



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

EVALUACIÓN DEL RETORNO ASOCIADO AL PROGRAMA DE FIDELIZACIÓN DE UN RETAIL FINANCIERO

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

SEBASTIÁN ALFREDO MORENO HASELBAUER

PROFESOR GUÍA:
CAROLINA SEGOVIA RIQUELME

MIEMBROS DE LA COMISIÓN:
ERICK MÉNDEZ GUZMÁN
PABLO MARÍN VICUÑA

SANTIAGO DE CHILE
2018

**RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE:** Ingeniero Civil Industrial
POR: Sebastián Alfredo Moreno Haselbauer
FECHA: 06/11/2018
PROFESOR GUÍA: Carolina Segovia Riquelme

EVALUACIÓN DEL RETORNO ASOCIADO AL PROGRAMA DE FIDELIZACIÓN DE UN RETAIL FINANCIERO

En Chile el número de tarjetas de débito y crédito se ha triplicado en 7 años alcanzando casi 39 millones en febrero de 2017 con 20 emisoras en el país. Como consecuencia, se tiene un escenario muy competitivo en donde predomina la visión de futuro: la rentabilidad es el fruto de la relación con el cliente en el largo plazo, que debe desarrollarse y mantenerse. Por esto, las empresas invierten parte de su capital en programas de fidelización. Aunque como los resultados de estos son complejos de medir, no siempre se conoce cuál es el retorno asociado a la inversión y si cumple su objetivo de fidelizar.

El presente trabajo tiene como objetivo evaluar el retorno económico del programa de fidelización de un *retail* financiero a través de la estimación del efecto causal en la relación del cliente con la empresa en el largo plazo. Para cumplir este objetivo se propone una interesante metodología basada en la combinación de dos modelos: Customer Lifetime Value (CLV) para estimar el valor monetario de los clientes en el largo plazo y Propensity Score Matching para inferir el efecto causal de participar en el club de puntos.

Los resultados del estudio indican que cerca de un 16% de la cartera participa activamente del club de puntos y el efecto promedio causado por esta activación corresponde a un incremental en el valor de los clientes de 15 M\$ en la siguiente rentabilidad anual y de 62 M\$ en el CLV como efecto de largo plazo, implicando crecimientos de 12,5% y 6,2% respectivamente. Este se explica por un aumento de 41 M\$ en el gasto anual con la tarjeta junto con una mayor preferencia por los negocios del Holding, alzas de 2,7% y 1,6% en las tasas de consumo de créditos y avances respectivamente, y un aumento de 0,5 canjes anuales revelando el deseo de los clientes a continuar participando del programa.

Se profundiza el estudio analizando el efecto en 5 segmentos de clientes y se demuestra que el programa impacta de distintas maneras según el perfil del consumidor. En particular, se encuentran mayores incrementales en dos segmentos de menor consumo y menor valor rentable inicial. Por lo tanto, se recomienda focalizar la gestión para activar en el club de puntos a estos grupos específicos, por presentar un mayor potencial económico para el negocio. Al activar el 10% de estos clientes en forma aleatoria se estima producir un beneficio valorizado en 21.267 MM\$ a 11 años.

Se concluye que el programa de lealtad impacta positivamente en la relación de los clientes con la empresa y su estructura actual contribuye a desarrollar el valor de estos y a mantener fidelizados a los más rentables. Como próximos pasos, se recomienda diseñar campañas para promover la activación de los clientes en el club de puntos priorizando a los clientes de mayor efecto y potencial de crecimiento.

AGRADECIMIENTOS

Este trabajo marca el fin mi etapa universitaria y el inicio de una etapa llena de desafíos profesionales y personales. Me siento muy orgulloso de haberme formado en esta honorable escuela de ingeniería y universidad.

Gracias papás y familia. Nada de esto hubiese sido posible sin el respaldo de ustedes. Gracias por el hogar que han construido, por inculcarme sus valores, por darme siempre las mejores oportunidades y por ser un apoyo incondicional. Por eso y mil cosas más, gracias.

Gracias Brenda por estar siempre presente, en las buenas y en las malas, por escucharme (aunque no entendieras nada de mi memoria) y ser un apoyo fundamental en este proceso.

Gracias compañeros y amigos de la universidad por todos los buenos momentos vividos durante esta etapa. En especial, gracias Pache, Vicky, Tommy, Emily y Pelao por toda la amistad y el apoyo entregado durante estos años. ¡Grande Los Tacas!

Gracias compañeros y amigos que encontré en la empresa donde hice mi trabajo de título. Son un increíble equipo, atentos, dispuestos a ayudar y siempre con una sonrisa. Aprendí mucho de cada uno de ustedes. En especial, gracias Cris por ser un excelente guía en este trabajo; gracias Juan, Diego, Cris A. y Álvaro, excelentes *partners* de memoria; gracias Jose, Connie y JP por todas las oportunidades, enseñanzas y el apoyo brindado.

Gracias profesores de la sección de Marketing Cuantitativo por guiarme durante este proyecto, por sus críticas constructivas y enseñanzas de cara a la vida laboral. Gracias al resto de profesores de la universidad y a todo el personal de la Escuela de Ingeniería.

¡Gracias totales!

Sebastián Alfredo Moreno Haselbauer

TABLA DE CONTENIDO

I.	INTRODUCCIÓN	1
1.1	Motivación	1
1.2	Empresa.....	2
1.3	Programa de fidelización.....	4
1.4	Justificación del proyecto	5
II.	OBJETIVOS	9
2.1	Objetivo general.....	9
2.2	Objetivos específicos	9
III.	MARCO TEÓRICO	10
3.1	Valor del cliente	10
3.1.1	Customer Lifetime Value	11
3.2	Causalidad	14
3.2.1	Métodos de <i>matching</i>	15
3.2.2	Variables instrumentales	17
3.3	Escalado de variables	18
3.3.1	Escalado por rango	18
3.3.2	Normalización	18
IV.	METODOLOGÍA.....	19
4.1	Selección de datos.....	19
4.2	Preprocesamiento.....	20
4.3	Transformación	20
4.4	Data Mining.....	21
4.4.1	Modelamiento del Customer Lifetime Value	21
4.4.2	Estimación del efecto causal	24
4.4.3	Evaluación e interpretación	27
4.4.4	Recomendaciones.....	28
V.	ALCANCES Y RESULTADOS ESPERADOS	29
VI.	DESARROLLO METODOLÓGICO.....	30
6.1	Preparación de los datos	30
6.1.1	Selección de los datos.....	30
6.1.2	Transformación de variables.....	31
6.1.3	Valores faltantes o erróneos	33
6.1.4	Outliers.....	34
6.2	Modelamiento del Customer Lifetime Value.....	35
6.2.1	Análisis descriptivos rentabilidad	35
6.2.2	Modelo de rentabilidad	39
6.2.3	Resultados modelo de rentabilidad.....	42
6.2.4	Análisis descriptivos retención	46
6.2.5	Modelo de retención	48
6.2.6	Resultados modelo de retención	50
6.2.7	Resultados CLV.....	54
6.3	Estimación de efecto causal	57
6.3.1	Definición de activos.....	57
6.3.2	Efecto promedio: estratificación tradicional.....	58

6.3.3	Propensity Score	59
6.3.4	Matching.....	61
6.3.5	Efecto promedio	65
6.3.6	Segmentación	68
6.3.7	Efecto por segmento	76
6.4	Recomendaciones	83
6.4.1	Cálculo del potencial económico	83
6.4.2	Acerca de la estructura del programa	84
6.4.3	Acerca de la gestión por segmentos.....	85
VII.	CONCLUSIONES	89
7.1	Sobre el Customer Lifetime Value.....	89
7.2	Sobre el efecto del programa de fidelización	92
7.3	Limitaciones y trabajos futuros.....	94
VIII.	BIBLIOGRAFÍA.....	96
IX.	ANEXOS.....	98
9.1	Anexos A: Variables.....	98
9.2	Anexos B: Modelo de retención	100
9.3	Anexos C: Modelo de propensión	102
9.4	Anexos D: Matching.....	107
9.5	Anexos E: Segmentos.....	112

ÍNDICE DE TABLAS

Tabla 1. Detalle de categorías y sus diferencias	3
Tabla 2. Tabla de niveles de canje y equivalencia a pesos.....	4
Tabla 3. Ejemplos de variables generadas	32
Tabla 4. Descriptivos rentabilidad a 3 años (2015-2017)	35
Tabla 5. Rentabilidad a 3 años (2015-2017) por categoría (2014).....	37
Tabla 6. Rentabilidad a 3 años (2015-2017) por consumo de productos financieros (2014)	38
Tabla 7. Desempeño de modelos de rentabilidad	39
Tabla 8. Variables modelo de rentabilidad	40
Tabla 9. Descriptivos pronósticos de rentabilidad a 3 años (2018-2020)	42
Tabla 10. Diferencia de rentabilidad a 3 años (2018-2020) según gasto en comercio externo (2017)	43
Tabla 11. Pronóstico de rentabilidad a 3 años (2018-2020) por Categoría	44
Tabla 12. Pronóstico de rentabilidad a 3 años (2018-2020) por Productos financieros .	45
Tabla 13. Pronóstico de rentabilidad a 3 años (2018-2020) por Máxima cuota	45
Tabla 14. Variables modelo de retención	49
Tabla 15. Desempeño modelos de retención (Criterio AIC)	49
Tabla 16. Caracterización perfiles de mayor y menor duración.....	52
Tabla 17. Probabilidades de supervivencia abierto por variables relevantes	53
Tabla 18. Descriptivos Customer Lifetime Value (2018-2028)	54
Tabla 19. Descriptivos CLV (2018-2028) por Categoría (2017)	55
Tabla 20. Descriptivos CLV (2018-2028) por Canje (2017).....	56
Tabla 21. Resultados con muestra estratificada.....	59
Tabla 22. Variables más relevantes (81%) para modelo de propensión	60
Tabla 23. Resumen diferencias de medias estandarizadas según modelo de propensión	61
Tabla 24. Sensibilidad según K vecinos más cercanos.....	62
Tabla 25. Diferencias de covariables pre y post matching (14 variables más relevantes)	64
Tabla 26. Resultados pre y post matching	66
Tabla 27. Efectos estimados con estratificación versus Propensity Score Matching	67
Tabla 28. Elección de método de segmentación y número de segmentos	68
Tabla 29. Quintiles de propensión.....	69
Tabla 30. Diferencias de medias estandarizadas por segmento	70
Tabla 31. Descriptivos de activos por segmento (variables diferenciadoras modelo de propensión)	71
Tabla 32. Promedios de activos por segmento (otras variables).....	72
Tabla 33. Resultados segmento 1	77
Tabla 34. Resultados segmento 2	78
Tabla 35. Resultados segmento 3.....	79
Tabla 36. Resultados segmento 4	80
Tabla 37. Resultados segmento 5.....	81
Tabla 38. Potencial económico de activar a los clientes en el club de fidelización	84
Tabla 39. Propensión a activarse en el club de puntos por segmento (2017)	85
Tabla 40. Beneficio marginal por gestionar por segmento	88

ÍNDICE DE ILUSTRACIONES

Ilustración 1. Tarjetas de débito y crédito* vigentes por año (a febrero)	1
Ilustración 2. Participación por tarjetas de crédito vigentes (2018-03)	1
Ilustración 3. Distribución de la cartera según GSE	3
Ilustración 4. Dispersión entre costo del programa y rentabilidad (2016).....	5
Ilustración 5. Inversión promedio en puntos y rentabilidad promedio por quintiles de rentabilidad.....	6
Ilustración 6. Inversión promedio en puntos y rentabilidad promedio por categoría.....	6
Ilustración 7. Flujo del proceso KDD	19
Ilustración 8. Metodología estimación incremental del valor asociado a puntos	21
Ilustración 9. Periodos de observación para estudiar causalidad.....	24
Ilustración 10. Proceso iterativo matching.....	25
Ilustración 11. Ilustración matching y segmentación	27
Ilustración 12. Detección de outliers con distancia de Mahalanobis.....	34
Ilustración 13. Rentabilidad a 3 años (2015-2017)	35
Ilustración 14. Rentabilidad a 3 años (2015-2017) versus rentabilidad anterior (2014)..	36
Ilustración 15. Box-plot de rentabilidad a 3 años (2015-2017) por categoría (2014)	37
Ilustración 16. Rentabilidad a 3 años (2015-2017) por de productos financieros.....	38
Ilustración 17. Rentabilidad a 3 años (2015-2017) por puntos canjeados y tasa de canje (2014).....	39
Ilustración 18. Pronóstico de rentabilidad a 3 años (2018-2020).....	42
Ilustración 19. Pronóstico de rentabilidad a 3 años (2018-2020) por Categoría (2017) ..	44
Ilustración 20. Pronóstico de rentabilidad a 3 años (2018-2020) por Tiene canje (2017)	46
Ilustración 21. Curva de retención (Kaplan-Meier) cliente promedio	47
Ilustración 22. Curva de retención (Kaplan-Meier) por categoría	48
Ilustración 23. Curva de retención (Kaplan-Meier) por canjes último año	48
Ilustración 24. Histograma de duración cartera 2017-12.....	51
Ilustración 25. Distribución del Customer Lifetime Value (2018-2028)	54
Ilustración 26. Distribución CLV (2018-2028) por Categoría (2017)	55
Ilustración 27. Distribución CLV (2018-2028) por Canje (2017)	56
Ilustración 28. Criterio de actividad club de puntos (2015-2016).....	58
Ilustración 29. Histograma Propensity Score sin y con matching.....	63
Ilustración 30. Efecto en rentabilidad 2017 y CLV 2018-2028 de activarse en el club de puntos	76

I. INTRODUCCIÓN

1.1 Motivación

Las tarjetas como medio de pago se han masificado enormemente en los últimos años. En febrero de 2010 la cantidad de tarjetas vigentes y afiliadas a la red Transbank eran aproximadamente 12,9 millones (sin considerar las de uso exclusivo en casas comerciales). En los siguientes 7 años este número se triplicó alcanzando 38,5 millones de tarjetas. En particular, las de crédito pasaron de ser 4,6 millones a 17,2 millones, con un crecimiento del 272%. Por lo que hoy son un medio de pago muy común en la población chilena.



Ilustración 1. Tarjetas de débito y crédito* vigentes por año (a febrero).

*Tarjetas de crédito de uso universal, no considera de uso exclusivo en casas comerciales

Fuente: SBIF

Existen 20 emisoras de tarjetas de crédito en el país. De estas, 8 concentran el 86% de las tarjetas vigentes:

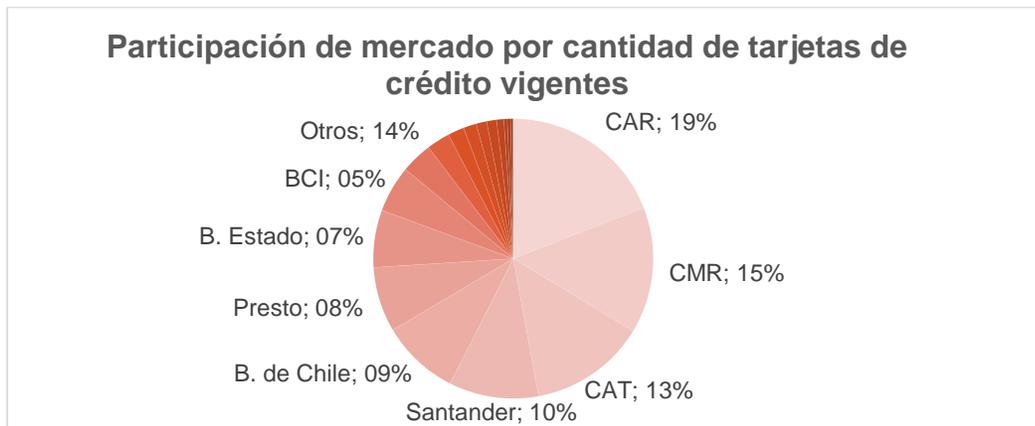


Ilustración 2. Participación por tarjetas de crédito vigentes (2018-03)

Fuente: SBIF

Resultado de lo anterior, actualmente es muy común que los consumidores posean en sus billeteras más de una tarjeta de débito, crédito y de distintas emisoras. Este escenario altamente competitivo ha derivado en una disputa constante entre las empresas por rentabilizar a sus clientes a partir de desarrollar y mantener relaciones de largo plazo. Como consecuencia, las empresas ofrecen beneficios atractivos a sus clientes y los clubes de fidelización¹ son un ejemplo común de esto.

Existen dos pilares para focalizar los programas de lealtad: (i) Maximizar frecuencia/gasto de los clientes y/o aumentar el número de productos que consumen; o (ii) aumentar la permanencia y mantener la base de clientes. El foco que se elija depende del negocio y de la estrategia deseada, aunque se puede abordar ambos focos simultáneamente. De cualquier forma, se estará aumentando el potencial de ingresos de la compañía, además de proporcionar otros beneficios indirectamente: imagen de marca, recolección de datos, etc. (Uncles, 2003)

Las emisoras de tarjetas de crédito pueden llegar a invertir bastante dinero en sus clubes de fidelización. El retorno esperado es una cartera de clientes leales. Pero normalmente la utilidad de esta inversión es incierta, principalmente porque medir el impacto de un programa de lealtad no es directo. Esto ocurre en la empresa en donde se enmarca este trabajo. El 9% de los costos totales son destinados al programa y no se tiene una noción de cuál es su impacto en términos económicos. En efecto, al no medir, se pierde conocimiento y oportunidades de llevar a cabo una mejor gestión.

