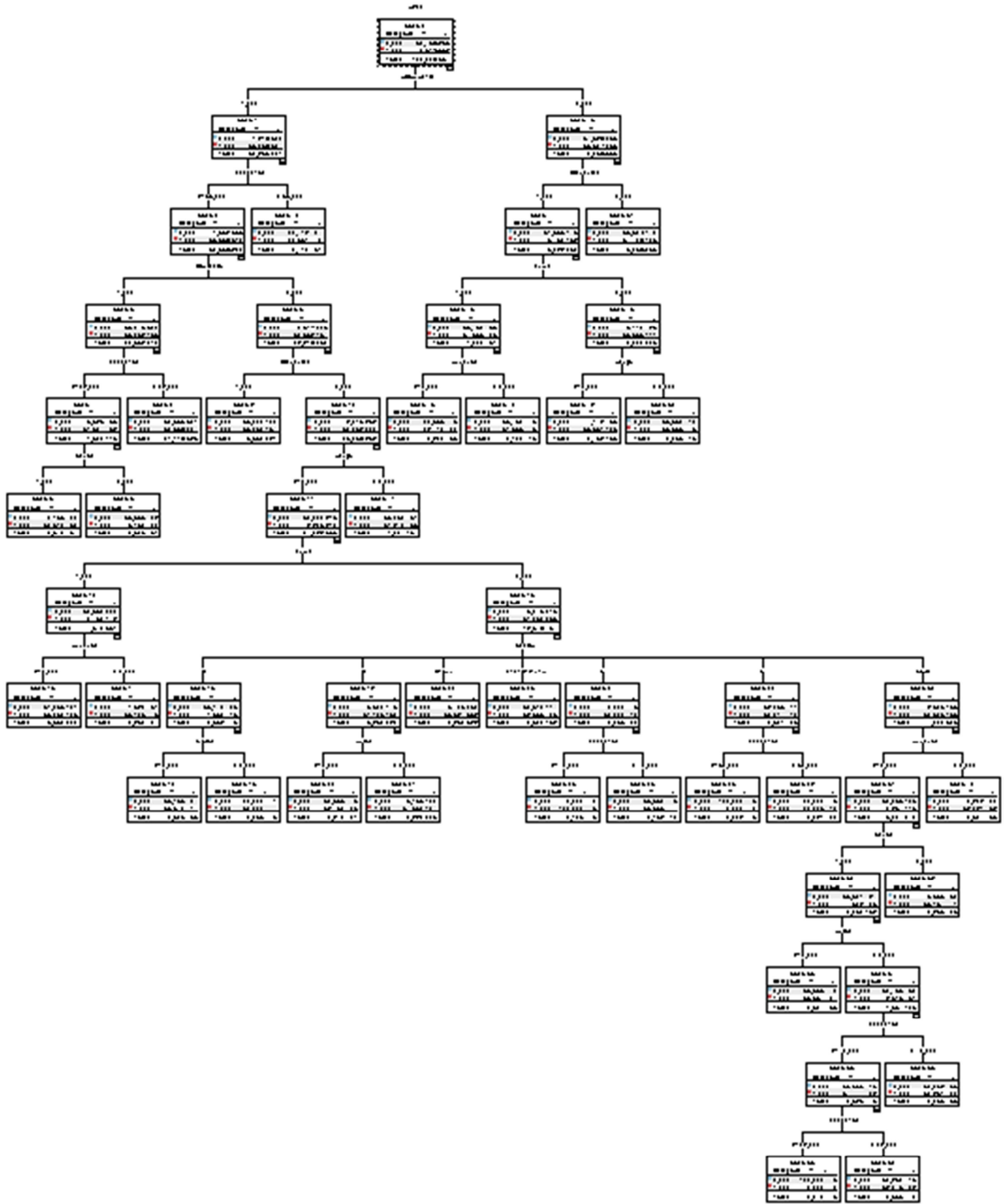
Figura 69: Árbol flexible protección



Fuente: "Elaboración propia"

