



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

**IDENTIFICACIÓN DE CLIENTES CON PATRONES DE ALTA INTERACCIÓN
CON LOS DRIVERS DE UNA TARJETA DE CRÉDITO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

CRISTÓBAL ENRIQUE GALLARDO MESA

PROFESOR GUÍA:
ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN:
PABLO MARÍN VICUÑA
JUAN PABLO ROMERO GODOY

SANTIAGO DE CHILE
MARZO 2016

**RESUMEN DE LA MEMORIA PARA OPTAR AL
TITULO DE INGENIERO CIVIL INDUSTRIAL
POR: CRISTÓBAL GALLARDO MESA
FECHA: 03/03/2016
PROF.GUÍA: ALEJANDRA PUENTE**

**IDENTIFICACIÓN DE CLIENTES CON PATRONES DE ALTA INTERACCIÓN
CON LOS DRIVERS DE UNA TARJETA DE CRÉDITO**

El mercado del plástico crediticio se caracteriza por ser altamente competitivo y relevante al transar junto al débito, montos cercanos al 14% del PIB chileno, lo que ha generado constantes ofertas hacia los usuarios para incentivar el uso y gasto. Particularmente en la empresa a trabajar se ofrecen como incentivos (1) compras con cuotas, (2) acceso a promociones y (3) el canje en un club de puntos, caracterizados internamente como los *drivers* de la tarjeta de crédito.

Estos *drivers* son relevantes de gestionar por que un 30% de las transacciones son realizadas con cuotas o descuentos, así como que el 35% de los clientes llega a niveles de canje, aumentando en los últimos años. Sin embargo, la empresa no conoce que tipo de cliente usa la tarjeta por un determinado incentivo.

Por esta razón se generaron KPI que midieran la intención de uso de *drivers*, perfilando cerca de 300 mil clientes de alta interacción con cuotas y promociones, y cerca de 430 mil con el club de puntos, distribuidos en un diagrama de Venn.

Con esta distribución se realizaron modelos de caracterización, obteniendo un mejor desempeño en puntos al capturar el 32% de los clientes en el mejor decil, seguido por cuotas con un 24% y en peor desempeño promociones con un 19%.

Los principales resultados apuntan a perfiles distintos, donde (1) puntos son clientes con características de poder económico, uso de la página web y acumulación en rubros de la competencia directa, ejemplificado en que alguien que acumula en estos rubros es 1,83 veces más probable a tener una alta interacción que el resto. De manera contraria, (2) cuotas apunta a un bajo poder económico, nulo ingreso a la web y edad superior a 35 años, mientras que (3) promociones son menores de 30 años, con poder económico medio y un uso oportunista de la tarjeta.

Posteriormente se extendió el análisis de puntos hacia principales líneas de premios, obteniendo además de perfiles más específicos, el efecto “círculo virtuoso” que los clientes canjean premios muy correlacionados a los rubros de acumulación.

Finalmente se propuso gestionar a 660 mil clientes de baja interacción que el modelo señaló con los mejores *lifts*. Con esta acción se espera generar un gasto incremental de \$12 MM, que es un 6,35% del gasto anual. Además se propone una gestión al largo plazo de los clientes, que en concordancia con los resultados, incentiven a los clientes con estos *drivers* según su etapa de vida y evolución económica, primero ofreciendo promociones y posteriormente puntos o cuotas.

Agradecimientos

Con este trabajo culmina un enorme proceso de seis años del que estoy profundamente agradecido de haber vivido, por sobre todo por las personas que de distinta manera me acompañaron entregándome momentos de alegría, compromiso, tristezas y aprendizaje que me hicieron crecer como persona.

En primer lugar quiero agradecer a mis papás, Eduardo y Paula, por el cariño, apoyo y confianza que me han entregado durante toda mi vida. A mis hermanos Sebastián, Felipe y Pablo que siempre han sido un impulso para seguir adelante con la alegría que me caracteriza. Aunque no lo demuestre, mi familia ha sido un pilar fundamental.

Estoy eternamente agradecido de mis amigos del colegio y de la vida por todos los enormes momentos que hemos vivido y aguantarme en los momentos que “desaparecí” por la universidad y el trabajo de título. Me encantaría nombrarlos uno por uno con las vivencias que hemos tenido, pero son tantos que no quiero dejar a alguien fuera.

Quiero agradecer a todas las personas que conocí durante los seis años de universidad. Me llevo muchos buenos recuerdos de amigos como Yeipi, Eyal, Haye, Benja, Nicole, Pedro, Vale, Coni, Pivo, “Los Ryders”, etc. que vivieron conmigo distintas etapas de estos seis años. Especialmente quiero agradecer a Seba, Martín, Chicho, Raúl y Pipe por lo grandes momentos que vivimos dentro y fuera de cada ramo, haciendo que cada reunión de trabajo fuera una junta más de amigos.

Mención especial a Martín y su desempeño en Taller III, del que estamos todos infinitamente agradecidos.

Por otro lado darle las gracias a mis profes guías Ale y Pablo, por todas las enseñanzas que me dieron en este largo proceso. Cada una de nuestras reuniones me dejaba reflexionando por horas, lo que me sirvió mucho para aprender y crecer como profesional.

Finalmente quiero darle mis gracias al equipo de BI que desde el primer minuto me entregaron total disposición a ayudarme y me acogieron como si fuera uno más del grupo. Gracias Juan Pablo y José Antonio por la constante paciencia que me tuvieron este año y sus constantes feedbacks y recomendaciones. De verdad, me siento un afortunado de haber trabajado con ustedes.

Tabla de Contenido

1. Introducción	5
2. Descripción y Justificación	8
3. Objetivos	11
4. Alcances	11
5. Marco conceptual	12
6. Metodología	15
7. Análisis descriptivo	18
8. Cuantificación de interacción con drivers	22
9. Clientes de Alta Interacción	29
10. Importancia de Variables	34
11. Análisis Exploratorio	37
12. Generación de Modelo de Propensión	40
12.1 Club de Puntos	40
12.2 Compras con cuotas	45
12.3 Compras con promociones	49
13. Conclusión del perfil de clientes en cada driver	52
14. Extensión interacción con líneas de premios	54
15. Modelos líneas de premios	57
15.1 Seguros (SOAP)	57
15.2 Ferretería	60
15.3 Viajes	62
15.4 Electro doméstico cocina	64
15.5 Gift card	65
16. Desempeño Modelos Canjes de Líneas	66
17. Perfilamiento de Líneas de Premios	68
18. Líneas de Acción y Aplicación Comercial	71
19. Conclusión	76
20. Limitaciones y Trabajo Futuro	78
21. Bibliografía	79
22. Anexos	80
Anexo A. Descripción Drivers	80
Anexo B. Descripción Variables	82

Anexo C. Distribución Clientes con Drivers.....	85
Anexo D. Exploratorio Drivers.....	87
Anexo D. Importancia Variables y Desempeño Modelos	92

Índice de tablas

<i>Tabla 1: Descripción variables continuas de interacción.....</i>	<i>22</i>
<i>Tabla 2: Correlación entre tasas de interacción.....</i>	<i>28</i>
<i>Tabla 3: Tasa promedio y tasa de corte del mejor decil para cada quintil.....</i>	<i>30</i>
<i>Tabla 4: Tasa promedio y tasa de corte mejor quintil para cada quintil.....</i>	<i>31</i>
<i>Tabla 5: Ejemplo de la tabla de contingencia.....</i>	<i>34</i>
<i>Tabla 6: Importancia variables para la alta interacción con el club de puntos.....</i>	<i>35</i>
<i>Tabla 7: Importancia variables para la alta interacción con las cuotas.....</i>	<i>36</i>
<i>Tabla 8: Importancia variables para la alta interacción con promociones.....</i>	<i>36</i>
<i>Tabla 9: Importancia de las variables puntos según Árbol tipo CART.....</i>	<i>37</i>
<i>Tabla 10: Resultados coeficientes regresión logística binaria.....</i>	<i>41</i>
<i>Tabla 11: Ajuste y predicción modelo de alta interacción.....</i>	<i>42</i>
<i>Tabla 12: Desempeño modelo árbol decisión en alta interacción con los puntos.....</i>	<i>43</i>
<i>Tabla 13: Curva de ganancia árbol decisión en alta interacción con puntos.....</i>	<i>45</i>
<i>Tabla 14: Coeficientes regresión logística para alta interacción con las cuotas.....</i>	<i>46</i>
<i>Tabla 15: Resultados regresión logística alta interacción promociones.....</i>	<i>49</i>
<i>Tabla 16: Descripción de líneas de premios involucrados.....</i>	<i>54</i>
<i>Tabla 17: Coeficientes regresión logística - Seguro (SOAP).....</i>	<i>57</i>
<i>Tabla 18: Coeficientes regresión logística - Ferretería.....</i>	<i>60</i>
<i>Tabla 19: coeficientes regresión logística viajes.....</i>	<i>62</i>
<i>Tabla 20: Coeficientes regresión logística – Electro cocina.....</i>	<i>65</i>
<i>Tabla 21: Coeficientes regresión logística - gift card.....</i>	<i>66</i>
<i>Tabla 22: Gasto anual promedio perfiles de clientes.....</i>	<i>73</i>
<i>Tabla 23: Ganancia por incremento del gasto de los clientes target.....</i>	<i>75</i>

1. Introducción

El mercado de las tarjetas de crédito se caracteriza por el alto número de emisores bancarios como no bancarios (los cuales están relacionados a las empresas de *retail*) que compiten en un mismo mercado. Producto de esto, éstas son unos de los principales medios de pagos de las personas, llegando a transar en conjunto con las tarjetas de débito, montos totales que representaron cerca del 14,64% del PIB Chileno en el año 2013¹.

Un ejemplo de las características de este mercado, es que en abril del año 2015 alcanzaron a circular cerca de 18,8² millones de tarjetas vigentes en el país, entre las que resaltan Falabella (18%), Ripley (15%) y Cencosud (11%), señalando lo común que es para una persona poseer más de una tarjeta en su poder.

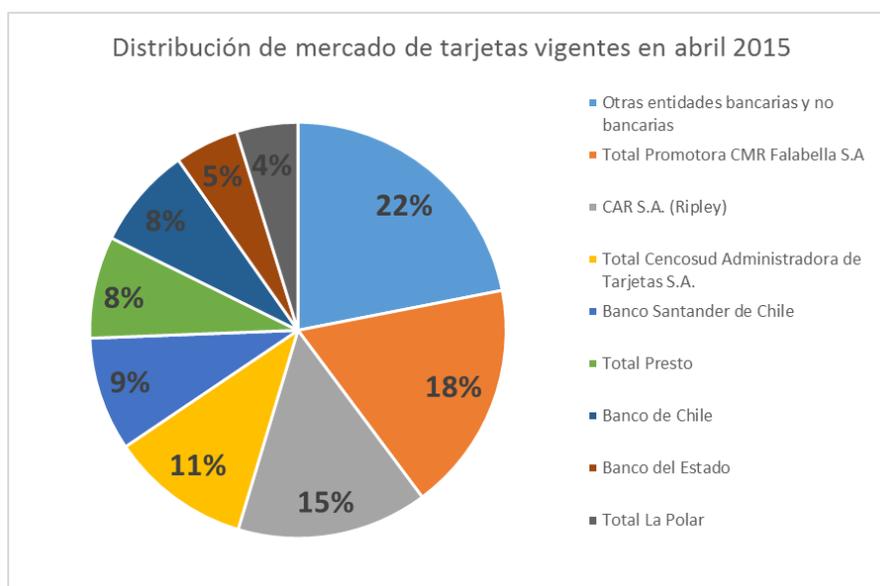


Ilustración 1: Participación de mercado de las compañías emisoras de retail de acuerdo al número de tarjetas vigentes. Fuente: elaboración propia con datos del Sbif.

Cómo consecuencia, tanto las empresas bancarias como no bancarias ofrecen distintos premios o beneficios por el uso de su tarjeta, generando una estrategia que incentive la lealtad entre el cliente y la tarjeta. Medidas que se ven reflejadas en la principalidad y gasto que tiene el cliente con dicha tarjeta.³

De esta manera, algunas de las ofertas más comunes que ofrecen las tarjetas de crédito son la posibilidad de acceder a descuentos en el precio de compra y el

¹ "Actualidad del mercado de las tarjetas y transacciones electrónicas en Chile", Javier Etcheverry, 18 de marzo de 2015.

² Reporte mensual mercado bancario y mercado no bancario, Superintendencia de Bancos e Instituciones Financieras (SBIF), Abril 2015.

³ La principalidad refleja el porcentaje de veces que el cliente prefiere la tarjeta de una empresa por sobre total de veces que elige una tarjeta. "Ser la tarjeta principal".

participar en clubes de lealtad que premian al cliente por el constante uso de la tarjeta, entre otras ofertas.

Ejemplos de estos casos son las alianzas de LANPASS con Santander o Banco de Chile con Delta Airlines para acumular millas de viaje, o bien los clubes de lealtad “CMR Puntos”, “Ripley Puntos” y “Puntos Más” que ofrecen las compañías de *retail*, permitiendo al cliente canjear sus puntos acumulados en un catálogo distribuido en distintos niveles de canje de acuerdo a la cantidad de puntos requeridos.

Particularmente el trabajo a realizar se llevará a cabo con una tarjeta de crédito no bancaria, vinculada a un holding de *retail* del país, la cual incentiva su uso con tres principales ofertas, denominados internamente como *drivers*⁴.

1. Deuda: Préstamo de dinero por medio de compras con cuotas, avances en efectivo y créditos de consumo.
2. Promociones: Compras con descuentos sobre el valor del producto por el uso de la tarjeta como medio de pago.
3. Club de puntos: Acumulación de puntos por el uso de la tarjeta, que posteriormente se pueden canjear en un catálogo de premios.



Ilustración 2: Principales drivers que incitan el uso de la tarjeta de crédito

Deuda

Driver vinculado con el préstamo de dinero hacia clientes que en un determinado momento no tienen dichos montos monetarios. A pesar de que este ítem se enfoca exclusivamente en permitir al cliente realizar transacciones con dinero prestado, éste préstamo se puede otorgar a través de tres herramientas distintas.

⁴ La terminología *driver*, es una denominación interna que la empresa le otorga a las ofertas que entrega la tarjeta a los clientes.

- Avances en efectivo: Acción en que el cliente puede retirar dinero en efectivo de su cupo de crédito en distintos puntos de venta o cajeros automáticos. Éste método está enfocado hacia bajos montos, desde los 5.000 pesos hasta el cupo disponible del cliente⁵
- Créditos de consumo⁶: Acción en que el cliente puede adquirir dinero prestado a través de un crédito exclusivo al de la tarjeta. Al ser un crédito éste está enfocado en montos mayores de dinero.
- Compras con cuotas: Tipo de movimiento donde el cliente puede postergar parte del pago hacia el futuro, particionado el pago en distintas etapas desde 2 a 36 meses.

Promociones

Driver que consiste en compras con descuentos sobre el valor real del producto, por el uso de la tarjeta como medio de pago. Debido a que el descuento se realiza sobre el producto a vender, es que este sólo se puede llevar a cabo en los comercios internos del *retail*.

Club de Puntos

Driver relacionado en premiar a los clientes por el uso de la tarjeta de crédito, por medio de un catálogo distribuido en 14 distintos niveles de canje desde los 5.000 a 240.000 puntos⁷, ofreciendo premios asociados a los distintos comercios del holding como lo son:

- Electrodomésticos, dispositivos tecnológicos y productos del hogar asociados al comercio de multitienda del holding
- Herramientas de ferretería asociado al comercio de herramientas y construcción del holding.
- Licores y alimentos asociados al comercio de supermercado del holding
- Seguros de circulación asociado al comercio de seguros del holding
- Viajes dentro y fuera del país asociado al comercio de viajes del holding
- Descuentos en futuras compras
- Experiencias como la realización de trekking, canopy y spa entre otros.

De este modo el club de puntos categoriza a los clientes en tres distintas categorías, normales (si en el año anterior acumuló menos de 15.000 puntos), premium (si acumuló entre 15.000 y 50.000 puntos) y elite (Más de 50.000), los cuales acumulan puntos de igual manera, que si no son canjeados luego de un año

⁵ El principal motivo de los avances en efectivo, es poder extraer dinero de su cupo por montos en efectivos que el cliente puede ocupar en cualquier comercio, el cual está supeditado al análisis de riesgo realizado al entregar la tarjeta.

⁶ El crédito de consumo se diferencia en que no todos los clientes pueden acceder, debido a que el préstamo necesita estar pre aprobado por un análisis de riesgo.

⁷ Anexo A niveles de canje de premios.

para los clientes normales y dos años para los clientes premium y elite, estos expiran sin derecho a renovación.

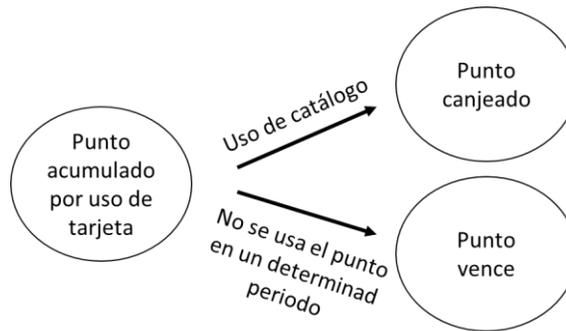


Ilustración 3: Representación de la "vida" de un punto

2. Descripción y Justificación

Actualmente en la empresa a trabajar existen 2,9 millones de clientes con su tarjeta activa para el uso, los cuales pueden interactuar con estos tres *drivers*. De estos, la penetración que ha tenido el uso de *drivers* ha sido relevante en cuanto al número de transacciones, número de clientes y montos asociados.

Particularmente en el caso de las compras con cuotas, en el año 2014 hubo 29,6 millones de transacciones que su pago fue realizado con cuotas, lo que representa un 25,9% de las transacciones del año. Hecho que reafirma su tendencia constante al observar que las transacciones con cuotas en cada mes representan entre un 25% y 30% del total de transacciones⁸.

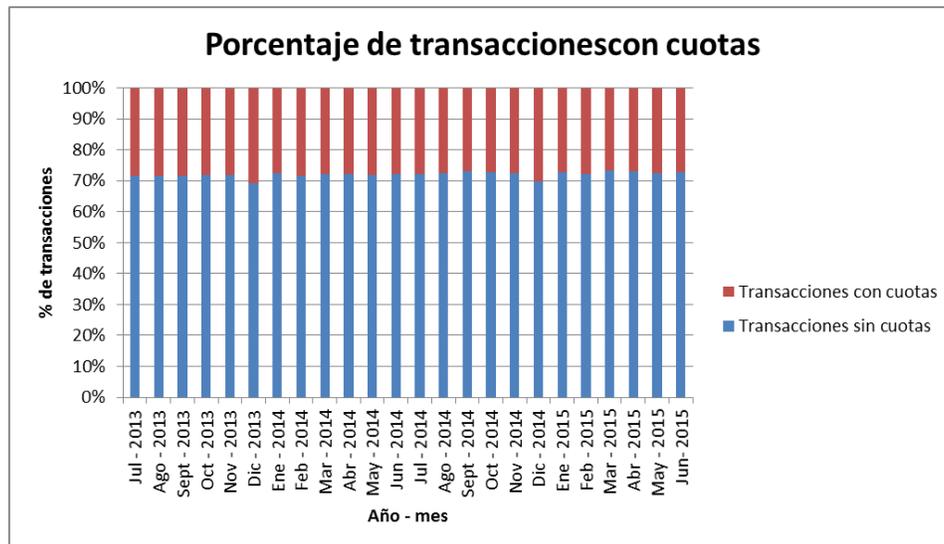


Ilustración 4: Porcentaje transacciones con y sin cuotas
Fuente: Elaboración propia con datos de la empresa

⁸ Anexo A. Desglose de transacciones con cuotas, según el número de cuotas a pagar.

Por otro lado, al observar el número de transacciones con la tarjeta de crédito en el comercio de multitienda, se cuantificaron cerca de cinco millones de transacciones que incluían algún descuento en el precio del producto, lo que representa un 30% del total de transacciones realizadas con la tarjeta de crédito.⁹

En el caso del club de puntos, se puede observar que en los últimos tres años sólo un 22% de los clientes que acumulan puntos, canjean al menos un punto acumulado. Hecho que se debe en gran medida a la necesidad de alcanzar como mínimo 5.000 para llegar a niveles de canje.

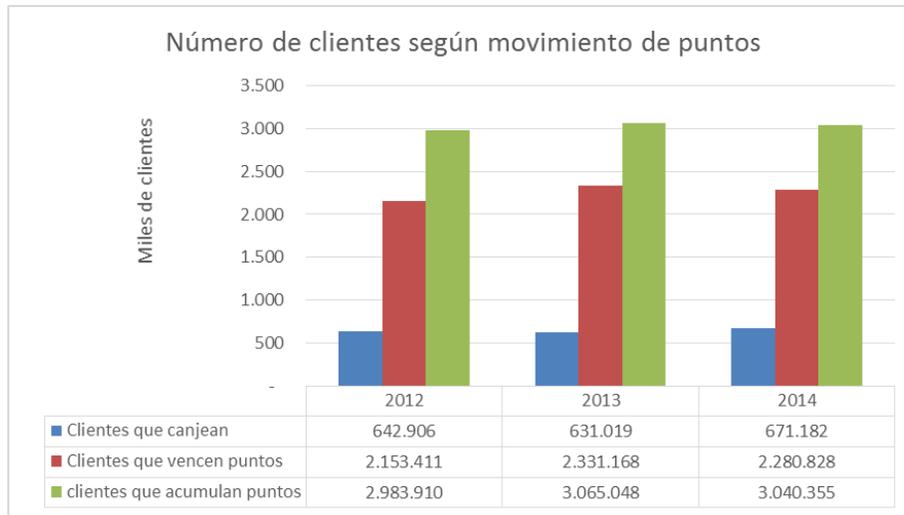


Ilustración 5: Número de clientes con acumulación, vencimiento y canje de puntos

A pesar de esto, un hecho importante que resalta la interacción con el club de puntos, es que en los últimos años se ha percibido una tendencia creciente sobre el número de clientes que en algún momento del año logra acumular al menos la cantidad mínima de puntos requeridos para poder realizar un canje.

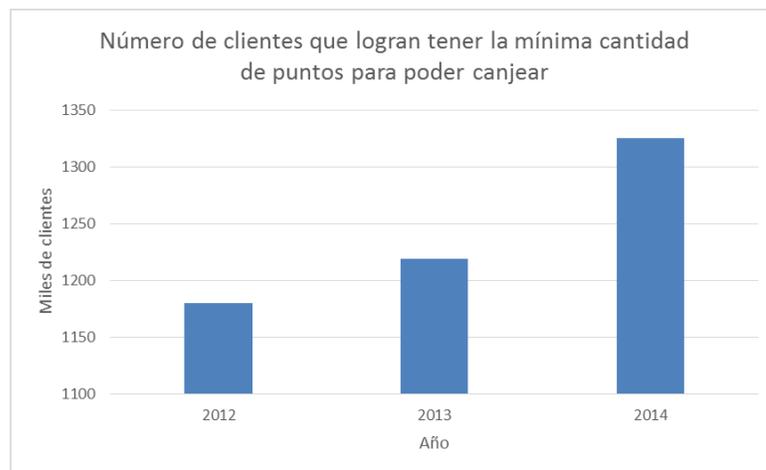


Ilustración 6: Número de clientes que en algún momento del año alcanzan los 5.000 puntos

⁹ Anexo A. Porcentaje de transacciones con descuentos y descuentos adicionales por el uso de la tarjeta.

Debido a estas razones, la empresa realiza distintas campañas de incentivo al uso de la tarjeta por los beneficios que esta ofrece a los actuales clientes y los cerca de 20.000 clientes que mensualmente realizan apertura de cuentas.

Sin embargo la principal problemática que tiene la empresa, es que a pesar de tener estos *drivers* como “herramientas” para estimular el uso de la tarjeta, se desconoce si existe un perfil de clientes que use la tarjeta por un determinado *driver* para así poder gestionar de manera eficiente a “clientes similares” que actualmente tienen un bajo uso y gasto de la tarjeta.

Más aún, esta problemática adquiere mayor importancia si se considera que el catálogo que ofrece el club de puntos está compuesto por una variedad de premios perteneciente a distintos comercios del holding, los cuales pueden atraer de manera distintas a los clientes con el club de puntos.

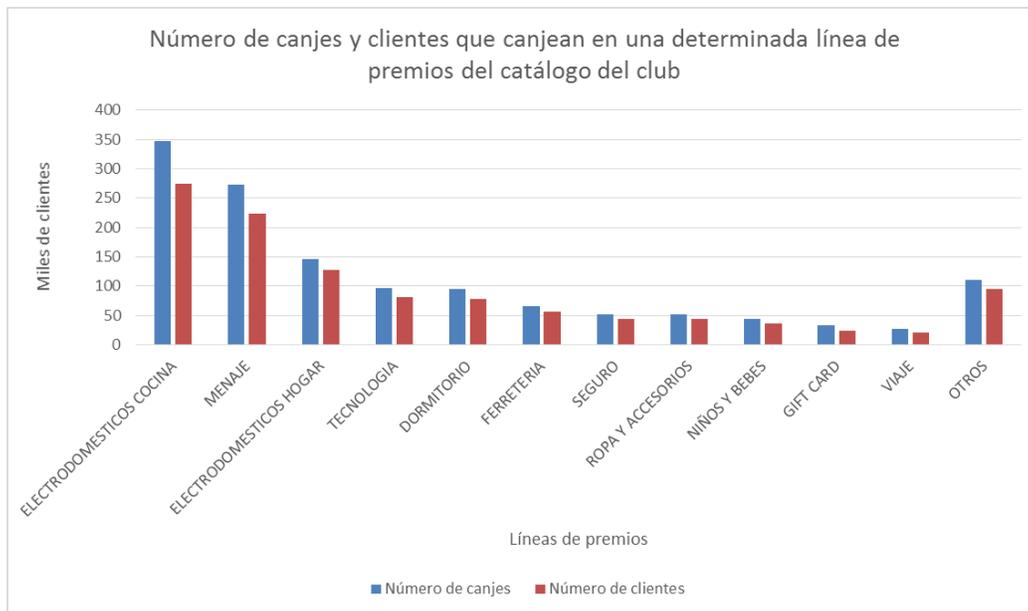


Ilustración 7: Canjes y clientes que canjean en una determinada línea de premio del club de puntos.

En consecuencia, el principal desafío que tiene este trabajo es poder identificar dentro del universo de clientes, quienes actualmente tienen un comportamiento de alta interacción con cada uno de los *drivers*, para así construir el “camino” que lleve a clientes con características similares, desde un estado de baja a alta interacción.

3. Objetivos

Objetivo General

Identificar a los clientes más propensos a tener una alta interacción con los *drivers* de la tarjeta de crédito.

Objetivos Específicos

1. Definir el criterio para determinar a un cliente de alta interacción con cada *driver*
2. Determinar el perfil de clientes de alta interacción para cada uno de los *drivers* de la tarjeta de crédito
3. Identificar las variables más importantes que determinan a los clientes considerados de alta interacción
4. Generar un modelo de propensión a tener una alta interacción con cada uno de los *drivers*
5. Proponer líneas de acción en torno a los resultados obtenidos

4. Alcances

- La interacción estudiada con cada driver, será únicamente considerando la data observada por transacciones con la tarjeta y movimientos realizada con el club de puntos, dejando de lado los efectos externos como opiniones, percepciones, *share of voice* u otros ámbitos complementarios a las transacciones.
- El estudio de las promociones será llevada a cabo únicamente en la multitienda del holding, donde los descuentos son identificables y las compras por boleta son de pocos sku, haciendo más verídica la asociación a una intención de uso.
- El estudio de modelos de alta interacción con el *driver* deuda será únicamente con la intención de comprar con cuotas, debido a que el uso de avances y créditos de consumo está asociado a temas de riesgos que incluyen información confidencial externa al estudio.
- Si bien el análisis se puede extender para los tres *drivers*, únicamente se extenderá para el club de puntos ya que es el *driver* del que la empresa tiene menor información y debido a la gran variedad de premios distintos que existen.

- No se realizarán experimentos con los resultados obtenidos.

5. Marco conceptual

A. DEFINICIÓN CONTEXTO

- Retail financiero: Vinculación del *retail* con el negocio financiero a través de tarjetas de crédito que los clientes pueden usar dentro del holding como fuera de este, según el tipo de tarjeta que estos posean.
- Metodología RFM: Metodología descriptiva sobre el comportamiento transaccional que los clientes tienen en este caso con la tarjeta de crédito. Esta descripción se hace en base a tres indicadores principales.
 - Recency: Meses transcurridos desde el último uso de la tarjeta
 - Magnitud: Monto promedios de las transacciones realizadas
 - Frecuency: Número de meses que el cliente realiza una transacción sobre el máximo de meses posibles

B. PROGRAMAS DE LEALTAD

Los programas de lealtad son herramientas que se han incorporado en distintos mercados para generar una recompensa a los clientes que logren un alto uso de los productos, ofreciendo variados premios según el comercio de la empresa.

En este contexto los clubes de lealtad, se basan en el concepto CRM (Customer Relationship Management), la cual se enfoca en gestionar la relación que la empresa tiene con el cliente para producir efectos positivos a largo plazo como:

- Disminución de costos, producto de que retener clientes es menos costos que adquirir uno nuevo
- El uso y gasto tiende a acelerar en el tiempo
- Clientes satisfechos actúen como puente gratuito de marketing al hacer referencias
- Clientes leales son menos sensibles a cambios de precio.¹⁰

C. MODELO REGRESIÓN LOGÍSTICA

El modelo de regresión logística busca estimar la respuesta que tiene una persona en torno a un set de opciones, estimando la probabilidad de que dicho cliente, según sus características, elija cada una de las opciones.

¹⁰ "Customer Relationship Management, perspectives from the marketplace", Simon Knox, Adrian Payen, Stan Maklan, Joe Peppard, Lynette Ryals, 2003.

Dicho esto, el modelo busca P_{ij} , que es la probabilidad de que el cliente i elija la opción j perteneciente al set de opciones. En este sentido la probabilidad de que un cliente elija una opción queda denotada por la utilidad que le generará elegir dicha opción, la cual está determinada por una parte determinística y una fracción aleatoria.

$$u_{ij} = v_{ij} + \epsilon_{ij}$$

Ecuación 1: utilidad compuesta por utilidad observable y utilidad no observable que entrega la opción j al cliente i

Donde el término v_{ij} es la fracción determinística de la utilidad generada por las características observables, mientras que ϵ_{ij} representa la utilidad aleatoria generada por características no observable, la cual está distribuida a través de una doble exponencial.

Por lo tanto, la probabilidad de que un cliente elija la alternativa j , depende de la probabilidad que esta utilidad sea mayor a la que entrega cualquier otra opción, como se ilustra a continuación.

$$P_{ij} = \int 1(v_{ij} + \epsilon_{ij} > v_{ik} + \epsilon_{ik})f(\epsilon_i)d\epsilon_i$$

$$P_{ij} = \int 1(\epsilon_{ij} - \epsilon_{ik} > v_{ik} - v_{ij})f(\epsilon_i)d\epsilon_i$$

Es a través de cálculos matemáticos, asociando la distribución del error como una distribución Error tipo I, que se obtiene la formula cerrada de la probabilidad de elección de la alternativa j por parte del cliente i .

$$P_{ij} = \frac{e^{\beta' * x_i}}{\sum_j e^{\beta' * x_j}}$$

Ecuación 2: Cálculo probabilidad que el cliente i elija la opción j

Donde β son los parámetros a estimar que determinan la probabilidad de elegir una determinada opción y x es el vector de características observables.