1.2 Empresa

La empresa es una institución financiera no bancaria encargada de administrar la tarjeta de crédito de un *holding* chileno. El grupo posee negocios en el sector financiero y de *retail*. La presencia en varios países de Latinoamérica lo han posicionado como un importante conglomerado a nivel regional.

La administradora de la tarjeta (en adelante “la empresa”) nace con su principal producto, la tarjeta de crédito. Esta institución no bancaria ha ido incorporando nuevos productos a lo largo de su historia: créditos de consumo, avances en efectivo y pagos automáticos, ampliando así su oferta.

¹ Los clubes de fidelización (o programas de fidelización) premian a los consumidores con productos o servicios a modo de recompensa por su lealtad hacia la empresa. Cada vez que el cliente utiliza los servicios de la empresa acumula puntos que luego puede canjear por premios. De esta forma las empresas incentivan a los clientes a preferirlos frente a la competencia.

Respecto a la tarjeta de crédito en sí, la empresa ofrece distintos tipos. La primera históricamente fue de uso exclusivo en la casa comercial y comercios asociados. Luego, se incorporaron las tarjetas Visa y MasterCard, abriéndose la posibilidad de pagar en cualquier comercio del mercado. Este hito marcó un nuevo rumbo para la empresa, la cual actualmente tiene una cartera de 2,8 millones de clientes.

Los clientes de la empresa se concentran mayormente en los segmentos socioeconómicos medios de la población. Más de la mitad de la cartera proviene del segmento D (53%), seguido por el C3 (23%). Correspondiendo el cliente promedio a un perfil que históricamente ha tenido pocas oportunidades en la banca tradicional.

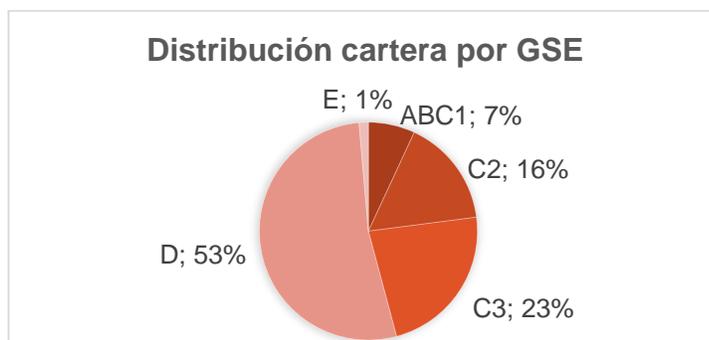


Ilustración 3. Distribución de la cartera según GSE
Fuente: Elaboración propia con datos de la empresa

La empresa separa a sus clientes en tres categorías: Categoría Normal, Premium y Elite. Los clientes pueden ascender de categoría al pagar mayor costo de mantención o en forma de premio, como un reconocimiento por su alto gasto con la tarjeta. Mientras más exclusiva es la tarjeta, mejores beneficios posee.

	Normal (85,4%)	Premium (11,4%)	Elite (3,1%)
Costo admin.	UF 0,105	UF 0,17	UF 0,35
Admin. gratis (ascenso categoría)		Al acumular 15.000 puntos en 12 meses	Al acumular 50.000 puntos en 12 meses
Duración puntos	1 año	2 años	2 años
Otros beneficios	Seguro accidentes	Mayor cobertura	La mejor cobertura
		Doble acumulación mes de cumpleaños	Doble acumulación mes de cumpleaños
		Otros	Otros

Tabla 1. Detalle de categorías y sus diferencias
FUENTE: Sitio web de la empresa. <Citado el 19/11/2017>

La cartera se compone principalmente de clientes categoría normal (85,4%), seguido por los Premium (11,4%) y Elite (3,1%). Esto es coherente con la distribución socioeconómica, ya que como es de esperar, estas categorías corresponden a clientes con mayor potencial de gasto.

1.3 Programa de fidelización

El programa de puntos es uno de los pilares fundamentales del marketing de la empresa, utilizándose como un atractivo importante para fidelizar a los clientes. En el 2016 significó una inversión de 24 mil millones de pesos, representando el 9% de los costos totales de la organización.² En lo siguiente, se detallan los antecedentes necesarios para comprender cómo funciona el club de puntos y sus condiciones.

El flujo de los puntos inicia en la acumulación, cada vez que el cliente utiliza su tarjeta de crédito acumula puntos en su cuenta. Por cada \$300 que el cliente gasta con la tarjeta acumula 1 punto y si está comprando dentro del *holding* acumula el doble. De esta forma, se incentiva a los clientes a preferir los negocios del grupo.

Para canjear, los clientes deben llegar a niveles de canje los que pueden convertir sus puntos a giftcard (para uso en el *holding*) o canjear directamente productos o experiencias en la página web, en donde, existen oportunidades de canjes desde 1.000 puntos. La equivalencia entre puntos y pesos del premio es creciente, a mayor nivel de canje mayor retorno marginal para el cliente.

Nivel (puntos)	Monto premio (\$)	Tasa equivalencia (\$/punto)
5.000	8.000	1,60
9.000	15.000	1,67
12.000	20.000	1,67
18.000	35.000	1,94
24.000	70.000	2,92
36.000	120.000	3,33
48.000	200.000	4,17
60.000	250.000	4,17
90.000	450.000	5,00
120.000	700.000	5,83
240.000	1.500.000	6,25

Tabla 2. Tabla de niveles de canje y equivalencia a pesos
FUENTE: Sitio web de la empresa. <Citado el 19/11/2017>

Por último, los puntos poseen una duración determinada desde su fecha de acumulación, y si no son canjeados vencen. Como se exhibe en la Tabla 1. Detalle de categorías y sus diferencias, esta duración es de 1 año para la categoría Normal y de 2 años para los

² Fuente: Estado de Resultados 2016

Premium y Elite. Como beneficio adicional, estos últimos acumulan doble puntaje durante su mes de cumpleaños.

Los incentivos están establecidos para motivar a los clientes a aumentar su gasto con la tarjeta. Además de los canjes, se les premia con ascensos a categorías con mejores beneficios y costos de administración gratis. Así la empresa desarrolla la lealtad de los clientes para que utilicen la tarjeta como su medio de pago principal.

1.4 Justificación del proyecto

Dada la descripción del programa de lealtad en la sección anterior, parece ser que los más beneficiados con el club son los clientes de mayor potencial de gasto. Esto se confirma en la memoria de (Gallardo, 2016), quien identificó que los clientes de alta interacción con el programa de puntos corresponden a un perfil con alto cupo, con cuentas adicionales y que utiliza los canales digitales. Pero esto no necesariamente significa clientes rentables.

Al graficar la dispersión entre rentabilidad y costo de fidelización por cliente para el año 2016 se puede observar que existe una correlación levemente negativa. Evidentemente esta es una observación preliminar de la que no se puede concluir nada concreto. Sin embargo, es una primera señal que motiva a estudiar en mayor profundidad el valor que brinda el programa de fidelización para la empresa.

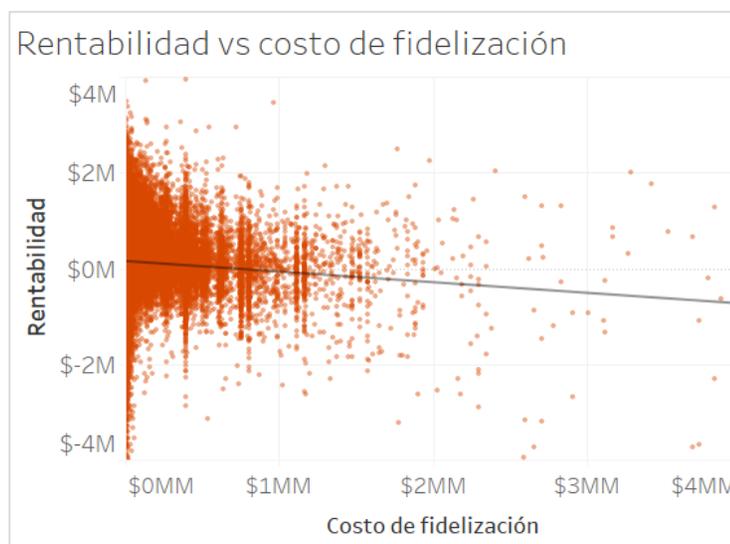
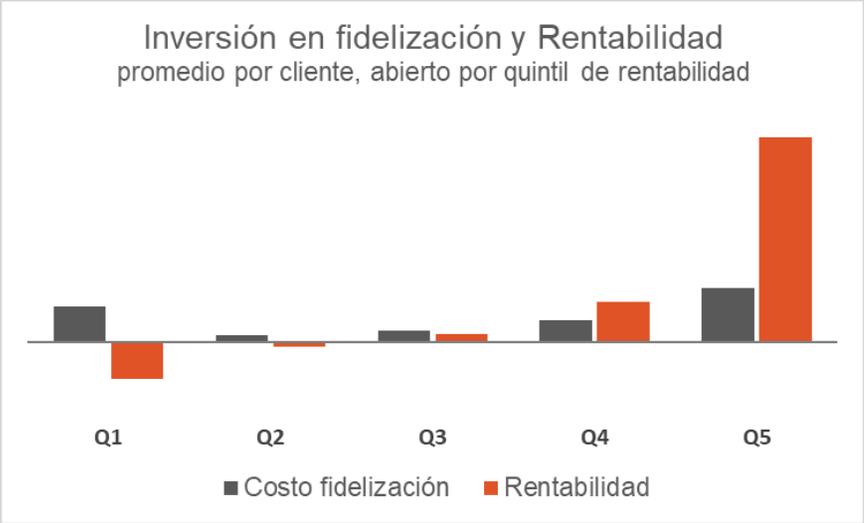


Ilustración 4. Dispersión entre costo del programa y rentabilidad (2016)

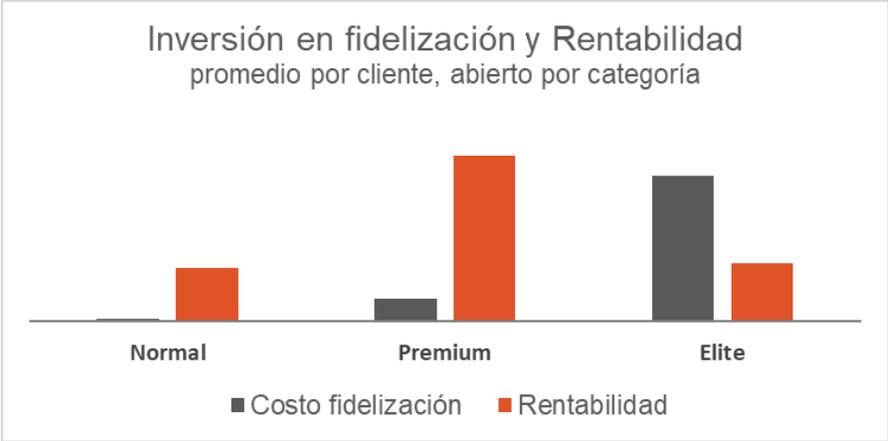
En la Ilustración 5 se presentan los costos promedios en fidelización y la rentabilidad promedio por cliente abierto por quintiles de rentabilidad, en donde los quintiles 1 y 5 son los de menor y mayor rentabilidad respectivamente. El costo y la rentabilidad están en diferentes escalas para facilitar la visualización del gráfico, como también por confidencialidad. Lo importante de este gráfico es analizar la relación entre el costo de

fidelizaci3n y la rentabilidad de los clientes. Los quintiles 4 y 5 se llevan parte importante de la inversi3n en fidelizaci3n y, a su vez, son muy rentables para la empresa, lo cual es bueno para el negocio. Por otro lado, existe un 20% de la cartera que tambi3n significa gran parte del costo del programa y, en cambio, no son rentables para la empresa. Entonces, surge la pregunta de si se est3n enfocando correctamente los recursos del programa en los clientes de mayor valor para la empresa o no.



Ilustraci3n 5. Inversi3n promedio en puntos y rentabilidad promedio por quintiles de rentabilidad
 Fuente: datos de la empresa 2016
 (Costo y rentabilidad en distintas escalas para visualizaci3n y confidencialidad)

Un caso particular de lo anterior es analizar el costo promedio en fidelizaci3n y la rentabilidad promedio por categor3a de clientes, ya que como se ha explicado anteriormente, las dos categor3as m3s exclusivas poseen beneficios diferenciadores en el programa de puntos.



Ilustraci3n 6. Inversi3n promedio en puntos y rentabilidad promedio por categor3a
 Fuente: datos de la empresa 2016
 (Costo y rentabilidad en distintas escalas para visualizaci3n y confidencialidad)

Se observa nuevamente algo alarmante. Se invierte desproporcionadamente en la categoría Elite y su rentabilidad no lo justifica, más bien, es relativamente similar a la rentabilidad promedio de la categoría Normal. Se reafirma entonces la necesidad de evaluar el foco del programa y su retorno asociado.

Lo anterior es preocupante para la empresa, sobre todo porque el programa es una importante inversión (24.000 MM\$, mencionado anteriormente en Programa de fidelización) y un pilar fundamental del marketing de la empresa. Si se desconoce cuál es el retorno que brinda a la organización, se pierden oportunidades de una mejor gestión. Según lo observado en los gráficos anteriores existe una inversión desproporcionada en clientes de baja rentabilidad, pero ¿y si son clientes fidelizados que se mantendrán en el largo plazo? Quizás esa baja rentabilidad subsidia una relación de largo plazo, ¿y qué hay del quintil de rentabilidad negativa? ¿Por qué se invierte tanto en fidelización en ellos? Surgen varias dudas respecto a la efectividad que tiene el club de puntos y si está diseñado en función de los clientes de mayor valor para la empresa.

Se plantea la necesidad de evaluar cuál es el valor que aporta el club, qué efecto tiene en el comportamiento de los clientes y a quién debe estar enfocado. El presente trabajo pretende hacerse cargo de esta necesidad a través de la evaluación del retorno asociado al programa de fidelización mediante la estimación de su efecto en el valor de los clientes para la empresa. De esta forma, se busca evaluar en función de su objetivo, que es fidelizar a los clientes para rentabilizarlos en el largo plazo.

En la literatura existen bastantes estudios acerca del efecto asociado a metas de canje y programas de puntos en el comportamiento de corto plazo de los clientes (Carboni, 2012) (Ríos, 2013) (Troncoso, 2016), analizándose principalmente aceleraciones en la frecuencia de compra y en el gasto. Sin embargo, existe un vacío acerca del impacto de estos en el largo plazo. Desde esta perspectiva surge una oportunidad para el negocio, medir el valor que genera el club de fidelización a nivel de la relación cliente-empresa en el largo plazo. Así mismo, se puede evaluar el programa acorde a su objetivo de fidelización desarrollando a la vez una noción de cuál es su retorno monetario asociado. Esto representa un gran desafío, ya que como actividad de marketing masivo que busca desarrollar lealtad, los resultados no son observables ni medibles directamente. Se deben realizar estimaciones con métodos especializados en la reducción de sesgos.

El desarrollar este conocimiento para la empresa define una base sobre la cual trabajar para gestionar el valor de la cartera a través del programa de puntos. Conocer cuánto valor aporta a la organización el programa de puntos a nivel de segmentos, por ejemplo, puede apoyar la gestión de los clientes a partir de desarrollar su lealtad, se puede optimizar el uso de recursos del programa según el efecto esperado en el comportamiento de cada segmento y/o también se puede llegar a replantear su estructura misma o de los beneficios asociados a los distintos tipos de clientes. En fin, cuantificar el retorno asociado significa construir los cimientos para hallar y trabajar por oportunidades que mejoren la gestión del programa de puntos y la fidelización de los clientes.

La contraparte de este proyecto es la Gerencia de Business Intelligence. Esta se encarga de estudiar el comportamiento de los clientes y sus preferencias en base a los datos con el fin de generar conocimiento entendible y gestionable por las demás áreas para, así, mejorar los resultados de la organización. Por tanto, este proyecto es relevante y valioso para la organización y esta gerencia en particular, dado que calza con sus herramientas y objetivos planteados.

En el siguiente apartado, capítulo II, se definirá el objetivo general y los objetivos específicos que establecerán la línea de trabajo del proyecto. En el capítulo III se presentará el marco teórico con los conceptos y métodos existentes en la literatura que se pueden utilizar para abordar el problema. En el IV se presentará la metodología elegida para el desarrollo del trabajo en base a los capítulos anteriores. En V se presentarán los alcances y resultados esperados del trabajo. En el capítulo VI se exhibirá el desarrollo metodológico y los resultados obtenidos, junto con recomendaciones tácticas para potenciar el valor de los clientes. Finalmente, en el capítulo VII se exhiben las conclusiones del trabajo, junto con una breve reflexión sobre las limitaciones del proyecto y propuestas para trabajos futuros.