## 6.5.2 APV Flexible

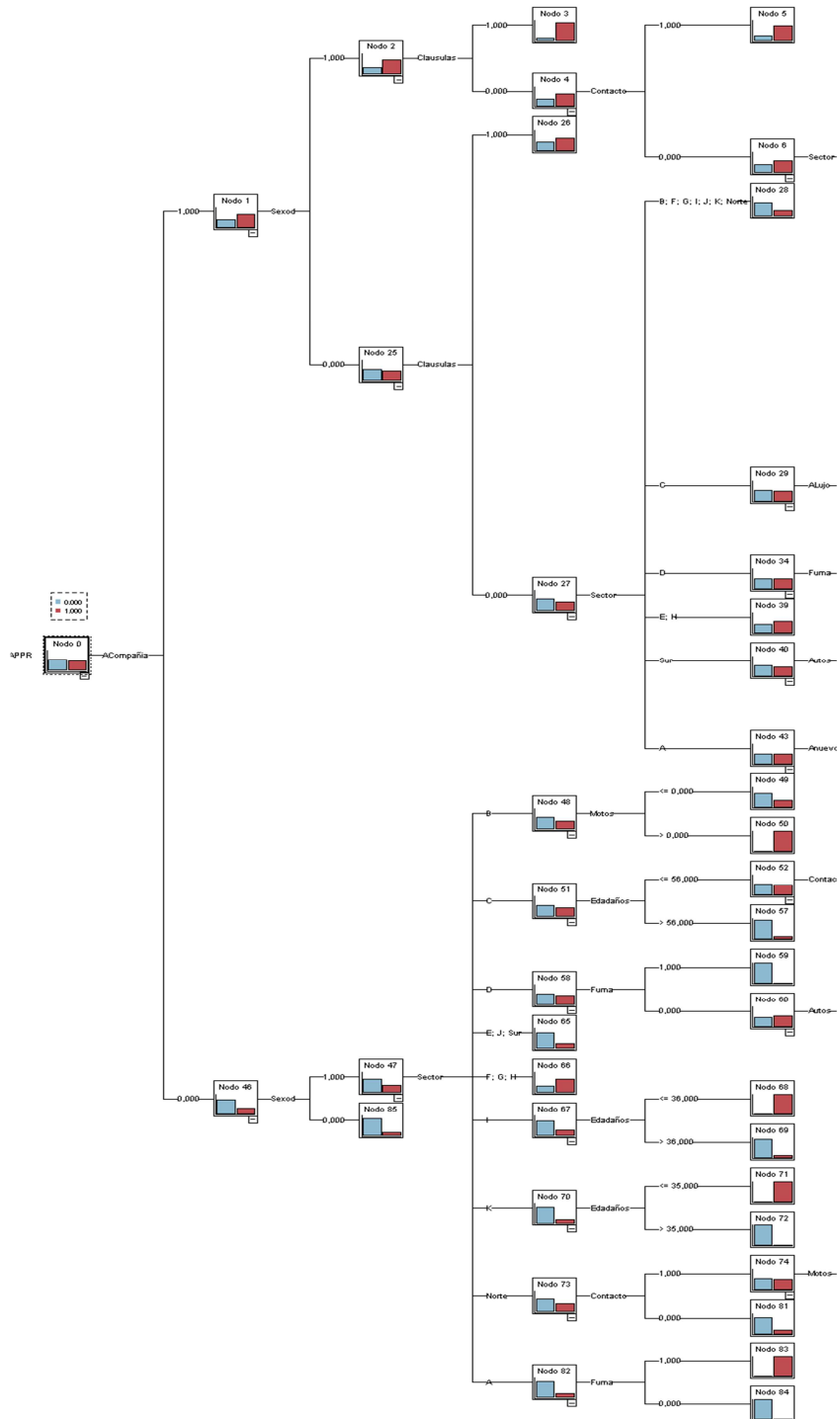
Figura 70: Árbol APV-Flexible



Fuente: "Elaboración propia"

### 6.5.3 APV- Protección

Figura 71: Árbol APV-Protección



Fuente: "Elaboración propia"

## 6.6 ANEXO F – Anova tenencia de productos

Tabla 41: Anova por posesión de productos

ANOVA DE TENENCIA DE PRODUCTOS						
		Suma de cuadrados	gl	M.C.	F	Sig.
Sexod	Inter-grupos	152,047	6	25,341	108,383	2,690E
	Intra-grupos	11489,419	49140	,234		
	Total	11641,466	49146			
Clausulas	Inter-grupos	135,158	6	22,526	275,717	,000
	Intra-grupos	4014,778	49140	,082		
	Total	4149,936	49146			
Contacto	Inter-grupos	61,456	6	10,243	41,492	9,312E-51
	Intra-grupos	12130,670	49140	,247		
	Total	12192,126	49146			
CobOnco	Inter-grupos	1171,894	6	195,316	1053,880	,000
	Intra-grupos	9107,116	49140	,185		
	Total	10279,010	49146			
CobAccidentes	Inter-grupos	4385,661	6	730,943	4789,446	,000
	Intra-grupos	7499,523	49140	,153		
	Total	11885,183	49146			
CobGM	Inter-grupos	403,628	6	67,271	522,020	,000
	Intra-grupos	6332,534	49140	,129		
	Total	6736,162	49146			
CobInv	Inter-grupos	900,098	6	150,016	683,170	,000
	Intra-grupos	10790,580	49140	,220		
	Total	11690,678	49146			
CobSalud	Inter-grupos	698,464	6	116,411	785,276	,000
	Intra-grupos	7284,598	49140	,148		
	Total	7983,061	49146			
Autos	Inter-grupos	113,109	6	18,851	16,713	2,304E-19
	Intra-grupos	55428,797	49140	1,128		
	Total	55541,906	49146			
ALujo	Inter-grupos	9,335	6	1,556	16,797	1,809E-19
	Intra-grupos	4551,883	49140	,093		
	Total	4561,218	49146			
SumaDepanua l	Inter-grupos	1,351E+07	6	2,252E+06	817,579	,000
	Intra-grupos	1,354E+08	49140	2754,976		
	Total	1,489E+08	49146			
Valor_Cliente	Inter-grupos	7,134E+06	6	1,189E+06	340,797	,000
	Intra-grupos	1,714E+08	49140	3488,897		
	Total	1,786E+08	49146			
Edad	Inter-grupos	21371,838	6	3561,973	38,459	6,868E-47
	Intra-grupos	4,551E+06	49140	92,618		
	Total	4,573E+06	49146			