D. ÁRBOLES DE DECISIÓN

La metodología de árboles de decisión corresponde a modelos de *Machine Learning* que permiten predecir tanto variables categóricas como continuas en función de variables explicativas, a través de la sucesiva división de la base de datos en grupos disjuntos de datos. Este modelo va adquiriendo la forma de un árbol compuestos por nodos parentales divididos en nodos hijo, producto del valor que toma una variable (ramas).

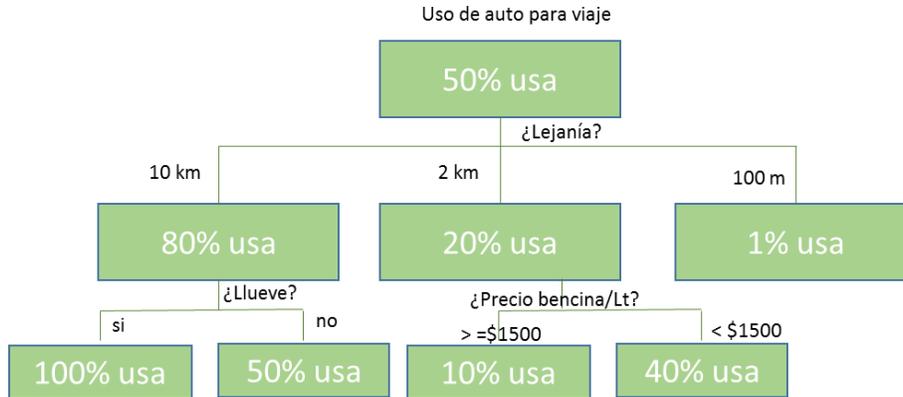


Ilustración 8: Ejemplo del funcionamiento del algoritmo árbol decisión¹¹

La realización de este proceso se puede llevar a cabo por distintos algoritmos que pueden generar arboles con distinta cantidad de nodos hijos.

CHAID

Los arboles tipo CHAID son un algoritmo que a partir de un nodo parental se pueden generar dos o más nodos hijos únicamente a través de variables categóricas (Si es una variable nominal, se generan distintas categorías), cuyo criterio de división se realiza a través de un test chi-cuadrado que determina la mejor división.

CART

El algoritmo de los árboles de clasificación y regresión (CART) funciona sobre una variable dependiente categórica, la cual se divide el conjunto de datos perteneciente al nodo padre en sólo dos nodos hijos.

El criterio de Split o división del nodo padre en los nodos hijos se hace a través del criterio de ganancia de información, el cual consisten en la mejora que produce un determinado corte de división. Para esto se calcula la entropía (medida de dispersión del conjunto de datos) definido para una elección binaria como:

$$Entropía (S) = -p_1 \log_2(p_1) - p_0 \log_2(p_0)$$

Ecuación 3: Calculo entropía generada por una determinada variable

En el que p_1 representa la fracción de ejemplos denotados en la clase 1 y p_0 la fracción de ejemplos perteneciente a la clase 0 del conjunto de ejemplo S.

¹¹ Ilustración propia

Por medio de este valor, se mide para cada una de las variables independientes, la ganancia de información que genera al dividir el conjunto de datos en dos grupos por medio de la fórmula:

$$Ganancia(S, F) = Entropia(S) - \sum_{v \in \text{Valores}(F)} \frac{|S_v|}{|S|} Entropia(S_v)$$

Ecuación 4: Ganancia generada por entropía

Donde S_v es el subconjunto de datos cuando se tiene el valor v de la variable F . En base este criterio, el árbol elige a la variable que genera la mejor ganancia de información al tomar un determinado valor.

6. Metodología

Para llevar a cabo el trabajo se cuenta con la base de datos de la tarjeta de crédito, la cual cuenta con información transaccional, movimientos de puntos, transacciones crediticias e información demográfica de cada uno de los clientes.

A través de esta información se hará uso de estas base de datos por medio de dos etapas, el pre procesamiento y selección de los datos, y posteriormente el análisis de estos mismos. En este sentido se procederá a realizar la metodología KDD ilustrada por cuatro principales etapas.

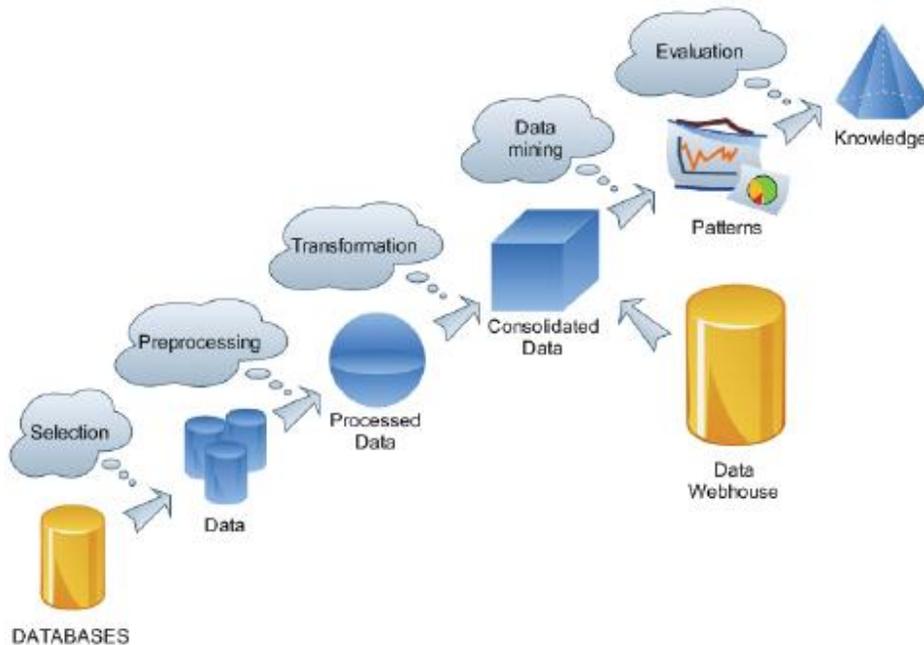


Ilustración 9: Etapas Proceso KDD
Fuente: Chapter 1, curso Web Intelligence, FCFM, Universidad de Chile

6.1 Limpieza y Selección

Esta etapa consiste en la selección de datos necesarios para realizar el estudio, seleccionando aquella información que será ocupada dentro de la gran base, así como la eliminación de aquella data que genere ruido (Por falta de información o información inconsistente).

Producto de la enorme cantidad de datos, este proceso consiste en consultas en el lenguaje SQL para obtener la información necesaria de las distintas bases y su posterior consolidación en una única base analítica.

De esta manera el trabajo comenzó realizando las consultas en siete base de datos distintas sobre la información que cada cliente tiene en temas demográficos, de contrato y transaccionales en el periodo de estudio acordado con la empresa, el cual consisten en un año móvil entre Julio de 2014 y Junio 2015.

Cabe destacar que la información extraída de cada una de estas tablas se llevó a cabo a nivel de cliente titular, ya que este es el único que puede interactuar con el club de puntos y que su información demográfica es más completa y verídica que los clientes adicionales.

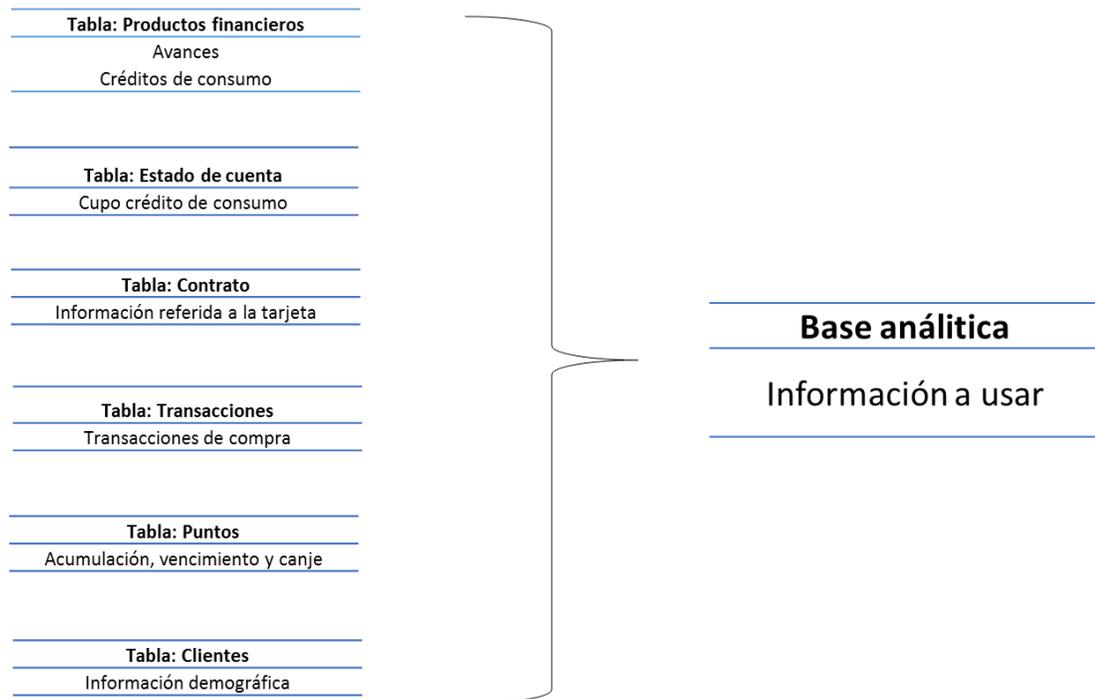


Ilustración 10: Formación base analítica a través de los datos conseguidos de siete bases de datos distintas

La información que entrega cada una de estas bases de datos se caracteriza por entregar la siguiente información.

- Productos financieros: Tabla donde se extrajo por cliente, los montos de avances en efectivo y créditos de consumo realizados en el periodo de estudio.
- Estado de cuenta: Tabla donde se extrajo la información de cupo que el cliente tuvo para créditos de consumo en el periodo de estudio.
- Contrato: Tabla donde se puede rescatar la información contractual de los clientes (Número de tarjetas, tipo de tarjetas, antigüedad del cliente), así como el cupo que estos clientes tienen en sus tarjetas de crédito. En este caso si un cliente tiene más de una cuenta, el cupo del cliente se considera como la suma de los cupos.
- Transacciones: Tabla donde se obtuvo el monto y transacciones totales que un cliente realiza en los comercios internos y externos al holding.
- Puntos: Información del cliente que entrega la cantidad de puntos acumulados, vencidos y canjeados.
- Cientes: Tabla que muestra la información demográfica de cada uno de los clientes.

6.2 Pre procesamiento

Luego de consolidar la información a utilizar en una única base analítica para el universo de 2,95 millones de clientes titulares que actualmente tienen su tarjeta activa para el uso, comenzó el proceso de filtrado.

Dado que el periodo de estudio, se realizó entre Julio de 2014 y Junio de 2015, es que se consideró únicamente a los clientes que tuvieron apertura de cuenta al menos dos meses antes del estudio (Mayo 2014), disminuyendo la base analítica a 2.557.397 clientes (86,7% de los clientes).

Por otro lado, en conjunto con la empresa, se decidió descartar del análisis a todos los clientes que no tuvieron transacciones durante el periodo de estudio, debido a que estos son ya considerados clientes fugados, y el análisis se estudiará únicamente para determinar dentro de los clientes que interactuaron con la tarjeta de crédito, quienes tienen un comportamiento denominado de alta interacción.

Al realizar este filtro se obtuvieron 2.058.820 (69,7% del total) clientes en el universo a estudiar para medir los clientes considerados de alta interacción con cada uno de los *drivers*.

6.3 Transformación

Una vez obtenida la cantidad de clientes necesarios para el estudio, junto con su información básica necesaria, se realizó el proceso de la transformación de ciertas variables para obtener información complementaria para el análisis posterior.

Como es necesario medir la interacción que cada cliente tuvo con cada uno de los *drivers*, es que fue necesario crear en conjunto con la empresa, variables que midieran la interacción a través de tasas (KPI) de comportamiento.

Por otro lado se realizó la creación de variables que no se encontraban en la base, pero que son obtenidas por fórmulas que incluyen variables existentes en la base de datos, como la antigüedad del cliente, zona geográfica de residencia, etc.

Posteriormente con todas las variables formuladas en la base analítica se procede a comparar la interacción de los clientes tienen con cada uno de los *drivers*, determinando en conjunto con la empresa, el criterio de calificación para pertenecer al perfil de alta interacción con cada uno de los *drivers*, generando la variable independiente si el cliente tiene o no una alta interacción con un determinado *driver*.

Finalmente se identificará y agruparan los distintos canjes de premios en principales línea de premios que logren determinar si un cliente tuvo una alta interacción con el club, producto del canje efectuado en una determinada línea.

6.4 Data Mining

Con la base de datos construida con las variables independientes y variables dependientes a necesitar, se realizará el estudio estadístico para determinar la propensión que cada cliente tiene a pertenecer al perfil de clientes considerado de alta interacción con cada uno de los *drivers* (debido a la interacción que este tuvo), a través de un modelo binomial de regresión logística y árbol de decisión.

Luego para el caso del club de puntos, se desglosarán las principales líneas de premios relacionados a comercios específicos del holding, para extender el modelo desde el “pertenecer al perfil de alta interacción con el club de puntos” hacia “pertenecer al perfil por el canje en una determinada línea de producto”.

Finalmente al comparar el desempeño de los modelos generados, se procederá a generar lineamientos de acciones hacia clientes que tienen un bajo comportamiento de uso, pero con características similares a algún perfil de alta interacción

7. Análisis descriptivo

Al obtener la información necesaria de los 2,05 millones de clientes entre Julio de 2014 y Junio de 2015, se pudo agrupar la información obtenida de estos clientes en tres líneas distintas, las cuales son los datos demográficos, transaccionales e interacción con los *drivers*.

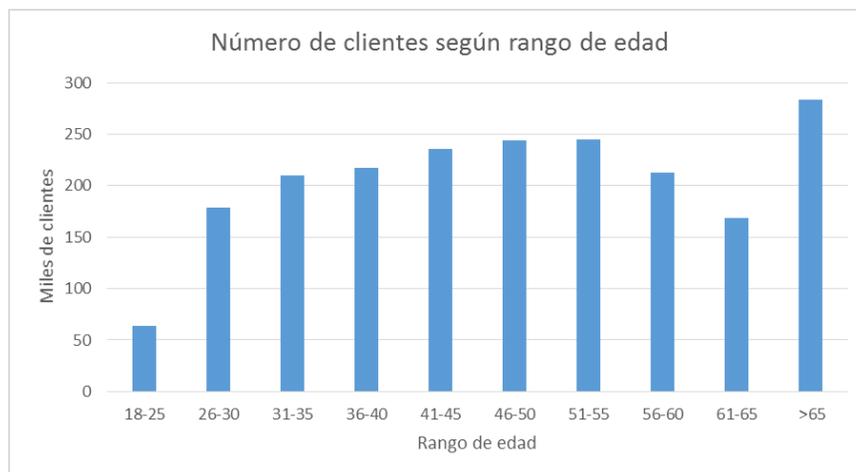
7.1 Demográfica naturales y de contrato

- Género

Variable categórica donde 51,4% son femeninos y 48,6% masculinos.

- Edad del cliente

Variable nominal la cual para comprender su distribución en la población, se puede categorizar en intervalos de edad.



*Ilustración 11: Distribución de clientes según rangos de edad.
Fuente: Elaboración propia con datos de la empresa*

- Zona geográfica

Variable categórica sobre la residencia del cliente. Para esta variable se acordaron en conjunto con la empresa las zonas a formar dentro del país, así como las comunas incluidas.

- Zona norte grande: XV, I y II regiones del país
- Zona norte chico: II y IV región del país
- Zona región centro: V y VI región del país
- Zona sur: VII, VIII, IX y XIV regiones del país
- Zona austral: X, XI y XII regiones del país
- Zona exterior RM: considerando las comunas periféricas de la región como Melipilla y Curacaví.
- Zona centro RM: comuna de Santiago
- Zona Norte RM: comunas del norte de la RM como Recoleta e Independencia
- Zona Sur RM: comunas del sur de la RM como San Miguel, La Granja y Puente Alto.
- Zona Oriente RM: comunas del sector oriente de la RM como Vitacura, La Reina y Peñalolén.

- Zona Poniente RM: comunas del sector poniente de la RM como Estación Central, Maipú y Quinta Normal entre otros

Se puede observar de esta agrupación que la mayoría de los clientes pertenecen a la región metropolitana, donde en estas predominan clientes del sector oriente y sur de Santiago. Por otro lado en las regiones predominan los clientes de la zona sur del país y centro del país.

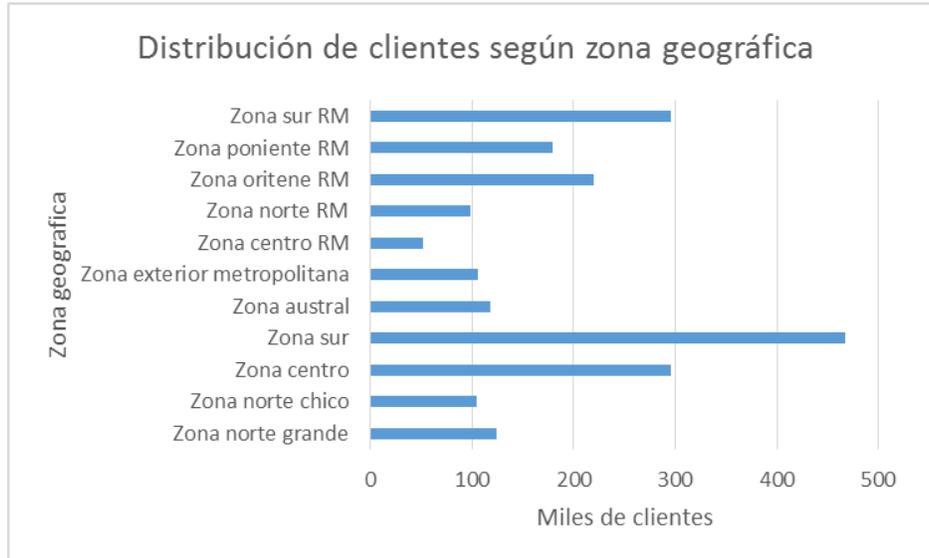


Ilustración 12: Distribución clientes según zona geográfica.
Fuente: elaboración propia con datos de la empresa.

- Clasificación del cliente (Elite – Premium – Normal)

Variable categórica sobre la clasificación que otorga el club de puntos de la empresa, el 87,3% de los clientes son normales, 11,4% premium y el 1,3% elite.

- Tipo de tarjeta

Variable binaria que indica si la tarjeta del cliente es una tarjeta normal cerrada (27,1%) o tarjeta visa (72,9%).

- Tenencia de un cliente adicional

Variable binaria que indica si el cliente tiene un adicional (27,4%) o no.

- Cupo del cliente

Variable nominal, que si observamos la distribución de los clientes según tramos, se puede apreciar que un 63,3% de los clientes tiene un cupo menor o igual a 1.000.000 pesos, 22,5% entre 1 y 2 millones de pesos y un 14,2% superior a los 2.000.000 pesos. Por lo tanto mayoría de los clientes tiende a tener cupos bajos.

- Tenencia de un hijo

Variable binaria que tiene la empresa sobre la tenencia de un hijo por parte del cliente. En esta, se puede observar que un 20,9% tiene al menos un hijo.

- Antigüedad del cliente

Variable nominal que indica cuantos años han pasado desde la apertura de cuenta del cliente desde el periodo de estudio.

- Días de looks

Variable nominal que hace referencia a la cantidad de días que el cliente realizó un “log - in”¹² en la página de internet.

7.2 Transaccional

- Tenencia de un contrato PAT (Pago automático)
- Variable *dummy*¹³ sobre la realización de al menos una transacción de compra en un determinado mes
- Variable *dummy* sobre la realización de al menos una transacción o acumulación de puntos en un determinado comercio del holding. (Los cuales son el comercio de multitienda, construcción y supermercado)
- Variable *dummy* sobre la realización de transacciones o acumulaciones de puntos en un determinado rubro fuera de la empresa. Estos son los rubros usado por la empresa para sus estudios internos, entre los que se encuentran.
 - Educación
 - Supermercado
 - Multitienda
 - Telefonía
 - Clínicas
 - Auto y accesorios
 - Restaurants
 - Viajes
 - Combustible
 - Farmacias
 - Mejora hogar

7.3 Interacción con drivers

- Monto total de avances

¹² Log-in refiero al ingreso a la página web, identificándose a través de un inicio de sesión.

¹³ Las variables *dummy* consisten en variables categóricas que sólo toman valor 1 ó 0, usualmente para verificar la tenencia o no de una característica

- Monto total crédito consumo
- Cupo crédito consumo
- Número de transacciones
- Número de transacciones con cuotas
- Número de transacciones con descuentos
- Puntos acumulados
- Puntos vencidos

Variable	Mínimo	Q1 ¹⁴	Q2	Q3	Q4	Máximo	Media	Desviación estándar
Monto total avance	0	0	0	0	100.000	33.645.935	105.654	309.853
Monto total crédito consumo	0	0	0	0	0	14.887.296	187.134	615.827
Total transacciones en multitienda	0	1	3	6	11	993	7	10
Total transacciones	1	4	11	22	47	6.348	36	66
Total de puntos acumulados	0	995	2.364	4.494	8.896	1.509.740	6.445	13.204

Tabla 1: Descripción variables continuas de interacción

8. Cuantificación de interacción con drivers

Para poder cuantificar la interacción que los clientes tienen con cada uno de los drivers que la tarjeta ofrece, se procedió a generar en concordancia con la empresa un KPI para cada interacción que el cliente pueda tener con la tarjeta, el cual pudiera determinar de manera más precisa la intención del cliente por poder interactuar sobre el máximo posible.

De esta manera los KPI's a generar varían de un *driver* a otro según el tipo de intención por interactuar y las variables que determinan esta interacción.

8.1 Tasa de canje

Para medir si la interacción de los clientes fue alta con el club de puntos, el indicador a usar es el de la cantidad de puntos canjeados sobre los puntos acumulados. Con esto se entenderá, además de si el cliente interactuó con el club, si esta interacción fue máxima según los recursos (puntos) que este tenía en su

¹⁴ Columnas que indican los cuatro valores de cortes para dividir el universo de clientes en cinco grupos iguales.

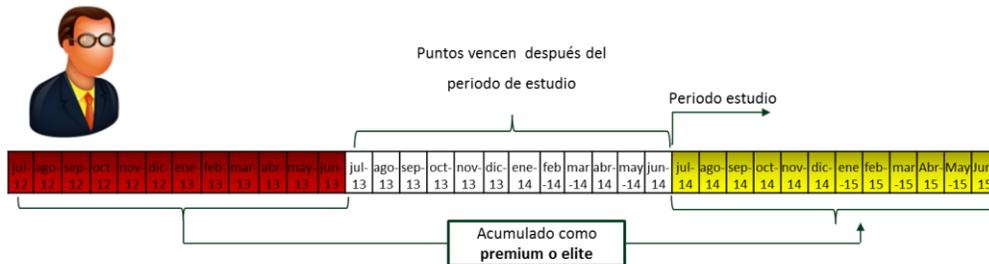
Anexo A: distribución de avances y créditos de consumo sólo para quienes hacen al menos un movimiento

poder. Es por eso un cliente que vence muy pocos puntos sobre los acumulados tendrá un mayor valor en el KPI.

Sin embargo, para el cálculo de esta tasa hay que tener en cuenta que los clientes normales vencen sus puntos un año después, mientras que los clientes premium y elite los vencen dos años después.

Por lo tanto, para calcular los puntos acumulados que realmente debería vencer en el periodo de estudio, es que se contabilizaron únicamente los puntos acumulados un año antes por clientes normales y dos años antes por un premium o elite.

Ejemplo 1: Cliente Premium - Elite



Ejemplo 2: Cliente Normal que en abril – 2014, cambia a Premium

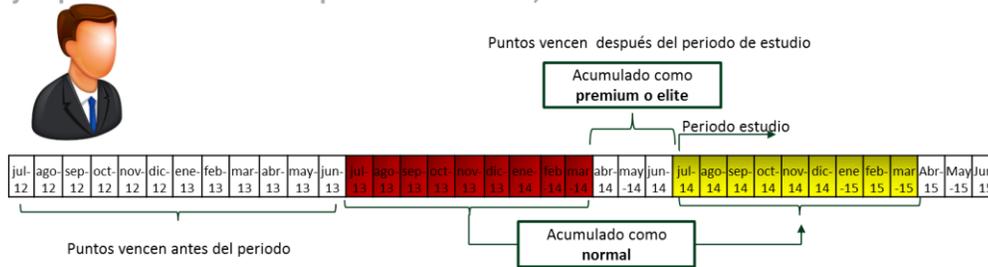


Ilustración 13: Esquema ejemplo del cálculo de puntos utilizados para el cálculo de la interacción, debido a su vencimiento efectivo en el espacio de tiempo a evaluar.

Con los puntos acumulados se calculó la tasa de canje de cada cliente, la cual es el complemento del total de puntos vencidos sobre los acumulados (conocida como breakage), producto de la no existencia de renovación de puntos.

$$1 - \frac{\text{Total puntos vencidos}}{\text{Total puntos acumulados}}$$

Ecuación 5: Tasa de canje de un cliente

Los principales resultados obtenidos es la identificación de una distribución “polarizada en los extremos”, donde se tiende a vencer mucho o muy poco, hecho que es transversal para distintos rangos de puntos acumulados.

Además se observa que a medida que se acumulan más puntos, el porcentaje de clientes que vence gran cantidad de sus puntos disminuye y el porcentaje de clientes que canjean la mayoría de sus puntos aumenta.

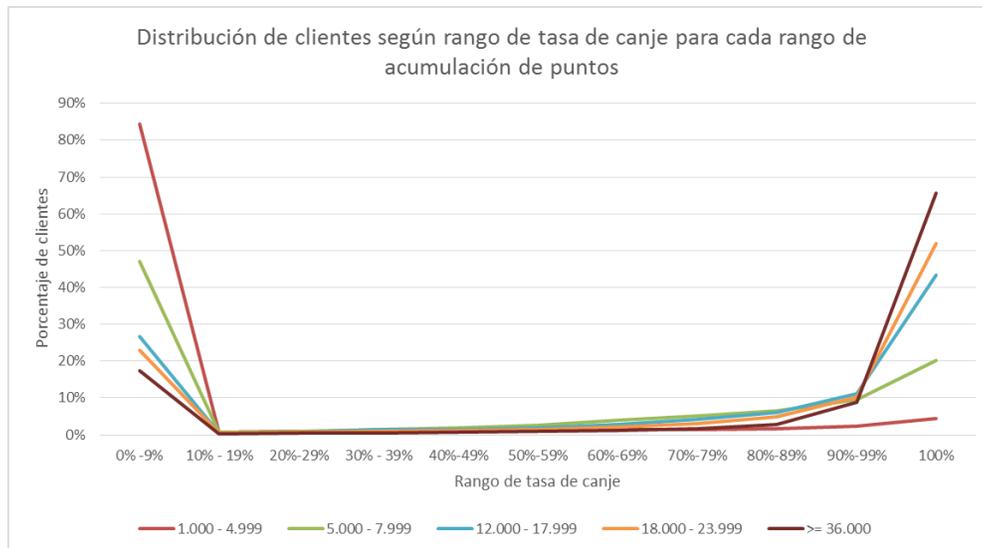


Ilustración 14: Número de clientes según el rango de interacción con el club de puntos.
Fuente: Elaboración propia con datos de la empresa.

8.2 Avance en efectivo

Para cuantificar la intención del cliente por interactuar con el pedido de avances en efectivo, se utilizó como indicador el monto total de avances que el cliente realizó en el periodo de estudio sobre el cupo que este tenía en el periodo de estudio. De esta manera se observará si la interacción fue alta o no en comparación a su máximo posible (cupó).

$$\frac{\text{Monto total en avances}}{\text{Cupo de crédito}}$$

Ecuación 6: Tasa uso avances

8.3 Crédito de consumo

De manera similar al avance, se utilizó el monto total que el cliente cobró como crédito de consumo sobre el cupo que este tuvo en su línea especial de crédito de consumo, para no solo ver si interactuó, sino si esta fue sobre el máximo posible que pudo haber interactuado.

$$\frac{\text{Monto total en crédito de consumo}}{\text{Cupo línea crédito consumo}}$$

Ecuación 7: Tasa uso crédito de consumo

Los principales resultados en torno a la interacción con estos dos préstamos de dinero, son que la cantidad de clientes que realizan avances es cercana al doble

de clientes que realizan créditos de consumo, con 588.784 y 295.145 clientes respectivamente. De estos, se ve que en los avances es más común tener una tasa de interacción más baja que las tasas de créditos de consumo, principalmente debido a los montos asociados al tipo de transacción, donde los créditos de consumo se caracterizan por ser de montos más altos.

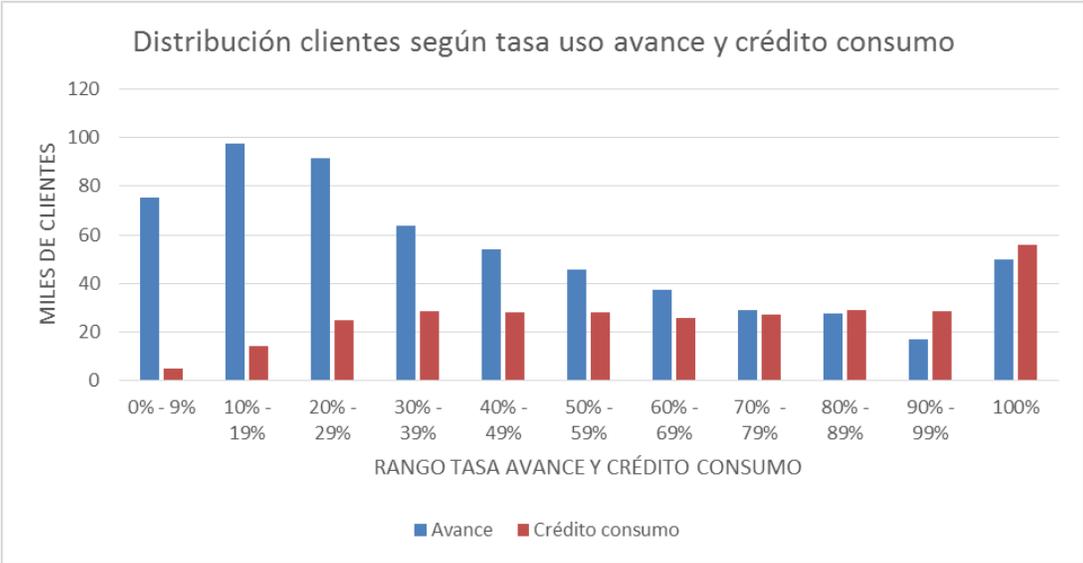


Ilustración 15: Distribución de clientes según interacción con avances y créditos de consumo.
 Fuente: Elaboración propia con datos de la empresa.