II. OBJETIVOS

2.1 Objetivo general

Evaluar el retorno económico asociado al programa de fidelización de un *retail* financiero a través de la estimación del efecto causal que provoca el club de puntos en la relación del cliente con la empresa en el largo plazo, para así identificar y cuantificar oportunidades que potencien el valor de la cartera.

2.2 Objetivos específicos

- Estimar el valor de los clientes para la empresa a través del modelamiento del Customer Lifetime Value a nivel de individuo.
- Evaluar el impacto causal del programa de fidelización en el valor de los clientes a nivel de segmentos (nueva segmentación en base a propensión a activarse en el club de puntos) utilizando métodos de *matching* para minimizar sesgos de autoselección.
- Realizar recomendaciones tácticas para incrementar el valor de la cartera, utilizando como base el potencial económico de activar a los distintos segmentos en el programa de lealtad.

III. MARCO TEÓRICO

El presente capítulo pretende introducir los conceptos relevantes para el desarrollo del trabajo y examinar los distintos métodos presentes en la literatura que se plantean como opciones para abordar el problema. En vista del primer objetivo, en 3.1 se presentan enfoques utilizados en trabajos anteriores como alternativas propuestas para medir el valor del cliente; en 3.2 se presentan opciones disponibles en la literatura para estudiar causalidad y abordar el segundo objetivo, estimar el efecto que causa el programa de fidelización en los clientes; finalmente, en 3.3 se detallan transformaciones de variables que pueden aportar al modelamiento.

3.1 Valor del cliente

En memorias anteriores se han estudiado los efectos de los programas de lealtad, estimando efectos en gasto, frecuencia de compra y número de productos contratados (Ares, 2008) (Carboni, 2012) (Ríos, 2013) (Troncoso, 2016). Como resultado general, se observan aumentos en el gasto y aceleración en la frecuencia de compra al estar próximos al canje, luego del canje los clientes reinician su patrón de compra junto con un aumento en la propensión de consumir nuevos productos con la empresa. Estos estudios se han enfocado principalmente en estudiar el corto plazo, pre y post canje. A diferencia de estos trabajos, este proyecto busca evaluar el impacto del programa en consistencia con su objetivo, fidelizar a los clientes y rentabilizarlos en el largo plazo. Por lo tanto, se propone estudiar el impacto en un periodo prolongado y utilizando un criterio de actividad en el club de puntos más riguroso que un único canje, ya que no necesariamente implica que el cliente haya actuado bajo la influencia del programa de fidelización.

Como variables a medir, se proponen dos presentes en la literatura: Customer Profitability (Rentabilidad) y Customer Lifetime Value (CLV). Ambas son ampliamente recomendadas en la literatura para la toma de decisiones y evaluación de actividades de marketing, ya que son métricas monetarias directamente relacionadas con los resultados de la empresa. (Pickering, 1999) describe ambas métricas y su diferencia de la siguiente manera:

- La rentabilidad por cliente es un dato que consolida ganancias y pérdidas producidas por este en un periodo único, por ejemplo, en un año. Permite la gestión comparando entre clientes o periodos anteriores.
- El CLV, en cambio, es una estimación de las rentabilidades futuras que brindará el cliente durante el tiempo que dure su relación con la empresa. Corresponde al valor presente del cliente y se basa en modelos predictivos a partir de la data histórica.
- Un cliente que actualmente tiene rentabilidad negativa puede ser muy valioso para la empresa. (Pickering, 1999) ilustra esto mediante el siguiente ejemplo: un joven profesional que no es rentable hoy para la empresa es probable que en el futuro consuma créditos para comprar su auto, casa o financiar sus inversiones y se convierta en un cliente muy rentable. En fin, los comportamientos de compra y

rentabilidad de los clientes cambian en el tiempo y, a diferencia del Customer Profitability, el CLV se hace cargo de esto y considera al cliente como un activo.

Para este trabajo se utilizarán ambas métricas, la rentabilidad por cliente y el CLV. La primera corresponde a un dato de la empresa y la segunda se debe modelar en base a rentabilidades pasadas. Medir el impacto del programa de puntos a través de estas métricas es una novedosa y potente forma de evaluar cuál es su retorno monetario asociado, más allá del cambio en frecuencia y gasto de los clientes que ya ha sido bastante estudiado y, en el negocio de las tarjetas de crédito, no necesariamente implica mayor rentabilidad.

3.1.1 Customer Lifetime Value

El CLV es el valor presente de las rentabilidades del cliente y se formula así³:

$$CLV_i = \sum_{t=0}^T \frac{\pi_i^t r_i^t}{(1+d)^t} \quad (1)$$

donde

π_i^t = utilidad de servir al cliente i en el periodo t

r_i^t = retención en el periodo t (probabilidad de que el cliente esté activo)

d = tasa de descuento o costo de capital de la firma

T = periodo de estimación del CLV

Para estimar el CLV, en la literatura existen diferentes métodos según el tipo de negocio. En este caso corresponde presentar los principales métodos para relaciones contractuales de cobro mensual.

i. Modelos RFM

Este método se basa en agrupar a los clientes en segmentos de valor según las variables *Recency* (tiempo desde la última compra), *Frequency* (cantidad de compras por periodo) y *Monetary Value* (monto gastado por el cliente). Se utilizan modelos de propensión para clasificar a los clientes dentro de estos grupos, de manera de predecir en qué segmento de valor estarán en el próximo periodo. Como ejemplo, si cada variable RFM se categoriza en 5 niveles, se tendrán 125 segmentos de valor como combinación de estos niveles (5x5x5). Este método se suele utilizar para la priorización de clientes en campañas de marketing.

³ Fuente: (Gupta S. L., 2003)

Este método es de fácil aplicación, pero presenta varias limitaciones: clasifica sólo para el siguiente periodo, no explica el comportamiento latente de los clientes y, al ser un clasificador, no pronostica el dinero futuro. (Fader P. H., 2005), (Kumar, 2006)

ii. Modelos probabilísticos

Se trata el comportamiento de los clientes como componentes separadas: frecuencia, gasto y tiempo de vida en la empresa. Utilizando este enfoque, se modelan las componentes bajo el supuesto de que siguen distribuciones conocidas. El método de Pareto/NBD (Schmittlein, 1987) es el más popular y se utiliza para relaciones no contractuales. Lo equivalente y recomendado para el caso de tarjetas de crédito sería utilizar el modelo BB/EG (Fader P. H., 2015), por su naturaleza contractual de facturación mensual. La probabilidad de compra en cada periodo se modela con una distribución Beta-Binomial, mientras que el gasto por ocasión de compra se modela con una Gamma-Gamma y la duración de vida (retención) del cliente se modela con una Gamma-Exponencial. Luego, los resultados de los modelos se combinan para consolidar la estimación los flujos futuros de cada cliente.

Se debe destacar que para modelar la duración existe censura en los datos, ya que para muchos clientes no se sabe aun cuándo terminará su relación con la empresa. Frente a esto, los modelos probabilísticos son los más atractivos dado que abordan esta censura y también permiten realizar estimaciones más allá del horizonte observado. La distribución Exponencial estima la probabilidad de abandono asumiendo que se mantiene constante. En cambio, Weibull asume que esta probabilidad depende del tiempo (Morrison, 1980).

iii. Modelos econométricos

Se modela la rentabilidad utilizando regresiones con variables explicativas, como, por ejemplo, una regresión lineal.

Los métodos anteriormente mencionados para modelar la retención de los clientes asumen que la relación con el cliente muere una vez que este se define como fugado. A diferencia de esto, (Pfeifer, 2000) propone utilizar cadenas de Márkov para modelar el comportamiento del cliente con transición entre estados de actividad e inactividad. Esto lo hace un método atractivo, pero no permite incluir heterogeneidad.

iv. Métodos de persistencia

Los métodos de persistencia también se enfocan en modelar los componentes del CLV por separado. Estos métodos están en la categoría de modelos dinámicos, series de tiempo multivariadas y modelos autorregresivos. Se utilizan principalmente para estudiar cómo afectan intervenciones de marketing y promociones en los distintos componentes

del valor del cliente. La limitante principal de estos métodos es que requieren de bastante historia para llegar a predecir en el largo plazo. (Gupta S. D., 2006)

v. Modelos de Machine learning

Los métodos de aprendizaje supervisado utilizan variables explicativas para pronosticar la rentabilidad futura de los clientes. Árboles, Random Forest o Redes neuronales, por ejemplo, son métodos de regresión populares en Machine Learning.

Para modelar la retención es posible utilizar modelos de propensión como Redes neuronales o Árboles de decisión. Aunque estos métodos no son recomendados cuando existe gran cantidad de periodos en los que puede ocurrir la fuga.

vi. Memorias anteriores

(Urzúa, 2007) estima el CLV para un supermercado mayorista y compara tres modelos: Pareto/NBD, árboles de regresión y MLP. Con información de un año estima los montos y la frecuencia futura para el siguiente año. Como conclusión, obtiene mejor precisión con los modelos de Machine Learning y, en particular, destaca el árbol de regresión por su fácil aplicación e interpretación.

(Reuse, 2011) desarrolla un CLV enfocado en predecir el monto de gasto en un año por los clientes del programa de lealtad de supermercado, considerando la retención de un 100%. Para esto compara los resultados obtenidos con regresiones lineales y con un modelo jerárquico bayesiano. En particular, destaca el último modelo con la incorporación de heterogeneidad y mejores resultados.

(Rodríguez, 2013) utiliza una regresión lineal para estimar la rentabilidad de 4 años de los clientes nuevos en un banco, utilizando variables sociodemográficas y transaccionales de los primeros 3 meses. Luego de estimar la rentabilidad, calcula una rentabilidad promedio por año para cada cliente. Por separado, pronostica la duración de los clientes con una regresión lineal. Finalmente, consolida la estimación del CLV al multiplicar ambos modelos.

vii. Métricas de desempeño

Para definir qué modelo se desempeña mejor, se presentan las siguientes métricas.

MAPE

Error absoluto porcentual promedio. Se utiliza normalmente para medir el error en pronósticos de variables continuas, en este caso para el modelo de rentabilidad.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{|Y_i|} \quad (3)$$

donde,

Y_i = Valor observado para cliente i

\hat{Y}_i = Valor pronosticado para cliente i

n = Cantidad de observaciones

WAPE

Error absoluto porcentual promedio ponderado. Es una modificación mejorada del MAPE, ya que soluciona los problemas que este último presenta: i) se indefine cuando el resultado real es 0 y ii) cuando los resultados reales son cercanos a 0 el error porcentual se vuelve muy volátil, distorsionando la magnitud del error promedio.

$$WAPE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{\sum_{i=1}^n |Y_i|} \quad (4)$$

donde,

Y_i = Valor observado para cliente i

\hat{Y}_i = Valor pronosticado para cliente i

n = Cantidad de observaciones

AIC

El criterio de información de Akaike (AIC) es una métrica de calidad de ajuste relativo a otros modelos. Permite comparar entre modelos probabilísticos, pero por sí solo no aporta información. En el CLV se utiliza para la elección del modelo de retención que consiste en probabilidades de supervivencia distribuidas por periodos.

$$AIC = 2k - 2 \ln(L) \quad (5)$$

donde,

k = cantidad de variables

L = Máxima verosimilitud

3.2 Causalidad

Causalidad se refiere a la relación entre causa y efecto producida por algún acontecimiento, proceso o fenómeno. En el contexto del trabajo, por ejemplo, si un cliente canjea sus puntos para obtener alguna recompensa existe la posibilidad de que este consumidor cambie su comportamiento posterior como consecuencia de esta vivencia. En este caso, el canje corresponde al acontecimiento que causa un efecto en su gasto y frecuencia, por ejemplo. Entonces, el fin de esta sección es presentar herramientas que permiten estudiar la existencia de efectos causales y cuantificar los impactos asociados.

Para inferir causalidad, lo más recomendado es realizar experimentos aleatorios. Sin embargo, en este caso no es factible. Quitar la participación en el programa a ciertos clientes puede significar problemas legales, éticos y reputacionales para la marca. Esto, junto con la gran cantidad de data histórica que posee la empresa son razones suficientes para descartar la opción experimental y optar por el estudio observacional.

La complejidad del enfoque observacional es que, al ser un experimento natural, se puede observar el comportamiento de un cliente en tratamiento, pero jamás se sabrá cómo se hubiese comportado el mismo cliente sin el tratamiento. Comparar clientes tratados y no-tratados directamente puede estar gravemente sesgado por la autoselección. A continuación, se describen métodos existentes en la literatura que permiten abordar este problema minimizando estos sesgos.

3.2.1 Métodos de *matching*

La técnica de *matching* se basa en el supuesto de independencia condicional. En donde, un individuo A en una situación de control C y el mismo individuo en una situación experimental T se comportan de manera diferente únicamente por efecto del experimento T. Como observar ambos escenarios no es posible, se identifican clones (individuos similares) con el fin de recrear la aleatorización del experimento y generar dos grupos comparables – tratamiento y control. (Stuart, Matching methods for causal inference: A review and a look forward, 2010).

Para la estimación se debe tener en cuenta que el objetivo del estudio es estimar el efecto promedio del tratamiento en los tratados (ET) y no el efecto promedio del tratamiento (EP). La diferencia recae en que el EP corresponde al efecto a nivel poblacional, cuando toda la población participa del tratamiento; mientras que el ET es el efecto promedio en los sujetos que participan del tratamiento y, por lo tanto, se puede considerar como efecto potencial para clientes similares. El ET se estima comparando clientes tratados con sus clones de control, mientras que estimar el EP es más complejo porque hay individuos sin clones y se debe extrapolar su potencial comportamiento.

Implementar este método involucra 4 pasos (Stuart, Matching methods for causal inference: A review and a look forward, 2010):

- Definir métrica de similitud para encontrar sujetos parecidos.
- Implementar algoritmo de *matching* en base a la métrica escogida.
- Revisar la calidad de la muestra emparejada e iterar en los pasos anteriores hasta alcanzar un balance adecuado.
- Analizar las diferencias en la variable resultado entre sujetos de tratamiento y control emparejados, e inferir el efecto causal del tratamiento.

En lo siguiente se describen alternativas para ejecutar estos 4 pasos recién mencionados, considerando que el objetivo es estimar el efecto promedio en los tratados.

i. Similitud

Una forma de medir similitud entre dos individuos es con la distancia de Mahalanobis, esta mide la distancia de los vectores de sus covariables en un valor escalar. Funciona bien con pocas covariables, menos de 8 (Rubin D. , 1979) (Zhao, 2004), pero no tan bien cuando las covariables no distribuyen normal o son muchas (Gu, 1993).

(Rubin D. R., 1983) propone reducir el número de covariables a una única dimensión, el puntaje de propensión a participar del tratamiento (en inglés Propensity Score). Para esta propensión se puede utilizar cualquier modelo de clasificación. Finalmente, la distancia entre dos individuos se define como la diferencia entre sus puntajes de propensión. Este método es muy utilizado en la literatura, especialmente en medicina y políticas públicas.

ii. Matching

Definida la métrica de similitud entre clientes, el siguiente paso es emparejar. Es decir, buscar para cada cliente de tratamiento su cliente más similar entre los candidatos de control. El método más popular es utilizar vecinos más cercanos. La forma más simple es emparejar 1:1 en donde a cada sujeto tratado se le busca un sujeto de control. También se puede emparejar k:1, con k clientes de control por tratado. En adición, se puede exigir una similitud mínima entre clientes para evitar que se asignen clientes muy distantes, esto se conoce como radio de Caliper. Finalmente, se puede considerar la opción de reemplazo, que permite que cada sujeto de control pueda ser emparejado con más de un sujeto tratado.

La selección del número de vecinos depende de la razón entre candidatos de control sobre los tratados y también de los resultados posteriores, ya sea de la calidad del emparejamiento (más adelante) como también del porcentaje de tratados que encontraron sus vecinos y, por lo tanto, serán utilizados para el análisis. Por otro lado, (Austin, 2011) estudia el radio de Caliper óptimo y recomienda utilizar el correspondiente a 0,2 veces la desviación estándar del logit de la propensión, ya que minimiza el sesgo. Por último, en cuanto a la opción de reemplazo, esta no se recomienda a menos que la razón entre candidatos de control sobre tratados sea baja, por ejemplo, 2:1 o 1:1.

iii. Calidad del emparejamiento

Para saber si el emparejamiento es adecuado, se debe revisar el balance de las covariables entre los sujetos emparejados a nivel muestral. Las distribuciones para cada variable deben ser similares. Si esto no es así, entonces se debe iterar en el modelo de propensión y/o el método de *matching* hasta obtener un mejor balance.

Es útil realizar una revisión gráfica de las variables, ya que permite una visualización rápida sobre el balance. Luego, se debe revisar la diferencia absoluta de medias estandarizada para el puntaje de propensión y las covariables. Se recomienda que esta diferencia esté bajo 0,1, con un máximo de 0,25. (Stuart, Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research, 2013).