Fuente: "Elaboración propia"

## 6.7 ANEXO G – Análisis de varianza de segmentos

Análisis de varianza de los distintos segmentos encontrados a partir de la tenencia y propensión de compra de los productos.

Tabla 42: ANOVA segmentos

ANOVA						
		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Sexod	Inter-grupos	870,091	7	124,299	567,050	0
	Intra-grupos	10771,375	49139	,219		
	Total	11641,466	49146			
Fuma	Inter-grupos	232,382	7	33,197	190,638	3,9886E-280
	Intra-grupos	8556,998	49139	,174		
	Total	8789,381	49146			
Contacto	Inter-grupos	139,453	7	19,922	81,222	7,1315E-118
	Intra-grupos	12052,673	49139	,245		
	Total	12192,126	49146			
Clausulas	Inter-grupos	408,611	7	58,373	766,677	0
	Intra-grupos	3741,325	49139	,076		
	Total	4149,936	49146			
Autos	Inter-grupos	534,171	7	76,310	68,169	1,98688E-98
	Intra-grupos	55007,735	49139	1,119		
	Total	55541,906	49146			
ALujo	Inter-grupos	12,963	7	1,852	20,007	5,45082E-27
	Intra-grupos	4548,256	49139	,093		
	Total	4561,218	49146			
Motos	Inter-grupos	28,121	7	4,017	35,354	1,28658E-49
	Intra-grupos	5583,800	49139	,114		
	Total	5611,921	49146			
Valor_Cliente	Inter-grupos	4,217E+06	7	602382,045	169,765	2,5859E-249
	Intra-grupos	1,744E+08	49139	3548,337		
	Total	1,786E+08	49146			
Edadaños	Inter-grupos	606674,016	7	86667,717	1073,830	0

	Intra-grupos	3,966E+06	49139	80,709		
	Total	4,573E+06	49146			
ACompañía	Inter-grupos	4660,235	7	665,748	4384,513	0
	Intra-grupos	7461,304	49139	,152		
	Total	12121,538	49146			
StanSumaDeanual	Inter-grupos	322,904	7	46,129	46,427	4,65125E-66
	Intra-grupos	48824,096	49139	,994		
	Total	49147,000	49146			
SumaDeCapitalaccidentes	Inter-grupos	1,136E+12	7	1,623E+11	456,163	0
	Intra-grupos	1,749E+13	49139	3,559E+08		
	Total	1,862E+13	49146			
SumaDeCapitalfallecimiento	Inter-grupos	1,897E+11	7	2,710E+10	129,010	6,0042E-189
	Intra-grupos	1,032E+13	49139	2,101E+08		
	Total	1,051E+13	49146			
SumaDeCapitalenGM	Inter-grupos	2,103E+12	7	3,004E+11	165,799	1,8803E-243
	Intra-grupos	8,904E+13	49139	1,812E+09		
	Total	9,114E+13	49146			
SumaDeCapitaleninvalides	Inter-grupos	3,125E+11	7	4,464E+10	233,440	0
	Intra-grupos	9,398E+12	49139	1,912E+08		
	Total	9,710E+12	49146			

Fuente: "Elaboración propia"

## Test de significancia bilateral para los diferentes segmentos

Tabla 43: Comparación de medias para los segmentos

Comparaciones de medias de columnas								
Número inicial de casos								
	1	2	3	4	5	6	7	8
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Sexod	B		A B H	A B C E F H	A B C F H	A B C H	A B C D E F H	A B
Clausulas	B H		A B D F G H	A B F G H	A B C D F G H	A B G H	A B H	
Contacto	B F G H		B F G H	B F G H	A B C D F G H		B F	B F