8.4 Transacciones con cuotas

Para identificar si un cliente tiene una alta interacción se cuantificó como indicador de intención, el total de las transacciones de compra con cuotas con la tarjeta de multitienda sobre el total de transacciones intencionales que este realizó con la tarjeta (sin considerar PAT). Hecho que al igual que antes verá no solo ve si el cliente interactuó una vez con las cuotas, sino el real uso de las cuotas sobre el total de veces que pudo hacerlo.

$$\frac{\text{Total de transacciones con cuotas}}{\text{Total de transacciones}}$$

Ecuación 8: Tasa compra con cuotas

Los principales resultados indican que en pocas transacciones no se observa una tendencia clara, mientras que a mayores transacciones esta sigue una tendencia cercana al 10% con un posterior decaimiento.

También se puede observar que en números de transacciones muy altos, la mayoría de los clientes tiene tasas muy bajas, lo que puede indicar que este *driver* no es tan atractivo para quienes realizan muchas compras.

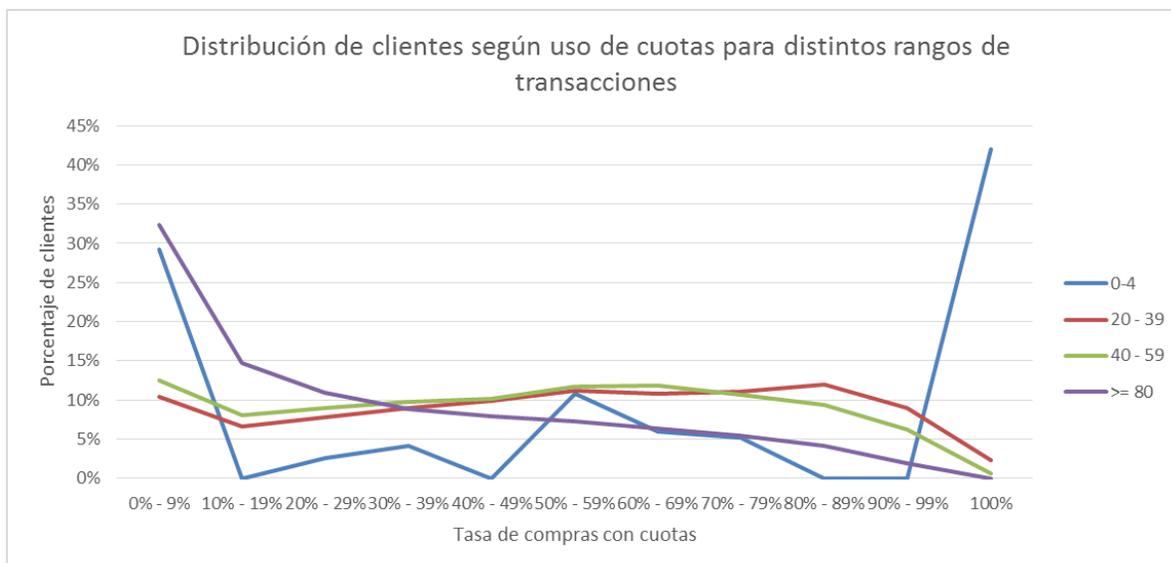


Ilustración 16: Distribución de clientes según tasa de compras con cuotas, para cuatro niveles de transacciones totales.

Fuente: Elaboración propia con datos de la empresa.

8.5 Transacciones con descuentos

De manera similar a la compra con cuotas se cuantificó el total de transacciones que involucra descuentos sobre el total de transacciones. Para este *driver*, a diferencia de las compras con cuotas, se consideraron únicamente las compras en la multitienda interna del holding, debido a que fuera de esta los descuentos no son identificables.

$$\frac{\text{Total de transacciones con descuentos}}{\text{Total de transacciones}}$$

Ecuación 9: Tasa compra con descuentos

Al observar la distribución se puede identificar que la moda se ubica cercano al 30%, con una posterior tendencia al decaimiento. También se puede identificar de manera similar a la compras con cuotas, que para baja cantidad de transacciones, la tendencia no es clara.¹⁵

¹⁵ Anexo C. se puede identificar las matrices de doble entrada sobre el número de clientes según el número de transacciones o acumulación de puntos y la tasa de interacción para cuotas, promociones y puntos.

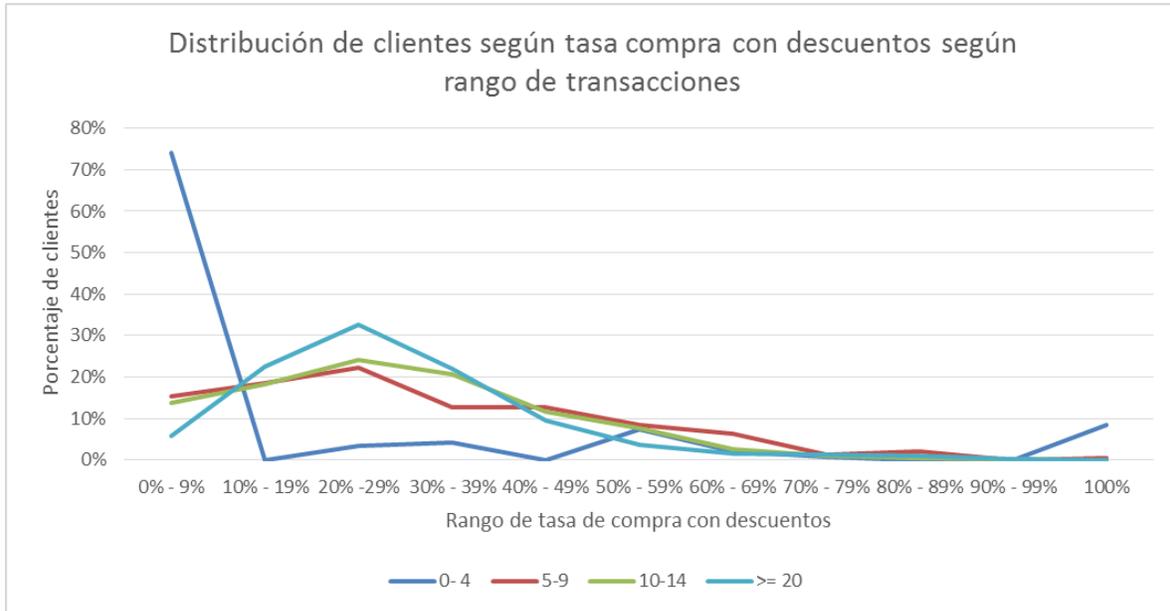


Ilustración 17: Número de clientes según rango de tasa de compra con descuento para cuatro tipos de transacciones totales.
Fuente: Elaboración propia con datos de la empresa.

Los resultados obtenidos de las distribuciones indican que todas las tasas tienen un comportamiento distinto. Para verificar esta diferencia se realizó un análisis de tendencia entre estas tasas, que pudiera señalar la existencia de una relación o grado de dependencia entre alguna de estas tasas de interacción.

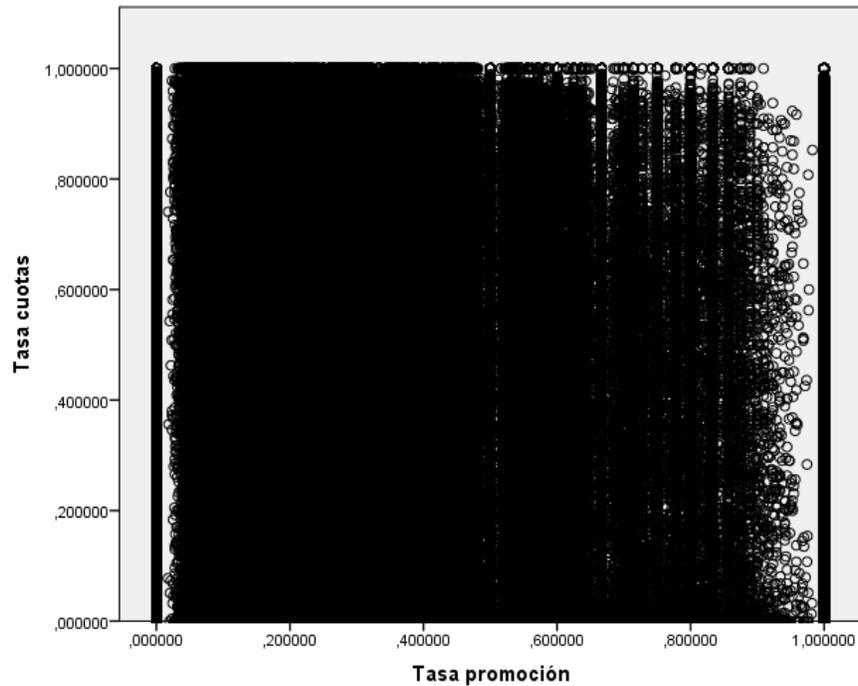


Ilustración 18: Relación entre tasa de promoción y tasa de cuotas

Los resultados del gráfico indican que prácticamente no existe una tendencia clara, sino más bien un efecto de total independencia al cubrir prácticamente todo el espectro de posibles combinaciones. Este resultado de independencia ocurre de manera similar, para el resto de las combinaciones de tasas.¹⁶

Finalmente se complementó el análisis de relación entre las tasas de interacción con una correlación lineal. Los resultados obtenidos son similares al análisis de tendencia al mostrar una correlación baja entre estas tasas, reafirmado la existencia de independencia y por lo tanto, la posible existencia de nichos de clientes distintos para cada uno de los *drivers*.¹⁷

Correlaciones					
Variable	Tasa de canje	Tasa de promoción	Tasa de cuotas	Tasa de avance	Tasa crédito de consumo
Tasa de canje	1	,093	-,120	,062	,179
Tasa promoción	-	1	-,060	-,019	,017
Tasa de cuotas	-	-	1	,068	,024
Tasa avance	-	-	-	1	,270
Tasa crédito de consumo	-	-	-	-	1

Tabla 2: Correlación entre tasas de interacción. Fuente: Elaboración propia con datos de la empresa.

Robustez Indicadores

Debido a que en la empresa nunca se habían generado estos indicadores, se realizó un análisis de robustez de la distribución de los clientes con respecto a estas tasas en cuatro ventanas de tiempo con seis meses de diferencia para saber si la distribución es estable en el tiempo y no por un hecho asociado a estacionalidad.

Los principales resultados obtenidos, son que a pesar de que el número de clientes ha aumentado notoriamente, los valores promedios de los indicadores generados y la distribución de los clientes en torno a la interacción medida por estos indicadores, como es el caso de la tasa de canje que en estas cuatro ventanas se ha mantenido cercana al 64%, lo cual es similar al cálculo de otras tasas de interacción.¹⁸ Cabe destacar que esto señala un “breakage” de 35%, lo cual es alto.

¹⁶ Anexo B. Resto de gráficos de relación de tasas.

¹⁷ Anexo B. Descriptivo de tasas de interacción
Anexo C. Matriz de doble entrada para cuotas

¹⁸ Anexo B. Gráfico de variación de la tasa de canje en el tiempo.

9. Clientes de Alta Interacción

Al observar la distribución de los clientes en torno a la interacción que tienen con los tres *drivers* de la tarjeta, se procedió determinar para cada driver a los clientes que cumplen con el perfil de alta interacción.

9.1 Club de puntos

En el caso de la interacción con el club de puntos, se observó que más del 10% de los clientes logra tener una tasa de canje del 100%, es por esto que en conjunto con la empresa se decidió considerar como clientes de alta interacción son todos aquellos que tuvieran una tasa de canje superior al 90%.

A través de los criterios antes expuestos, se perfilaron a 433.160 clientes de alta interacción con el club de puntos (21%)¹⁹ los cuales representan un 71% de los puntos canjeados y un 65% de los movimientos de canje.

9.2 Avances en efectivo

En el caso de los avances en efectivo, al haber pocos clientes que realizan un movimiento, se determinó en conjunto con la empresa que aquellos clientes de alta interacción, son los pertenecientes al mejor quintil de la tasa de avance. Con este criterio se determinaron 179.768 clientes, quienes representan un 42% de los montos totales emitidos en avances

9.3 Créditos de consumo

De manera similar a los avances en efectivo, se calificó al mejor quintil de clientes como los de alta interacción. Con este criterio se 59.180 clientes de alta interacción para crédito de consumo que representan el 27,7% de los montos totales de créditos de consumo realizados.

9.4 Compras con cuotas

En el caso de las compras con cuotas, el análisis varía respecto a los casos anteriores, ya que se mide el número de transacciones en vez de montos. Este hecho provoca que el número de transacciones realizadas con cuotas, dependa directamente de cuantas transacciones el cliente realizó.

De esta manera, junto con la empresa se decidió realizar una metodología que únicamente comparara a clientes con un similar número de transacciones, para así determinar de una manera más precisa la intención de realizar transacciones por la oferta de cuotas.

¹⁹ Anexo C. Matriz de doble entrada drivers.

Particularmente en este caso se observó que en promedio los clientes realizan 25,4²⁰ transacciones, por lo que se dividió la base de clientes en cinco grupos distintos de acuerdo al número de transacciones y la distribución de transacciones.

A partir de esta división, se tomó como criterio de alta interacción a los clientes pertenecientes al mejor decil de cada grupo, siempre y cuando tuvieran al menos tres transacciones.²¹

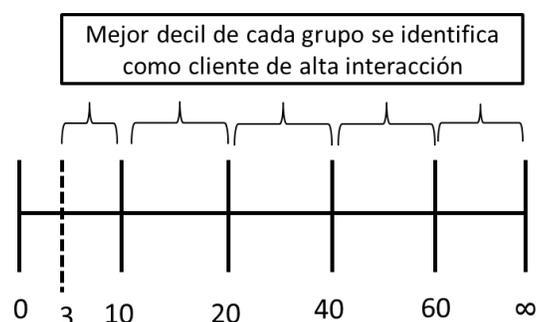


Ilustración 19: División de clientes en quintiles según transacciones totales

Como se mencionó anteriormente, al dividir la base y comparar a los clientes únicamente con aquellos con similar número de transacciones, se tiene una comparación más fidedigna sobre la intención de usar la tarjeta por el *driver*, ya que si se realizará únicamente la elección de las mejores tasas, el universo de clientes de alta interacción se concentraría en pocas transacciones.

De acuerdo a este criterio, y como se observó en la distribución de clientes, la tasa promedio y la tasa de corte para la elección de un cliente de alta interacción va disminuyendo, apaciguando el efecto del número de transacciones. Finalmente se identificaron 264.489 (12,8%) clientes de alta interacción con las compras con cuotas.

Quintil	I	II	III	IV	V
Tasa promedio	60,90%	56,80%	52,50%	47,60%	34,23%
Tasa de corte	100%	93,70%	90,40%	85,70%	76,60%

Tabla 3: Tasa promedio y tasa de corte del mejor decil para cada quintil de transacciones

9.5 Compras con promociones

De la misma manera que las compras con cuotas, la determinación de los clientes de alta interacción con la compras con promociones se llevó a cabo a través

²⁰ Este promedio se considera eliminando al 5% de los clientes ubicados en los extremos de la distribución para evitar el efecto de *outliers*.

²¹ Cabe destacar que en casos de empates, también se consideraron como clientes de alta interacción, ya que su tasa de interacción era igual que algunos clientes considerados de alta interacción.

de la división de clientes en cinco grupos iguales de acuerdo al número de transacciones realizadas en la multitienda del holding.

Consecuentemente con las compras con cuotas, se consideró a un cliente de alta interacción con las promociones a aquellos pertenecientes al mejor quintil de cada grupo, siempre y cuando hayan realizado al menos tres transacciones. De esta manera se identificaron 287.215 (13,77%) clientes de alta interacción con las promociones.²²²³

Quintil	I	II	III	IV	V
Tasa promedio	30%	28,51%	27,60%	27,34%	28,50%
Tasa corte	50%	44%	40%	37,50%	37,90%

Tabla 4: Tasa promedio y tasa de corte mejor quintil para cada quintil de transacciones

Distribución universo de clientes

Finalmente con la identificación y perfilamiento de los clientes de alta interacción para cada uno de los drivers, se procedió a construir un diagrama de Venn con los clientes de alta interacción con promociones, puntos y cuotas^{24 25}.

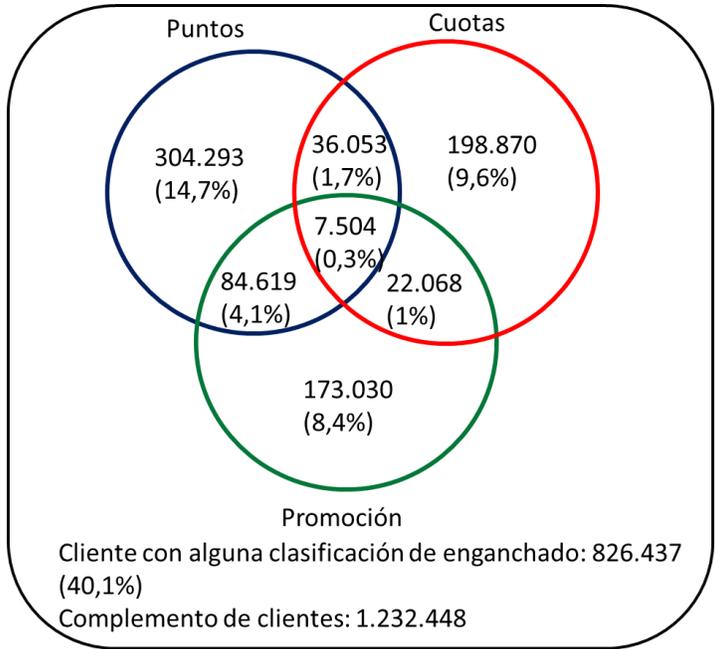


Ilustración 20: Diagrama de Venn de la distribución de clientes del universo a trabajar

²² Anexo C. Tabla de tasa promedio y tasa de corte para el caso de promociones. Esta tiene un valor más constante por quintil de trabajo, alcanzando cerca de un 30%.

²³ La razón de considerar el mejor quintil de cada grupo, alude a que la empresa consideraban que el mejor decil de clientes eran muy pocos en comparación a los clientes de alta interacción de los otros drivers.

²⁴ Se construyó el diagrama sin considerar avances y créditos de consumo, ya que como se explicó en los alcances, el modelo de deuda solo considerará la interacción con las cuotas.

²⁵ Mapa de Venn es una representación gráfica en el ámbito de las matemáticas para identificar conjuntos, intersecciones, etc.

A partir de estos resultados, se puede observar que si bien existen clientes que pertenecen a más de un perfil de alta interacción, la mayoría sólo pertenece a un solo perfil.

Por otro lado, al observar el gasto que representan cada uno de estos perfiles del total anual, los de alta interacción con el club de puntos explicaron el 51% del gasto, con promociones el 19% y con cuotas el 8%.

Finalmente en base a la distribución generada por el diagrama de Venn, se propuso en conjunto con la empresa la hipótesis que los clientes de alta interacción con cada uno de estos *drivers* tienen características distintas, representado por un determinado perfil.

Caracterización RFM

Luego de identificar a los clientes pertenecientes a cada uno de los perfiles de alta interacción, se realizó un análisis RFM sobre estos grupos para comprender la similitud y diferencia que estos grupos tienen en torno a variables transaccionales de corto plazo, las que difieren de características demográficas u otras que expliquen un fenómeno al “largo plazo”.

Al realizar la frecuencia de pago (Meses en que se realiza transacción), se observó que los comportamientos son bastantes similares al concentrar a la mayoría de los clientes con transacciones en los doce meses de año.

Sin embargo es importante destacar el “peak” que tienen el perfil de cuotas en los tres y doce meses. Hecho que se debe a la construcción del perfil que admite mínimo tres transacciones, concentrando un nicho en tres transacciones y otro que es fiel a comprar los doce meses con cuotas.

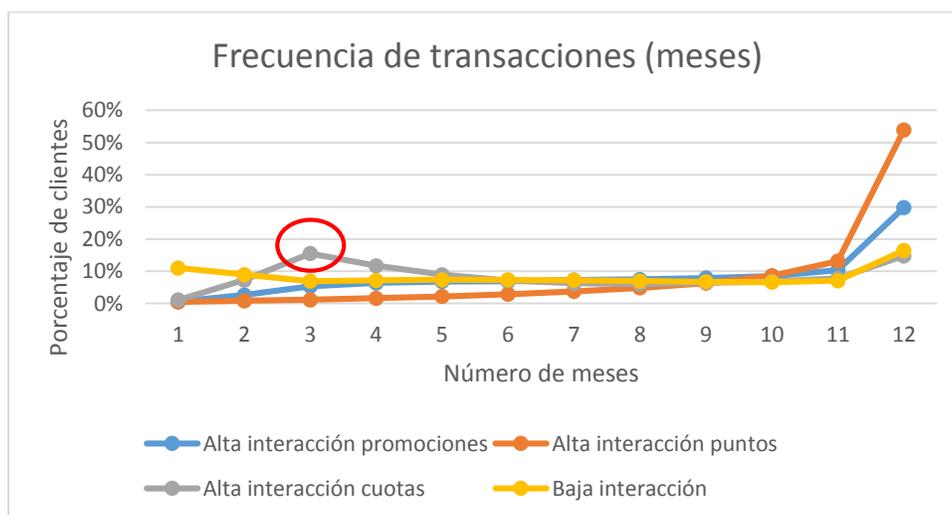


Ilustración 21: Frecuency de compra perfiles de clientes

En el caso del *recency* de compra, se observa que el comportamiento de los cuatro grupos de clientes es similar al comprender a la mayoría de los clientes con compras realizadas en el último mes de estudio y muy pocos clientes en meses posteriores.

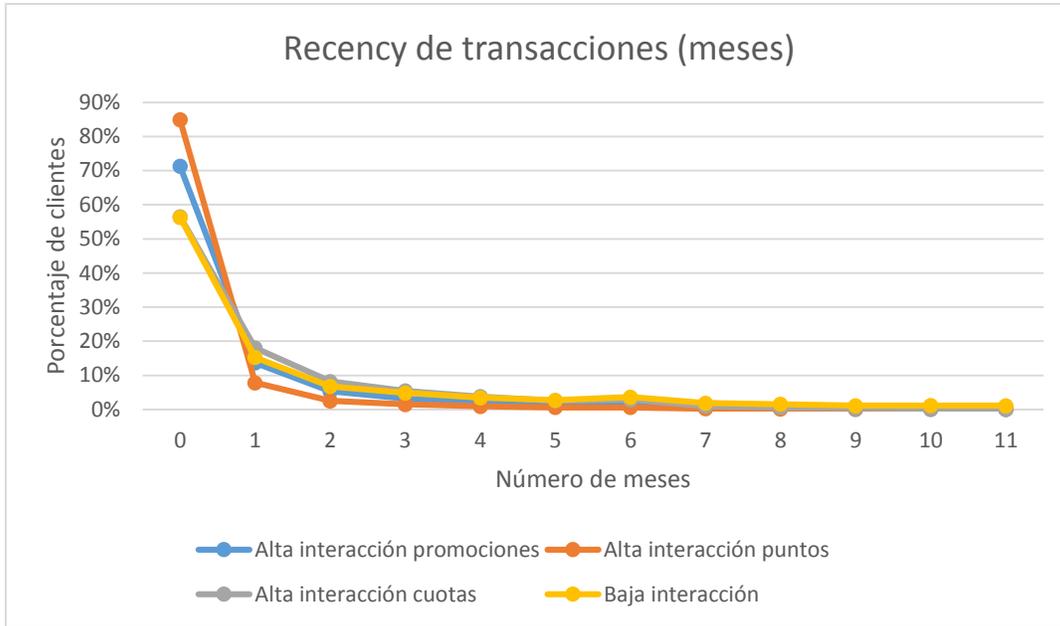


Ilustración 22: Recency compra perfiles de clientes

Finalmente en el caso del ticket promedio, el comportamiento de ticket promedio entre grupos se observa nuevamente similar.

- Baja interacción: \$49.876
- Alta interacción cuotas: \$48.9436
- Alta interacción promoción: \$48.002
- Alta interacción puntos: \$45.174

En conclusión con esta sección, se puede observar lo comparable que son los grupos de clientes y el efecto similar que hay entre estos grupos si el análisis es realizado desde una perspectiva del “corto plazo” con variables muy relacionadas, incluso constructoras de la variable dependiente.

Sin embargo este hecho no necesariamente se puede trasladar hacia un análisis de “largo plazo” donde el perfilamiento de los clientes ocurra con variables demográficas y transaccionales que no “construyan” la variable dependiente. De esta manera el análisis a seguir buscará la existencia de distintos perfiles a “largo plazo” con variables que no tengan incidencia en la construcción del efecto a evaluar.

10. Importancia de Variables

Posterior a generar la variable dependiente (alta o baja interacción con un determinado *driver*), se generó como paso previo al modelo de clasificación supervisada, un proceso para determinar las variables más importantes dentro de la gran cantidad de variables independientes recopiladas.

En consecuencia al trabajo previo, se utilizó cómo universo de variables para cada *driver*, todas las variables demográficas y transaccionales descritas anteriormente para la formación de la base analítica.

Para este proceso se tomaron en cuenta dos criterios de importancia de variables. Por un lado se crearon tablas de contingencia y por otro se generó un árbol tipo CART que logra crear un ranking de las variables más importantes.

10.1 Tabla de contingencia

Las tablas de contingencia son una metodología que permite ver el grado de importancia que tiene una determinada variable, por medio de la proporción de casos positivos y negativos que existen en la muestra, cuando esta variable toma un determinado valor.

Posteriormente se compara la proporción de casos positivos para cada uno de los escenarios o valores de la variable, con respecto a la proporción poblacional de casos positivos, obteniendo un valor denominado ODD Ratio el cual indica cuanto más probable es observar un caso positivo cuando la variable toma ese determinado valor.

Este valor de ODD Ratio se puede apreciar cuando el cliente tiene un cupo mayor o igual a 1.500.000 pesos, este es 1,75 veces más propenso a tener una alta interacción con el club de puntos.

Cupo	Cientes alta interacción puntos	Cientes baja interacción puntos	Total clientes	Cientes Alta int/Total	ODD Ratio	LN(ODD)	LN(ODD)	Percentil	peso
0 - 399.999	18.609	474.381	492.990	3,77%	0,180	-1,716	1,716	23,95%	41,10%
400.000 - 899.999	102.015	530.713	632.728	16,12%	0,768	-0,265	0,265	30,73%	8,13%
900.000 - 1.499.999	118.826	290.804	409.630	29,01%	1,381	0,323	0,323	19,90%	6,42%
>= 1.500.000	193.016	330.456	523.472	36,87%	1,755	0,563	0,563	25,43%	14,31%
Total general	432.466	1.626.354	2.058.820	21,01%	1,000	-	-	100,00%	69,96%

Tabla 5: Ejemplo de la tabla de contingencia para la formación de la importancia de las variables

Finalmente comparando el ODD Ratio con el percentil de clientes de la población que tienen dicha característica, se genera un porcentaje el cual indica la importancia de la variable para predecir el comportamiento de la variable dependiente.

Realizando este proceso para cada una de las variables en los tres *drivers* de estudio, se generaron rankings de importancia para cada uno de los *drivers* de estudio. En el caso de la interacción con puntos, se observa particularmente la

importancia que tiene el cupo y el haber interactuado en determinados rubros fuera del holding, como lo son las multitiendas y supermercados.

Variables interacción puntos	Peso
Cupo	69,96%
Acumular en rubro multitiendas	55,87%
Acumular en rubro farmacia	44,84%
Acumular en supermercado	44,74%
Tenencia visa	38,54%
Meses look	32,24%
Tenencia de hijo	31,01%
Tenencia adicional	26,51%
Antigüedad	26,20%
Edad	8,91%
Acumular en rubro telefonía	7,30%
Genero	6,43%
Dos contratos	6,32%
Zona geográfica	6,20%

Tabla 6: Comparación importancia variables para la alta interacción con el club de puntos

En el caso de la alta interacción con la compra con cuotas, se observa la importancia de realizar una transacción en el área de la educación, la actividad que el cliente tiene en cuanto al log-in de la página, el cupo que este tiene y la variable demográfica de la residencia geográfica del cliente.

Variable cuotas	Peso
Transacción Educación	33,00%
Días look	23,7%
Cupo	20,54%
Zona geográfica	17,94%
Transacción en rubro viaje	17,30%
Ticket promedio	17,29%
Transacción en rubro restaurant	17,16%
PAT	16,18%
Transacción Supermercado interno	14,72%
Edad	13,82%
Tenencia visa	12,50%
Transacción en rubro telefonía	12,36%
Categoría cliente	12,19%
Transacción en rubro supermercado	10,88%
Pide avance	7,44%
Género	6,78%
Adicional	6,32%

Transacción comercio interno construcción	6,16%
Antigüedad	6,10%
Monto avance	5,57%
Transacción multitienda interna	4,14%
Dos contratos	1,50%
Tiene hijos	0,97%

Tabla 7: Importancia variables para la alta interacción con las cuotas

Finalmente en el caso de la alta interacción con las promociones, se puede observar que a diferencia de los puntos, la edad del cliente tiene una mayor importancia a la hora de describir una alta interacción, dejando al género nuevamente como una variable poco relevante.

Promoción	Peso
Transacción en diciembre	27,98%
Transacción en noviembre	23,66%
Transacción en rubro multitienda	18,03%
Transacción en electro hogar	16,28%
Edad	16,10%
Transacción en rubro supermercado	14,50%
Cupo	13,98%
Meses look	13,58%
Tiene hijos	13,08%
Categoría club de puntos	12,39%
Visa	11,34%
Ticket promedio	10,30%
Tiene adicional	5,63%
Zona geográfica	4,5%
Transacción fuera de comercio multitienda	4,16%
Dos contratos	2,72%
Antigüedad	1,30%
Genero	1,27%
Antigüedad	1,30%
Genero	1,27%

Tabla 8: Importancia variables para la alta interacción con promociones

10.2 Árbol tipo CART

Para contrastar la importancia antes generada, se realizó un análisis de un árbol tipo CART el cual entrega la importancia que tiene cada variable a la hora de generar el modelo. Los principales outputs que entrega este método son el peso o importancia de la variable, así como la importancia normalizada.

En el caso de la interacción con el club de puntos, se puede observar nuevamente que la tendencia de las variables importantes es similar a la entregada el método de tablas de contingencia.