Diferencia de medias estandarizada

Esta métrica permite evaluar el balance de las covariables, entre el grupo de tratamiento y control, de manera estándar para todas las variables.

$$SMD = \frac{|\bar{X}_c - \bar{X}_t|}{\sigma_{c,t}} \quad (6)$$

donde,

\bar{X}_c = promedio grupo control

\bar{X}_t = promedio grupo tratamiento

$$\sigma_{c,t} = \text{desviación estándar combinada} = \sqrt{\frac{(\sigma_c^2 + \sigma_t^2)}{2}}$$

La extensión de la diferencia de medias estandarizada para variables binarias y categóricas se encuentra en (Yang, 2012).

iv. Estimación del efecto

Una vez verificado el balance de las covariables, los clientes son comparables y se puede inferir causalidad. Para afirmar la existencia de un efecto, se debe comprobar una diferencia significativa entre los grupos.

3.2.2 Variables instrumentales

Las variables instrumentales es un método de regresión lineal que surge como alternativa a los métodos de *matching* para solucionar la endogeneidad. Realizar una estimación consistente e insesgada se puede abordar por medio de la incorporación de instrumentos que cumplan supuestos de: validez (correlación nula con el error) y relevancia (alta correlación con el tratamiento). Luego, para obtener los estimadores se recrea una cadena de causalidad utilizando el método de mínimos cuadrados de dos etapas.

Este método se usa normalmente para modelar fenómenos económicos y sociales. El problema es que posee una gran dificultad para levantar los instrumentos, ya que deben cumplir los supuestos de manera estricta, siendo el de exclusión el más complejo, porque

no es testeable y debe ser fundamentado con razonamiento teórico. Por otro lado, para su aplicación en este trabajo es probable que se tenga que recolectar las variables instrumentales para cada cliente. (Ríos, 2013) propone levantar variables de preferencia y gustos sobre los clubes de fidelización mediante encuestas. Recolectar estas variables es costoso y poco factible para este trabajo.

3.3 Escalado de variables

Para que los modelos funcionen mejor y para facilitar la interpretación de las variables, cuando se tienen distintas escalas es útil escalar las covariables continuas. Esto significa comprimir sus escalas en un rango definido. A continuación, se presentan los dos tipos de escalados más comunes en la literatura.

3.3.1 Escalado por rango

También se conoce como Escalado [0,1], porque se comprime la escala de la variable en el rango [0,1].

$$X_{escalado[0,1]} = \frac{X - X_{mínimo}}{X_{máximo} - X_{mínimo}} \quad (7)$$

Todas las variables quedan con el mismo rango, lo que facilita la interpretación y comparación de los coeficientes. Sin embargo, como el rango se define según límites empíricos, mínimo y máximo, en variables estables el ruido puede verse amplificado.

3.3.2 Normalización

Esta transformación se interpreta como la desviación típica desde el promedio.

$$X_{normalizado} = \frac{X - X_{promedio}}{\sigma_x} \quad (8)$$

La normalización soluciona el problema del escalado por rango. No obstante, es sensible a valores extremos que pueden distorsionar el promedio, lo que se puede evitar tratando los *outliers* previamente.

IV. METODOLOGÍA

La metodología propuesta se basa en el proceso KDD (*Knowledge Discovery in Databases*) propuesto por (Fayyad, 1996). Corresponde a un flujo de etapas diseñado para la extracción de conocimiento útil de grandes bases de datos y es probablemente la metodología más reconocida en la disciplina.

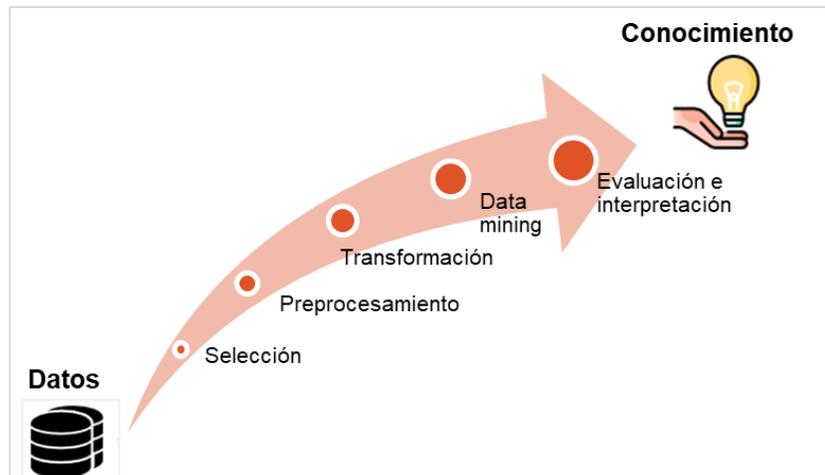


Ilustración 7. Flujo del proceso KDD

La figura ilustra el flujo de trabajo en líneas generales. Cabe señalar que lo más relevante de la metodología se encuentra en la etapa de Data Mining (4.4), en donde se propone una singular combinación de modelos. Sin más preámbulos, en la presente sección se explica con mayor detalle cada una de estas etapas en función de alcanzar los objetivos planteados.

4.1 Selección de datos

Esta etapa incluye el levantamiento de las distintas fuentes de información y los datos a utilizar durante el trabajo. En este caso, dado que se pretende desarrollar modelos de rentabilidad y retención para elaborar el Customer Lifetime Value, como también con modelos de propensión a ser activos en el programa de fidelización para la etapa de causalidad (explicados más adelante 4.4 Data Mining), se deben seleccionar las tablas de datos de la empresa que dispongan información relevante para desarrollar estos modelos. Como, por ejemplo, tablas que contengan información de contratos, de consumo de productos de la empresa, información transaccional de la tarjeta de crédito, información sociodemográfica del cliente, información de la rentabilidad por cliente, información sobre acumulación y canje de puntos, etc.

4.2 Preprocesamiento

Primero, se realizan análisis exploratorios de los datos para desarrollar un entendimiento del asunto y levantar primeras hipótesis. Se presentan estadísticos descriptivos como medias, medianas y gráficos. El análisis de correlaciones permite identificar las variables más relevantes con respecto a las variables a explicar en los modelos y también entrega conocimiento acerca de qué variables entregan la misma información.

En el proceso, es posible encontrarse con campos de datos vacíos o erróneos. Estos registros se deben descartar o corregir si es posible, para así no ensuciar el modelamiento. Para imputar valores nulos o incorrectos, si estos son relativamente pequeños se utiliza la mediana porque no altera de forma relevante el modelamiento. Sin embargo, si la variable cuenta con un porcentaje relativamente alto de nulos es recomendable descartarla.

En segundo lugar, se realiza un análisis de *outliers* para identificar y descartar observaciones atípicas que puedan ensuciar la calibración de los modelos. Los métodos multivariados son preferibles cuando se posee gran cantidad de variables, además permiten identificar casos extremos que de manera univariada pueden pasar desapercibidos. Mahalanobis es un método multivariado en donde se calcula la distancia de cada observación respecto al centro de la muestra. Para clasificar a los *outliers* se define una distancia máxima y los que superan esta distancia entran en esta categoría. Para definir este criterio, (Hadi, 1992) recomienda utilizar el percentil 0,999 o 0,975 de una distribución Chi-cuadrado con grados de libertad igual a la cantidad de variables utilizadas, ya que así distribuye la distancia por construcción.

4.3 Transformación

Luego de pre-procesar los datos, se debe realizar la transformación de variables. Esto corresponde a la creación de nuevas variables que sinteticen la información útil entregada desde los datos crudos, con el fin de reducir la dimensionalidad y facilitar el objetivo del trabajo. Por ejemplo: agregar el gasto por cliente a nivel mensual, calcular la frecuencia de compra, definir variables binarias, etc.

Para apoyar el proceso creativo de la generación de variables, es de gran utilidad plantear hipótesis sobre qué comportamientos de los clientes pueden explicar diferencias entre clientes en la variable a modelar. Por ejemplo, que los clientes más rentables suelen pagar en cuotas, por lo tanto, hay que generar variables asociadas al uso de cuotas.

Para las variables continuas se prueban dos transformaciones: el escalado [0,1] y la normalización (3.3 Escalado de variables). Estas permiten estandarizar las escalas de las variables y ajustarlas a distribuciones similares, lo que puede facilitar la interpretación y el modelamiento. La decisión de cuál transformación aplicar depende del modelo a realizar y los resultados, el no aplicar estas y mantener la escala original también es una

opción. Por lo tanto, para cada modelo se prueba la variable original y las dos transformaciones mencionadas, se selecciona finalmente con la que se obtienen mejores resultados.

4.4 Data Mining

Luego de pre-procesar y trabajar los datos, es turno de desarrollar el modelamiento y llevar a cabo los métodos planteados en el marco teórico para abordar el problema en función de los objetivos. Lo siguiente es la explicación del proceso, en donde se combinan dos modelos: CLV para estimar el valor del cliente y Propensity Score Matching para la evaluación de impacto en experimentos no controlados. Con el segundo modelo se estima el incremental de rentabilidad asociado al efecto del club de puntos, y con el primer modelo se obtiene una nueva métrica para estimar el efecto en el largo plazo.

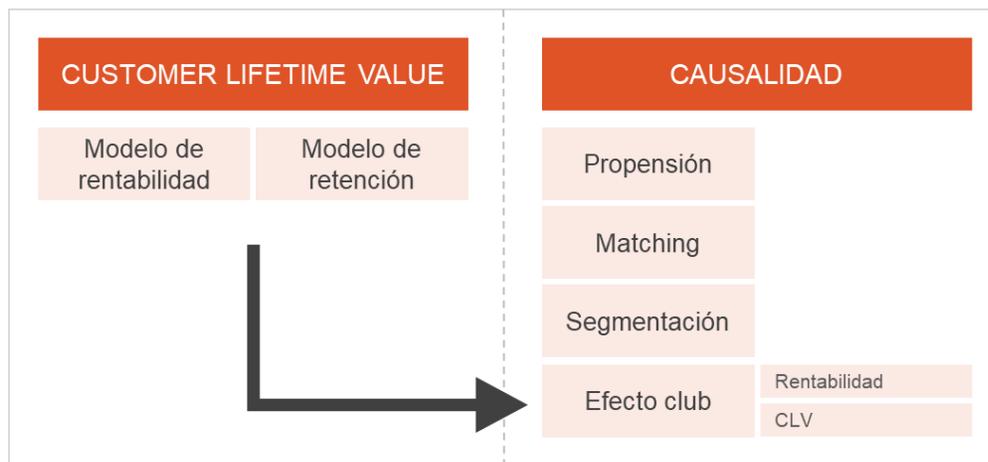


Ilustración 8. Metodología estimación incremental del valor asociado a puntos

4.4.1 Modelamiento del Customer Lifetime Value

El modelo del CLV consolida en una sola variable cuánto valen los clientes para la empresa, es comparable entre clientes, interpretable de forma monetaria y tiene visión de largo plazo. El fin de este modelo es brindar una métrica coherente con el objetivo del programa de fidelización para así medir su retorno asociado en el largo plazo.

Para estimar el CLV, se modelan las componentes de rentabilidad y retención por separadas, a nivel de cliente. En lo siguiente se describe la metodología para el desarrollo de los modelos y la fórmula que une ambas componentes.

i. Modelo de rentabilidad

Se utiliza información de un año (2014) para explicar el valor presente de las rentabilidades de los 3 años siguientes (2015-2017 en valores nominales) a nivel de

cliente. Se prefiere este periodo de covariables, porque es representativo del comportamiento del cliente con la tarjeta. Para el entrenamiento del modelo se particiona la base aleatoriamente en calibración (70%) y testeo (30%), ambas particiones están en el mismo horizonte de tiempo. Se prueban distintos algoritmos para seleccionar el que mejor se ajuste al problema:

- Modelo “ingenuo”: asume que la rentabilidad del cliente hoy se mantendrá constante en el futuro.
- Regresión lineal: minimiza los errores cuadrados.
- Árbol de regresión CHAID: mapea las observaciones y busca relaciones entre las variables explicativas y el resultado, utilizando la detección automática de interacciones de Chi-cuadrado.
- Red neuronal MLP (Perceptrón multicapa): sistema compuesto de muchos operadores simples procesando en paralelo, cuyas funciones se calibran de manera automática para minimizar el error total.

Estos modelos propuestos son seleccionados en base a los trabajos anteriores (Urzúa, 2007), (Rodríguez, 2013). El primero corresponde a un modelo simple que se utiliza como base para estudiar el desempeño de los más complejos. El resto de modelos provienen de distintas familias de algoritmos proporcionando variedad para abordar el problema. La elección del modelo final se basa en el menor WAPE en la partición de testeo.

ii. Modelo de retención

La retención de los clientes está sujeta a la definición de fuga establecida por la empresa. En base a esta, se modelan las probabilidades de supervivencia (de no fugarse) para cada año futuro por cliente.

Para entrenar el modelo de retención se posee un horizonte de observación de aprox. 7 años (2011/01 - 2018/05). Cada cliente con al menos una transacción dentro de este periodo es una observación. Cada observación posee un tiempo de vida, desde su primera transacción hasta la última, y un estado: fugado o censurado. Además, cada cliente posee características e información de su comportamiento que pueden aportar al modelo y, por ende, se rescatan a través de covariables. Estas covariables corresponden al último año móvil de cada cliente, es decir, se observa su consumo de productos, gasto, entre otras variables, durante los 12 meses previos a su última transacción. Recordando que esta última transacción puede corresponder al momento de fuga o no, entonces, se espera que el comportamiento del último año para cada cliente ayude a explicar su probabilidad de continuar en la empresa en los siguientes periodos. Este método de modelamiento con tiempos hasta el evento y covariables dinámicas se conoce como análisis de supervivencia.

Se prueban distintos modelos probabilísticos disponibles en la librería Survival de R (Therneau, 2015) y comúnmente utilizados en análisis de supervivencia: exponencial, weibull, lognormal y loglogístico. Al ser modelos paramétricos, en donde se ajustan distribuciones conocidas, se emplea toda la base para entrenamiento y se opta por el de mejor calidad de ajuste, en base al criterio AIC. Además, se realiza un test de log verosimilitud de Chi-cuadrado para comprobar que el modelo seleccionado es tienen mejor ajuste que el modelo nulo.

iii. Estimación del CLV

Se consolida el pronóstico de la rentabilidad a 3 años en valor presente con la estimación de retención del cliente, extrapolando la rentabilidad promedio por anual hacia el futuro. Así se obtiene el Customer Lifetime Value por cliente.

$$CLV_i = VP3(Rent_i) + \sum_{t=4}^T \frac{R_{i,t} * \overline{VP(Rent_t)}}{(1 + d_t)^t} \quad (9)$$

donde

$VP3_i$ = valor presente de las rentabilidades a 3 años para el cliente i

$R_{i,t}$ = retención cliente i en el periodo t

$\overline{VP_i}$ = rentabilidad anualizada (promedio) para el cliente i

T = plazo de estimación

d = tasa de descuento para el año t

En la ecuación (9) el elemento de la izquierda representa al valor presente de las rentabilidades a 3 años por cliente, resultante del modelo de rentabilidad. Este considera implícitamente la posibilidad de fuga de un cliente dentro de este plazo de 3 años, ya que en el entrenamiento de este modelo no se exige la permanencia de los clientes, más bien los que se fugan poseen rentabilidad cero en los años correspondientes, por lo tanto, el modelo considera internamente la fuga como un camino probable para clientes de características y comportamientos similares. Luego, esta rentabilidad estimada (a 3 años) se anualiza y se obtiene una rentabilidad promedio por año, la cual se asume se mantendrá constante hacia el futuro. En el lado derecho de la ecuación, a partir del cuarto año, se suma la rentabilidad anualizada ponderada por la probabilidad de supervivencia de cada año futuro. Así se construye una esperanza de los flujos futuros del cliente.

Como se ha mencionado anteriormente, conocer el impacto de activarse en el club sobre el CLV no es trivial. Inferir causalidad desde los coeficientes de la regresión o desde una comparación directa entre activos e inactivos sería inadecuado. La autoselección de los clientes causa estimaciones sesgadas. Para realizar un contraste fiable, se debe tratar la muestra y generar grupos comparables.

4.4.2 Estimación del efecto causal

Una vez modelado el valor de cada cliente, se procede a estimar el efecto causal del club de puntos en rentabilidad y CLV como impacto de corto y largo plazo respectivamente. Dado que el programa de puntos es un experimento natural no controlado, existen sesgos de autoselección, por lo que se decide utilizar Propensity Score Matching para minimizar estos sesgos.

El horizonte de observación a estudiar contempla tres periodos: en el 2015-2016 se identifica a los clientes activos y no activos del programa (2 años debido a duración de los puntos). El año anterior se utiliza para identificar a clientes similares y así recrear los grupos de tratamiento (activos) y control (no activos) comparables. Finalmente, en el año posterior se contrastan los resultados de ambos grupos, con lo cual se puede inferir causalidad.



Ilustración 9. Periodos de observación para estudiar causalidad

i. Definición de activo

Para evaluar el programa de puntos en coherencia con su objetivo, no es suficiente estudiar el impacto de un canje ocasional, sino que es necesario definir un criterio de actividad de acuerdo con variables de interacción al club de puntos. Dado que esta clasificación es ambigua, se realiza un análisis gráfico y/o con k-medias para definir el corte óptimo.

ii. Propensity Score

Para encontrar clientes similares, entre activos e inactivos, se requiere una métrica de similitud. Se utiliza la distancia entre puntajes de propensión, porque un modelo de propensión a ser activo en el programa reduce la enorme cantidad de variables que poseen los clientes a una única dimensión. Esto facilita la tarea del *matching* priorizando los aspectos del cliente más relevantes en relación con el club de puntos y, además, entrega información importante como propensión en sí.