Autos	B H		B H	B H	A B C D F G H	B H	A B C D F H	B
ALujo	B F H		B	B C F H	A B C F H		A B C F H	
ACompañía	H	A E G H	A B D E G H	A E G H	A G H	A B C D E G H	A H	
SumaDepanual	C D E H	C D E				C D E	A B C D E F H	C D E
Valor_Cliente	F	A C F	F	A B C E F H	A C F		A B C D E F H	A B C E F
Edadaños	B C D F G H	C		B C F	A B C D F G H	C	B C D F	B C D F G

Fuente: "Elaboración propia"

## 6.8 ANEXO H - Análisis de segmentación

Tabla 44: Datos de Segmentos 1

	Número inicial de casos							
	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5	Grupo6	Grupo7	Grupo8
	Media	Media	Media	Media	Media	Media	Media	Media
CobOnco	5%	31%	28%	44%	15%	17%	36%	66%
CobAccidentes	28%	63%	66%	82%	45%	0%	79%	92%
CobGM	8%	20%	0%	34%	35%	10%	19%	13%
CobInv	31%	48%	52%	41%	37%	41%	34%	26%
CobSalud	2%	39%	33%	26%	4%	14%	26%	0%
PGM	5%	0%	100%	0%	20%	43%	3%	4%
Pinv	1%	22%	36%	25%	13%	46%	7%	2%
POnco	75%	20%	17%	9%	72%	19%	24%	0%
PSalud	96%	0%	1%	26%	90%	3%	55%	100%
Paccidentes	67%	0%	1%	7%	51%	100%	13%	0%

Fuente: "Elaboración propia"

Tabla 45: Datos de segmentos 2

	Número inicial de casos							
	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5	Grupo6	Grupo7	Grupo8
	Media	Media	Media	Media	Media	Media	Media	Media
CobOnco	5%	31%	28%	44%	15%	17%	36%	66%
CobAccidentes	28%	63%	66%	82%	45%	0%	79%	92%
CobGM	8%	20%	0%	34%	35%	10%	19%	13%
CobInv	31%	48%	52%	41%	37%	41%	34%	26%
CobSalud	2%	39%	33%	26%	4%	14%	26%	0%
PGM	5%	0%	100%	0%	20%	43%	3%	4%
Pinv	1%	22%	36%	25%	13%	46%	7%	2%
POnco	75%	20%	17%	9%	72%	19%	24%	0%
PSalud	96%	0%	1%	26%	90%	3%	55%	100%
Paccidentes	67%	0%	1%	7%	51%	100%	13%	0%

Fuente: "Elaboración propia"

Tabla 46: Datos de segmentos 3

	Número inicial de casos							
	Grupo1	Grupo2	Grupo3	Grupo4	Grupo5	Grupo6	Grupo7	Grupo8
	Media	Media	Media	Media	Media	Media	Media	Media
SóloAPV	40%	26%	18%	0%	0%	51%	8%	4%
SóloProtección	7%	7%	4%	0%	0%	10%	0%	6%
SóloFlexible	46%	53%	64%	100%	100%	28%	62%	79%
APFL	3%	6%	10%	0%	0%	6%	30%	5%
APPR	1%	3%	1%	0%	0%	2%	0%	1%
FLPR	2%	4%	3%	0%	0%	2%	0%	4%
Osoloapv	3%	15%	12%	0%	0%	9%	0%	3%
Osoloprotec	5%	0%	13%	0%	0%	7%	100%	0%
Osoloflex	5%	8%	6%	0%	0%	9%	0%	3%

Fuente: "Elaboración propia"

Tabla 47: Test de significancia bilateral

TSC_4404					
	1	2	3	4	5
	(A)	(B)	(C)	(D)	(E)
Sexod		A	A B	A B C	A B C
Clausulas		A C E		A B C E	
Contacto	C D E	A C D E			
Autos	C E	A C E		C E	C
ALujo		A C E		A C E	
ACompañía	B		A B E	A B C E	A B
SumaDeannual	C E	A C E	E	A B C E	
Valor_Cliente	C E	A C D E		A C E	C
Edadaños	C D E	A C D E			

Fuente: "Elaboración propia"

Tabla 48: Error en encuesta para segmentos

Error encuesta	
Grupo 1	13%
Grupo 2	13%
Grupo 3	20%
Grupo 4	20%
Grupo 5	41%
Grupo 6	26%
Grupo 7	18%
Grupo 8	22%

Fuente: "Elaboración propia"

## 6.9 ANEXO I- Limpieza de datos

Tabla 49: Criterios de limpieza de Datos

Criterios de limpieza	
Datos incompletos	Aquellos clientes que tenían al menos un dato sin valor
Datos fuera de rango	Valor del cliente sobre 1500 UF
	Rut sobre 20 millones, que son empresas

Fuente: "Elaboración propia"