Variable independiente	Importancia	Importancia normalizada
Acumular en rubro supermercado	,041	100,0%
Acumular en rubro Farmacia	,040	97,2%
Acumular en rubro multitienda	,031	74,4%
Días de looks	,028	67,6%
Cupo	,027	66,9%
Meses de Looks	,023	55,9%
Tenencia de Hijo	,019	46,2%
Acumular en comercio interno construcción	,010	25,3%
Monto avance	,010	23,8%
Tenencia adicional	,008	18,6%
Acumular en comercio interno supermercado	,007	17,5%
Antigüedad	,007	16,2%
Acumular en rubro telefonía	,007	16,0%
Acumular en comercio interno multitienda	,006	15,4%
Tiene PAT	,005	11,7%
Tenencia dos tarjetas	,004	10,4%
Tenencia de tarjeta visa	,002	4,3%
Genero	,001	1,8%
Edad	,001	1,5%
Zona geográfica	,000	,6%

Tabla 9: Importancia de las variables para alta interacción con puntos según Árbol tipo CART

Realizando este análisis para las promociones y compras con cuotas se puede observar que en ambos casos ocurre la misma consigna que la interacción con los puntos, obteniendo resultados similares en cuanto a la importancia de las variables.²⁶

11. Análisis Exploratorio

En el caso de la interacción con el club de puntos se observó entre otras variables, el efecto que tiene el cupo sobre la probabilidad de pertenecer al perfil de alta interacción. En este caso la tendencia obtenida fue que a medida que aumenta el cupo de los clientes hay una mayor proporción de clientes que pertenecen al perfil de alta interacción.

²⁶ Anexo E. Importancia variables en modelo de promociones y cuotas según el modelo árbol CART.

La razón de este comportamiento se puede deber a que esta variable entrega una noción del poder adquisitivo que el cliente tiene, por lo que este tendrá mayores posibilidades de poder llegar a canjear sus puntos y evitar vencer gran parte de estos. De manera similar fue la tendencia de las principales variables obtenidas del estudio como la importancia de acumulación en ciertos rubros, comportamiento de ingreso a la página y tenencia de hijo que pueden apuntar a un mayor involucramiento y gasto con el club.²⁷



Ilustración 23: Tasa alta interacción según rango de cupo

En el caso del análisis con respecto a la alta interacción con las cuotas, se estudió el caso de la variable zona geográfica, la cual es una de las variables más representativas. Los principales resultados arrojan un comportamiento donde los clientes que viven en las zonas asociadas a un mayor poder económico (oriente y centro RM) reflejan una menor tasa de clientes con alta interacción, mientras que la zona poniente y norte reflejan una mayor tasa dentro de la Región Metropolitana.

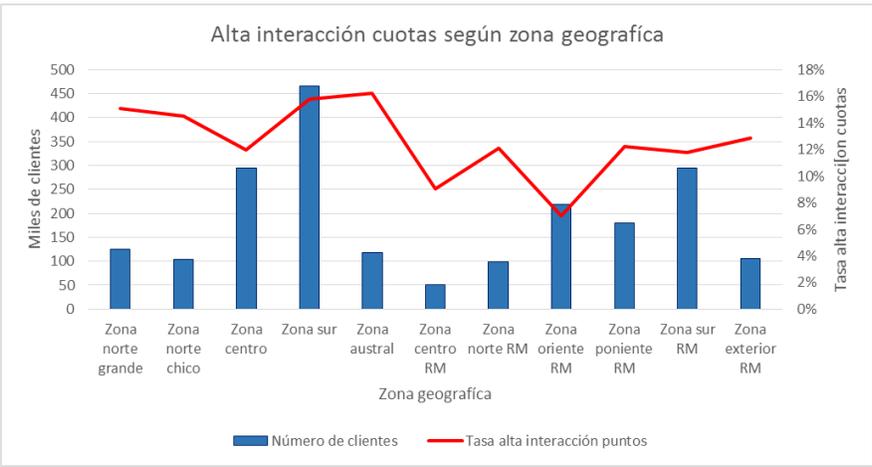


Ilustración 24: Clientes de alta interacción con cuotas según la zona geográfica de residencia

²⁷ Anexo D. análisis exploratorio de variables acumulación en rubros, ingreso a la página e hijos con el club.

Al realizar el análisis de cuotas con otras variables importantes como el cupo y categoría del cliente se observa que características asociadas a un menor poder económico (bajo cupo y menor categoría) tienen una mayor tasa de clientes que pertenecen al perfil de alta interacción, así como una baja interacción con la página y mayor edad.²⁸

Por último, en el caso de la interacción con las promociones se estudió entre otras variables, el efecto etario producto de la importancia que tiene esta variable a diferencia del resto de los *drivers*. En este análisis se puede observar que en los rangos de personas más jóvenes (menor edad) tienden a haber una mayor proporción de clientes de alta interacción con las promociones.

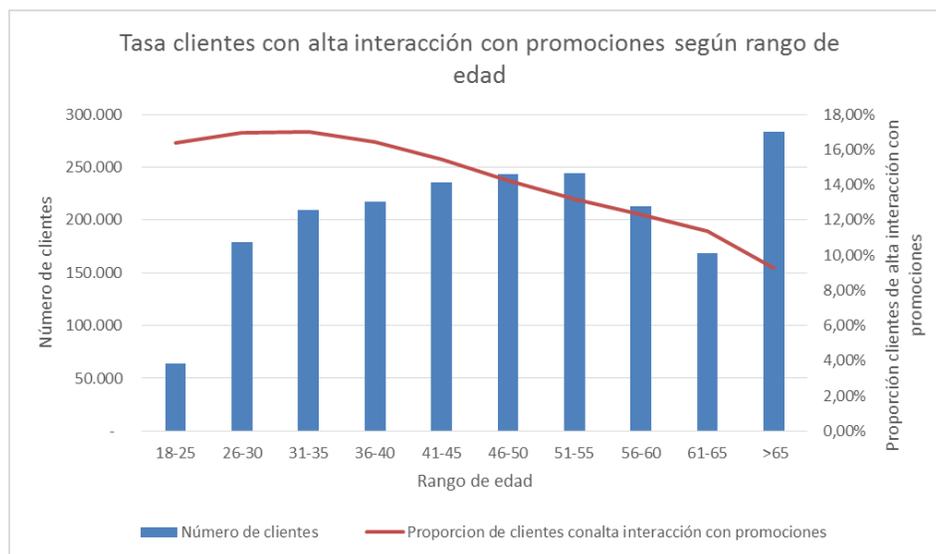


Ilustración 25: Clientes de alta interacción con promociones según rango de edad

Este efecto se puede deber a que producto de la renta y necesidad de las personas en etapas más jóvenes, el poder comprar “ahorrando” cobra mayor importancia que en etapas más longevas. De manera similar se observó en otras variables importantes el efecto positivo del cupo y categoría del cliente, así como el uso de la tarjeta en diciembre o para comprar electrodomésticos.²⁹

Discusión sobre resultados

Por medio del análisis exploratorio realizado con los tres principales *drivers*, se observó que en general estos apuntan a clientes distintos. Por un lado los clientes del perfil de puntos apunta a características de asociadas a un mayor poder económico y uso en rubros particulares, mientras que cuotas apunta a características que asociadas a un bajo poder económico, poco uso y una mayor edad.

²⁸ Anexos D. Análisis exploratorio variables con respecto a la alta interacción con cuotas.

²⁹ Anexos D. Análisis exploratorio variables con respecto a la alta interacción con promociones.

Más aún los clientes de promociones apuntan a edades más jóvenes que cuotas, y si bien tienen poder económico, este es de menor valor que los de puntos.

12. Generación de Modelo de Propensión

Con la información de las posibles variables que puedan explicar la pertenencia al perfil de interacción con el club de puntos, compras con cuotas y compras con promociones, y la importancia que representan cada una de estas variables en explicar este fenómeno, se procedió a realizar un modelo de propensión.

Sin embargo a la hora de realizar estos modelos predictivos, un problema que podemos identificar es el desbalanceo de las clases a predecir. Este hecho se debe a que a la hora de elegir a los clientes del perfil de alta interacción, estos fueron una cantidad menor del universo de clientes.

Debido a este inconveniente la literatura recomienda dos opciones. Por un lado realizar un modelo ponderado que resuelva este efecto del desbalanceo, y por otro lado, el balanceo de la base de entrenamiento. En este caso, el método elegido es el del balanceo de la base producto de su fácil y rápida interpretación al balancear la base en 50/50.

De esta manera, para la realización del modelo se separó la base en un 80% de base de entrenamiento del modelo (El cual está balanceada) y un 20% para la validación de los resultados (la cual está desbalanceada de acuerdo a proporción original).

12.1 Club de Puntos

12.1.1 Regresión logística

Para la realización de este modelo, no se consideraron variables que afectarán directamente la construcción del KPI de interacción, como lo son la categoría del cliente en el club, variables RFM y puntos acumulados.

Por otro lado, cabe destacar que si se ocupó la variable *dummy* de acumulación o no en determinados rubros fuera del holding y comercios internos, debido a que prácticamente todos los clientes acumulan algún punto y no hace alusión a montos, si no el uso de la tarjeta.

Finalmente se realizaron distintas iteraciones de modelos con la transformación de ciertas variables como el cupo, probando con intervalos de montos o el logaritmo que hace alusión a efectos mayores en cupos de menor monto.

<i>Variable</i>	Coeficiente	Significancia	Exp(coef)
<i>Acumular en rubro supermercados</i>	0,637	0,000	1,890
<i>Acumular en rubro multitienda</i>	0,775	0,000	2,170
<i>Logaritmo del cupo</i>	0,894	0,000	2,444
<i>Tenencia visa</i>	0,239	0,000	1,270
<i>Número de hijos</i>	0,579	0,000	1,784
<i>Días de look</i>	0,017	0,000	1,017
<i>Tenencia adicional</i>	0,355	0,000	1,426
<i>Antigüedad (0- 4 años)</i>	-0,053	0,000	0,948
<i>Antigüedad (5-9 años)</i>	0,063	0,000	1,065
<i>Antigüedad (10 - 14 años)</i>	0,048	0,000	1,049
<i>Constante</i>	-13,509	0,000	0,000

Tabla 10: Resultados coeficientes regresión logística binaria

Los principales resultados obtenidos, son el efecto positivo que tiene el acumular puntos en rubros que son directamente la competencia del holding (tanto es su comercio de supermercado como de multitiendas). Hecho que es de gran relevancia si se considera que estas entidades son la competencia directa del holding.

Por otro lado se puede evidenciar el efecto positivo que tiene el tener un mayor cupo o una tarjeta visa que es concordante con el análisis exploratorio, así como el tener actividad con la página web al registrar días de “log - in”.

En el caso de la antigüedad del cliente, se indica que los clientes más nuevos son los menos propensos, lo que se puede deber a que llevan menos tiempo inserto en los beneficios del club. Sin embargo el mejor segmento es aquel con antigüedad entre cinco y diez años, mostrando un “apogeo” de atracción al club.

Finalmente existe un efecto positivo en tener un cliente adicional, lo que se puede deber a la existencia de dos canales de acumulación de puntos para un mismo pozo, así como el tener un hijo que puede incrementar el gasto con la tarjeta por posible nuevas necesidades.

Desempeño del modelo

Para medir la calidad del modelo en torno al ajuste y poder predictivo, se observa que el modelo logra explicar entre un 28% y 38% de la variabilidad que tiene la variable dependiente, así como el predecir correctamente el 74% de los casos.

-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke	AIC	BIC	Predicción global validación	Predicción alta interacción validación
726731,857	0,285	0,380	726.753,857	726.879,775	73,1%	74%

Tabla 11: Ajuste y predicción modelo de alta interacción

Sin embargo la forma elegida para evaluar el modelo, es el método de curva de ganancia y áreas bajo la curva de la curva ROC. Estos métodos consisten en evaluar cuantos clientes que realmente tienen una alta interacción, se logran identificar en los niveles de mayor probabilidad otorgado por el modelo.

De esta manera se ordenaron por su probabilidad pronosticada, viendo en cada decil de clientes, cuantos realmente tienen una alta interacción en comparación a la población real de clientes de alta interacción.

percentil	Probabilidad de corte	Probabilidad promedio	total clientes	alto nivel	Alta interacción capturada	tasa real
10%	,801	88,1%	41.203	26.678	30,8%	64,75%
20%	,674	80,8%	82.410	44.763	51,6%	54,32%
30%	,564	74,5%	123.611	57.595	66,4%	46,59%
40%	,471	68,8%	164.816	66.688	76,9%	40,46%
50%	,386	63,6%	206.018	73.598	84,9%	35,72%
60%	,303	58,7%	247.221	78.810	90,9%	31,88%
70%	,226	54,1%	288.421	82.588	95,2%	28,63%
80%	,153	49,7%	329.623	85.038	98,1%	25,80%
90%	,081	45,5%	370.984	86.366	99,6%	23,28%
100%	-	41,4%	412.036	86.714	100,0%	21,05%

A partir de esta tabla, se puede graficar la ganancia que entrega el modelo en comparación a tomar la decisión “lanzando una moneda”, es decir con 50% de probabilidad de calificar al cliente. Los principales resultados indican que al observar el decil de clientes con mejor probabilidad, se logró identificar cerca de un 31% de los clientes que realmente tienen una alta interacción con el club de puntos.³⁰

Hecho complementario es la curva ROC generado por la probabilidad pronosticada la cual tiene un área 81,1% sobre los casos correctamente clasificados.

³⁰ Anexo E. Curva ROC de desempeño

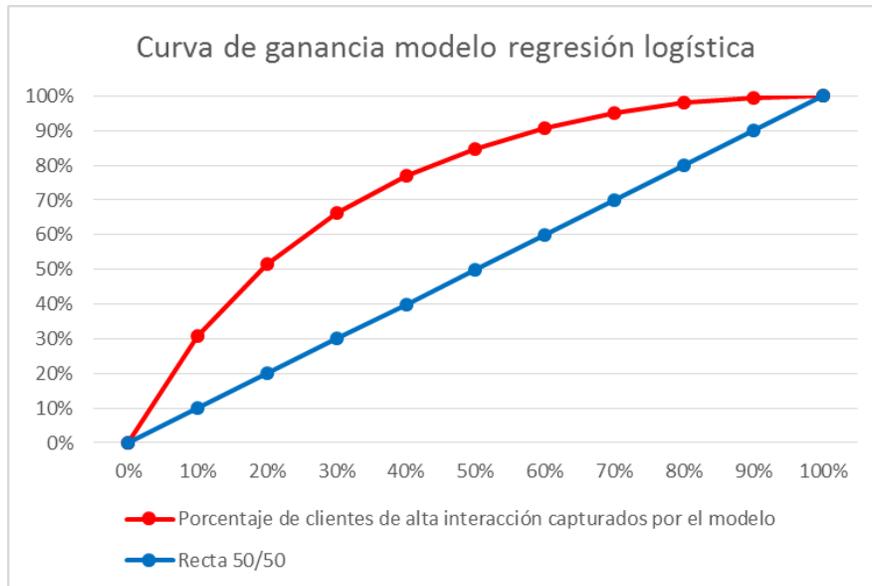


Ilustración 26: Curva ganancia información clientes de alta interacción

12.1.2 Árbol decisión

Para contrastar el modelo de regresión lineal, se generó un modelo de árbol de decisión con las mismas variables y bases que el modelo de regresión lineal (entrenamiento balanceado y test del 20% desbalanceado). Luego de probar el modelo con árboles tipo CART, CHAID y QUEST se obtuvo que el árbol de mejor desempeño fue el árbol CART. Los principales resultados son similares a la regresión logística, logrando acertar a cerca de tres cuartos de los clientes y un mejor desempeño en calificar a los clientes de alta interacción.

Muestra		Pronosticado		
		Baja ³¹ interacción	Alta interacción	Porcentaje correcto
Entrenamiento	Baja interacción	253.514	92.537	73,3%
	Alta interacción	88.884	256.868	74,3%
			Porcentaje Global	73,8%
Contraste	Baja interacción	238.300	87.022	73,3%
	Alta interacción	22.338	64.376	74,2%
			Porcentaje Global	73,5%

Tabla 12: Desempeño modelo árbol decisión en alta interacción con los puntos

La manera que árbol de decisión se construyó, fue eligiendo como primera variable discriminante el cupo, generando un propensión del 30,4% para cupo superiores 606.000 pesos. Posteriormente la segunda variable discriminante es el

³¹ Cabe destacar que en ese caso el corte con probabilidad 0,5 producto del balanceo 50/50

haber acumulado en el rubro de las multitiendas, resultando en el nodo una propensión del 46,7%, reflejado en un lift del 2,2% para ese nodo.

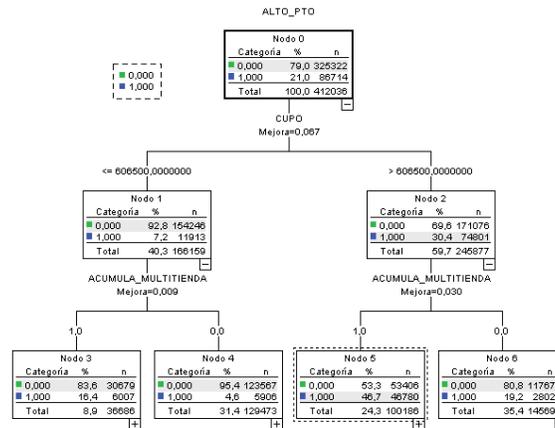


Ilustración 27: Ramas superiores árbol decisión muestra de validación - puntos

Luego en niveles inferiores el árbol crece incorporando la acumulación en el rubro de supermercados, caracterizando en caso posterior, el efecto de tener hijos y los días que el cliente ingresa a la página, obteniendo en el nodo más propenso una probabilidad de 73,3% (lift 3,49%)

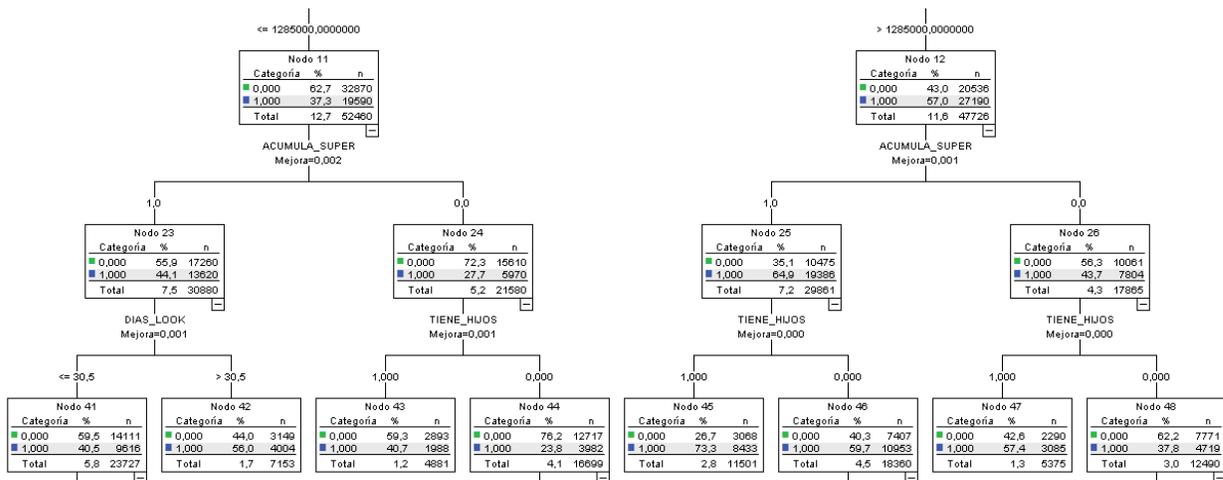


Ilustración 28: Ramas inferiores árbol decisión muestra de validación - puntos

Evaluando nuevamente el modelo a través de la curva de ganancia, se puede observar que la curva de ganancia de información es similar a la entregada por el modelo de regresión logística, donde el mejor percentil de clientes logra capturar el 30% de los clientes de alta propensión.

Al igual que la regresión logística, la curva ROC generada fue de un 81,1% de casos correctamente especificados.³²

³² Anexo E. Curva ROC modelo árbol decisión.

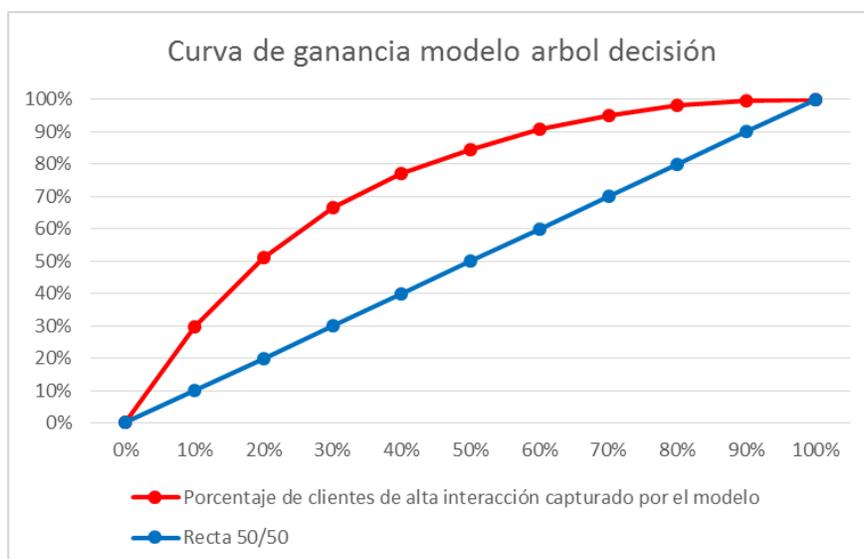


Tabla 13: Curva de ganancia árbol decisión en alta interacción con puntos

12.2 Compras con cuotas

En el caso de la interacción con cuotas, si se incorporó como posible variable la categoría del cliente, ya que este no afecta necesariamente comprar siempre con cuotas o no.

Si bien no se ocuparon variables de “corto plazo” como frequency o recency, se decidió incorporar el ticket promedio (monto promedio de compras) ya que este no “construye” el KPI de alta interacción y puede entregar nociones de comportamiento que a la empresa le interesa.

12.2.1 Regresión logística

Al igual que el modelo de puntos, se realizaron distintas iteraciones con distintas transformaciones de las variables obteniendo los siguientes resultados.

Variable³³	Coefficiente	Significancia	Exp(coef)
<i>Realizar transacción en rubro educación</i>	-0,571	0,000	0,565
<i>Realizar transacción en rubro restaurant</i>	-0,230	0,000	0,794
<i>Realizar transacción en rubro viajes</i>	-0,397	0,000	0,672
<i>Realizar transacción en rubro supermercado</i>	0,347	0,000	1,013
<i>Logaritmo del ticket promedio</i>	0,227	0,000	1,254
<i>Zona norte Grande</i>	0,150	0,000	1,162
<i>Zona norte chico</i>	0,115	0,000	1,122
<i>Zona centro</i>	-0,141	0,000	0,868
<i>Zona sur</i>	0,165	0,000	1,179

³³ Los niveles bases son Zona exterior RM para zona y Cliente normal para categoría del cliente

<i>Zona austral</i>	0,224	0,000	1,252
<i>Zona norte RM</i>	-0,037	0,075	0,964
<i>Zona sur RM</i>	-0,064	0,000	0,938
<i>Zona centro RM</i>	-0,281	0,000	0,755
<i>Zona oriente RM</i>	-0,597	0,000	0,550
<i>Zona poniente RM</i>	0,027	0,139 ³⁴	1,027
<i>Logaritmo del cupo</i>	-0,230	0,000	0,794
<i>Días de Looks</i>	-0,022	0,000	0,978
<i>Categoría premium</i>	0,034	0,006	1,035
<i>Edad</i>	0,013	0,000	0,397
<i>Constante</i>	0,286	0,000	1,415

Tabla 14: Coeficientes modelo regresión logística para alta interacción con las cuotas

Los principales resultados tienen sentido con el análisis exploratorio, al indicar que el efecto “poder económico” tiene gran relevancia en pertenecer a este perfil. Este hecho se refleja con el efecto negativo que tiene el tener un mayor cupo o el vivir en la zona centro u oriente de la Región Metropolitana.

Por otro lado se puede apreciar el efecto positivo que tiene la edad del cliente y el efecto negativo que tiene el ser un cliente “activo” en términos de Ingreso a la página de internet, mostrando un perfil de clientes con menor interés por las ofertas de la tarjeta (puntos, descuentos, etc.) en la página y que posiblemente necesiten el préstamo de dinero por su etapa de vida.

Por último es importante resaltar el efecto negativo que tiene el hacer transacciones en rubros asociados a educación, restaurants y viajes, lo que se puede justificar por el hecho que estos rubros están asociados a pagos de habitualidad (donde el pagar con cuotas no es un incentivo a ocupar la tarjeta) y pagos que en general tienen un gran valor monetario.

Hecho contrario a las compras en supermercados, que dado su necesidad en el día a día y valores de los productos, pueden ser rubros que atraen más a ocupar la tarjeta por esta oferta.

Al estudiar el modelo a través de la curva de ganancia que este genera, se puede observar que si bien no es tan bueno como el modelo de punto, este logra capturar el 24% de los clientes en el mejor decil.

³⁴ Si bien la zona poniente y norte de la RM no son significativas al 95%, se mantuvieron ya que de todas maneras entregan información al modelo.

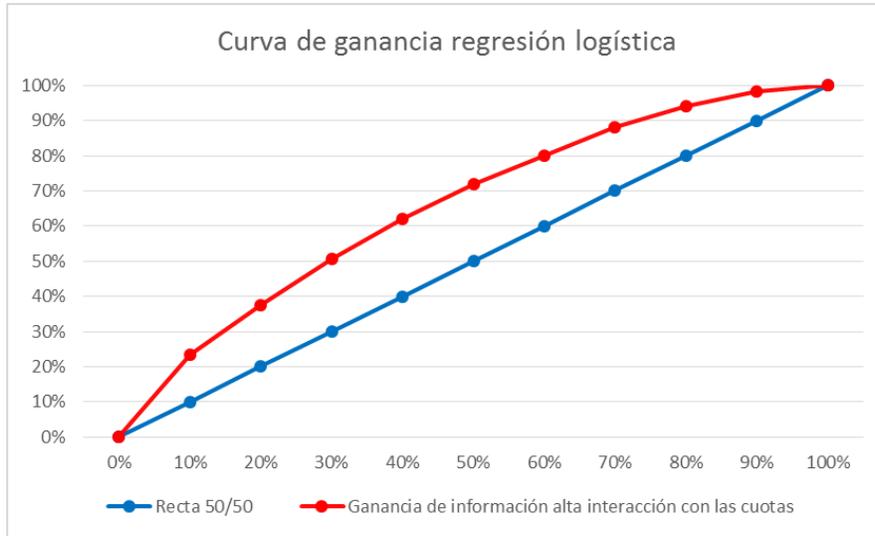


Ilustración 29: Curva de ganancia alta interacción cuotas (regresión logística)

12.2.2 Árbol decisión

Al generar el modelo árbol de decisión sobre la variable “pertenencia al perfil de clientes de alta interacción con cuotas”, se identificó que la primera variable de corte es el haber realizado una transacción en el rubro de Educación, obteniendo un resultado similar a la regresión, con un efecto negativo.

Luego el árbol se dividió por el número de días que el cliente ingresa a la página de internet y la zona geográfica donde este vive. De esta manera se puede observar nuevamente la característica de “inactivos” con las página de internet que tienen los clientes más propensos, así como el no ser residente en la región centro del país, zona centro y oriente de la Región Metropolitana

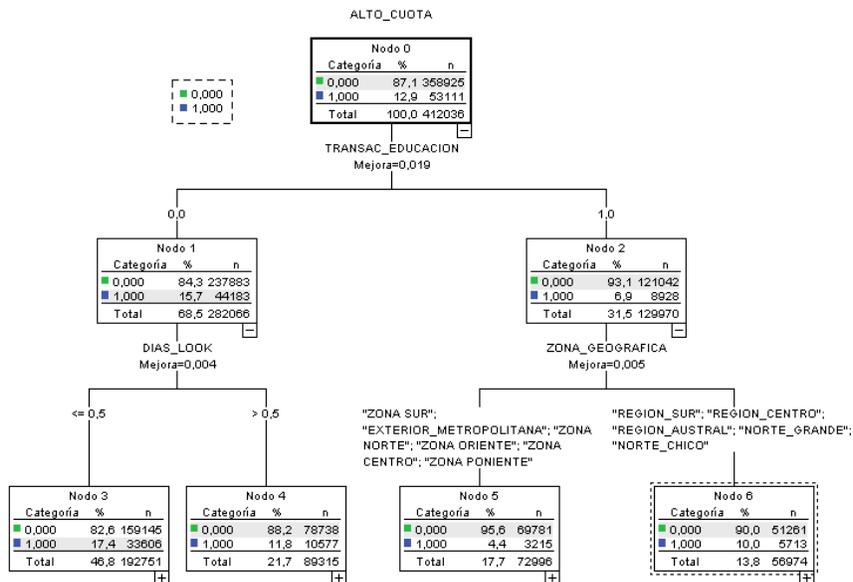


Ilustración 30: Ramas superiores árbol decisión - cuotas.

Posteriormente el árbol crece en sus ramas inferiores más propensas, incorporando información de edad, zona geográfica del cliente y cupo con las mismas tendencias observada en la regresión logística sobre el efecto positivo que tiene ser un cliente más longevo y el efecto negativo de tener un cupo más alto, llegando en este caso a una propensión de 22,9% (lift 1,8).

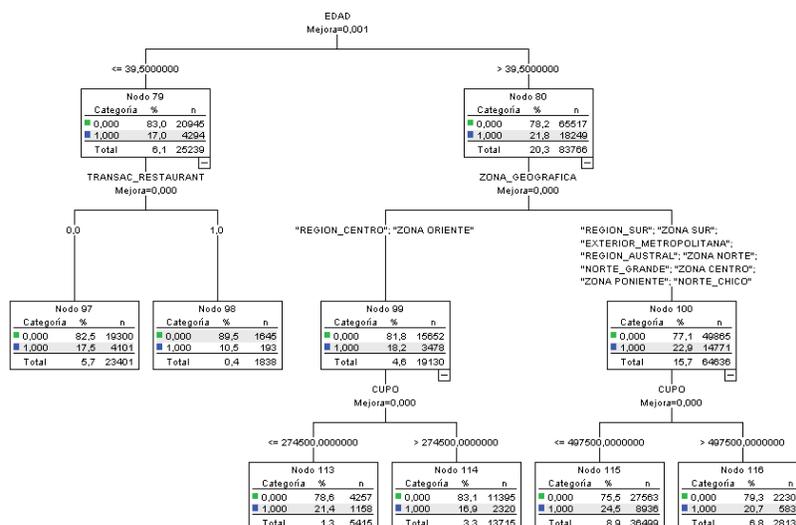


Ilustración 31: Ramas inferiores modelo árbol decisión - cuotas

Los resultados obtenidos son levemente superiores al de la regresión logística al captar un 24,1% de los clientes de alta interacción con las cuotas en el decil más propenso. Lo que es complementario a la curva ROC de área bajo la curva de 71,1%.³⁵

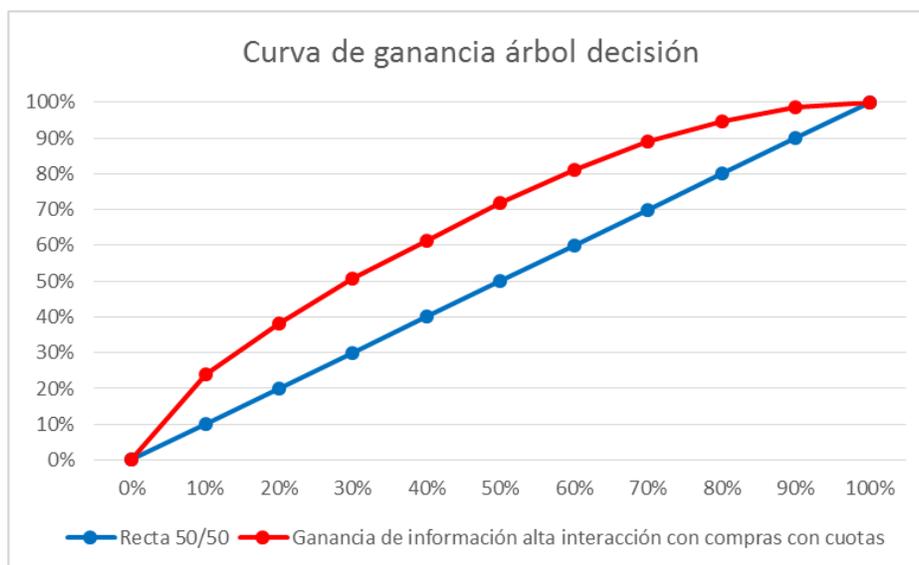


Ilustración 32: Curva de ganancia de información modelo árbol decisión (cuotas)

³⁵ Anexo E. Curva ROC modelo de cuotas.