El modelo se construye explicando la variable de actividad en el club de puntos (2015-2016) en base a las variables observadas el periodo anterior (2014). Se prueban los modelos: árbol C5, árbol CHAID y regresión logística. La elección del modelo se realiza en base al balance de las covariables como resultado del *matching* (más adelante). Se debe destacar que el objetivo de este modelo es el emparejamiento y no su uso predictivo, por lo tanto, no se utiliza partición de testeo y no es relevante la cantidad de variables.

iii. Matching

En esta etapa se realiza la asignación de clientes clones. A cada cliente activo se le asigna el cliente de mayor similitud del grupo inactivo. Existen distintos parámetros que se pueden fijar para el algoritmo de *matching*, el número de vecinos más cercanos, el radio de Caliper y la opción de reposición o no. Estas opciones deben ser analizadas para utilizar la opción que más se adapte al problema. El algoritmo es implementado en R con la librería Matching desarrollada por (Sekhon, 2011).

Una vez realizado el *match*, se revisa el balance de las covariables para asegurar que los grupos son comparables. Es decir, comprobar que, para cada covariable del modelo de propensión los grupos activos e inactivos seleccionados poseen una distribución similar. Para esto, se utiliza la diferencia de medias estandarizada, la que debe ser idealmente menor a 0,1 y como máximo 0,25. Si no se encuentra un balance adecuado, entonces se debe iterar entre el modelo de propensión y los parámetros del *matching*.



Ilustración 10. Proceso iterativo matching

iv. Segmentación

Como resultado del paso anterior, se obtienen grupos de control y tratamiento comparables, tal como si se tratase de un experimento controlado con grupos aleatorios. Entonces, es posible estimar un efecto promedio del programa. Sin embargo, aporta mayor valor para la gestión el conocer efectos diferenciados por segmentos de clientes. De esta manera, se pueden comprender distintos efectos según tipos de clientes y sus potenciales resultados al activarse en el club de lealtad.

Esta etapa consiste en realizar una segmentación de los clientes (emparejados) en base a los puntajes de propensión a participar en el programa. En teoría, esta agrupación resulta en clientes con características y comportamientos similares, ya que la propensión resume esta información. Este método se llama Subclasificación con Propensity Score.

Se prueban 3 métodos de segmentación: cuantiles, rangos equidistantes o k-medias, todos en base a la propensión de los tratados, y también se prueban distintos números de segmentos. Al igual que en el *matching* general, se debe revisar el balance de las covariables, ahora para cada segmento. Por lo tanto, la elección del método y el número de segmentos está sujeta a obtener el balance más adecuado posible dentro de cada clúster. (Rosenbaum, 1984) recomienda utilizar quintiles del *Propensity Score*, lo que elimina el 90% del sesgo en todas las covariables relevantes, aunque si la base es muy grande o existen variables no balanceadas se puede aumentar este número.

v. Caracterización de segmentos

Como se menciona más arriba, la segmentación por el puntaje de propensión agrupa clientes que comparten características y comportamientos similares. Por lo tanto, para comprender el efecto del programa en los distintos segmentos e interpretar por qué ocurre lo observado, es importante caracterizar y entender cómo son estos segmentos. Esta caracterización se realiza por medio del análisis descriptivo de variables relevantes.

vi. Estimación del efecto

Finalmente, se estima el efecto causal del programa de lealtad como la diferencia entre clientes activos e inactivos dentro de cada segmento. Es en este punto, entonces, en donde se contrasta el resultado en la rentabilidad real de 2017 y en el CLV estimado en base al comportamiento de ese año, como efecto de corto y largo plazo respectivamente. Para inferir causalidad, debe existir una diferencia significativa, lo cual debe testearse mediante un test de medias.

La siguiente figura ilustra la metodología de identificación de clones y la segmentación en base a propensión, con el fin de inferir efectos causales según tipos de clientes.

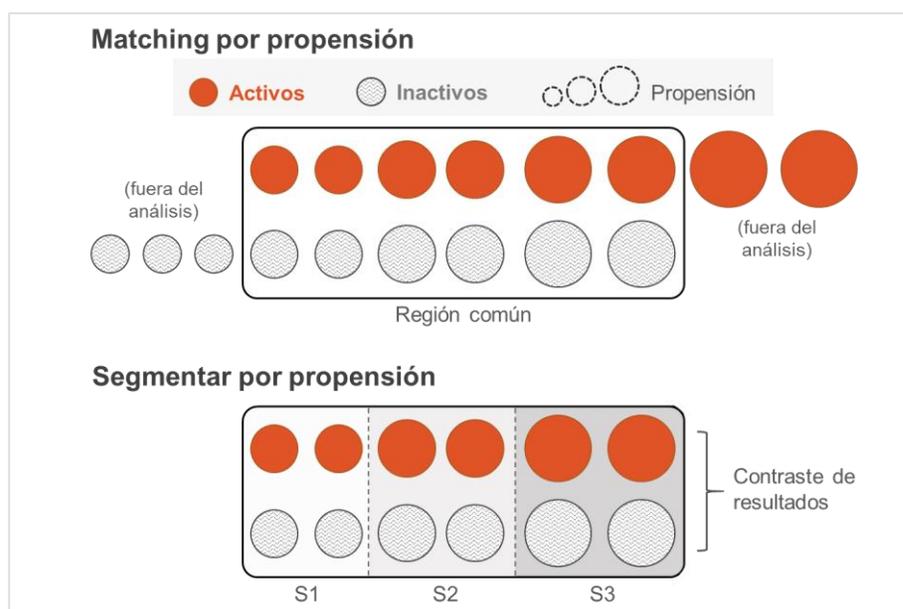


Ilustración 11. Ilustración matching y segmentación

4.4.3 Evaluación e interpretación

Con los resultados obtenidos en las etapas anteriores se puede conocer qué tipos de clientes participan del programa y, por ende, quiénes son más propensos a hacerlo de acuerdo con sus características y comportamientos previos. Por lo mismo, se puede analizar si efectivamente el club de puntos está alcanzando o “fidelizando” a los clientes de mayor valor para el negocio, es decir, si los recursos del club de puntos están enfocados en los clientes que brindan mayor rentabilidad.

A partir de los clientes que se han activado en el club de lealtad y de sus pares de mayor similitud (clones), es posible estimar el efecto causal asociado a la participación en el programa de puntos. Es decir, cómo cambia el comportamiento de consumo de los clientes y sus resultados luego de haber participado en este club. Estos efectos observados se pueden entender como el retorno causado por el mismo programa, por lo tanto, puede considerarse como un efecto potencial para futuras gestiones de los clientes a través del sistema de puntos.

Además, se estudia la existencia de este efecto a nivel de segmentos. Por lo que, es posible que exista un efecto transversal para todos los clientes como también es posible que estos reaccionen de distintas maneras de acuerdo su perfil como cliente. La descripción previa de cada grupo y el análisis detallado del cambio de comportamiento por variable, son de vital importancia para comprender por qué estos perfiles reaccionan de esa forma. Esto permite dar una interpretación que genere conocimiento más allá de los mismos resultados numéricos del club de fidelización.

4.4.4 Recomendaciones

Para finalizar, corresponde mostrar la utilidad del conocimiento generado en las etapas previas. Primero y de acuerdo con los efectos estudiados a nivel de segmentos, se realiza un análisis del aporte del programa de puntos en el desarrollo del valor de los clientes. Este primer análisis permite evaluar la estructura del club y proporciona una visión general que, en caso de encontrarse con resultados negativos, puede sugerir cambios el mismo programa.

Luego, según los efectos estudiados y la cartera actual de clientes, se estima el potencial económico de llevar a cabo una gestión que active a estos en el club de lealtad. De acuerdo con este beneficio estimado, se proponen recomendaciones a nivel táctico que luego pueden materializarse, por ejemplo, a través de campañas focalizadas, en premios a los clientes, en incentivos de puntos, en ampliar la oferta de canjes en ciertos niveles de canje, etc. Sin embargo, el diseño de estas requiere de estudios de mercado y un conocimiento mayor de las necesidades de los segmentos, por lo tanto, queda fuera del alcance del proyecto y se propone para futuros trabajos.

V. ALCANCES Y RESULTADOS ESPERADOS

- Estudio de carácter observacional, no considera experimentos ni encuestas para obtener datos. El trabajo se basa en datos disponibles de la empresa, específicamente desde el 2014 hasta el 2017.
- Se espera modelar el Customer Lifetime Value para cada cliente como valor presente de las rentabilidades a 3 años extrapolada por la probabilidad de supervivencia durante los siguientes años, hasta un límite de 10 años. El modelamiento de la rentabilidad se basa en las rentabilidades históricas calculadas por la empresa.
- Se define a un cliente activo en el club de puntos según un criterio arbitrario en base a métricas de interacción con el club de puntos. Este criterio (de elaboración propia) busca identificar a quienes participan activamente del club y actúan bajo la influencia de este. Con esto se pretende realizar la separación, por ejemplo, entre un cliente que canjea 3.000 puntos en 3 canjes y vence sólo 300 puntos en un año versus un cliente que canjea 3.000 puntos una única vez y ha vencido cerca de 25.000 puntos, a pesar de que hayan canjeado la misma cantidad de puntos es posible discriminar de que el primer cliente posee un mayor grado utilización y de preocupación por sus puntos. Por lo tanto, se espera que la definición propuesta sea más rigurosa que un único canje, el cual puede darse de manera eventual en los clientes sin tener mayor compromiso con el club de puntos.
- Se espera obtener la propensión por cliente a activarse en el programa de puntos según el criterio y los periodos definidos para el estudio de causalidad. El foco de este modelo de propensión es realizar el *matching* de clientes, por lo tanto, su capacidad predictiva no es relevante para la metodología.
- Se espera obtener el efecto causal del programa diferenciado por segmentos. Este puede suponerse como un efecto potencial para clientes que aún no participan del programa. El trabajo se centra en estimar este efecto reduciendo los sesgos de autoselección, no obstante, los resultados finales no están exentos de error. Los estimadores sirven como orientación para la gestión y es probable que difieran respecto a los resultados reales observados en el futuro por las dinámicas del mercado y el negocio.
- Se espera que el conocimiento desarrollado sirva como base para la gestión del club de puntos. Por lo tanto, como primer análisis de los resultados, se realizan recomendaciones generales del tipo tácticas en base al potencial económico de cada segmento. Pero está fuera del alcance diseñar, testear o simular cambios en la estructura del programa, en los beneficios de los clientes o de campañas y acciones operativas.

VI. DESARROLLO METODOLÓGICO

6.1 Preparación de los datos

6.1.1 Selección de los datos

El universo de clientes corresponde a 2,6 millones de clientes cuenta abierta al cierre del 2014 y que abrieron su cuenta desde antes del 2014. De esta manera, se posee un año de información del comportamiento de cada cliente, suficiente para rescatar posibles estacionalidades que puedan evidenciar los distintos tipos de clientes.

Se recopilan datos provenientes de distintas fuentes de la empresa. Dada la inmensa cantidad de columnas y variables disponibles, en lo siguiente se describen estas fuentes agrupadas por tema y se resume su información de mayor relevancia.

- Sociodemográficas: contiene el identificador del cliente, datos de edad, GSE, región de habitación, sexo, estado civil, entre otras. Esta fuente se actualiza mensualmente y contiene información de todos los clientes de la empresa y de quienes han comprado alguna vez en el Holding declarando su RUT.
- Transaccional tarjeta: se registran todas las transacciones realizadas con la tarjeta de crédito. Se tiene el detalle de fecha, monto, rubro, negocio y cantidad de cuotas utilizadas. Esta base contiene toda la historia de transacciones y se actualiza a nivel diario, por su naturaleza sólo considera clientes de la tarjeta de crédito.
- Transaccionales negocios: la empresa cuenta también con un historial de transacciones de los clientes que compran en los *retails* del Holding, registrando al cliente comprador, la fecha, el identificador del negocio, la tienda, los productos comprados, el medio de pago, etc.
- Contrato: esta tabla contiene información del contrato como la fecha de emisión (o antigüedad), el monto cupo, tipo de tarjeta y tenencia de clientes adicionales. Su actualización es de carácter mensual y contiene información histórica de todos los clientes que han sido titulares de la tarjeta de crédito.
- Estados de cuenta: esta tabla contiene información acerca de los montos transados, pagados, abonados y facturados por los clientes, datos acerca de los intereses, montos adeudado y atrasado, tanto con la tarjeta de crédito como de productos financieros. Esta tabla se actualiza diariamente incorporando toda la historia de estados de cuenta (mensuales) a nivel de cliente.
- Canales: registra fecha de ingresos a la página web (registrados con su usuario), llamados al Callcenter o atenciones en sucursales. Esta base está a nivel de cliente y se actualiza acumulativamente a nivel diario.

- Club de puntos: contiene información relativa a puntos, tanto el detalle de puntos acumulados por cada cliente de la empresa al realizar una transacción con la tarjeta, con su fecha de creación como de vencimiento. En la misma fuente se registran los puntos vencidos o canjeados por los clientes. Esta tabla es acumulativa y se actualiza a nivel diario, con identificador de cliente, fecha, cantidad de puntos y tipo de transacción (acumulación, vencimiento o canje).
- Rentabilidad: esta tabla contiene el dato de rentabilidad anual calculado por la empresa para cada cliente, con detalle de las cuentas de ingresos y costos. Por ejemplo, ingresos por términos de intereses, productos financieros o costos por riesgo, fidelización, etc. Esta tabla se actualiza manualmente a nivel anual y contiene las rentabilidades de los años 2014, 2015, 2016 y 2017 a nivel de cliente existente para cada año correspondiente.

Recalcando nuevamente la inmensa cantidad de variables y columnas que se pueden extraer desde cada fuente, no es necesario mencionar ni entrar en detalle en cada una de estas. Más bien, en la siguiente sección se presentan las variables recolectadas y otras que son generadas a partir de transformaciones, se exhiben algunos ejemplos relevantes y se referencia al total de variables a utilizar en anexos.

6.1.2 Transformación de variables

Existen varias hipótesis en la empresa acerca de los comportamientos que pueden diferenciar a los clientes rentables o duraderos de los que no. Lo siguiente son algunos ejemplos de hipótesis que dan origen a variables a rescatar:

- Clientes que utilizan la tarjeta sólo para descuentos dentro del *holding* probablemente son menos rentables, por lo que se crea la variable “porcentaje del gasto destinado a comercios externos”.
- Clientes que pueden ser poco rentables son quienes sólo utilizan la tarjeta en épocas de navidad y fiestas por las promociones en *retail*, por lo tanto, se crea la variable “porcentaje del gasto en trimestre X” para cada trimestre.
- Clientes que pueden ser muy rentables son quienes necesitan y consumen crédito, por lo tanto, se crean variables como la “cuota promedio ponderada”, “porcentaje de transacciones con cuotas”, “tiene crédito de consumo contratado”, etc.
- Clientes de riesgo muy alto pueden ser poco rentables en el futuro porque son propensos al cierre de la tarjeta, por lo que se crea la variable “meses con mora”.
- Clientes que no necesitan cuotas y pagan su tarjeta antes de la facturación para evitar costos de administración pueden ser clientes muy poco rentables. Por lo tanto, se crea la variable “porcentaje de estados de cuenta con pago prefacturación”.

Dado el proceso anterior de levantamiento de variables a través de hipótesis, se generan cerca de 120 variables presentadas en el Anexo 1. A modo de ejemplo, en la siguiente tabla se exhiben algunas de estas variables agrupadas por temas.

Sociodemográficas	Rentabilidad
Edad	Ingresos
Sexo	Costos
Zona geográfica	Porcentaje de los costos asociados al cliente por riesgo crediticio
GSE	Porcentaje de ingresos generados por el cliente por término de intereses
Contrato	Riesgo
Tiene adicionales	Puntaje score de riesgo
Antigüedad del cliente	Porcentaje del cupo disponible (sin utilizar)
Monto de cupo	Meses con mora
Productos	Transaccional
Categoría	Gasto
Número de avances últ. año	Recency
Monto en créditos últ. año	Días con compra
Tiene seguros	Número de rubros con compra
Tiene adicionales	Variación gasto semestral
Tiene pagos automáticos	Porcentaje del gasto en negocios externos
Puntos	Porcentaje de transacciones con cuotas
Stock de puntos	Cuota promedio ponderada
Cantidad de canjes últ. año	Canales
Puntos canjeados últ. año	Número de ingresos web
Tasa de canje	Número de interacciones con sucursales

Tabla 3. Ejemplos de variables generadas

En cuanto al programa de puntos, la empresa utiliza la variable Tasa de canje para medir el nivel de uso de los clientes. Esta variable es relevante para este proyecto y se calcula de la siguiente manera:

$$Tasa\ Canje_{[t_0,t_1]} = 1 - \frac{Ptos\ Vencidos_{[t_0,t_1]}}{Ptos\ a\ Vencer_{[t_0,t_1]}} \quad (10)$$

Indica porcentualmente cuántos puntos fueron canjeados de los puntos que iban a vencer en un determinado periodo. Una tasa de canje de 100% significa que el cliente canjeó todos sus puntos. Sin embargo, esta variable esta variable arrastra un comportamiento que puede estar rezagado, ya que solo se mide al momento de vencer si estos puntos vencieron o no, pero no se puede ver cuándo se originaron o canjearon estos puntos.

En vista de que la Tasa de canje posee la limitante recién mencionada, no es posible utilizarla para definir a un cliente como activo en el club de fidelización durante un periodo observado, porque puede indicar un comportamiento retrasado. Como alternativa a esta,

se crea la variable Tasa de vencimiento que indica el porcentaje de puntos vencidos durante un año sobre el stock de puntos al inicio del periodo. Esta variable es relevante para más adelante (6.3.1 Definición de activos) porque indica qué tan pendiente está el cliente de que sus puntos no caduquen durante un periodo de tiempo determinado.

$$Tasa\ Vencimiento_{[t_0,t_1]} = \frac{Ptos\ Vencidos_{[t_0,t_1]}}{Stock\ Ptos_{t_0}} \quad (11)$$

En el contexto de modelar la componente de rentabilidad para construir el Customer Lifetime Value, se crea la variable “Rentabilidad presente” (en pesos chilenos) para ser explicada por las covariables del año anterior. La Ecuación (12) presenta el cálculo, en donde se utilizan tasas de descuento entregadas por el área de Finanzas de la empresa.

$$VP(Rentabilidad_{2015,2017}) = \frac{Rent_{2015}}{(1 + 0,030)} + \frac{Rent_{2016}}{(1 + 0,033)^2} + \frac{Rent_{2017}}{(1 + 0,035)^3} \quad (12)$$

Por otro lado, todas las variables continuas de entrada para los modelos son transformadas mediante el escalado [0,1] y el normalizado estándar. La decisión final de utilizar o no las variables escaladas, y de qué transformación usar, depende de los resultados del modelamiento.

6.1.3 Valores faltantes o erróneos

En la variable edad existen valores incorrectos como “2013”. A pesar de ser una cantidad despreciable (<0.001%) se corrigen estos valores con la mediana, 45 años. Para la variable zona geográfica también se encuentran casos erróneos o nulos, una cantidad despreciable, pero que son imputados con la moda: Región Metropolitana. Variables como estado civil, GSE y sexo, poseen mayor cantidad de datos nulos, por lo que se crea la categoría “sin información”. La imputación y corrección de los datos puede ser innecesaria para efectos del trabajo, dado que son relativamente muy pocos casos y es de esperar que no tengan consecuencias en el modelamiento ni en los resultados. Sin embargo, se desarrolla el proceso con el fin de dejar automatizada la limpieza de las variables para asegurar el pronóstico del CLV (modelo) a toda la base de clientes, tanto durante el propio trabajo como para futuras ejecuciones de este modelo.

Por otro lado, variables del tipo gasto o puntos canjeados cuentan con observaciones negativas o nulas. Esto puede ocurrir por devoluciones, cambios o falta de transacciones durante el horizonte de tiempo considerado. Por razones lógicas, los valores negativos son truncados a 0 y los faltantes también son imputados a este valor.

6.1.4 Outliers

Para el análisis de valores atípicos, se utiliza la distancia de Mahalanobis con 69 variables continuas, filtrando las correlacionadas. Se obtiene distancia promedio respecto al centro igual a 72, mediana de 41 y cuartil superior de 62, lo que implica una gran concentración en valores bajo 72; mientras que la máxima distancia es de 2.236.071, evidenciando que existen casos muy extremos.

Para definir la distancia de corte se prueba el percentil 0,999 de la chi-cuadrado, pero con este método se obtiene una cantidad de aproximadamente 8% de outliers, lo cual es un porcentaje relativamente grande para descartar del modelamiento.

Como alternativa, se realiza un análisis univariado con el método de 3 desviaciones estándar para cada variable y, en particular, en la variable “Monto en créditos de consumo” se encuentra el mayor porcentaje de outliers con 2.82%. Entonces, se aplica este porcentaje para la eliminación de outliers, quedando el corte en una distancia de 201.

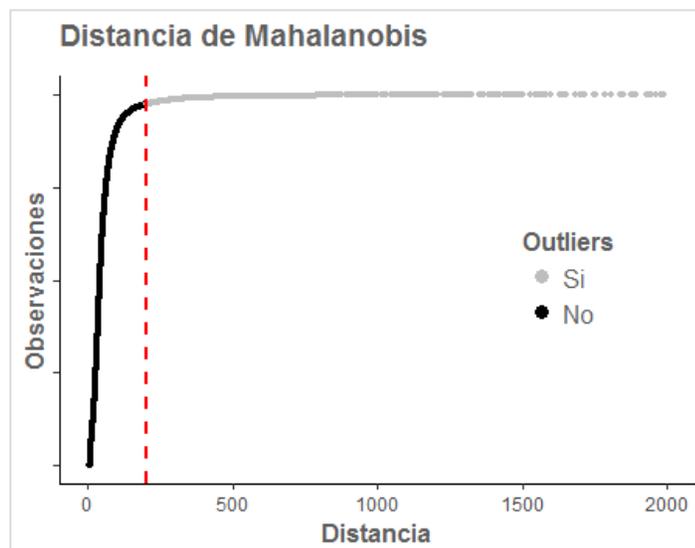


Ilustración 12. Detección de outliers con distancia de Mahalanobis
(Para la visualización se omiten valores mayores a 2.000 (0,001% observaciones))

6.2 Modelamiento del Customer Lifetime Value

6.2.1 Análisis descriptivos rentabilidad

En vista del primer modelo para el Customer Lifetime Value, se presentan descriptivos de rentabilidad a 3 años con los periodos de entrenamiento (2015-2017) y en valor presente 2014 (valores nominales en pesos). Además, se observa su relación con variables relevantes del 2014.

6.2.1.1 Rentabilidad a 3 años

En el siguiente gráfico se muestra la distribución de la rentabilidad presente a 3 años. Se observa una concentración entre los primeros tramos desde -75 M\$ a 100 M\$, con un *peak* en rentabilidad cercana a 0 correspondiente a clientes inactivos.



Ilustración 13. Rentabilidad a 3 años (2015-2017)

La siguiente tabla presenta estadísticos descriptivos de la rentabilidad a 3 años. El promedio es de 166 M\$, con una desviación típica más de 3 veces el promedio, lo que implica una gran dispersión. Por último, se entiende que la mitad de los clientes están en el rango de -29 M\$ y 256 M\$, pero mayormente concentrados a la izquierda.

Promedio (M\$)	Desv. est. (M\$)	Q1 (M\$)	Mediana (M\$)	Q3 (M\$)
166	519	-29	24	256

Tabla 4. Descriptivos rentabilidad a 3 años (2015-2017)

6.2.1.2 Rentabilidad a 3 años vs Rentabilidad anterior

El siguiente gráfico de dispersión entre la rentabilidad a 3 años (2015-2017) y la del año anterior (2014) muestra al total de clientes con cuenta abierta al inicio del 2015 y con al menos un año desde su apertura de cuenta, es decir, con al menos un año de tarjeta vigente. Se observa una concentración en las cercanías de la diagonal, que junto con un coeficiente de correlación de Pearson 0,49 implican que la rentabilidad anterior puede ser una variable relevante para pronosticar las rentabilidades futuras.

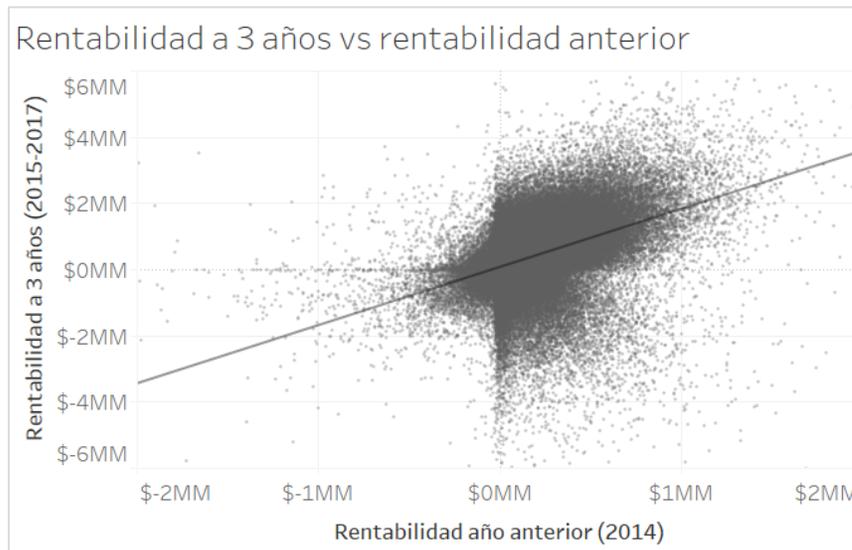


Ilustración 14. Rentabilidad a 3 años (2015-2017) versus rentabilidad anterior (2014)

6.2.1.3 Rentabilidad a 3 años vs Categoría/Gasto

La categoría es una variable muy relevante para el negocio, porque contiene mucha información acerca del perfil de los clientes, principalmente del gasto y los beneficios, por su propia construcción. En la Ilustración 15 se muestra un Box-plot de la rentabilidad a 3 años por categoría, en donde la categoría Normal ha sido abierta por quintiles de gasto y quintil 5 corresponde al de menor gasto.

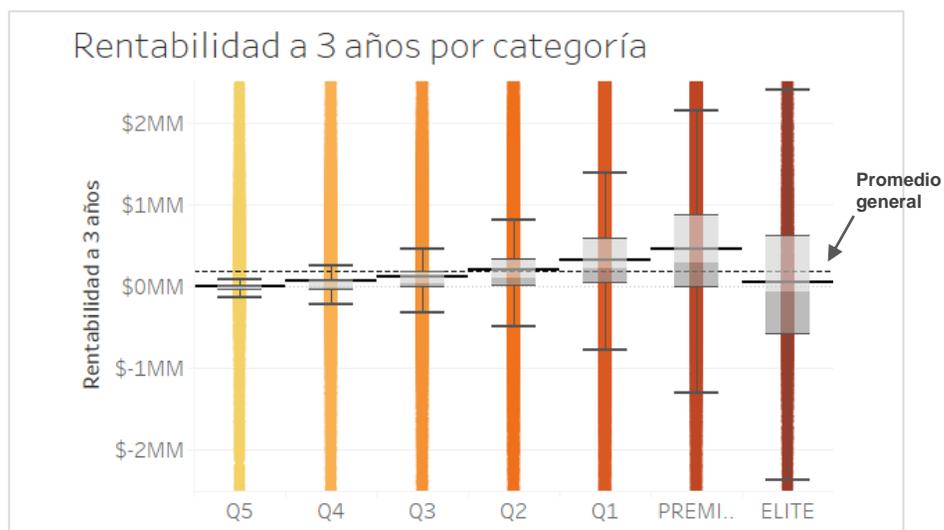


Ilustración 15. Box-plot de rentabilidad a 3 años (2015-2017) por categoría (2014)

Q5 (M\$)	Q4 (M\$)	Q3 (M\$)	Q2 (M\$)	Q1 (M\$)	Premium (M\$)	Elite (M\$)
-5	56	111	186	325	453	53

Tabla 5. Rentabilidad a 3 años (2015-2017) por categoría (2014)

Se evidencia que a mayor categoría aumenta el promedio de rentabilidad futura, a excepción de la categoría Elite que posee un promedio sólo mayor al quintil de menor gasto. A mayor categoría, también existe un aumento en la dispersión, lo que implica el aumento de casos muy positivos y negativos. Incluso, se puede observar que la categoría Elite está cargada hacia valores negativos. En conclusión, hay una aparente relación entre la categoría/gasto y la volatilidad de las rentabilidades futuras, que en promedio significa una correlación positiva, a excepción de los Elite.

6.2.1.4 Rentabilidad a 3 años vs Productos financieros

El consumo de productos financieros, avances y créditos de consumo, también es una variable relevante para el negocio, ya que significan mayores colocaciones, intereses y cuotas que la compra típica con tarjeta de crédito. Por lo tanto, es importante de analizar.

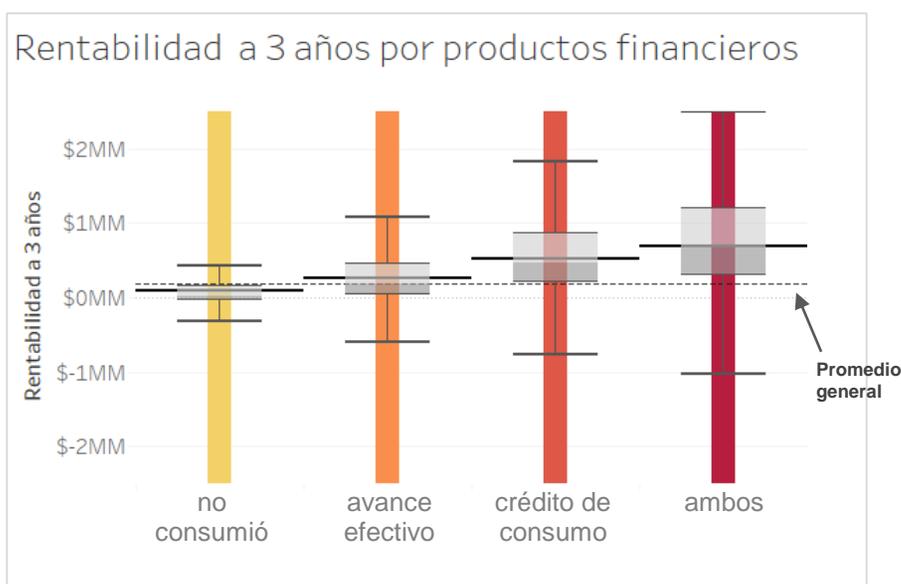


Ilustración 16. Rentabilidad a 3 años (2015-2017) por de productos financieros

No consume (M\$)	Sólo avance (M\$)	Sólo créd. de consumo (M\$)	Ambos productos (M\$)
71	266	528	691

Tabla 6. Rentabilidad a 3 años (2015-2017) por consumo de productos financieros (2014)

Del gráfico y la tabla anterior, se observa que el consumo de estos productos está vinculado con un aumento considerable de las rentabilidades a futuros. Aunque, igual pasa con el gasto, también aumenta la dispersión. En este caso, los cuartiles inferiores y la mediana se encuentran por sobre el 0 y, en especial, en los créditos de consumo estos valores están por sobre el promedio de la cartera. Por lo que es de esperar que estos productos impacten fuertemente en el valor futuro de los clientes.

6.2.1.5 Rentabilidad a 3 años vs Puntos

En vista del estudio del club de puntos, se puede observar en la Ilustración 17 que existe gran dispersión en cuánto a puntos canjeados y rentabilidad. Estas variables poseen una correlación de Pearson igual 0,06 que es despreciable. Se puede observar también que, a menores puntos canjeados existe mayor cantidad de clientes con tasa de canje⁴ baja (rojo), hay muchos clientes que han canjeado una cantidad relativamente baja de puntos y tiene oportunidad de canjear más. No se observa una relación clara entre la tasa de canje y la rentabilidad.

⁴ Cálculo de la tasa de canje detallado en 6.1.2

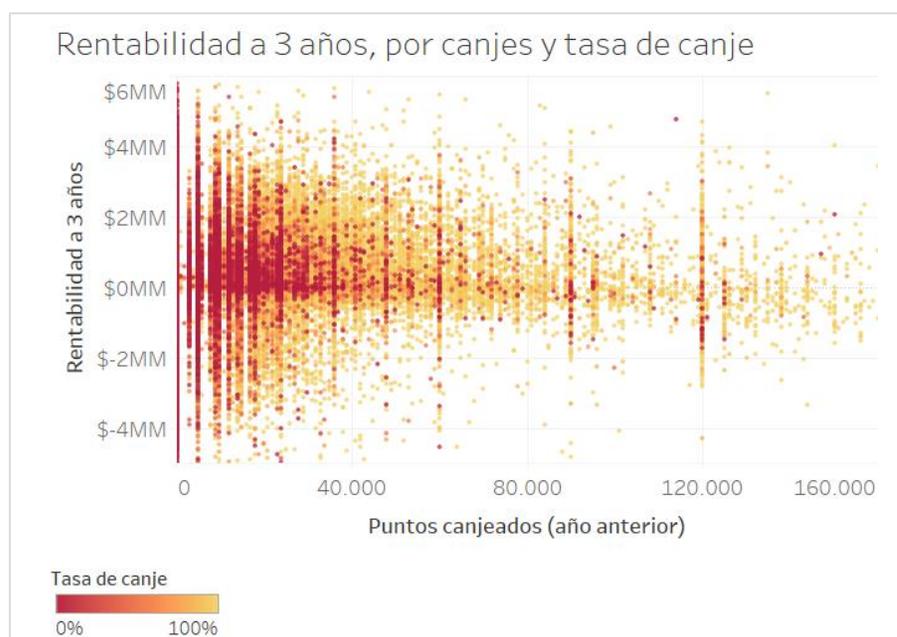


Ilustración 17. Rentabilidad a 3 años (2015-2017) por puntos canjeados y tasa de canje (2014)

6.2.2 Modelo de rentabilidad

Se modela la variable valor presente de la rentabilidad a 3 años con distintos métodos y se obtiene los siguientes errores:

Plazo	Modelo	MAE (M\$)	MAPE (%)	WAPE (%)
3 años	Ingenuo	190	147	76
	Reg. Lineal	187	188	75
	Árbol CHAID	170	138	68
	MLP	161	134	65

Tabla 7. Desempeño de modelos de rentabilidad

Los errores se encuentra dentro de lo esperado para modelos de pronóstico de rentabilidad a nivel de clientes, teniendo como referencia a trabajos anteriores (Urzúa, 2007), (Reuse, 2011) y (Rodríguez, 2013). El modelo de mejor desempeño es la red neuronal (MLP) con un 65% de WAPE, lo que significa que el pronóstico a nivel de cliente se equivoca en promedio en un 65% respecto al valor real.