12.3 Compras con promociones

12.3.1 Regresión logística

En el caso del modelo de alta interacción con las promociones, al tratarse únicamente de transacciones realizadas en la multitienda del holding, se utilizó información transaccional de los clientes en el resto de los comercios o rubros para identificar si el comportamiento fuera de este tenía efecto sobre la utilización de la tarjeta de crédito por los beneficios asociados a descuentos.

<i>Variable</i>	<i>Coficiente</i>	<i>Significancia</i>	<i>Exp (Coef)</i>
<i>Hacer transacción en Diciembre</i>	0,612	0,000	1,843
<i>Hacer transacción en noviembre</i>	0,269	0,000	1,309
<i>Hacer transacción en electro hogar</i>	0,687	0,000	1,987
<i>Edad</i>	-0,017	0,000	0,983
<i>Cliente Premium</i>	0,206	0,000	1,229
<i>Cliente Elite</i>	0,130	0,000	1,139
<i>Logaritmo del cupo</i>	0,138	0,000	1,147
<i>Transacciones fuera multitienda</i>	-0,051	0,000	0,951
<i>Transacciones en rubro multitiendas³⁶</i>	-0,013	0,536	0,987
<i>Constante</i>	-1,834	0,000	0,160

Tabla 15: Resultados modelo de regresión logística alta interacción promociones

Los principales resultados obtenidos por el modelo indican efectivamente que el ocupar la tarjeta fuera de este comercio afecta de manera negativa en el comportamiento “descuentero”, lo que puede reflejarse en que este tipo de clientes únicamente utiliza la tarjeta en este comercio por los descuentos.

Por otro lado existe un efecto positivo por realizar una compra en electro domésticos, hecho que para la empresa resulta acorde debido a que este tipo de productos se caracteriza por ofrecer descuentos adicionales al normal por el uso de la tarjeta.

Además cabe resaltar el hecho de ocupar la tarjeta de crédito por compras en Diciembre (cualquier rubro), reflejando que en el periodo de mayor necesidad de descuentos, el preferir ocupar la tarjeta como medio de pago es relevante.

Finalmente las únicas variables demográficas que el modelo incorporaba son la clasificación del cliente, donde un cliente premium es el más propenso, así como

³⁶ Si bien no es significativa, entrega cierta información del comportamiento.

la edad del cliente al clasificar a los clientes jóvenes como más propenso al comprar siempre con cuotas, lo que va en concordancia con el análisis exploratorio.

Al observar la curva de ganancia que entrega el modelo, se observa que su desempeño es similar al de las compras con cuotas, identificando al 19% de los clientes en el decil más propenso.

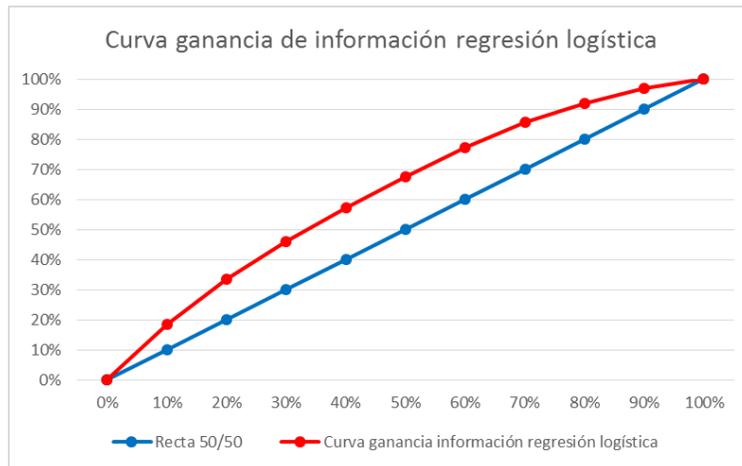


Ilustración 33: Ganancia de información alta interacción promociones

12.3.2 Árbol decisión

El modelo de árbol de decisión revela que la primera variable de corte es la compra en el mes de diciembre, seguido por la realización de alguna transacción en electrodomésticos. Lo que indica que a priori los clientes no son diferenciables por características demográficas, sino únicamente del uso de la tarjeta en ciertos lugares o momentos.

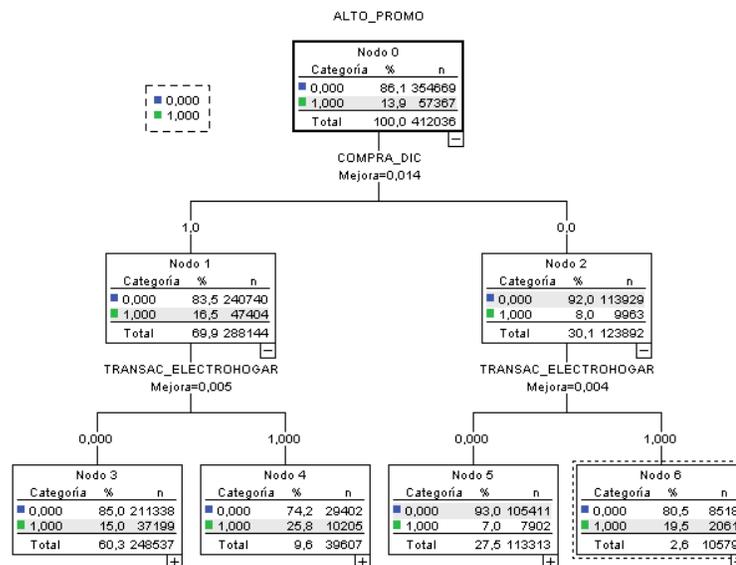


Ilustración 34: Ramas superiores árbol de validación - promociones

Sin embargo a medida que el árbol crece en sus ramas inferiores, se observa que las variables demográficas toman relevancia en la clasificación del cliente. De esta manera se observa que la edad indica que los clientes más jóvenes son más propensos a la reiterada compra con descuentos, así como el efecto positivo que tiene el ser un cliente de mayor categoría con el club de puntos y el tener un cupo superior a los 300.000 pesos.

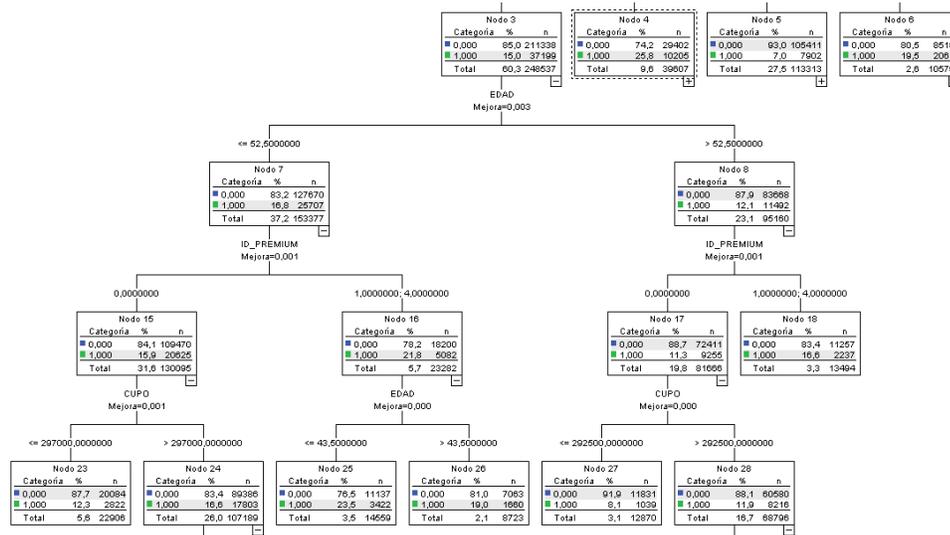


Ilustración 35: Ramas inferiores árbol de validación - promociones

De la misma manera que antes, al observar la curva de ganancia generada por el modelo, se presentan resultados similares a la regresión logística al identificar un 19% de los clientes de alta interacción con el mejor decil de clientes. Hecho complementario a un área bajo la curva del 65,7%³⁷

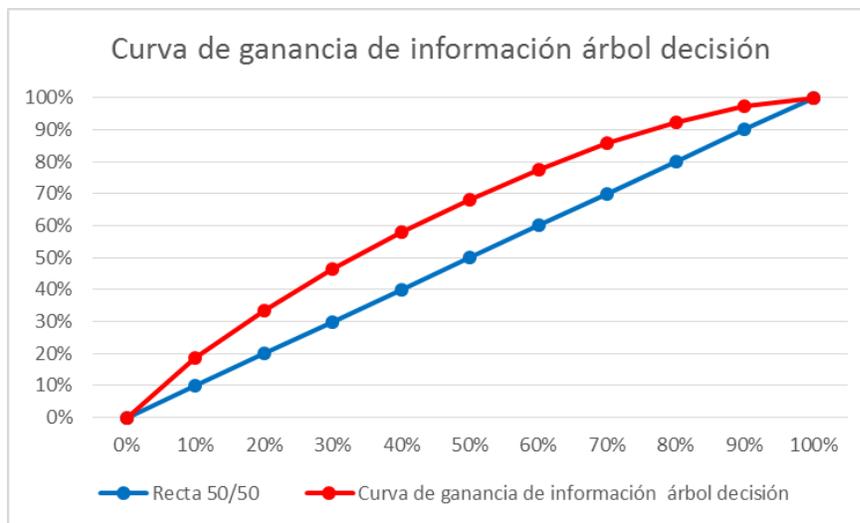


Ilustración 36: Curva de ganancia clientes alta interacción con promociones

³⁷ Anexo E. Curva ROC promociones.

13. Conclusión del perfil de clientes en cada driver

Luego de realizar los modelos de regresión lineal y árbol decisión para cada uno de los tres *drivers*, se pudo observar que el desempeño de ambos modelos es bastante similar en cuanto a las curvas de ganancias entregadas y áreas bajo la curva ROC.

De esta manera, ambos modelos entregaron resultados muy buenos para la interacción con el club de puntos, seguido del modelo de cuotas y con peor desempeño el de promociones. Este hecho muestra que si bien podrían haber características que describen al perfil de “descuenteros”, este perfil no está tan bien definido como el de los perfiles de puntos y cuotas.

Producto de la similitud del desempeño de los modelos, se decidió escoger el modelo árbol decisión para tomar medidas de acción, debido a su fácil comprensión por parte de la empresa y perfilamiento genérico que sobre los clientes. Finalmente los perfiles de clientes identificados para cada *driver* son los siguientes.

1. Club de puntos

- Clientes con cupo superior \$800.000 (valor alto)
- Tiene al menos un hijo y posee tarjeta visa
- Acumulación de puntos en multitiendas y supermercados fuera del holding que son competencia directa de los principales comercios del holding
- Ingresa a la página web
- Tiene cuenta adicional

Perfil “puntos adictos”

“Un cliente del perfil de alta interacción con los puntos se caracteriza por tener **poder económico (cupo alto)**, **tiene hijos** en su cuidado, **ingresar** durante distintos días a la página web y **acumular puntos en la competencia directa** del holding”

Al estudiar las características de los clientes más propensos con variables que no se incorporaron al modelo, se puede apreciar que estas son congruentes con el tipo de perfil

- Ticket promedio bastante alto (52.220 pesos) y una frecuencia de compra de los 12 meses
- Promedio de edad de 47 años
- Promedio de puntos acumulados es 25.522 puntos, lo cual es alto al alcanzar hasta el 5to nivel de canje.

2. Compras con cuotas

- No ingresa a la página web
- Cupo inferior a \$500.000
- Edad superior 35 años
- No viven en la región centro, zona centro RM y zona oriente RM
- No utilizan las tarjetas en rubros de viaje y educación

Perfil “cuoteros”

“Un cliente de alta interacción de cuotas se caracteriza por tener **bajo poder económico**, al tener un cupo bajo y no vivir en las zonas de mayor poder económico como oriente y centro RM, ser adulto mayor de **más de 35 años**, prácticamente **no usan la página web** y no ocupan la tarjeta en rubros de pagos **habituales** ni de **alto valor**”.

Al caracterizar este grupo con variables que no se incluyeron en el modelo se observó nuevamente la coherencia con el perfil y la diferencia que tiene con el perfil de puntos.

- Sólo un 1,5% es ABC1, mientras que un 7,2% es C2 y 14,3% es C3
- El 60% son clientes femeninos
- El promedio de acumulación de puntos es de 1.977 puntos, lo cual es bajo para poder interactuar con el club de puntos

3. Compras con promociones

- Compran en diciembre y áreas de electro domésticos
- Edad inferior a 35 años
- Cupo superior a \$600.000 (medio alto)
- Categoría premium o elite

Perfil “descuenteros”

“Un cliente de alta interacción de promociones se caracteriza por preferir la tarjeta en **productos y fechas** que hay mayores **descuentos**, ser de edad **joven** y tener un **poder económico medio alto** reflejado en el cupo y categoría”

Al caracterizar este grupo con variables que no se incluyeron en el modelo se observó:

- El 44% realizó una compra con descuento superior al normal (por usar la tarjeta)

- Hay una media de 4.400 puntos acumulados sin considerar el 5% de los casos extremos y 6.403 considerando todos los casos, dando a entender que un porcentaje no menor puede interactuar con el club de puntos
- La mayoría pertenece a la región centro y sur del país

14. Extensión interacción con líneas de premios

Una vez definido el principal perfil de los clientes que son más propensos a tener una alta interacción con cada uno de los *drivers* que ofrece la tarjeta de crédito, se decidió extender el modelo de alta interacción con el club de puntos hacia modelos de alta interacción con una línea de premios en particular.

La razón de esta extensión se debe a que como se mencionó anteriormente, el club de puntos de la tarjeta se caracteriza por ofrecer variados tipos de premios asociados a los distintos comercios del holding, por lo que agrupar a todos los cliente en un mismo “saco” de gestión puede ser menos efectivo.

Más aún, el poder identificar qué tipo de premio es más atractivo para un cliente y como afecta un determinado tipo de premio en el uso y gasto de los clientes en los comercios es muy relevante.

De esta forma se procedió a jerarquizar la base de premio del club de puntos en distintas líneas asociadas a comercios del *Holding* o premios que apunten a necesidades o intenciones similares, formando las siguientes líneas de premios.

Línea de premio	Canjes	% Canjes	Cientes que canjean	%Clientes	Puntos canjeados	% Puntos
Decoración	12.332	0,92%	11.285	1,44%	72.062.000	0,59%
Dormitorio	95.192	7,11%	81.122	10,33%	574.898.000	4,68%
Electrodoméstico alta inversión	4.353	0,33%	3.934	0,50%	242.429.000	1,97%
Electrodomésticos cocina	346.216	25,85%	302.986	38,58%	2.526.085.000	20,56%
Electrodomésticos hogar	145.403	10,86%	134.515	17,13%	1.073.961.000	8,74%
Experiencias	4.349	0,32%	2.935	0,37%	27.575.000	0,22%
Ferretería	64.772	4,84%	59.454	7,57%	610.639.000	4,97%
Gift card	24.223	1,81%	17.338	2,21%	220.896.000	1,80%
Gift card alta inversión	8.105	0,61%	6.072	0,77%	869.100.000	7,07%
Menaje	238.710	17,82%	212.153	27,02%	1.614.566.000	13,14%
Sabores	10.662	0,80%	7.995	1,02%	60.241.000	0,49%
Seguros (SOAP)	52.332	3,91%	44.147	5,62%	271.059.000	2,21%
Tecnología	96.896	7,23%	86.167	10,97%	1.426.345.000	11,61%
Viajes	26.962	2,01%	25.852	3,29%	1.007.175.000	8,20%
Otros	208.839	15,59%	188.259	23,97%	1.687.972.000	13,74%
Total	1.339.346	100,00%	1.184.214	150,80%	12.285.003.000	100,00%

Tabla 16: Descripción de líneas de premios en torno a cliente que canjean y puntos involucrados

Como se puede observar, existe una gran cantidad de líneas de premios, por lo que para la extensión del modelo únicamente se estudiarán las principales líneas que la empresa considera relevante por el número de canjes, puntos canjeados y adhesión a los principales comercios del holding. De acuerdo a este criterio las líneas a estudiar son:

- Electrodomésticos cocina

Caracterizado por maquinas eléctricas usada en la cocina como hervidores, cafeteras, tostadores, batidoras, cuchillos eléctricos y destapadores eléctricos entre otros productos, mostrando la gran diversidad que esta línea agrupa.

La elección de esta línea se debe a que presenta el mayor número de canjes y presenta similitud con las líneas de electro domésticos hogar y menaje. Además está relacionado al comercio de multitienda.

- Ferretería

Premios enfocados en herramientas como atornilladores, serruchos, guantes, huinchas, etc.

Si bien esta línea no representa gran cantidad de los canjes ni puntos invertidos, este está altamente vinculado al comercio de herramientas y construcción del holding al ofrecer productos que se venden en este comercio.

- Seguro

Enfocado en el canje de seguros obligatorios de autos (SOAP), camionetas y motos. Premios que se ofrecen únicamente en Marzo.

La elección de esta línea se debe a que únicamente se ofrece un producto, que a comparar con los premios en forma desagregadas, es el segundo premio más canjeado. Además está vinculado al comercio de seguro del holding.

- Viaje

Canje de viajes dentro y fuera del país, así como estadías en hoteles en regiones.

La elección de esta línea de premios se debe a la gran participación que representa en términos de puntos canjeados, además de ofrecer viajes asociados a uno de los comercios del holding.

- Gift card

Canjes en gift card de distintos valores que el cliente puede utilizar dentro de los tres principales comercios del holding (multitienda, construcción y supermercado). En particular se estudiará las gift card asociadas a montos inferiores a los \$60.000.

La elección de esta línea se debe a que esta ofrece premios que se pueden gastar en los distintos comercios internos.

Una vez seleccionados las principales líneas de premios a extender, es importante señalar que un cliente pudo haber canjeado los premios en más de una línea (35% de los clientes), por lo que no es posible la realización de un único modelo multinomial. Debido a esto se realizaron cinco modelos distintos con las mismas variables utilizadas en el modelo general de interacción con el club de puntos.

Para la formación de la variable independiente en cada uno de los cinco modelos, se consideró un caso positivo si el cliente pertenece al perfil de alta interacción con el club de puntos y además canjeó sus puntos en la línea de premios de estudio, mientras no cumpla estas dos condiciones el cliente será clasificado como un caso negativo.

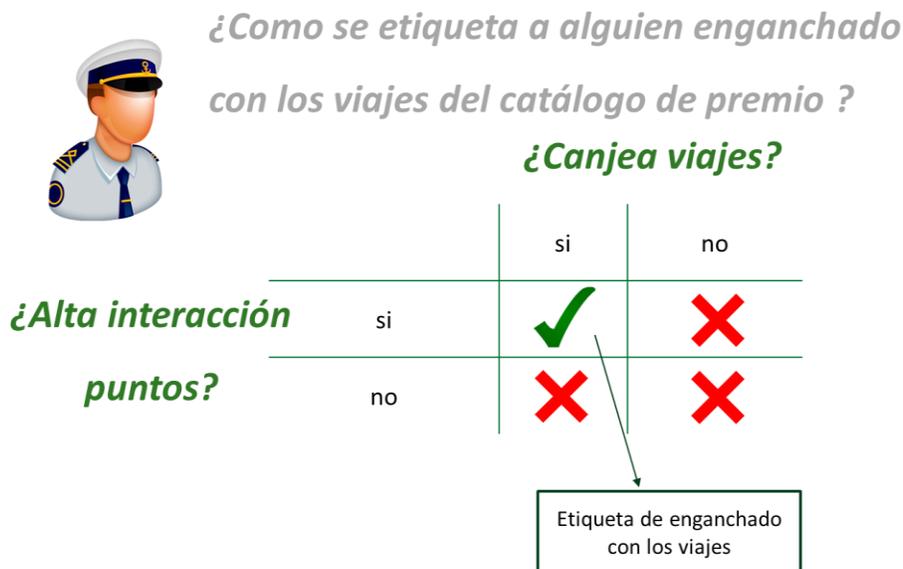


Ilustración 37: Ejemplo de criterios para pertenecer al perfil de alta interacción con una determinada línea de premio (viajes)

15. Modelos líneas de premios

Al igual que los modelos anteriores, se realizó un modelo de regresión lineal y árbol de decisión para comparar los principales resultados obtenidos.

15.1 Seguros (SOAP)

Regresión logística

Los principales resultados de la regresión logística indican que la variable cupo y días que el cliente realiza “log-in” siguen siendo de gran relevancia para que el cliente pueda interactuar con este premio, lo que puede ser una condición base del poder económico del cliente y la necesaria información con el catálogo debido a la estacionalidad del premio al ofrecerse únicamente en Marzo.

Por otro lado existe un efecto positivo en la edad, lo que se puede deber a que los temas de seguridad pueden ser más atractivos a personas más adultas que tengan autos, así como el efecto positivo que tiene el residir en la zona periférica de la región metropolitana, que puede hacer hincapié a clientes de menores recursos que aprovechan este canje para suplir los gastos obligatorios.

Finalmente un hecho relevante es la importancia que tiene el acumular los puntos en comercios relacionados al premio de canje (Seguro obligatorio de autos), como lo es el rubro de combustibles y autos, existiendo una cierta concordancia entre el premio canjeado y en que se ocupa la tarjeta.

<i>Variable</i>	<i>coeficiente</i>	<i>Significancia</i>	<i>Expr(coef)</i>
<i>Acumula en combustible</i>	0,899	0,000	2,456
<i>Acumular en autos y accesorios</i>	0,429	0,000	1,535
<i>Acumular en servicios y educación</i>	0,599	0,000	1,820
<i>Número días de looks</i>	0,031	0,000	1,031
<i>Logaritmo del cupo</i>	1,186	0,000	3,275
<i>Edad</i>	0,020	0,000	0,981
<i>Zona norte grande</i>	-0,663	0,000	0,515
<i>Zona norte chico</i>	-0,415	0,000	0,661
<i>Zona centro</i>	-0,172	0,000	0,842
<i>Zona sur</i>	-0,305	0,000	0,737
<i>Zona austral</i>	-0,358	0,119	0,699
<i>Zona norte RM</i>	-0,099	0,000	0,906
<i>Zona centro RM</i>	-0,318	0,000	0,728
<i>Zona oriente RM</i>	-0,458	0,000	0,633
<i>Constante</i>	-16,437	0,000	0,000

Tabla 17: Coeficientes regresión logística - Seguro (SOAP)

Árbol decisión

De la misma manera que antes, se generó un modelo de árbol de decisión el cual, a diferencia del modelo general, mostró una mayor importancia al ingreso a la página web al ser la primera variable discriminadora. Efecto que nuevamente se pueda deber por la estacionalidad del premio.

Luego si el cliente es alguien que se informó con la página, el árbol discrimina según el cupo del cliente que entrega cierta información de la capacidad de llegar a niveles de canje al cortar en \$635.000. En caso de que el cliente no ingrese a la página web, es relevante que este haya acumulado puntos en el rubro de combustible, el cual está altamente relacionado con el tipo de premio.

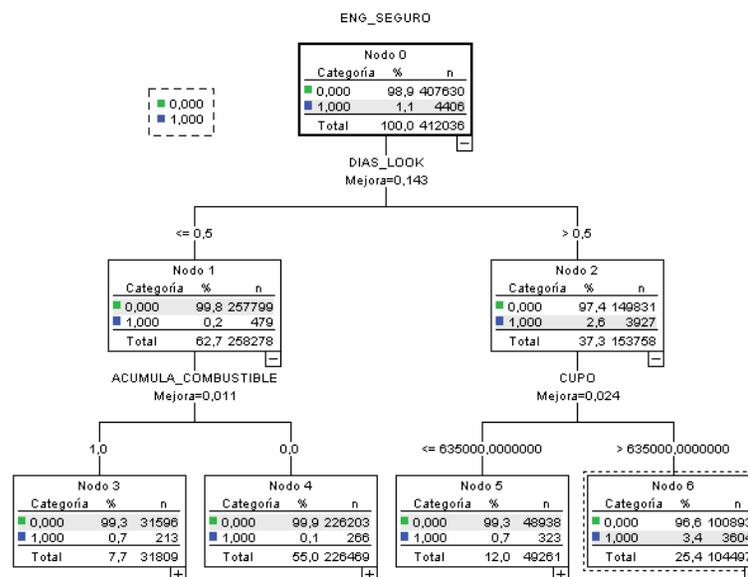


Ilustración 38: Ramas superiores árbol decisión - Seguro (SOAP)

Posteriormente en las ramas inferiores del árbol, se observa el rol que toman las variables que el modelo general tampoco tomaba en cuenta, como el no ser un cliente de la Zona Oriente o Norte Grande y el tener una edad superior a 30 años, lo que concuerda con la regresión logística.

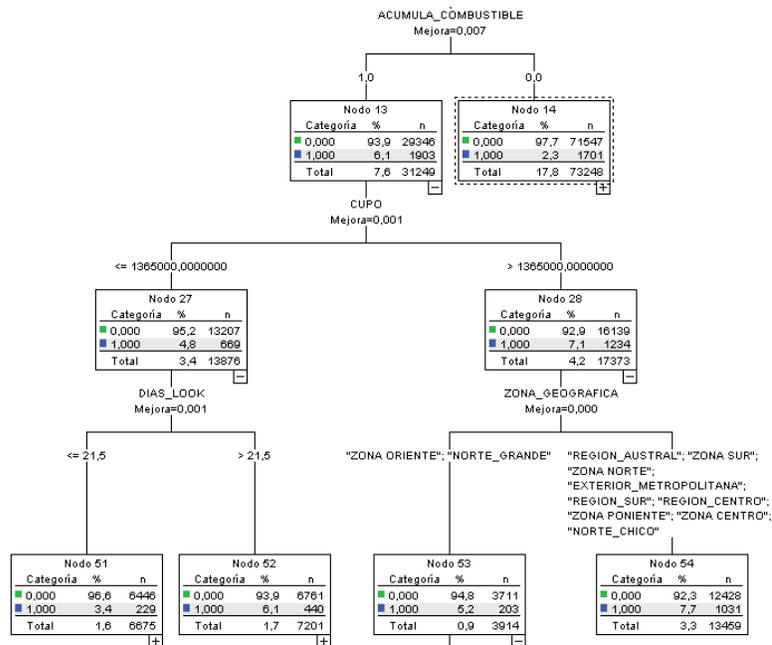


Ilustración 39: Ramas inferiores árbol decisión - Seguro (SOAP)

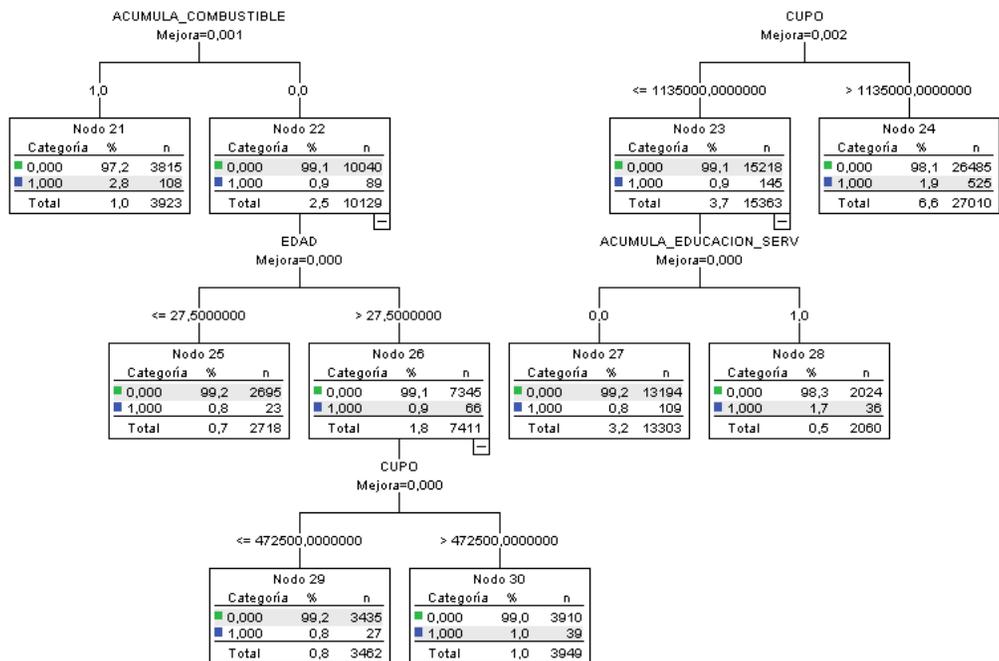


Ilustración 40: Ramas inferiores árbol decisión - Seguro (SOAP)

15.2 Ferretería

Regresión logística

La diferenciación que logra con el modelo global apunta en dos direcciones, por un lado resalta la importancia de ser un cliente de género masculino, mientras que por otro lado resalta nuevamente el hecho de haber acumulado los puntos canjeados (esfuerzo del cliente) en el comercio interno de construcción del holding o en los rubros de combustibles y autos, los que están altamente relacionados con premios de ferretería.

Es por esto que se destaca nuevamente el efecto “circulo virtuoso” que están teniendo los clientes al premiarse a sí mismos con premios relacionados a su “esfuerzo”.

Además a pesar de considerar las mismas variables base (cupo y días que el cliente realiza log – in), se puede diferenciar con que el efecto del cupo es inferior al modelo original, indicando que en este caso no es necesario el tener un cupo muy alto, sino que estos premios también apunta a segmentos de menor poder.

<i>Variable</i>	<i>coeficiente</i>	<i>Significancia</i>	<i>Expr(coef)</i>
<i>Logaritmo del cupo</i>	0,741	0,000	2,098
<i>Tenencia tarjeta visa</i>	0,647	0,000	1,909
<i>Acumular en comercio herramientas y construcción</i>	1,540	0,000	4,665
<i>Acumular en rubro de combustibles</i>	0,696	0,000	2,006
<i>Acumular en rubro de autos y accesorios</i>	0,584	0,000	1,794
<i>Número de días de looks</i>	0,017	0,000	1,017
<i>Género masculino</i>	0,715	0,000	2,044
<i>Constante</i>	-13,061	0,000	,000

Tabla 18: Coeficientes regresión logística - Ferretería

Árbol decisión

En el caso del árbol de decisión, las primeras variables discriminatorias son el cupo del cliente, dividiendo la base en \$600.000, señalando como más propensos a aquellos con mayor cupo. Luego en ambos casos actúa el haber acumulado en un rubro o comercio relacionado al premio (Comercio herramientas y construcción y rubro combustible), donde particularmente se ve que para ambas ramas, el no haberlo hecho disminuye considerablemente la propensión a canjear este tipo de premio.