En la siguiente tabla se exhiben las covariables de la red neuronal agrupadas por temas. Junto a cada variable (numérica) se señala el sentido en el que aporta, aunque como las redes neuronales son “cajas negras”, estos sentidos son estimados mediante su correlación bivariada con el resultado.

Sociodemográficas	cor*	Transaccional	cor*
GSE		Frecuencia de facturación	++
Productos		Meses con compra	++
Categoría_2		Recency en meses	--
Mono de cupo	+	Máximo monto por transacción	+
Porcentaje cupo disponible	-	Gasto 2do semestre	+
Monto en avances últ. año	++	Porcentaje del gasto en semestre 2	+
Número de créditos últ. año	++	Número de negocios (internos) con compra	++
Tiene adicionales	+	% pagado con la tarjeta en neg. internos	++
Se traspasó de tarjeta en últ. año	+	Gasto en negocio interno: tiendas p.d.	+
Tiene seguro cesantía	+	Gasto en negocio interno: mej. del hogar	+
Tiene seguro desgravamen	+	Gasto en negocio interno: supermercado	+
Tiene seguro otros	+	Compras en negocio interno: otros	+
Rentabilidad		Gasto en negocios externos	+
Ingresos últ. año	+++	Porcentaje del gasto en negocios externos	++
Costos últ. año	++	Rubros	
Riesgo		Número de rubros con compra	++
Tramo score de riesgo		Con compra en rubro salud	++
Meses en mora últ. año	+	Con compra en rubro supermercados	++
Canales		Con compra en rubro transporte	++
Ingresos web últimos 6 meses	++	Con compra en rubro tiendas p. d.	++
Activo sucursal en últimos 6 meses	+	Con compra en rubro seguros	+
Cuotas		Puntos	
Cuota promedio ponderada por montos	++	Tasa de canje	++
Máxima cuota	++	Puntos canjeados últ. año	+
Porcentaje de transacciones en cuotas	+	Stock de puntos	+
Porcentaje de intereses sobre gasto	+	Distancia porcentual al próximo nivel de canje dado su actual stock de puntos	--

*correlaciones variables numéricas: + (0; 0,3), ++ [0,3; 0,7], +++ [0,7; 1], - (-0,3; 0), -- (-0,7; -0,3], --- [-1; -0,7]

Tabla 8. Variables modelo de rentabilidad

De la tabla anterior, se destacan los siguientes puntos:

- Existe una fuerte correlación entre ingresos y rentabilidades futuras. Si el cliente hoy genera ingresos, existe una alta probabilidad de que lo siga haciendo en el futuro.
- Sorprendentemente, los costos también tienen una correlación media y positiva con las rentabilidades futuras. Probablemente porque clientes de baja actividad poseen costos fijos y generan ingresos despreciables, en cambio, clientes de alta actividad generan mayores costos, pero a la vez mayores ingresos.
- En cuanto a los productos del cliente, las variables más relevantes corresponden a monto en avances y número de créditos de consumo. Estas poseen una correlación media con las rentabilidades futuras del cliente, lo cual es lógico dado que son productos de gran cantidad de cuotas y mayores intereses.
- Respecto a la mora, existe una relación leve y positiva entre meses en mora y rentabilidades futuras, esta correlación no es fuerte porque la mora también conlleva costos asociados a riesgo y una mayor probabilidad del cierre de la tarjeta

- Respecto al pago en cuotas, evidentemente mientras mayor uso de estas, se espera mayor rentabilidad en el futuro. Indicadores como la cuota promedio ponderada o la máxima cantidad de cuotas con la que ha comprado el cliente, tienen una correlación de intensidad media con sus rentabilidades futuras.
- En cuanto al comportamiento transaccional del cliente, existen muchas variables que explican la rentabilidad. Por ejemplo,
 - a mayor porcentaje del gasto destinado a comercios externos se espera una mayor rentabilidad, ya que son clientes que le dan un uso general a la tarjeta;
 - de manera simultánea, si ha comprado en más negocios del *holding* se espera una mayor rentabilidad, ya que representa una fidelización con este grupo;
 - por otro lado, a mayor tiempo sin compra (*recency*) existe una menor rentabilidad futura. Esta variable está relacionada a una mayor probabilidad de fuga y/o a una menor frecuencia de compra, por lo que es esperable.
- Por último, en cuanto al programa de puntos existe una correlación positiva entre el uso de los puntos y las rentabilidades futuras. También a mayor distancia porcentual del próximo nivel de canje menores serán sus rentabilidades futuras. Es decir, si posee un stock de 0 puntos, le queda el 100% del trayecto para alcanzar el primer nivel de canje (5.000) puntos, lo cual está relacionado con una menor cantidad de rentabilidad para la empresa en los siguientes años.

Al parecer, la rentabilidad futura de los clientes sí está apalancada, en parte, por el programa de puntos, lo que se puede observar explícitamente en la correlación de variables de puntos con la rentabilidad, como también de forma implícita en las compras en negocios internos dada la doble acumulación de puntos en estos.

Finalmente, cabe destacar que se obtuvo mejores resultados con la normalización de las variables numéricas. Los siguientes puntos resumen lo obtenido.

- Con el escalado [0,1] el modelo está sujeto a los máximos del año utilizado para el entrenamiento. Al escalar las variables del año 2017 para realizar el pronóstico de rentabilidad 2018-2020, los nuevos máximos son mayores. Como resultado, se obtiene una distribución de rentabilidad pronosticada altamente concentrada en valores cercanos a cero.
- La opción de no escalar las variables también es sensible al crecimiento anual de las variables. Por ejemplo, alzas generalizadas en el gasto o en los ingresos a la web provocan un efecto de sobreestimación en los resultados, en donde, el promedio de rentabilidad pronosticada para 2018-2020 duplica las rentabilidades reales del 2015-2017.
- La normalización entrega los resultados que más hacen sentido a los expertos del negocio. La distribución de rentabilidad pronosticada para 2018-2020 (Ilustración

18, más adelante) se asemeja a la de 2015-2017, con cambios esperables y un crecimiento promedio de 19 M\$ en rentabilidad a 3 años, lo que resulta bastante más realista. En fin, este escalado se adapta mejor al modelado de rentabilidad porque considera tendencias a través de la estandarización.

6.2.3 Resultados modelo de rentabilidad

A partir del modelo elaborado, se realiza el pronóstico de rentabilidad a 3 años (2018-2021) para la cartera de clientes actual (a diciembre 2017). En este subapartado se presentan los resultados esperados, empezando por la distribución general de la cartera.

6.2.3.1 Pronóstico de rentabilidad



Ilustración 18. Pronóstico de rentabilidad a 3 años (2018-2020)

Promedio (M\$)	Desv. est. (M\$)	Q1 (M\$)	Mediana (M\$)	Q3 (M\$)
185	324	-13	85	285

Tabla 9. Descriptivos pronósticos de rentabilidad a 3 años (2018-2020)

Se observa una distribución de forma similar a la de los 3 años anteriores vista anteriormente, que va desde \$-500.000 a \$2.500.000, mayormente cargada hacia la izquierda. El promedio es de 185 M\$, con un crecimiento de 19 M\$ respecto a los 3 años anteriores. Por otro lado, la mediana también es mayor, con un crecimiento mayor, de 61 M\$ respecto al periodo previo.

La hipótesis detrás de esta distribución cargada hacia la izquierda es que, como la tarjeta está fuertemente asociada al *retail* y, al ser una cartera tan grande de clientes, es de esperar que parte de estos hayan abierto la tarjeta para aprovechar sus descuentos y ofertas en *retails* asociados, más no la usan como una tarjeta principal. La siguiente tabla

muestra las diferencias en rentabilidad entre clientes que gastan en comercios externos versus los que no, según el porcentaje del gasto destinado a negocios externos. Si bien esta comparación no es concluyente, sí demuestra una correlación que apoya la hipótesis.

Clientes que compran (casi) únicamente en el holding*	Porcentaje del gasto en negocios externos			
	0%	<= 10%	<= 20%	<= 30%
Porcentaje de clientes	34,6%	41,8%	47,5%	52,9%
Rentabilidad prom. a 3 años (M\$)	-0,6	29,5	50,9	70,7
Diferencia versus grupo que si compra en comercios externos (M\$)	-284,3	-267,5	-255,8	-243,5

*Criterio según porcentaje de gasto dedicado al comercio externo.

Tabla 10. Diferencia de rentabilidad a 3 años (2018-2020) según gasto en comercio externo (2017)

6.2.3.2 Pronóstico de rentabilidad por Categoría/Gasto

Al igual que en el análisis descriptivo de rentabilidad, es importante revisar cómo se distribuyen las rentabilidades por categoría, dado que es una variable relevante para el negocio y está estrechamente relacionada con el gasto y el uso de puntos. En la siguiente ilustración se observa que la categoría Normal está mayormente concentrada en rentabilidades bajo 250 M\$. Mientras que los Premium y Elite poseen una distribución similar y con distribuciones más altas. En los 3 grupos se observa una cola en el lado izquierdo de clientes no rentables, en Premium y Elite corresponde a una proporción menor que en los clientes Normales, ya que en este último grupo existen muchos clientes inactivos que utilizan la tarjeta eventualmente por descuentos sin generar ingresos o consumir productos financieros.

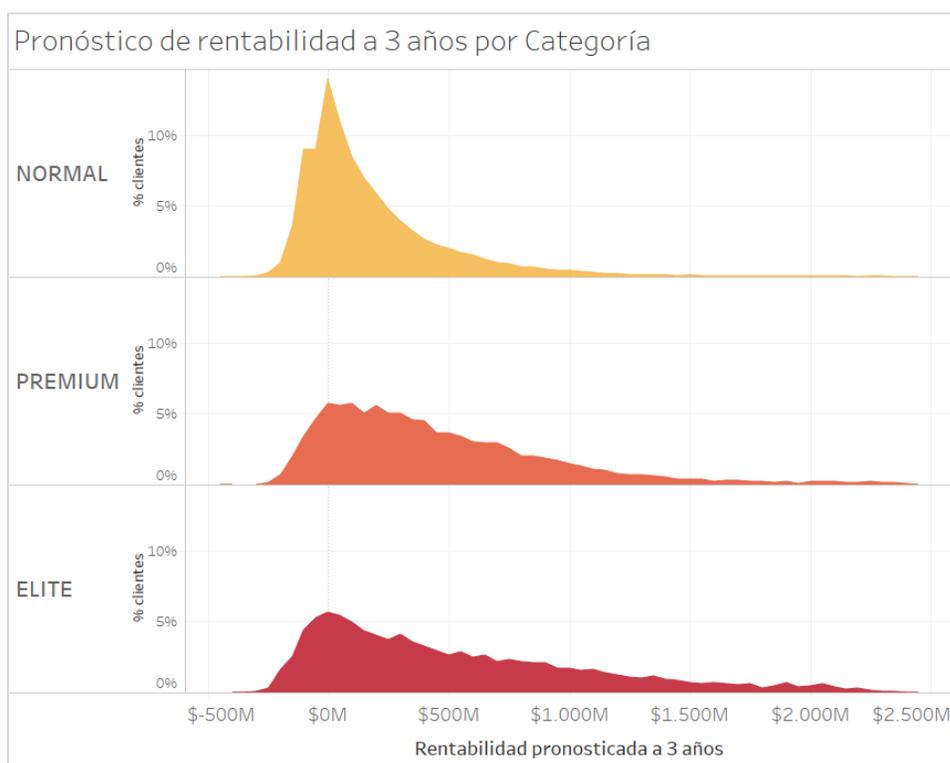


Ilustración 19. Pronóstico de rentabilidad a 3 años (2018-2020) por Categoría (2017)

Los promedios de rentabilidad a 3 años por categoría se muestran en la siguiente tabla. Se espera que los clientes Elite sean los más rentables en promedio durante los próximos 3 años, bastante distinto a lo observado en los 3 años anteriores. Esto se explica porque los Elite del año 2017 son muy diferentes a los Elite que existían en el 2014, ahora son mejores clientes, consumen más créditos de consumo, más avances, más cuotas y utilizan menos puntos que antes.

Categoría	N clientes	Rentabilidad promedio (M\$)
Normal	2.283.000	143
Premium	298.000	429
Elite	82.000	488

Tabla 11. Pronóstico de rentabilidad a 3 años (2018-2020) por Categoría

6.2.3.3 Pronóstico de rentabilidad por Productos financieros

Los ingresos que genera el cliente hoy están muy correlacionados con las rentabilidades futuras del cliente. En el grosor de la cartera estos ingresos provienen de dos aristas principalmente: productos financieros y compras en cuota con la tarjeta de crédito. La siguiente tabla describe los promedios entre clientes abierto por consumo de productos financieros en el último año. Al igual que lo observado en los descriptivos previos al modelamiento, estos productos marcan grandes diferencias, ya que generalmente son de altos montos, en alta cantidad de cuotas y con mayores intereses que la tarjeta misma.

Quienes tienen créditos de consumo se espera que renten a 3 años 576 M\$ más que los que no consumen productos financieros y quienes tienen avances en efectivo lo hacen en 425 M\$ más.

Tiene PPF*	% Clientes	Rentabilidad promedio (M\$)
No tiene	75,6%	76,7
Avance en efectivo	15,4%	501,6
Crédito de consumo	15,5%	652,7

Tabla 12. Pronóstico de rentabilidad a 3 años (2018-2020) por Productos financieros

*Consumo de crédito y avance no son excluyentes.

6.2.3.4 Pronóstico de rentabilidad por Cuotas

Siguiendo la línea del punto anterior, se analiza la rentabilidad pronosticada a 3 años según el uso de cuotas con la tarjeta de crédito. Se encuentra que con la variable máxima cuota se pueden crear reglas simples y que marcan grandes diferencias entre los resultados. La siguiente tabla exhibe los pronósticos promedios de los clientes separados según la máxima cuota con la que compraron en el último año. Se nota que el 47,5% de los clientes que ha comprado en 6 cuotas o más genera en promedio 281 M\$ más que quienes no utilizaron tal cantidad de cuotas durante el periodo.

Clientes con máxima cuota*	Tiene compra en cuota:		
	>=3	>=6	=>9
Porcentaje de clientes	78,8%	47,5%	22,8%
Rentabilidad prom. a 3 años (M\$)	278,5	371,9	472,8
Rentabilidad grupo con máx. cuota menor** (M\$)	24,7	91,3	151,4
Diferencia versus grupo con cuota menor** (M\$)	253,9	280,5	321,3

Tabla 13. Pronóstico de rentabilidad a 3 años (2018-2020) por Máxima cuota

*Con al menos una compra en el año en cuotas superior o igual a las analizadas.

**Clientes con al menos una compra en el año.

6.2.3.5 Pronóstico de rentabilidad por Canjes

Por último y en vista del análisis de causalidad, se observa la distribución de rentabilidad pronosticada abierta por clientes que canjearon y que no canjearon en el último año. Se evidencia que los primeros generan mayores rentabilidades para la empresa, a diferencia de los que no canjearon que están cargados hacia valores relativamente bajos o

negativos. Aproximadamente 752 mil clientes (28%) canjearon en el año 2017 y rentaron 370 M\$ en promedio, más de 3 veces el promedio de los que no canjearon (112 M\$).

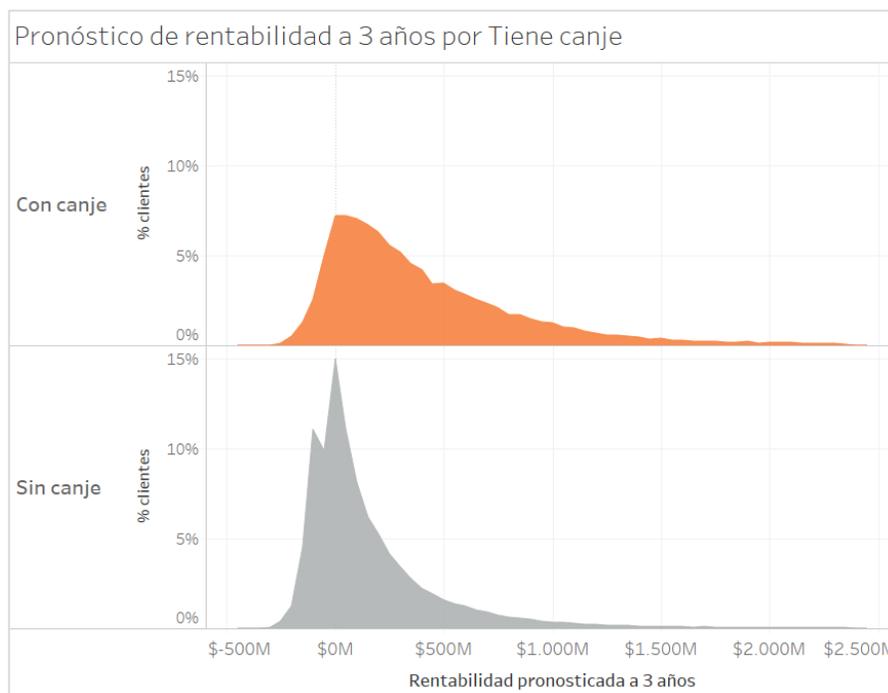


Ilustración 20. Pronóstico de rentabilidad a 3 años (2018-2020) por Tiene canje (2017)

6.2.4 Análisis descriptivos retención

La segunda componente del CLV es la retención. En esta sección se presentan algunos descriptivos en relación con la supervivencia de los clientes en la empresa.