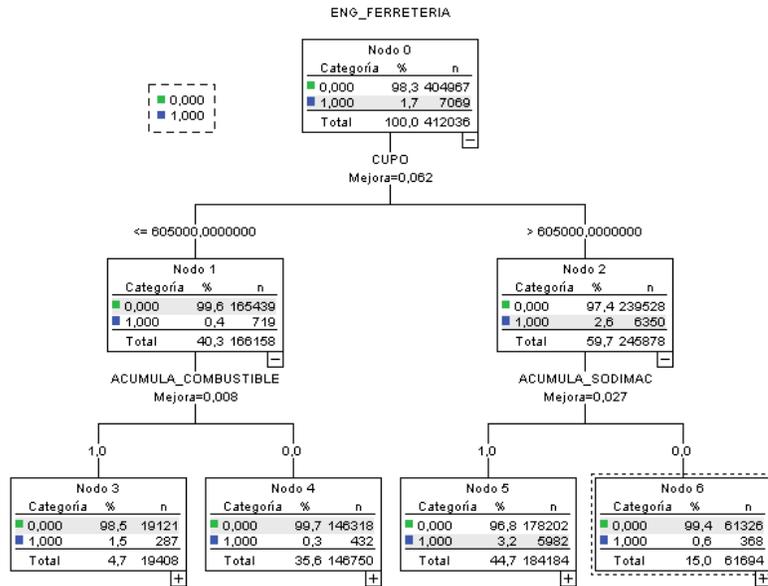


Ilustración 41: Ramas inferiores árbol decisión – Ferreteria

Posteriormente en los niveles inferiores más propensos, el árbol discrimina por el haber acumulado los puntos nuevamente en rubros semejantes al tipo de premio (Combustible y Auto y accesorios) junto a ser de género masculino, mostrando el efecto positivo que tienen estas características al llegar en el nodo más propenso a una probabilidad de 9,5% (lift 5,58).

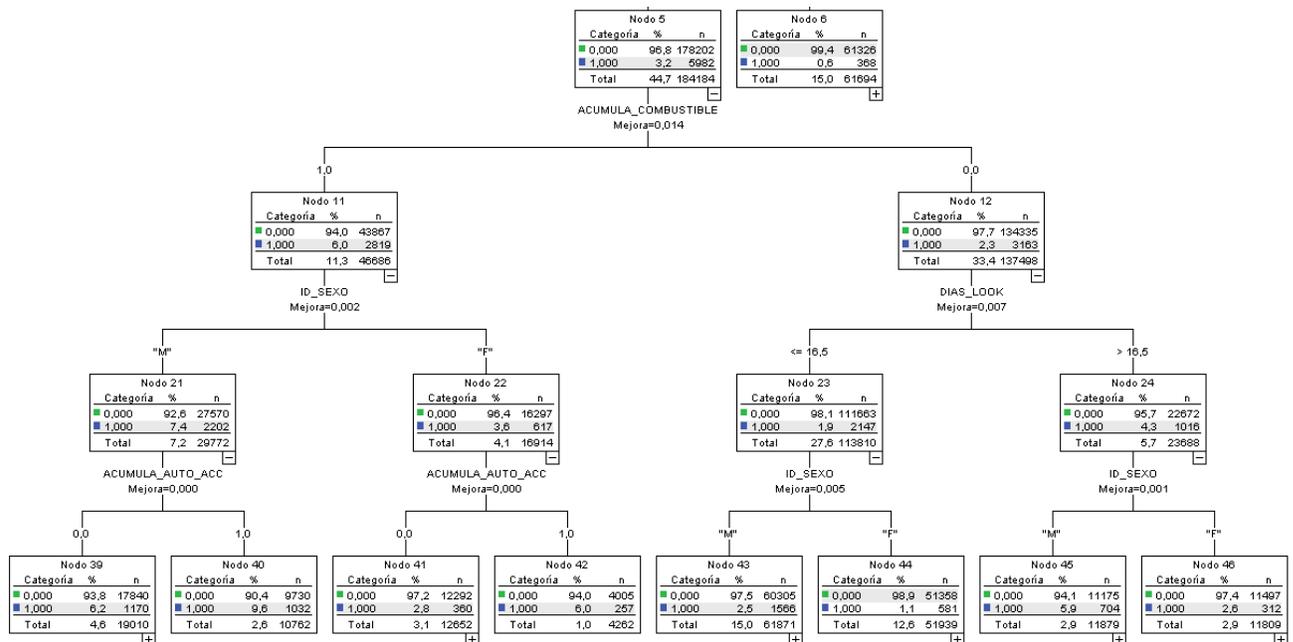


Ilustración 42: Ramas inferiores árbol decisión – Ferreteria

15.3 Viajes

Regresión logística

En el caso de viajes, la regresión logística muestra la gran importancia que genera el tener un cupo más alto al obtener un coeficiente mucho mayor al de los otros modelos. Este hecho resalta la gran importancia que tiene el poder económico en el canje de estos premios.

En esta misma línea el modelo además incorpora variables asociadas a la zona geográfica al haber un efecto negativo en residir en el Norte del país o la zona poniente de la región metropolitana, así como un efecto positivo en variables que pueden indicar un mayor gasto como la tenencia de tarjeta visa e hijos.

Por otro lado también se incorpora los efectos negativos de ser de género masculino y tener mayor antigüedad, lo que se puede justificar con que algunos premios de viajes incluyen estancias en hotel y spa, que pueden atraer más a este tipo de clientes.

Finalmente se observa el importante efecto positivo que tiene el haber acumulado los puntos en productos o servicios relacionados al premio, como en este caso en el rubro de “viajes”.

<i>Variable</i>	<i>coeficiente</i>	<i>Significancia</i>	<i>Expr(coef)</i>
<i>Acumular en rubro viajes</i>	0,948	0,000	2,579
<i>Logaritmo del cupo</i>	2,048	0,000	7,753
<i>Número de días de looks</i>	0,013	0,000	1,013
<i>Tenencia de hijo</i>	0,558	0,000	1,747
<i>Antigüedad</i>	-0,010	0,000	0,990
<i>Tenencia de tarjeta visa</i>	0,435	0,000	1,545
<i>Tenencia de cliente adicional</i>	0,209	0,000	1,233
<i>Zona norte Grande</i>	-0,372	0,000	0,689
<i>Zona norte chico</i>	-0,722	0,000	0,486
<i>Zona centro³⁸</i>	0,125	0,015	1,133
<i>Zona centro RM</i>	0,166	0,115	1,180
<i>Zona poniente RM</i>	-0,322	0,000	0,724
<i>Género masculino</i>	-0,530	0,000	0,588
<i>Constante</i>	-29,682	0,000	0,000

Tabla 19: coeficientes regresión logística viajes

Los resultados obtenidos son coherentes con la creencia de la empresa, debido a que estos premios son clasificados en mayor medida como “caros”, por lo

³⁸ No se incluyeron las zonas que no son significativas al 95%

que son utilizados por clientes que se caractericen con gastar más y tener un mayor poder económico.

Árbol decisión

Al observar el comportamiento del árbol de decisión, se evidencia el efecto que tiene el poder económico del cliente al dividir la base en los dos primeros niveles únicamente con la variable de cupo. En estas divisiones, los clientes más propensos son los pertenecientes a los grupos de mayor cupo, revelando la gran importancia de esta variable.

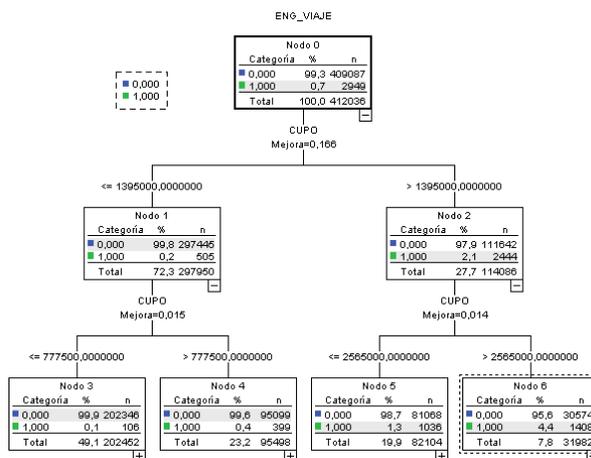


Ilustración 43: Ramas inferiores modelo árbol de decisión - viajes

Posteriormente para cada uno de los cuatro grupos actúa la variable de acumulación de puntos en el rubro de viajes, percibiendo nuevamente el efecto “circulo virtuoso” como en los otros modelos de líneas de premios.

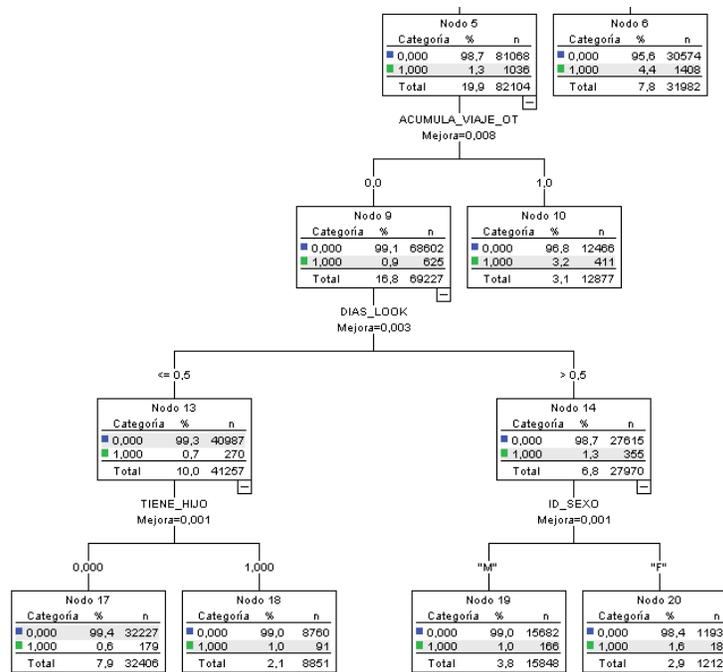


Ilustración 44: Ramas inferiores modelo árbol de decisión - viajes

Finalmente, en las ramas inferiores más propensas se puede observar en menor medida el efecto positivo que tiene el ser de género femenino, interactuar con la página de internet y el pertenecer a la zona oriente de la RM.

15.4 Electro doméstico cocina

Regresión logística

Los resultados obtenidos y las variables utilizadas por el modelo, señalan que este se comporta de manera similar al modelo general de interacción con el club, con la diferencia que estos clientes se caracterizan por haber acumulado en el rubro de farmacias, el cual está asociado a gastos de necesidad.

Por otro lado el modelo incorpora el efecto positivo de la antigüedad, lo que significa que los clientes más antiguos son más propensos a este tipo de canje, hecho que se puede deber a que ya establecidos en el tiempo es más fácil canjear.

Finalmente el modelo no incorpora la información del cupo y un efecto muy bajo del ingreso a la página, por lo que teniendo en cuenta la cantidad de canjes que tiene esta línea y la variedad de premios, no necesariamente son clientes poderosos o que constantemente estén informándose, sino que puede apuntar a clientes que únicamente canjean los premios más “fáciles” o comunes.

Variable	coeficiente	Significancia	Exp (coef)
Acumular en rubro de farmacias	0,801	0,000	2,227
Acumular en comercio interno de construcción	0,776	0,000	2,173

<i>Acumular en comercio interno de multitienda</i>	1,466	0,000	4,333
<i>Tenencia de tarjeta visa</i>	0,932	0,000	2,539
<i>Tenencia de cliente adicional</i>	0,403	0,000	1,497
<i>Número de días que realiza un look</i>	0,011	0,000	1,011
<i>Antigüedad</i>	0,034	0,000	1,035
<i>Constante</i>	-3,680	0,000	,025

Tabla 20: Coeficientes regresión logística – Electro cocina

Árbol decisión

Al realizar el modelo de árbol de decisión, se puede observar la importancia que tiene la variable de acumular puntos en un rubro de farmacias, al ser la primera variable discriminadora del modelo. Posteriormente en caso no acumular en farmacias es importante el tener que acumular en la multitienda del holding que vende los productos ligados esta línea premio. Hecho que muestra que si un cliente no acumula en ninguno de estos dos, prácticamente es imposible que canjee.

Por otro lado si el cliente acumula en farmacias, actúa a diferencia del modelo global, la variable de tener más de cinco años de antigüedad.³⁹ Este efecto da a entender que una característica de involucramiento con la tarjeta al llevar tiempo con ella y ocuparla en un rubro asociado a la necesidad, afecta de manera positiva.

Luego en ramas inferiores el modelo se comporta manera similar al modelo global, incorporando los mismos efectos de tenencia de una tarjeta visa, un adicional y el haber acumulado en los dos principales comercios del holding.

En consecuencia, este modelo es bastante similar al modelo global de puntos, lo que se puede deber a la gran cantidad de canjes que tiene esta línea y a su amplia oferta de distintos premios que pueden atraer a variados tipos de clientes que en resumen buscan no perder sus puntos con premios “estándares” o simples.

15.5 Gift card

Regresión logística

En el caso de canje de gift card, también se observa un comportamiento similar al modelo global, con la inclusión de la edad del cliente que actúa de manera negativa, indicando que los jóvenes son más propensos a este canje. Otro efecto es la importancia de haber acumulado puntos en el comercio de supermercado del holding que también entrega estos premios.

<i>Variable</i>	<i>coeficiente</i>	<i>Significancia</i>	<i>Exp(coef)</i>
<i>Edad</i>	-0,016	0,000	0,985
<i>Acumular en rubro farmacias</i>	0,812	0,000	2,253

³⁹ Anexo E. Árbol electrodomésticos.

<i>Acumular en rubro multitienda</i>	0,709	0,000	2,031
<i>Acumular en comercio interno multitienda</i>	1,484	0,000	4,412
<i>Acumular en comercio interno supermercado</i>	0,460	0,000	1,584
<i>Número días de looks</i>	0,020	0,000	1,020
<i>Cupo medio (400.000 – 899.999)</i>	1,038	0,000	2,823
<i>Cupo alto (900.000 – 1.499.999)</i>	1,540	0,000	4,663
<i>Cupo muy alto (>= 1.500.000)</i>	2,113	0,000	8,272
<i>Constante</i>	-3,297	0,000	0,037

Tabla 21: Coeficientes regresión logística - gift card

Árbol decisión

Al realizar el modelo árbol de decisión se observó que el comportamiento era bastante similar a modelo global, teniendo en este caso como variable predominante el haber acumulado en rubros de multitiendas y el tener una actividad con la pagina superior a los 10 días, dejando de lado la importancia del cupo que tiene el modelo global.⁴⁰

Debido a este hecho nuevamente ocurre un efecto similar al de premios de electrodomésticos, que producto de la amplia oferta, este no atrae a un cliente en particular, sino que aquel que principalmente ha ocupado la tarjeta en rubros muy asociados al tipo de premio, como lo son el rubro de las multitiendas y los comercios internos del holding, y que muestra comportamiento de ingreso a la página.

Finalmente en niveles inferiores, si bien es menor el efecto, se observa que un cliente menor a 27 años es más propenso al canje de estos premios, lo que según la empresa se debe a que los clientes más jóvenes usualmente ocupan la gift cards, ya que es un producto moderno.

16. Desempeño Modelos Canjes de Líneas

Nuevamente el modo de evaluar el desempeño de los cinco modelos realizados es la curva de ganancia, que indica cuantos clientes se logran capturar en los niveles más propensos otorgados por el modelo y la curva ROC.

Los principales resultados obtenidos muestran que los cinco modelos realizados tienen muy buenos resultados al captar cada uno de ellos al menos el 30% de los clientes de alta interacción en el decil más propenso. Este hecho se puede deber en gran medida al buen resultado que tuvo el modelo global, permitiendo tener ciertas características de base que ayudan a tener un modelo con resultados similares al modelo global.

^{20 y 40}: Anexo E. árbol modelos de gift card y electro domésticos cocina.

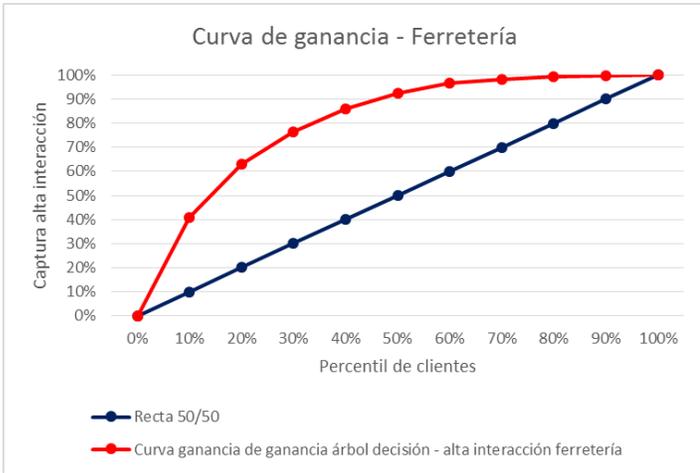


Ilustración 46: Curva ganancia - ferretería

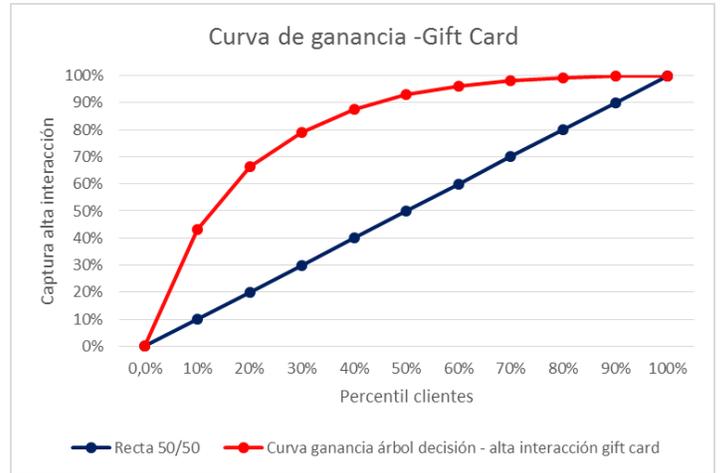


Ilustración 45: Curva ganancia - gift card

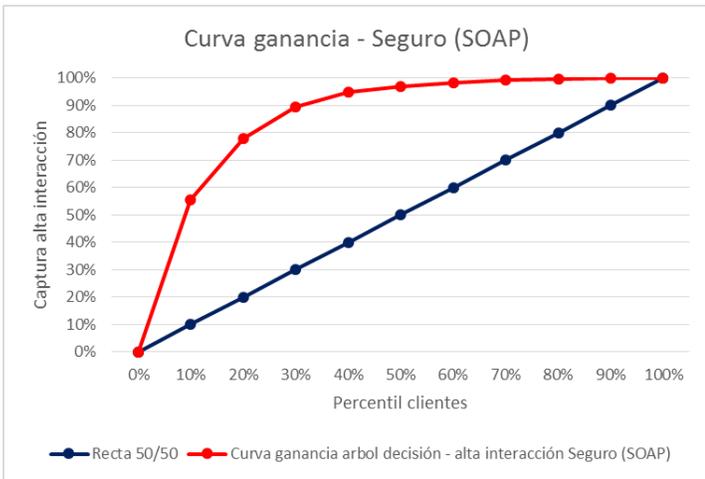


Ilustración 47: Curva de ganancia – seguro (SOAP)

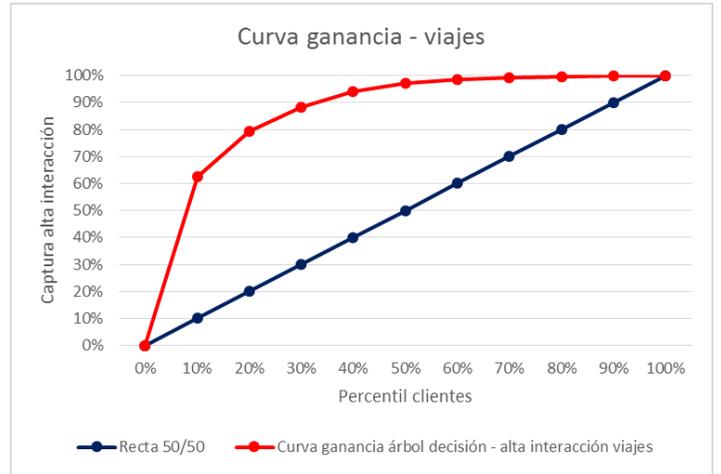


Ilustración 48: Curva de ganancia - viajes

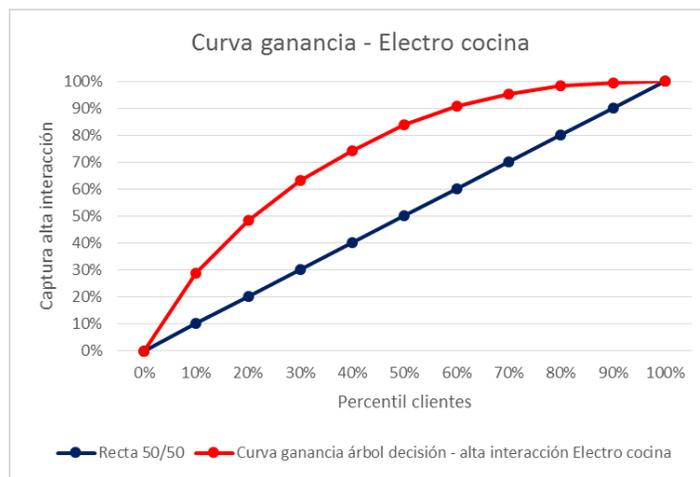


Ilustración 49: Curva ganancia - electro doméstico cocina

Sin embargo, como se puede observar, el hecho más destacable es el peor desempeño que tiene el modelo de “electrodomésticos de cocina” en comparación a los otros modelos de premios estudiados. Las principales razones que pueden explicar este efecto de poca diferenciación en esta línea es la cantidad de canjes que existen (20,5% de los canjes) y la variedad de premios que involucra, que provoca que este sea un destino de un canje simple o “facil” para no perder los puntos.

Por otro lado, la mayor diferenciación de los perfiles de viajes, seguro y ferretería antes obtenidos, provocaron que estos modelos tuvieran los mejores desempeños en las curvas de ganancias y curvas ROC con áreas superiores al 80%.⁴¹

17. Perfilamiento de Líneas de Premios

Uno de los problemas que tiene el modelo global de puntos es que la atracción por el club, incluía en un mismo grupo una gran gama de premios distintos que puede incentivar a clientes con características distintas.

Debido a esto, el modelo únicamente se enfocaba en información de uso de la página, poder económico del cliente por poder cajar y preferencia sobre la competencia directa, dejando de lado características demográficas del cliente o el uso de la tarjeta en productos o servicios semejantes al premio.

Es por esto que a través de los cinco modelos antes realizados se obtuvieron los siguientes perfiles para las cinco líneas de premios estudiadas, los cuales ahondan aún más en características específicas de clientes que no sólo le “atraigan” los puntos, sino una línea específica de premios.

- Seguro (SOAP)

“Se caracteriza por clientes que **necesariamente ingresan a la página web**, tienen un **cupo** cercano a los **\$1,2 M**, **mayores a 30 años** y que **acumulan en combustibles**”

Al caracterizar a los clientes más propensos con variables no ocupadas en el modelo, estos en promedio acumulan 11.653 puntos alcanzando únicamente a canjear en los dos primeros modelos y un 61,8% pertenece al segmento C2 o C3, lo que concuerda con un segmento de menor nivel económico.

Debido a estas características, se perfiló a estos clientes como “Adultos Tecnológicos”.

⁴¹ Anexo E. Curva ROC modelos de líneas de premios.

- Viajes

“Se caracteriza por ser con un **cupo superior a \$2 M**, ingresan cerca de **15 días** a la **página web**, **acumula** sus puntos en el rubro de **viaje**, son de género **femenino** y con **más de 10 años** con la tarjeta”.

Al observar las características de los clientes más propensos se observa un 64% de estos pertenecen al segmento ABC1, acumulan en promedio cerca de 18.000 punto, su ticket promedio es de \$78.000 y un 34% pertenece a la zona oriente RM. Resaltando nuevamente la característica de tener poder económico.

Se perfiló a estos clientes como “Mujeres con poder económico”

- Ferretería

“Se caracteriza por tener **cupo superior a \$400.000** (no necesariamente alto), género **masculino** y **acumular** puntos en el rubro de **combustible** y en el comercio de **herramientas y construcción**”

Al caracterizar a los clientes más propenso se observa que un 72% pertenece a los segmentos C3 y D, acumulan en promedio 5.142 puntos alcanzando a canjear en el nivel más bajo y tienen una antigüedad de sólo cinco años, lo que da referencia a un segmento de menor nivel.

Perfilados como “Maestros de construcción”

- Electrodomésticos cocina

“Se caracteriza por **acumular** en el comercio de **multitienda** y rubro de **farmacias**, tener una **antigüedad superior a los cinco años** y la tenencia de una tarjeta **visa**”

Al caracterizar a los clientes más propenso con otras variables, estos en promedio canjean 6.838 por lo que lo hacen en el nivel más bajo, el 75% canjeo previamente electrodomésticos y el 64% pertenece al segmento C2 y C3

Cómo se evidenció anteriormente, este perfil de clientes es bastante similar al perfil global, hecho que se debe a la similitud que tiene con los premios de menaje y tecnología que en conjunto con estos capturan más del 60% de los canjes, generando las principales características del modelo global.

Debido a esto, resalta nuevamente la consigna de que estos clientes no tienen un perfil porque apenas llegan a niveles de canje, premiándose con los premios más comunes y de mayor oferta en el nivel más bajo, lo que repiten durante cada año.

Producto de la poca diferenciación se perfilaron a estos clientes como “canjeadores simples”.

- Gift card

“Se caracteriza por **acumular** en el comercio de **multitienda** y **supermercado**, tener **menos de 30 años**, **ingresar** a la página **web** y un **cupo superior a \$600.000.**”

En este caso si bien la diferenciación del modelo global no es tan evidente, esta reside en ser un cliente más joven y el acumular en los comercios internos que están altamente asociados a las gift card.

Finalmente el decil más propenso se caracteriza por tener 10 años de antigüedad, al momento de canjear tienen acumulado 7.999 puntos, el 85% hizo un canje en el periodo anterior y mayoritariamente son del sector C3 y D (63 %).

Se perfiló a estos clientes como “jóvenes involucrados”

Discusión de perfiles

Los principales hallazgos obtenidos de las líneas de premios, es que el desempeño de los modelos mejora mucho en ciertos casos que el perfil es evidentemente diferenciable del resto, mientras que los menos diferenciables tienen un desempeño similar al modelo global.

Ejemplo de esto es ferretería y electrodoméstico cocina donde el primero se diferencia al apuntar al sector de hombres con menos recursos, se perfilaron como “maestros de construcción”, mientras que el segundo al no poder lograr una diferenciación se decidió perfilar como “canjeadores simples”.

Sin embargo el efecto más importante el modelo logró determinar es el efecto “circulo virtuoso” que ocurre en cada uno de los modelos. Este efecto muestra para los cinco casos, que el esforzarse por acumular puntos en un determinado rubro o comercio, se traduce en un canje en productos similares a estos rubros o comercios internos.

✓ Construcción	➡	Ferretería
✓ Combustible	➡	Seguro
✓ Viajes On Them	➡	Viajes
✓ Multitienda y supermercado	➡	Gift card
✓ Multitienda	➡	Electro cocina

Ilustración 50: Efecto círculo virtuoso en cada línea estudiada

Producto de este efecto y a la caracterización más específica de los clientes, que se pueden ofrecer líneas de acciones complementarias al estudio general de los *drivers* que permita generar acciones focalizadas que incentiven el uso y gasto con la tarjeta.

18. Líneas de Acción y Aplicación Comercial

Una vez identificados el perfil de clientes de alta interacción con la compra con cuotas, compra con promociones y canje de puntos, se procedió a proponer línea de acción que la empresa pueda ejecutar en forma inmediata de acuerdo a las necesidades que estime primordiales.

De esta manera en conjunto con la empresa se procedió a generar una estrategia, que en base a los resultados antes obtenidos, permitan generar un mayor uso de la tarjeta y como consecuencia, un mayor gasto.

18.1 Estrategia propuesta

La estrategia propuesta con la empresa, alude en gran medida a generar acciones fáciles de comprender y aplicar para evitar la no implementación por la confusión que pueden generar tantos resultados obtenidos.

Para evitar este inconveniente, la estrategia a utilizar consiste en dos etapas. La primera identificará los clientes a gestionar con un único *driver* y posteriormente para los clientes *target* del club de puntos, se identificará que línea de premio utilizar.

Primera etapa

Para la gestión de esta etapa, con la empresa se propuso gestionar únicamente a los 1.232.448 clientes de baja interacción⁴².

La razón de no considerar a los clientes que pertenecen al perfil de alta interacción con otro *driver*, es que estos ya tienen un buen comportamiento de gasto

⁴² Ver diagrama de Venn, capítulo 9.

con la tarjeta en comparación con los de baja interacción. Además si se focalizan las acciones únicamente a este universo de clientes, se estará gestionando al grupo de peor comportamiento que tiene oportunidad de crecimiento, y no a todos “a tontas y a locas”.

Con los resultados obtenidos sobre estos clientes, se propuso una gestión de acuerdo a dos reglas. Por un lado, por orden de la empresa no gestionar menos de 200.000 clientes⁴³ en cada *driver* y por otro lado gestionar a los que tengan las mayores probabilidades de respuesta (perteneciente a los “sets” de nodos más propensos).

Acorde a estas exigencias, se propusieron las siguientes condiciones en cada *driver*.

- Target punto
Clientes con un lift ≥ 1.2 , capturando 253.433 clientes.
- Target promociones
Clientes con un lift ≥ 1.19 , capturando 274.767 clientes
- Target cuotas
Clientes con un lift ≥ 1.55 , capturando 273.271 clientes.

De acuerdo a este criterio, puede haber clientes que sean *target* de más de un *driver*, por lo que se generó un diagrama de Venn sobre los clientes *target*.

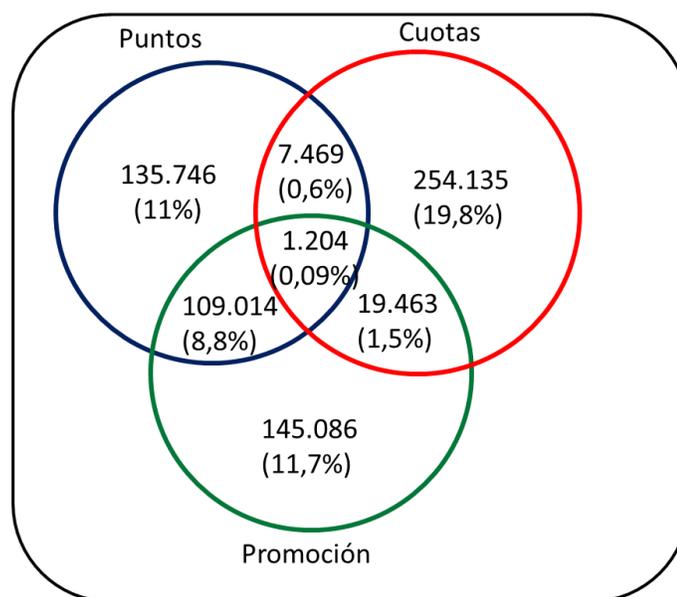


Ilustración 51: Clientes target según propensión a la alta interacción

⁴³ Esto se debe a que por política de la empresa, no se realizan campañas hacia grupos muy pequeños de clientes.

Como se puede observar, la mayoría de los clientes se identifican como *target* de un *driver*. Sin embargo hay un grupo de clientes que pertenece a más de un *driver*, los que sólo se gestionarán con un único *driver*.

Para identificar que *driver* utilizar en estos casos de intersecciones, se acordó con la empresa en seleccionar el *driver* que genere un mayor gasto promedio. De esta manera se calculó el gasto promedio que tienen los clientes que actualmente son calificados de alta interacción con cada *driver*, así como al grupo baja interacción.

Conjunto	promedio gasto anual ⁴⁴	Diferencia gasto con baja interacción
Alta interacción puntos	\$2.271.307	\$1.729.858
Alta interacción promociones	\$1.253.548	\$712.099
Alta interacción con cuotas	\$684.047	\$142.598
Baja interacción	\$541.449	-

Tabla 22: Gasto anual promedio perfiles de clientes

Por lo tanto se priorizará el gestionar con puntos, seguido de promociones y finalmente con cuotas. Con estos criterios, se propone gestionar a un cliente con un *driver*:

- 253.433 clientes con el club de puntos
- 164.549 clientes con las promociones
- 245.135 clientes con las cuotas

Segunda etapa

Posteriormente se ejecutaron los modelos de líneas de acción a los 253.433 clientes *target* del club de puntos, escogiendo como “herramienta” de gestión a la línea que entrega la mayor probabilidad.

Sin embargo cabe destacar que el premio asociado a Seguro se caracteriza por ser estacional (se ofrece en marzo), por lo que para los clientes que se les ofrezca seguro, adicionalmente se les ofrecerá el segundo de mayor propensión. De esta manera los clientes a gestionar con las determinadas líneas son:

- 31.831 (12,55%) con electrodomésticos
- 41.639 (16,42%) con premios de ferretería
- 53.955 (21,2%) con viajes
- 61.771 (24,37%) con gift card
- 62.237 (24,55%) con Seguro
 - 1.213 con gift card el resto del año
 - 5.955 con viajes el resto del año
 - 29.633 con electro domésticos el resto del año

⁴⁴ Sin considerar al 5% extremo por la posible presencia de *outliers*

- 25.436 con ferretería el resto del año

18.2 Beneficios comerciales

A partir de la estrategia propuesta, se pueden cuantificar los beneficios esperados en torno al incremento de gasto que los clientes tienen por estas estrategias de marketing. Estos beneficios se obtienen a través de la siguiente fórmula:

$$\sum (\text{Incremento gasto}_i * \text{probabilidad de respuesta}_i) - \text{costo gestión}$$

- Probabilidad de respuesta: aquella probabilidad de tener una alta interacción obtenida por el modelo
- Costo gestión: Al tratarse de clientes activos y no fugados, no hay costos de generar un plástico nuevo ni adquirir a los clientes. Por otro lado, si bien hay costos de acción (*mailing*), estos son muy inferiores al incremento anual por lo que no se tomarán en cuenta.
- Incremento gasto: Diferencia del gasto promedio que tiene un cliente del perfil target con respecto al gasto actual que este tiene. Sin embargo hay que tener en cuenta que existe la posibilidad que el cliente no tenga capacidad para llegar a tal nivel de gasto, por lo que se ocupará la variable de potencial de gasto que la empresa tiene calculada a través de la siguiente fórmula

$$\text{Incremento gasto}_i = \min\{\text{Gasto potencial}_i, \text{Gasto perfil target}\} - \text{gasto actual}_i$$

Por lo tanto, si el potencial es inferior al gasto promedio del perfil, solo habrá un incremento hasta el potencial de gasto del cliente.

Al realizar este cálculo, sólo el 2% de los clientes target de cada uno de los *drivers* tuvo un potencial inferior al gasto promedio del perfil target, lo que indica que es posible incentivarlos hasta el nivel propuesto.

Ejemplo

“El cliente *i* tiene un **potencial de gasto** de \$2.792.844, **gastó** en el periodo \$349.440 y forma parte del Target de puntos cuyo **gasto promedio** fue de \$2.271.307.

Su potencial es mayor que el gasto promedio del target, por lo que el incremento de gasto es el del target menos el gasto actual (\$2.271.307 - \$349.440 = \$1.229.157).

Si su probabilidad de alta interacción es de 0.2, entonces su incremento esperado es de \$245.831 anuales”

Aplicando la fórmula para cada uno de los *drivers*, se obtiene que la ganancia generada por el club de puntos es la mayor, seguido por las promociones y finalmente por la compra con cuotas. De esta manera la ganancia total generada en un año produce un incremento del gasto de un 6,35% del gasto anual.

Estrategia	Clientes target	Promedio ganancia	Min ganancia	Max ganancia	Ganancia total	% del gasto anual
Cuotas	522.829	\$66.584	-	\$138.492	\$34 MM	1,4%
Promoción	566.513	\$71.526	-	\$323.317	\$40 MM	1,7%
Puntos	418.251	\$457.657	-	\$1.294.644	\$191 MM	7,8%
				Total	\$266 MM	10,9%

Tabla 23: Ganancia por incremento del gasto de los clientes target

Resultado que a nivel mensual, produce un incremento de \$7.585 para el target de cuotas, \$9.075 para el target promoción y de \$37.755 para el target de puntos.

19. Conclusión

Actualmente la tarjeta de crédito incentiva el uso y gasto de ésta por medio de tres principales *drivers* que son las compras con descuentos, prestamos de dinero y poder usar un club de lealtad. Producto de estas tres ofertas, el estudio realizado consistió en determinar el perfil de clientes que ocupa la tarjeta por un determinado *driver*, identificando la diferencia entre estos y así generar una oferta más focalizada que incremente el uso de la tarjeta.

Los modelos generados indican que los perfiles de los clientes se determinan por las siguientes características:

1. Puntos: Clientes con un mayor involucramiento con la tarjeta al ocuparla reiteradas veces, incluso en la competencia directa del holding. Además tienen un alto poder económico (cupó) y necesidad de gasto (hijos), así como un ingreso reiterado a la web
2. Cuotas: Clientes que no ingresan a la página web, con un bajo poder económico, de mayor edad y que prácticamente no llegan a niveles de canje
3. Promociones: Clientes jóvenes que ocupan la tarjeta mayoritariamente en la multitienda, aprovechando las compras en ocasiones y productos que generalmente presentan descuentos mayores por el uso de la tarjeta.

Debido a esto, los perfiles de *drivers* apuntan a clientes distintos, por lo que se hace necesario definir estrategias diferenciadas para cada *driver* según las necesidades de los perfiles antes descritos. Hecho que reafirma la distribución actual del diagrama de Venn, la ausencia de una tendencia clara entre las tasas y una baja correlación lineal entre estas tasas.

Posteriormente al extender el modelo de puntos, se pudo obtener para la mayoría de los casos, un perfilamiento de clientes aún más específico como los “maestros de construcción” y “mujeres con gran poder adquisitivo” para ferretería y viajes respectivamente. Por otro lado, la caracterización no se pudo diferenciar en electrodomésticos de cocina por la gran variedad de premios que envuelve, provocando que sea un premio de canje “fácil” para no perder puntos.

Sin embargo el efecto más importante en esta extensión del modelo, fue el efecto “circulo virtuoso” que intervino de manera transversal para todas las líneas estudiadas, el cual señala que el cliente canjea premios altamente relacionados a los rubros o comercios en que este acumuló (esfuerzo).

Producto de los resultados obtenidos, y en pos de los perfiles obtenidos por el modelo, se le recomienda a la empresa que su gestión focalizada, este vinculada con la “etapa de vida” del cliente para generar una política y estrategia de crecimiento y desarrollo de este por una oferta dinámica en el tiempo que se adecúe a las necesidades e intereses de este.

Esta estrategia consistes que en edades muy jóvenes incentivar ocasionalmente al cliente con promociones y descuentos, ya que estos clientes ocupan mayoritariamente la tarjeta en la multitienda y ahorran con descuentos, generando una atracción inicial con los beneficios de la tarjeta.

Para edades superiores o con cinco años con la tarjeta, se hace necesario evaluar si el cliente tiene poder económico suficiente para canjear puntos, donde en caso positivo, incentivar con alguna de estas líneas de acuerdo a sus características.

La diferenciación radica en incentivar en virtud del “circulo virtuoso” ofreciendo bonos o doble puntaje a los clientes, en comercios semejantes a la línea de premios a gatillar el canje. Ejemplo: Un hombre con cupo de \$500.000, se le quiere incentivar con premios de ferretería, entonces se incentiva comprar en el comercio de construcción.

Finalmente si la evaluación del poder económica es negativa, se sugiere incentivar el uso con la promesa de poder realizar pagos con cuotas en rubros y comercios afines a este perfil, como el de los supermercados o farmacias, que son compras necesarias que el cliente tiene que realizar. Estrategia a realizar con la aprobación de capacidad de endeudamiento del cliente por parte del área de riesgo.

En resumen este trabajo muestra que es posible incentivar el uso y gasto de los clientes que tienen una baja interacción con la tarjeta a través de una gestión personalizada, tanto a nivel de *drivers* como de líneas de premios, en concordancia con sus necesidades específicas.

20. Limitaciones y Trabajo Futuro

El trabajo realizado cuenta con tres principales limitaciones, que podrían entregar un resultado mucho más preciso en el perfilamiento y posibles acciones.

La primera es haber realizado el estudio a nivel de cliente titular y no por rut comprador, debido a que la información de relacionada a puntos no tiene ese nivel de desagregación. Producto de esto, no se pudo generar acciones a nivel de “shopper”, quien es quién realmente impulsa la compra por sobre el comprador.

Otra limitación existente, es que si bien se determinó si un cliente tiene una alta o baja interacción con un determinado *driver*, no se realizó un estudio más exhaustivo sobre la comparación de interacción entre *drivers*, logrando determinar si la interacción con uno es mayor o no que con otro *driver*.

En tercer caso, el estudio únicamente utilizó información interna observable (base de datos), sin incorporar información complementaria como datos macroeconómicos sobre gastos en determinados rubros, deuda, etc. Así como estudios de mercados con información de preferencia, conocimientos y pensamiento que los clientes tienen con respecto a los *drivers* y tipos de premios.

La incorporación de estos puntos en análisis futuros puede ser de gran ayuda para un análisis más preciso y posiblemente con mejores resultados.

Como continuación del trabajo realizado, se proponen los siguientes lineamientos que puedan ahondar más el trabajo:

- Expansión del modelo de puntos, agregando líneas que por su reciente incorporación no pudieron ser evaluadas, como lo son las vinculadas al supermercado o experiencias
- Estudiar el efecto que tienen las líneas de acciones propuestas sobre los clientes target, evaluando a grupos en torno al incremento del uso de la tarjeta
- Extender los estados posibles (alta interacción o baja interacción) hacia distintos posibles estados de interacción que el cliente puede tener con el *driver*, para evaluar en el tiempo que acciones y características determinan que un cliente pase de un estado otro.

21. Bibliografía

1. Nicole Pereira, 2014. Identificación de clientes con patrón de consumo eléctrico fraudulento. Memoria para optar al título de Ingeniero Civil Industrial, Santiago, Universidad de Chile, FCFM.
2. Karem Padilla 2015. Identificación de clientes de alto valor para el desarrollo de alianzas de una empresa. Memoria para optar al título de Ingeniero Civil Industrial, Santiago, Universidad de Chile, FCFM.
3. Carlos Roco, 2010. Modelamiento predictivo para el aumento de consumo de tarjeta de crédito sobre el análisis de comportamiento transaccional de clientes en una institución financiera. Memoria para optar al título de Ingeniero Civil Industrial, Santiago, Universidad de Chile, FCFM.
4. María Constanza Rojas, Modelo predictivo de desafiliación de empresas para una caja de compensación de asignación personal. Memoria para opta al título de Ingeniero Civil Industrial, Santiago, Universidad de Chile, FCFM.
5. Gang Wu, Edward Y. Chang. "Class-Boundary Alignment for Imbalance Dataset" Department of Electrical & Computer Engineering, University of California, Santa Barbara, 2013.
6. Xu, R. y Wunsch II, D. C.: «Survey of clustering algorithms». *Neural Networks, IEEE Transactions on*, 2005, 16(3), pp. 645–678.
7. Praveen K Kopalle, Scott Neslin, Baohong Sun, Yacheng Sun, Vanitha Swaminathan "The Joint Sales Impact of Frecuency Reward and Customer Tier Components of Loyalty Programs" September 2011.
8. Pietro Zidda, Nathalie Demoulin, 2008. "On the impact of loyalty cards on store loyalty: Does the customers' satisfaction with the reward scheme matter?" *Journal of retailing and consumer services*, Volume 5 Issue 15, Pages 386–398
9. J. Sempere, «Aprendizaje de árboles de decisión, » Universidad Politécnica de Valencia, Valencia.
10. Kivetz, Ran, Oleg, Urminsky y Yuhuang Zheng (2006b). "The Goal-Gradient Hypothesis Resurrected: Purchase Aceleration, Ilusionary Goal Progress and Customer Retention". *Journal Marketing Research*, 43 (1), 39-58
11. Cristophe Croux, Nicolas Glady, "Predicting customer wallet without survey data", *Journal of service research*, vol. LXI, pp 219-231, 2009.
12. Simon Knox, Adrian Payen, Stan Maklan, Joe Peppard, Lynette Ryals, "Customer Relationship Management, perspectives from the marketplace", 2003.

22. Anexo

Anexo A. Descripción Drivers

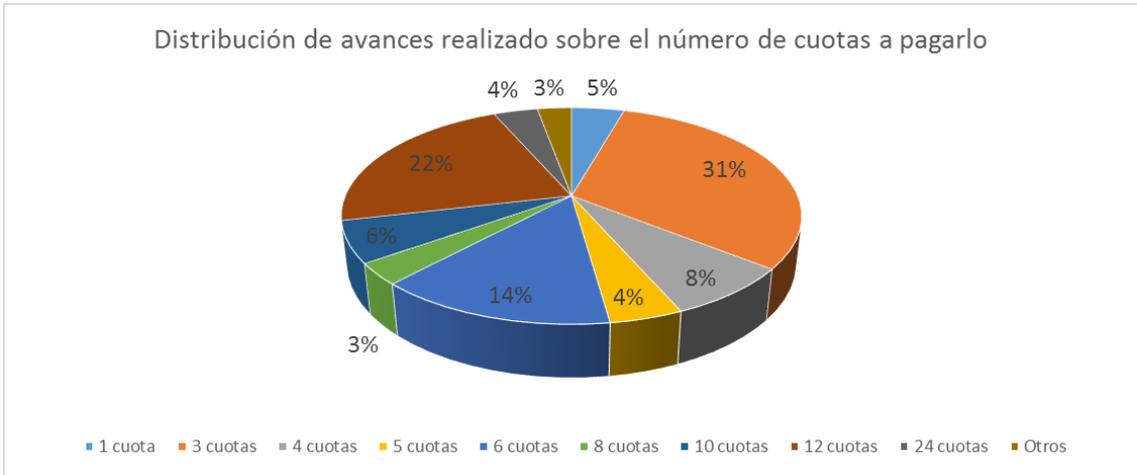
1. Niveles del catálogo de premios.

Descripción de los niveles de canjes y puntos requeridos en el canje para ese nivel. Como se puede observar el tramos inferior es de 1.000 puntos y 3.000 puntos, pero estos son niveles nuevos donde se ofrecen productos como descuentos o promociones donde se “baja” de nivel un premio de niveles mayores.

Nivel	Puntos requeridos
1	1.000
2	3.000
3	5.000
4	9.000
5	12.000
6	18.000
7	24.000
8	36.000
9	48.000
10	60.000
11	90.000
12	120.000
13	180.000
14	240.000

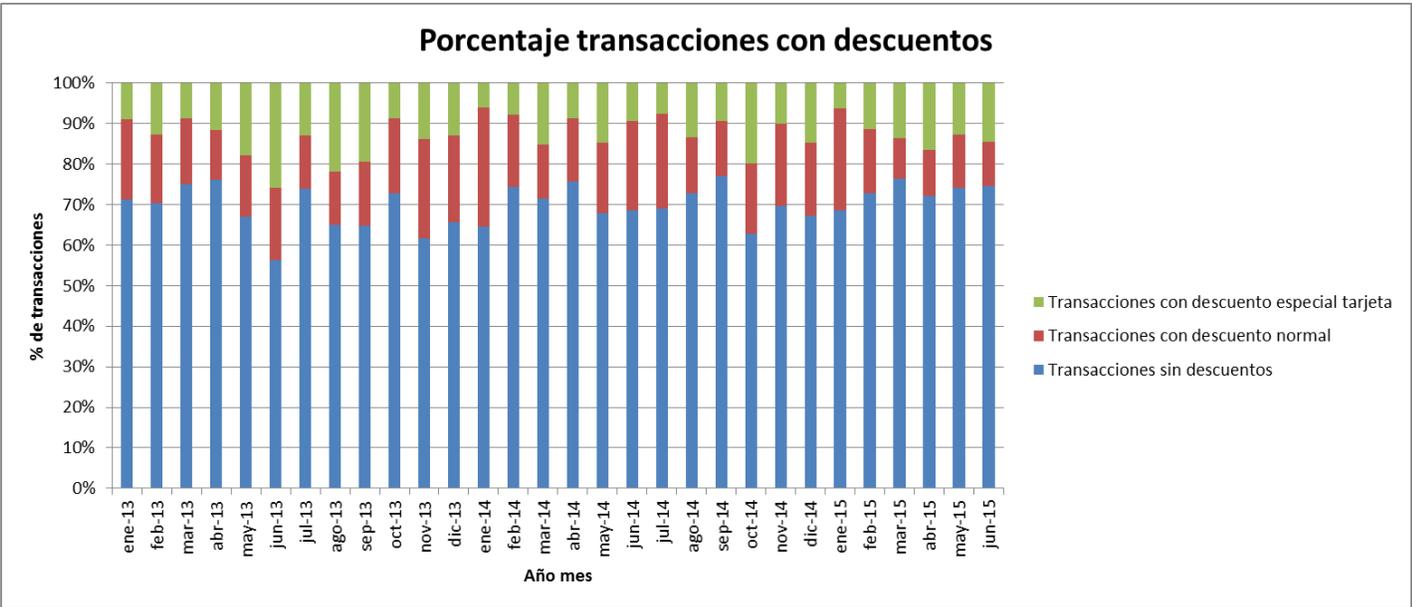
2. Distribución de pagos según cuotas

En los movimientos de avances en efectivo se puede observar que predominan los pagos con tres, seis y doce cuotas, evidenciando que el préstamo que los clientes piden es para posponer el pago cuando tengan el dinero



Transacciones con descuentos

Al observar la participación de las transacciones con descuento en el comercio de multi tienda, se observa que estas tienen una gran participación alcanzando alrededor de un 30% de las transacciones.



Anexo B. Descripción Variables

1. Distribución avances y créditos de consumos

Al realizar la distribución de los clientes según el monto de avances y créditos de consumo, únicamente considerando a aquellos que hayan realizado al menos un movimiento, se observa que en avances se concentra en montos bajos, mientras que en créditos de consumo más del 40% lo hace con valor superior a un millón de pesos.

Variable	Mínimo	Q1	Q2	Q3	Q4	Máximo	Media	Desv. típ.
Avance	\$1.000	\$84.000	\$170.000	\$300.000	\$561.000	\$33.645.935	\$377.191	490.239
Crédito consumo	\$29.978	\$500.000	\$1.000.000	\$1.000.000	\$2.000.000	\$14.887.296	\$1.331.289	1.083.868

2. Correlación entre variables independientes

Se observa una correlación baja entre las variables continuas que se puede usar en el estudio.

	Ticket promedio	Cupo	Días ultima compra	Días look	Edad	Antigüedad
Ticket promedio	1	,157**	,158**	-,050**	,013**	,040**
Cupo		1	-,083**	,109**	,125**	,486**
Días ultima compra			1	-,151**	,065**	-,028**
Días look				1	-,278**	-,067**
Edad					1	,407**
Antigüedad						1

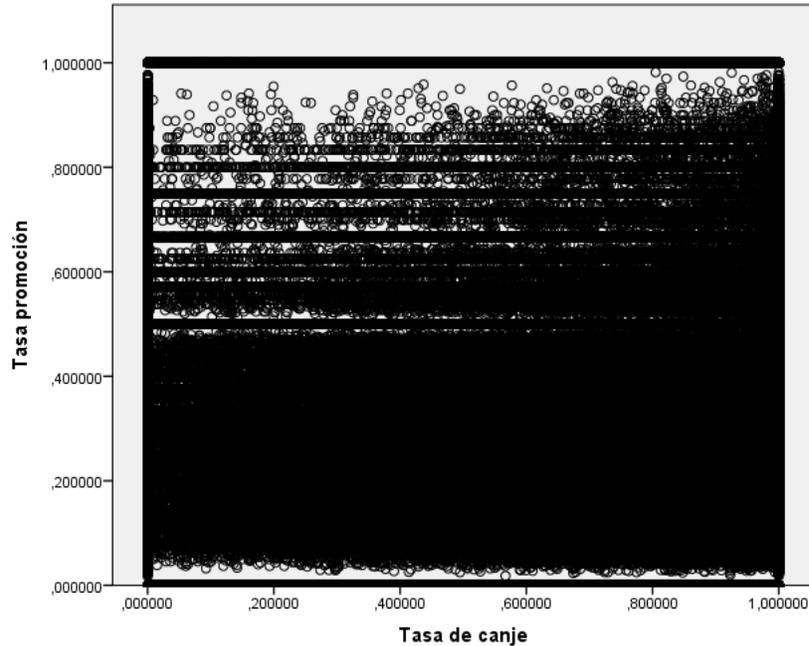
3. Descripción general de cada uno de los drivers

Se puede observar la gran varianza que tiene la acumulación de puntos. Por otro lado la media de la tasa de canje es un 29%, la de compra con cuotas de un 24% y de compra con promociones 26%, viendo que en general las tasas apuntan a un cuarto del máximo logro

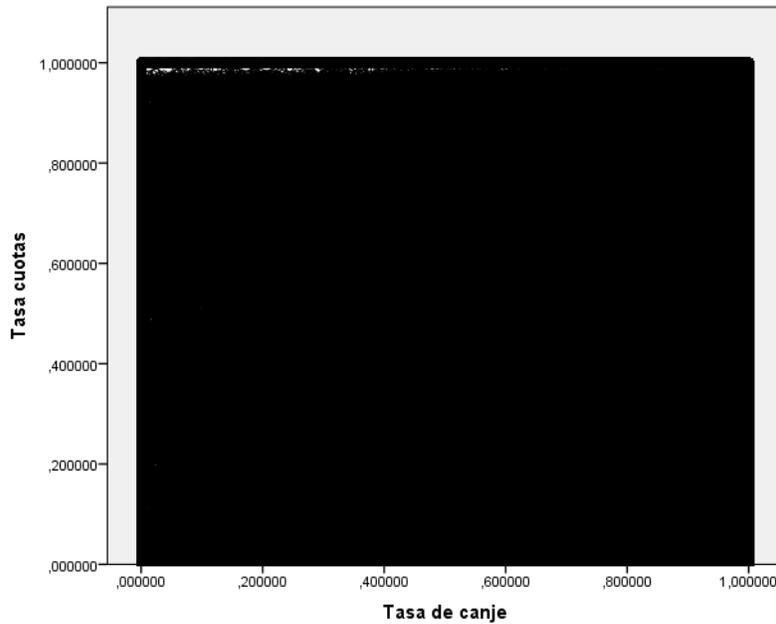
	Mínimo	Máximo	Media	Desviación estándar
Tasa Avance	0	10,7963	,102853	,2660633
Monto avance	0	33.645.935	86.737,48	283.267,179
Tasa crédito consumo	0	4	,076955	,2407398
Monto crédito consumo	0	14.887.296	153.532,56	562.381,191
Transacciones totales con tarjeta	0	6.348	41,73	66,954
Tasa compra con cuota	0	1	,24273	,258566
Tasa de transacciones con descuentos	0	1	,26283	,301730
Transacciones totales con tarjeta en holding	0	1.817	7,46	13,304
Monto Acumulado	0	1.509.740	5.293,17	12.070,786
Tasa de canje	0	1	,29236	,400747

4. Relación entre tasas

El análisis de tendencia con el resto de las combinaciones son similares al caso expuesto, donde no se observa una tendencia clara de a variación de una tasa cuando cambia el valor de otra.

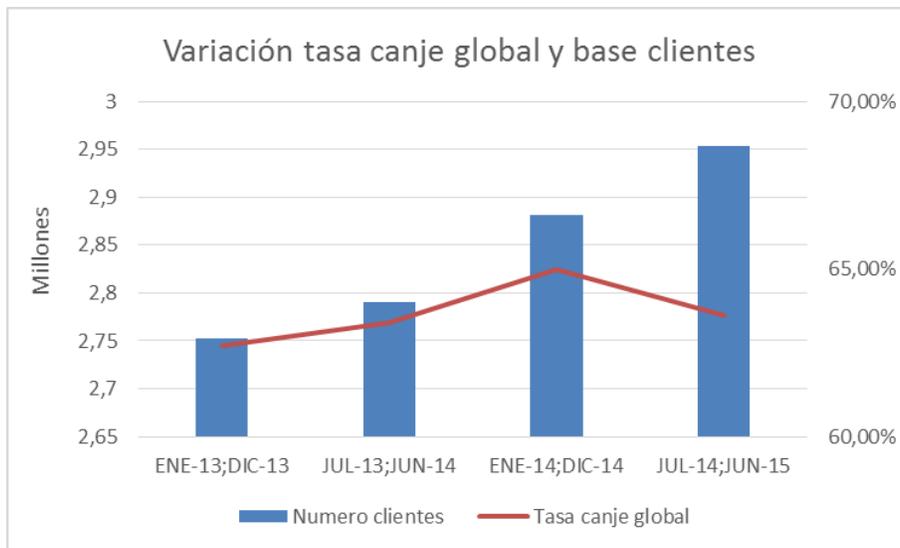


El caso de cuotas con puntos es el más extremo, al mostrar que prácticamente hay independencia entre estas variables.



5. Gráfico de robusted del indicador de la tasa de canje

A pesar de haber un aumento de clientes (esperable) la tasa de canje ha tenido una variación muy pequeña en dos años de diferencias, manteniendose cerca de un 64%



Anexo C. Distribución Clientes con Drivers

1. Matriz doble entrada indicadores

Puntos

Se puede observar que a medida que la acumulación aumenta, la cantidad de clientes con tasa 0 disminuye. Por lo tanto existe una correlación muy alta entre la acumulación y tasa de canje.

Total puntos acumulados	Rango Tasa de canje											Total	%	
	0%	0% - 9%	10% - 19%	20% - 29%	30% - 39%	40% - 49%	50% - 59%	60% - 69%	70% - 79%	80% - 89%	90% - 99%			100%
0 - 999	406.740	455	399	405	410	377	377	376	304	297	378	3.232	413.750	20,1%
1.000 - 4.999	714.280	5.989	7.518	8.669	9.662	11.252	13.085	15.517	17.167	19.508	24.535	44.410	891.592	43,3%
5.000 - 8.999	111.891	2.132	3.324	4.564	7.040	10.473	16.225	22.327	26.528	30.720	38.564	73.878	347.666	16,9%
9.000 - 11.999	25.727	688	1.105	1.636	2.680	4.160	4.889	5.926	8.961	13.002	18.737	44.856	132.367	6,4%
12.000 - 17.999	19.083	636	932	1.449	2.360	2.550	3.704	5.698	7.806	10.779	17.668	62.989	135.654	6,6%
18.000 - 23.999	6.484	208	342	582	647	890	1.143	1.578	2.361	3.750	6.986	32.814	57.785	2,8%
24.000 - 35.999	4.141	146	257	292	448	494	672	887	1.416	2.090	4.435	28.561	43.839	2,1%
>= 35.999	2.447	76	121	149	172	251	337	450	685	1.118	3.650	26.776	36.232	1,8%
% del total	62,7%	0,5%	0,7%	0,9%	1,1%	1,5%	2,0%	2,6%	3,2%	3,9%	5,6%	15,4%	-	-
Total general	1.290.793	10.330	13.998	17.746	23.419	30.447	40.432	52.759	65.228	81.264	114.953	317.516	2.058.885	100,0%

Cuotas

A medida que se llega a mayo cantidad de compras es más “difícil” llegar a niveles de tasas de compra con cuotas más altos

# Transacciones	Rango tasa compras con cuotas											Total	%	
	0%	0% - 9%	10% - 19%	20% - 29%	30% - 39%	40% - 49%	50% - 59%	60% - 69%	70% - 79%	80% - 89%	90% - 99%			100%
0-4	120.943	-	-	10.777	17.171	-	44.946	24.777	21.235	-	-	174.143	413.992	20,1%
5 - 19	61.245	15.894	40.602	48.090	42.328	47.259	66.826	82.150	75.096	122.456	44.123	102.796	748.865	36,4%
20 - 39	19.113	22.766	26.785	31.568	36.032	39.753	45.106	43.519	44.268	48.094	36.052	9.105	402.161	19,5%
40 - 59	7.604	15.442	15.152	16.558	17.893	18.467	21.233	21.298	19.428	16.898	10.949	930	181.852	8,8%
60 - 69	2.328	5.741	5.307	5.461	5.637	5.717	6.231	5.962	5.562	4.295	2.642	109	54.992	2,7%
70 - 79	1.732	4.843	4.193	4.325	4.349	4.308	4.540	4.256	3.918	3.087	1.701	43	41.295	2,0%
>= 80	11.646	58.224	31.728	23.490	19.156	17.043	15.668	13.747	11.773	8.924	4.261	68	215.728	10,5%
% del total	10,9%	6,0%	6,0%	6,8%	6,9%	6,4%	9,9%	9,5%	8,8%	9,9%	4,8%	13,9%	-	-
Total general	224.611	122.910	123.767	140.269	142.566	132.547	204.550	195.709	181.280	203.754	99.728	287.194	2.058.885	100,0%

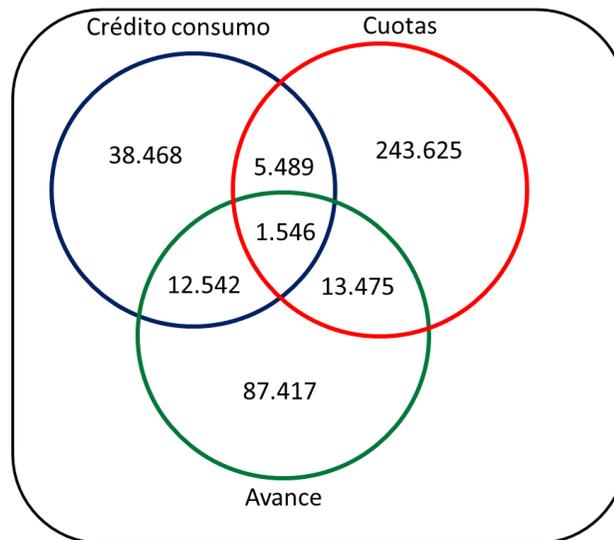
Promoción

Nuevamente el tener mayores transacciones dificulta el tener una mayor tasa de compra con promoción. Además la mayoría se concentra en un rango cercano al 30%.

# Transacciones	Rango tasa compra promociones												Total
	0%	0% - 9%	10% - 19%	20% - 29%	30% - 39%	40% - 49%	50% - 59%	60% - 69%	70% - 79%	80% - 89%	90% - 99%	100%	
0 - 4	693.775	-	-	52.988	67.263	-	117.963	33.980	13.321	-	-	133.789	1.113.079
5 - 9	68.764	-	84.064	100.875	57.742	57.717	38.446	28.803	5.358	8.840	-	1.744	452.353
10 - 14	10.635	19.356	39.874	52.239	44.778	25.140	16.567	5.784	2.095	774	259	82	217.583
15 - 19	2.000	6.965	26.380	34.275	22.553	13.276	5.642	1.904	490	316	76	22	113.899
>= 20	573	8.782	36.257	52.786	35.703	15.366	5.843	2.567	2.150	1.526	392	26	161.971
% del total	37,68%	1,70%	9,06%	14,24%	11,08%	5,42%	8,96%	3,55%	1,14%	0,56%	0,04%	-	-
Total general	775747	35.103	186.575	293.163	228.039	111.499	184.461	73.038	23.414	11.456	727	135.663	2.058.885

2. Distribución driver deuda

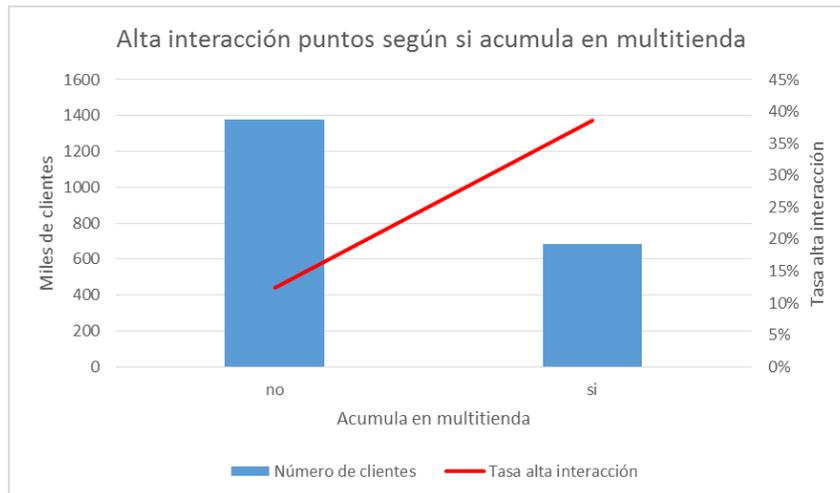
Mapa de Venn para el driver deuda, mostrando la distribución de clientes según compras con cuotas, crédito de consumo y avances en efectivo.



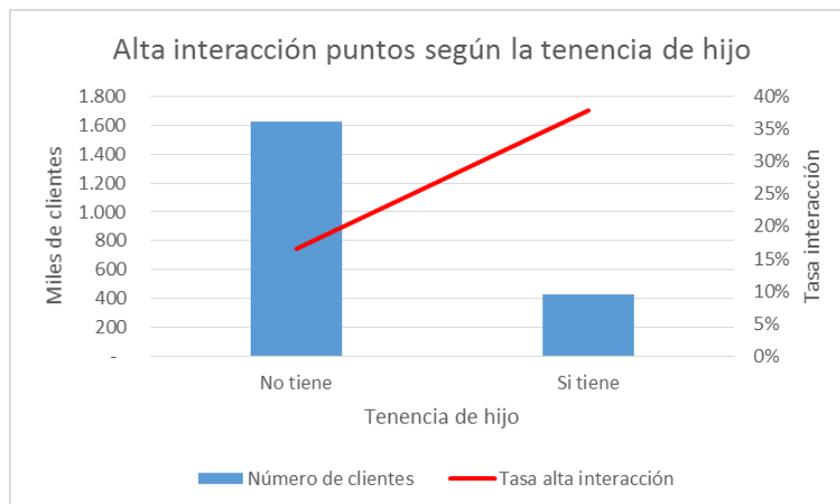
Anexo D. Exploratorio Drivers

1. Puntos

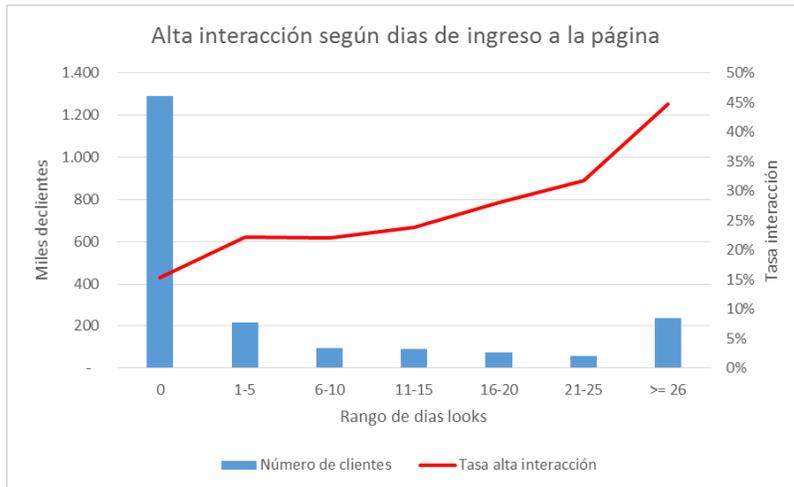
Tasa según si acumula en el rubro de multitienda o no. Se observa un notorio efecto positivo.



El tener un hijo al cuidado, también tiene un efecto positivo desde una tasa del 20% al 35%

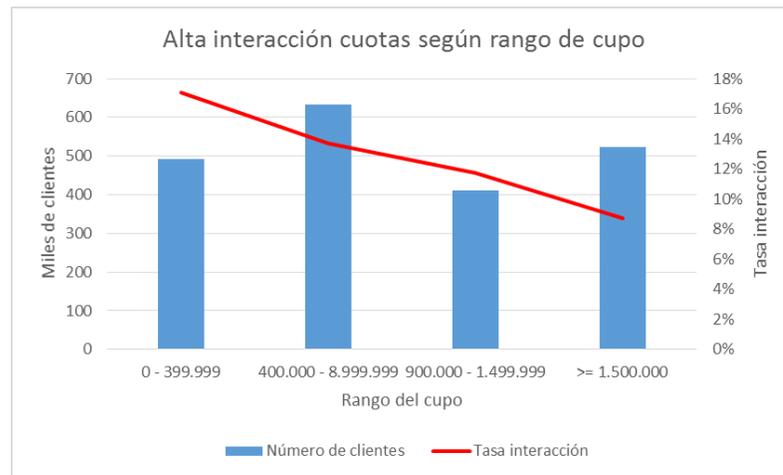


Finalmente se puede ver un efecto positivo ascendente a medida que el cliente tiene una mayor cantidad de días de ingreso a la página de internet

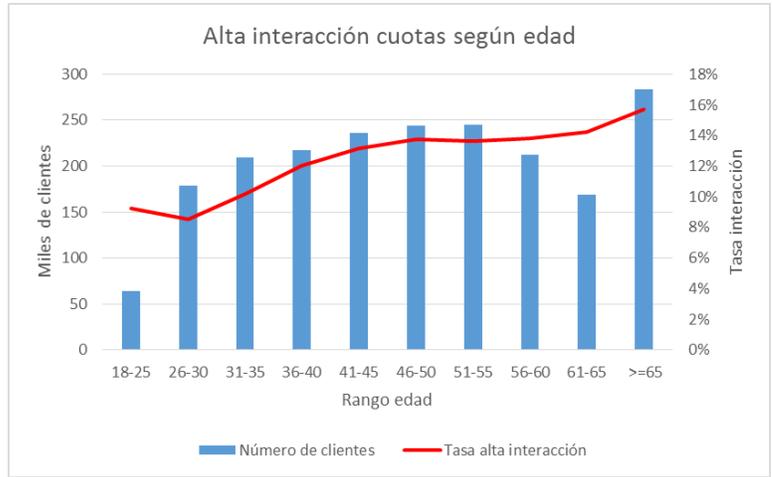


2. Cuota

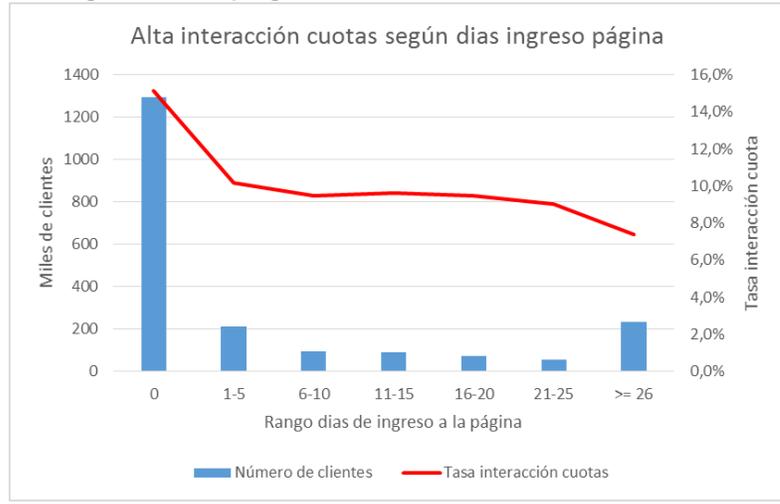
Como se podía prever un cliente con menor cupo es más propenso a la compra con cuotas, por lo tanto hay un efecto negativo.



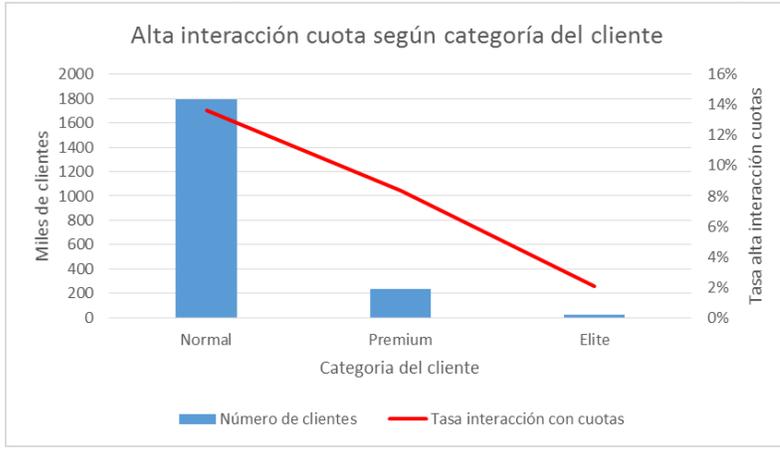
Por otro lado, la alta interacción con las cuotas toma una tendencia positiva a medida que aumenta la edad del cliente, teniendo un “peak” en el mayor rango de edad.



En el caso de la interacción según el número de días de ingreso a la página, se observa a diferencia de los puntos, una tendencia negativa al tener el “peak” en aquellos clientes sin ingreso a la página.

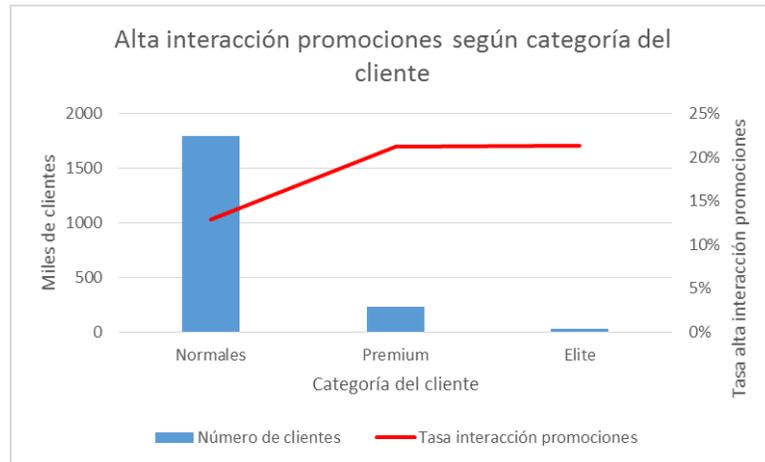


En el caso de la categoría del cliente se observa una mayor tasa para clientes normales, asociados a un menor poder económico e interacción con puntos.

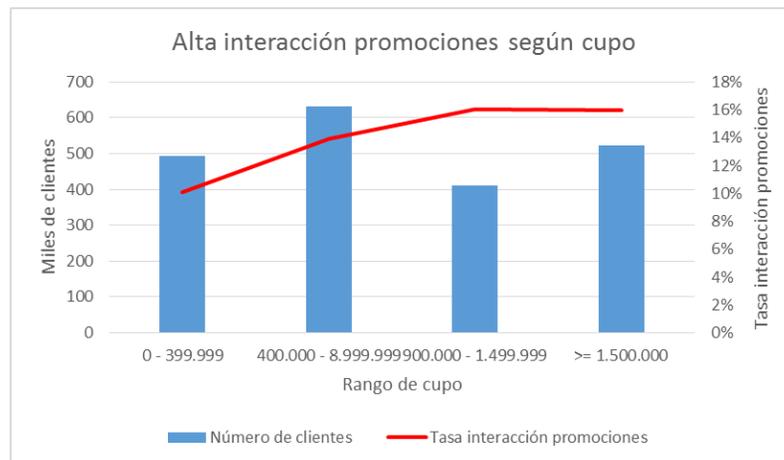


3. Promoción

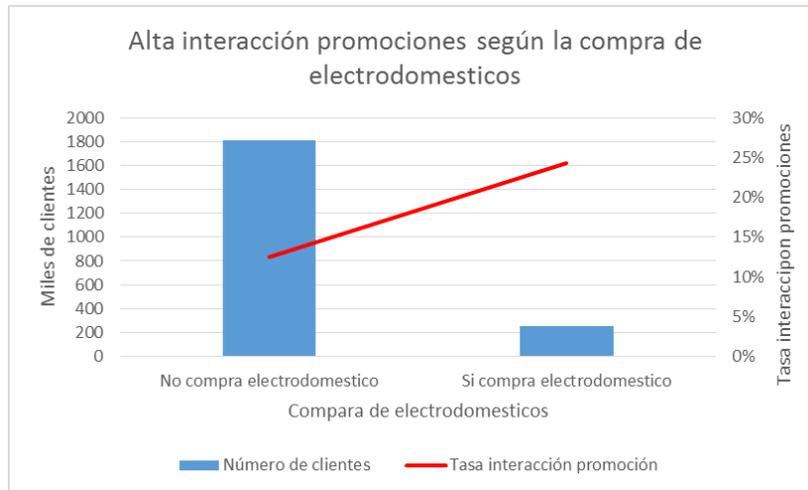
Al observar la alta interacción con las promociones según la categoría del cliente, se aprecia que los clientes normales son menos propensos en comparación a los premium y elite del club de puntos.



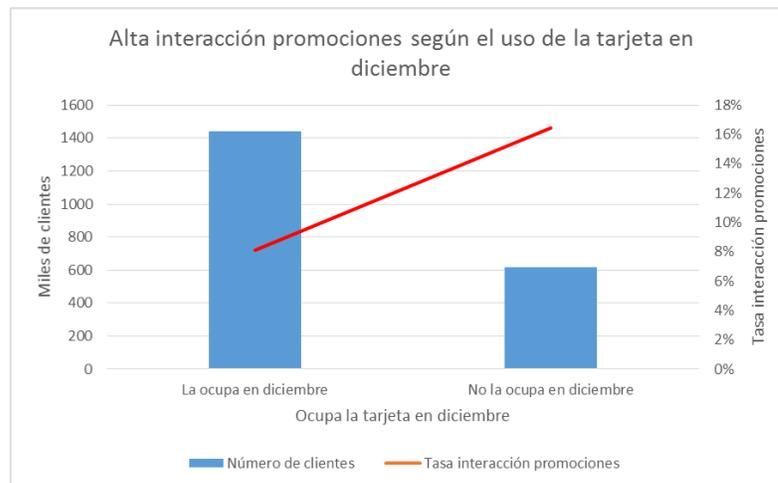
En el caso del cupo, se observa que la menor tasa de alta interacción se da en los rangos muy bajos y luego en rangos altos se mantiene estable.



En el caso del haber realizado una compra de un electrodoméstico, se observa un efecto positivo. Principalmente porque esta categoría es más común que además de ofrecer descuentos normales, ofrece descuentos adicionales por el uso de la tarjeta.



Finalmente se puede observar el comportamiento de que los clientes que prefieren realizar sus transacciones en cualquier rubro en el mes de diciembre, tienen una mayor tasa. Lo que quiere decir que quienes optan por el uso de esta tarjeta como medio en este mes tiene una tendencia a comprar siempre con promociones.



Anexo D. Importancia Variables y Desempeño Modelos

1. Importancia variable para promoción y cuotas

Importancia variable modelo cuotas árbol CART

Importancia de las variables independientes

Variable independiente	Importancia	Importancia normalizada
Realizar transacción en educación	,003	100,0%
Realizar transacción en restaurant	,002	64,2%
Ticket promedio	,002	64,0%
Realizar transacción en viajes	,002	57,2%
Cupo	,001	38,3%
Edad	,001	34,6%
Zona geográfica	,001	32,0%
Categoría del cliente	,001	24,3%
Transacción en supermercado	,000	13,6%
Tenencia tarjeta visa	,000	10,9%
Antigüedad	5,293E-05	1,6%
Tenencia hijo	4,906E-07	,0%

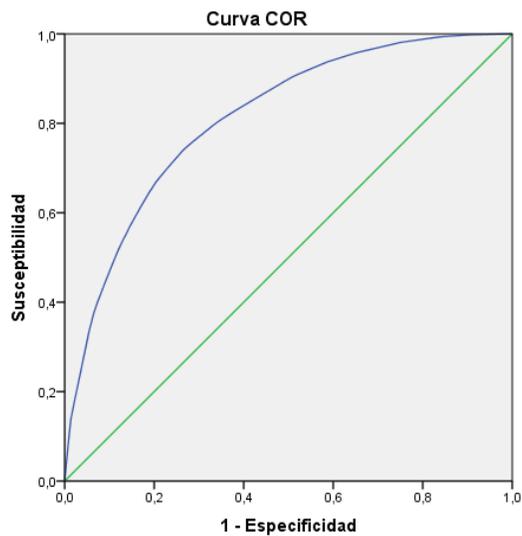
Importancia variable modelo promoción árbol CART

Variable independiente	Importancia	Importancia normalizada
Realizar transacción en diciembre	,003	100,0%
Realizar transacción en electrodomésticos	,003	86,6%
Realizar transacción en noviembre	,002	75,1%
Edad	,001	27,6%
Ticket promedio	,001	22,2%
Categoría	,001	19,5%
Tenencia hijo	,000	8,0%
Cupo	,000	7,3%
Realizar transacción fuera de la multitienda	,000	5,8%
Antigüedad	,000	4,1%
Tenencia adicional	6,334E-05	2,2%
Zona geográfica	2,774E-07	,0%

2. Curvas ROC desempeño

Drivers

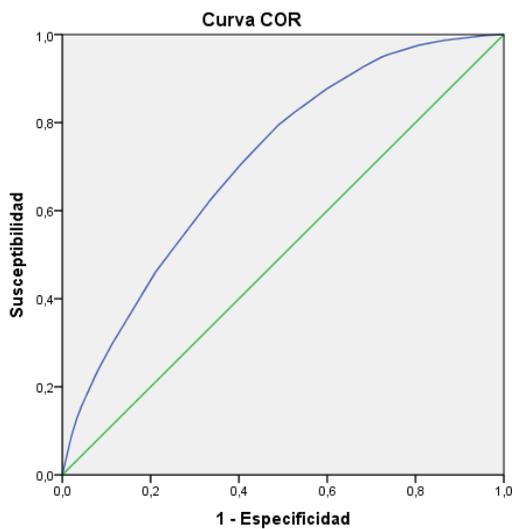
Modelo de puntos



Los segmentos diagonales son producidos por los empates.

Área: 81,1% (Árbol decisión)

Modelo de cuotas

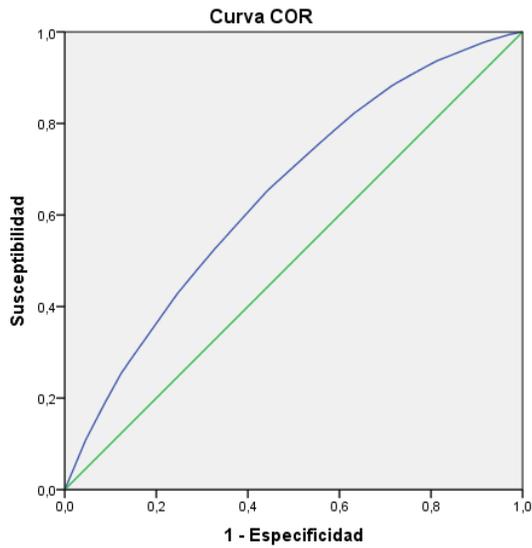


Los segmentos diagonales son producidos por los empates.

Área: 71,1% (Árbol decisión)

-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke	AIC	BIC
545455,126	,094	,125	545499,126	545740,169

Modelo de promociones



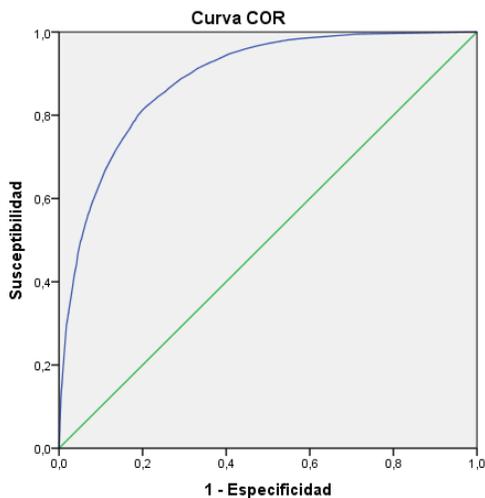
Los segmentos diagonales son producidos por los empates.

Área 65,7% (Árbol decisión)

-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke	AIC	BIC
604971,108	0,066	0,088	616120,883	616396,997

Líneas de premios

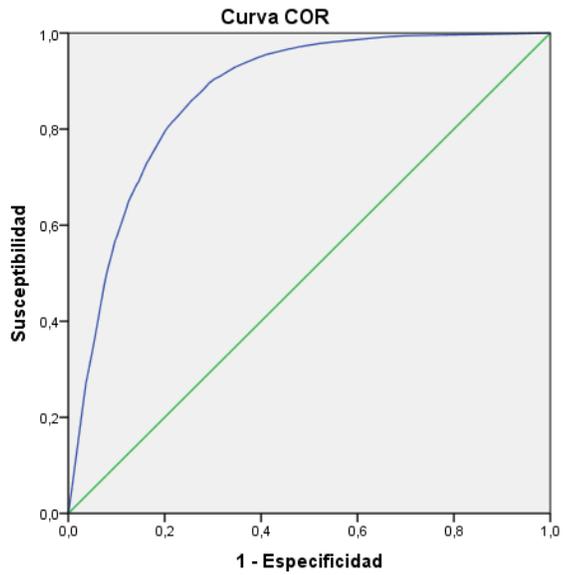
Viajes



Los segmentos diagonales son producidos por los empates.

Área: 88,8% (Árbol decisión)

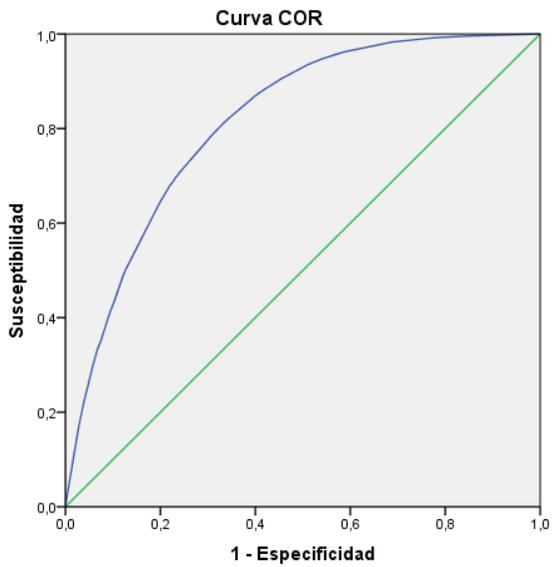
Seguro (SOAP)



Los segmentos diagonales son producidos por los empates.

Área 87,3% (Árbol decisión)

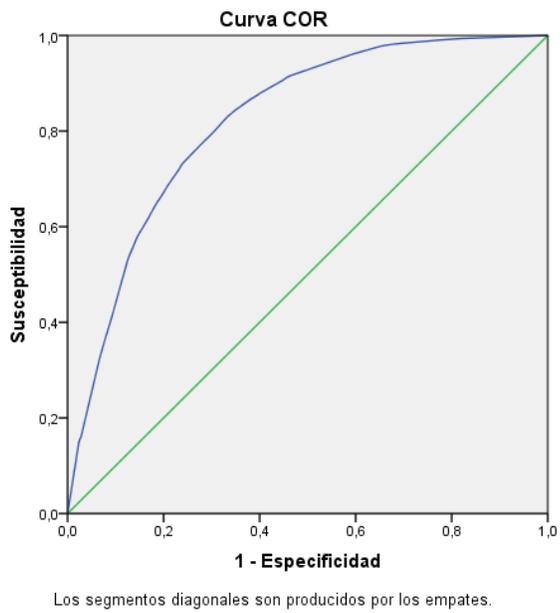
Ferretería



Los segmentos diagonales son producidos por los empates.

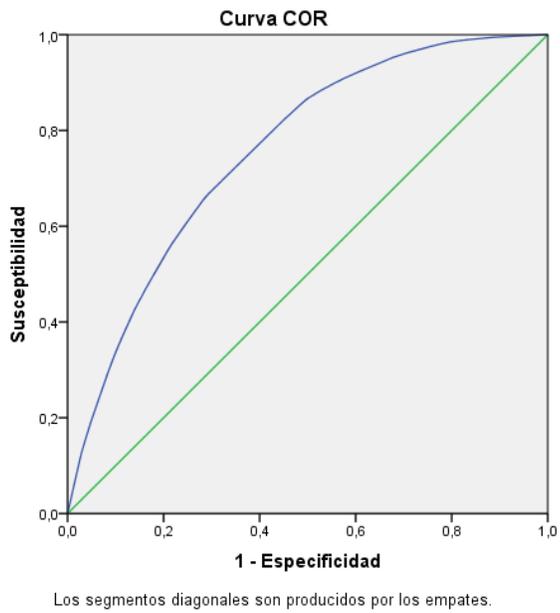
Área 81,4% (Árbol decisión)

Gift card



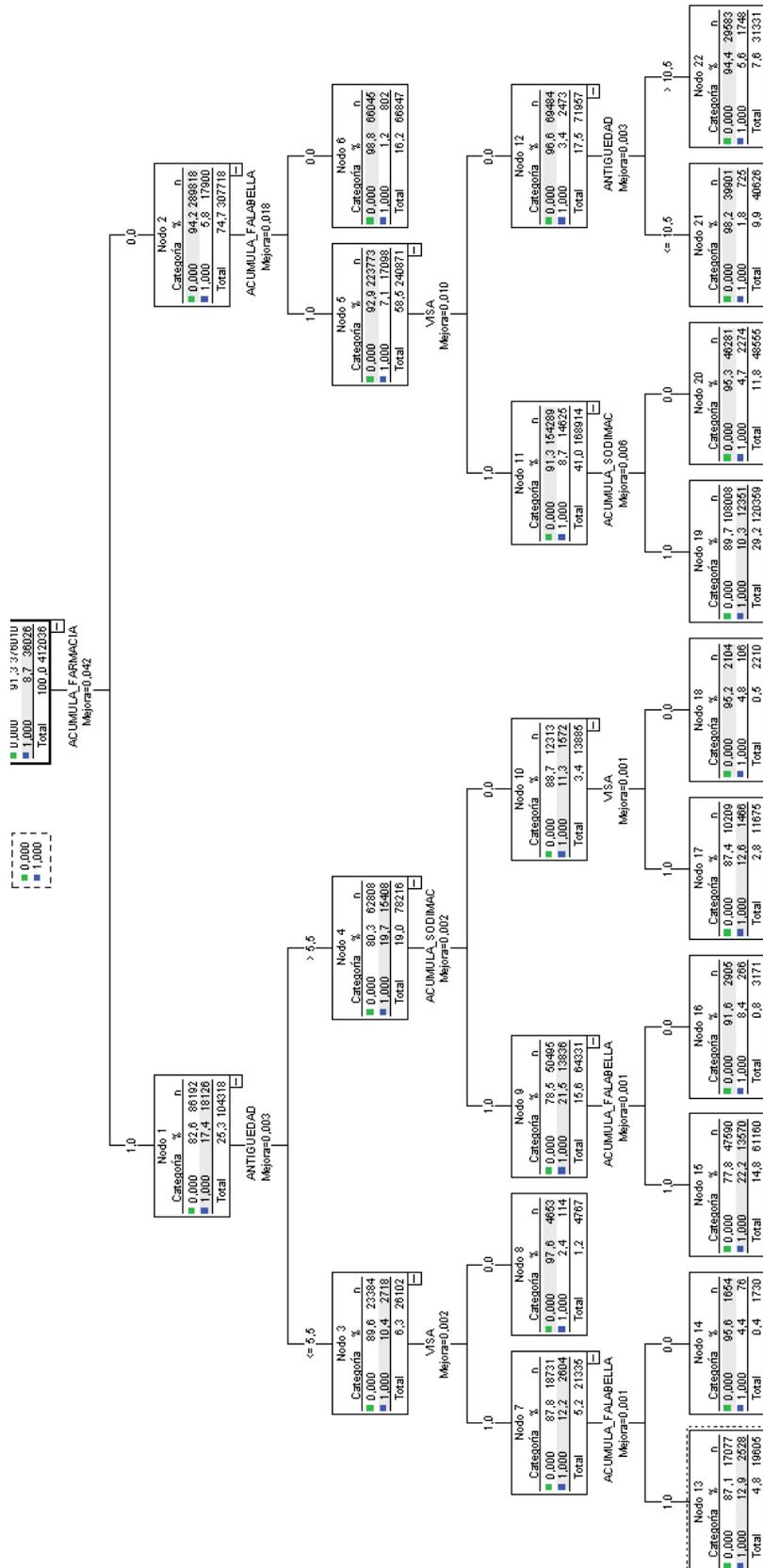
Área 82,1% (Árbol decisión)

Electro domestico cocina



Área 75,8% (Árbol decisión)

3. Árbol decisión electrodoméstico cocina



4. Árbol decisión gift card