Cabe destacar que el criterio de fuga definido por la empresa es de 12 meses sin compras, pagos, ni consumo de avances o de créditos de consumo. Los clientes que superan este tiempo de inactividad son definidos como fugados. En caso de tratarse de un proyecto que tiene como objetivo prevenir la fuga de los clientes, parecer ser más razonable utilizar un *recency* menor (e.g. 3 meses) y/o personalizado (e.g. con *recency/frequency*) para así agilizar la gestión del cliente antes de que sea demasiado tarde. En cambio, en este trabajo lo que se busca es identificar la fuga (cuasi) permanente de los clientes y el considerar un periodo de 12 meses es suficientemente amplio para abarcar a quienes compran de manera estacional, pero con regularidad (al menos una vez al año), y suficientemente pequeño para excluir de las proyecciones financieras a clientes efímeros. A pesar de no presentarse una justificación cuantitativa, este criterio posee cierta lógica, es comprensible y avalado por los expertos del negocio para abordar el objetivo.

En base al criterio mencionado en Anexo 2 se ilustra la permanencia de los clientes en la empresa con una muestra de 100 individuos. Se identifica la primera transacción de cada cliente como momento de inicio y la última transacción como momento de fuga o censura

según sea el caso. Si el cliente supera el tiempo de inactividad de 12 meses es clasificado como fugado, mientras que el resto de casos corresponden a clientes que aun no se sabe cuándo se fugarán, por lo que son observaciones censuradas. Para estudiar los tiempos de permanencia con presencia de datos censurados es necesario señalar que existe un campo de la estadística llamado análisis de supervivencia que se especializa en este tipo de análisis. En particular, se decide utilizar el estimador de Kaplan-Meier como herramienta exploratoria, ya que permite representar gráficamente la supervivencia de los individuos a través de una función escalonada teniendo en cuenta la censura.

6.2.4.1 Retención

El siguiente gráfico muestra la curva de retención para un cliente promedio (del horizonte observado). Al primer año el porcentaje de fugados es de 2,5%. Se evidencia que la pendiente disminuye su inclinación entre los años 5 y 7. Esto sugiere de que a mayor permanencia los clientes disminuyen su probabilidad de fuga y tienden a mantenerse en la empresa. Finalmente, un cliente promedio tiene un 92% de probabilidad de mantenerse activo hasta el séptimo año según el criterio de fuga definido.

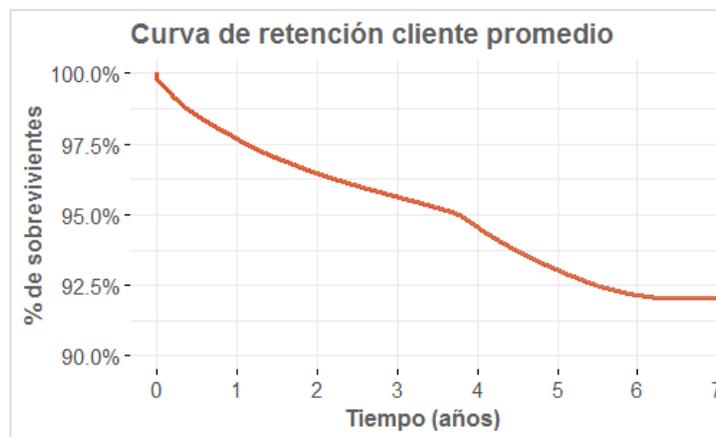


Ilustración 21. Curva de retención (Kaplan-Meier) cliente promedio

6.2.4.2 Retención vs Categoría/Gasto

En términos de categoría (relacionada con el gasto por construcción), las categorías más exclusivas o de mayor gasto posee una mayor retención esperada, sobre el 97% a 7 años. Si bien la categoría Elite en promedio es poco rentable, son clientes que se mantienen en el largo plazo, por ende, pueden ser clientes de muy alto valor para la empresa, lo que se espera observar con el modelo de CLV. Por otro lado, los clientes del quintil 5 y 4 son

clientes de retención relativamente baja que por construcción son clientes que utilizan poco la tarjeta y brindan muy baja rentabilidad.

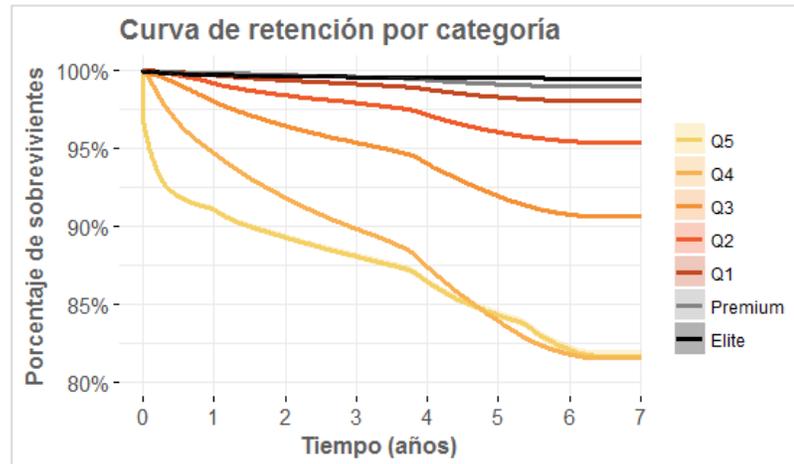


Ilustración 22. Curva de retención (Kaplan-Meier) por categoría

6.2.4.3 Retención vs Canjes

Al abrir la curva de retención por canjes, se puede observar que los clientes que canjean son más propensos a mantenerse en la empresa durante mayor tiempo. Se nota una gran caída en la retención de los clientes que canjearon 3 o más veces en su último año. Esto puede darse porque los clientes liquidaron sus puntos y abandonaron la tarjeta, o por algún evento en particular que haya ocurrido en el periodo de observación.

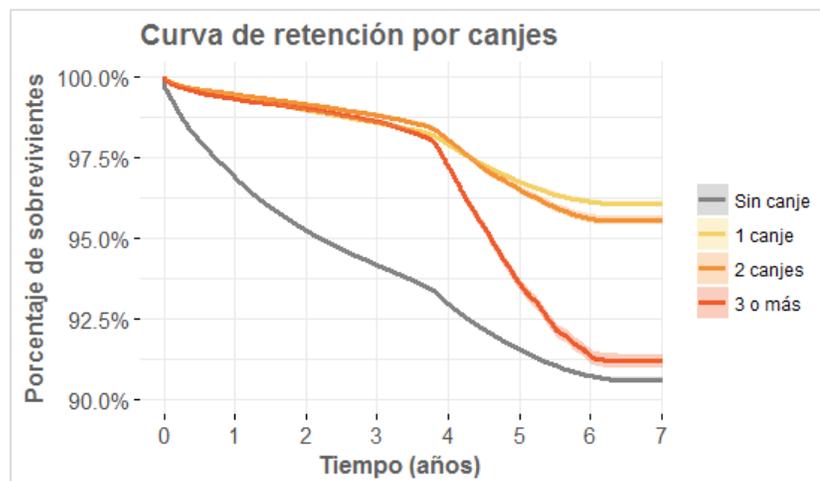


Ilustración 23. Curva de retención (Kaplan-Meier) por canjes último año

6.2.5 Modelo de retención

El modelo de retención consiste en ajustar distribuciones de probabilidad conocidas a estas curvas de retención vistas en los descriptivos. Se incorpora heterogeneidad

mediante las covariables del cliente en su última transacción observada. De esta manera, se pueden estimar las probabilidades de retención para el futuro de los clientes a partir de su actividad y características actuales.

Las variables del modelo se muestran en la siguiente tabla. Estas fueron seleccionadas al demostrar significancia (al 0,05) en el análisis de varianza (ANOVA), además de resultar en el modelo de mejor calidad de ajuste al minimizar el AIC. Este último criterio busca maximizar la verosimilitud penalizando por el número de parámetros, evitando así la complejidad y el sobreajuste del modelo (más detalle en 3.1.1: vii).

Variables	Variables
Categoría_2	Ratio deuda / cupo
Tramo edad	Gasto en negocio interno: supermercado
GSE	Gasto en comercios externos
Tiempo activo (años)	Stock de puntos
Tiene seguro desgravamen	Número de canjes últ. año
Tiene seguro cesantía	Categoría_2=Q5 * Con canje últ. año
Tiene adicionales	Categoría_2=Q4 * Con canje últ. año
Tiene crédito de consumo	Categoría_2=Q3 * Con canje últ. año
Tiene avances	Categoría_2=Q2 * Con canje últ. año
Tramo de monto cupo	Categoría_2=Q1 * Con canje últ. año
Tramo % del cupo disponible	Categoría_2=Premium * Con canje últ. año
Meses en mora	Categoría_2=Elite * Con canje últ. año
Tramo score de riesgo	

Tabla 14. Variables modelo de retención

En el siguiente cuadro se exhiben los resultados para los modelos probados. Estos resultados corresponden al ajuste de los modelos al total de la base como entrenamiento. Como primera observación, se evidencian claras mejoras en el ajuste al incorporar heterogeneidad mediante las covariables de los clientes. Se tiene que el modelo con mejor calidad de ajuste (menor AIC) es la distribución de Weibull con covariables. Este criterio es relativo a los otros modelos y por si solo no aporta información.

Modelo	AIC sin heterogeneidad	AIC con heterogeneidad
Exponencial	910.175	636.567
Weibull	892.690	604.703
Log-normal	894.906	629.312
Log-logístico	892.718	609.022

Tabla 15. Desempeño modelos de retención (Criterio AIC)

El test de log verosimilitud de Chi-cuadrado para el modelo Weibull resulta en un estadístico χ^2 de 891.998 en 55 grados de libertad, con un p-valor igual a 0. Por lo que el modelo es significativamente mejor que el modelo nulo (sólo con el intercepto).

Los coeficientes del modelo se encuentran en Anexo 3. Si bien no es posible interpretar los coeficientes directamente, la transformación $\exp(-\beta/\gamma)$ permite analizar el efecto de las variables, donde β es el coeficiente de la variable y γ el parámetro de escala (Klein, 2003). Esta transformación se conoce como Ratio de tiempo esperado (ETR, por las siglas en inglés) y se interpreta así, ETR=1,06 significa que cada unidad de X aumenta en un 6% el tiempo de vida esperado. De estos *ratios* se destaca lo siguiente:

- Por cada 100 M\$ gastados en el supermercado del *holding* y en negocios externos, aumentan la esperanza de vida en la empresa en 8% y 3% respectivamente.
- Por cada año activo en la empresa, los clientes aumentan en un 97% su tiempo de vida esperado en la empresa.
- Los clientes Premium y Elite se espera que se mantengan en la empresa un 148% más de tiempo que los clientes Normales del quintil de más bajo gasto.
- En relación con el programa de fidelización, cada canje disminuye en un 9% el tiempo de vida esperado, quizás porque los clientes liquidan los puntos y dejan la empresa o les cuesta retomar la motivación de utilizar la tarjeta. Sin embargo, hay que ser precavido en la interpretación de este coeficiente, ya que esta variable está relacionada con otras del modelo y existe una especie de efecto combinado que se debe tener en cuenta.
 - A mayor gasto, el cliente posee mayor esperanza de vida, pero también posee mayor propensión de canjear.
 - A mejor categoría, el cliente posee mayor esperanza de vida, pero también posee mayor propensión de canjear.

Dada esta última observación, en el siguiente sub-apartado se presentan análisis descriptivos a cerca de la retención esperada de la cartera de clientes de diciembre 2017 para facilitar la interpretación misma del modelo.

6.2.6 Resultados modelo de retención

El modelo de retención entrega una curva de probabilidades para cada cliente. Esto complica la visualización gráfica de los resultados. Por lo que, a continuación, se presenta análisis particulares para comprender mejor el modelo y los resultados para la cartera de diciembre de 2017.

6.2.6.1 Duración

El modelo de retención no pronostica el tiempo hasta la fuga, sino que las probabilidades de supervivencia para cada año futuro, entonces no es posible estimar cuándo va a fugarse un cliente. Teniendo esto en cuenta, el siguiente gráfico presenta un histograma

de duración de los clientes actualmente activos en la empresa, sujeto al supuesto de que el cliente se fuga en el primer periodo que posea probabilidad de supervivencia menor a 50%. Este valor se conoce como la mediana del tiempo de vida y es normalmente utilizado para entregar un tiempo de vida esperado.

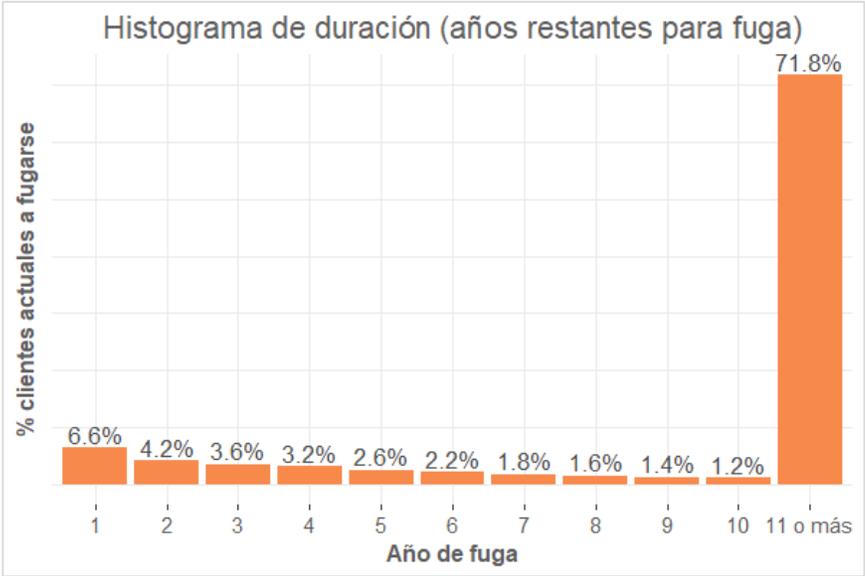


Ilustración 24. Histograma de duración cartera 2017-12

*Año de fuga: primer periodo del cliente con probabilidad de supervivencia menor a 50%

**Se omiten 307 mil clientes ya fugados

De 2,66 millones de clientes se omiten 307 mil clientes que ya están en estado fugado. Respecto a los clientes censurados, se espera que un 6,6% de los clientes se fuguen de la empresa en el primer año, este porcentaje es decreciente en los siguientes periodos porque los clientes que llevan más tiempo tienen mayor probabilidad de mantenerse. Finalmente, se espera que alrededor de 1,69 millones de clientes de diciembre 2017 (71,8%) se mantengan por 11 años o más. Recordando nuevamente que esta estimación está sujeta a la probabilidad de supervivencia mencionada anteriormente.

6.2.6.2 Perfiles de mayor y menor duración

En Anexo 4 se aprecian las probabilidades de supervivencia promedio agrupadas por deciles y se observa que 6 de estos tienen más de 80% de probabilidades de mantenerse en la empresa durante los próximos 10 años, mientras que hay 3 deciles con probabilidad casi nula (<2%) y un decil intermedio con una probabilidad de supervivencia de 40%. Con el fin de conocer qué características y comportamientos presentan los clientes de mayor y menor retención, se selecciona y caracteriza al decil de mayor y menor probabilidad de supervivencia respectivamente, según las probabilidades estimadas para un plazo de 10 años. Para llevar a cabo esta caracterización, se exhibe el *Lift* (promedio del decil sobre el promedio de la cartera) de las covariables del modelo en que se presentan mayores diferencias. Esto permite comparar directamente las características entre ambos segmentos como también respecto al cliente promedio de la empresa.

Variable	Lift respecto el promedio de la cartera	
	Decil de menor duración	Decil de mayor duración
Gasto en negocio interno: supermercado (100 M\$)	0,2	11,0
Gasto en negocios externos (100 M\$)	0,1	18,3
Número de canjes	0,2	3,7
Stock de puntos	0,3	9,6
Tiene crédito de consumo =1	0,1	2,0
Tiene adicionales =1	0,6	2,8
GSE		
ABC1	1,2	2,3
C2	1,0	1,4
C3	0,8	1,0
D	0,7	0,9
E	1,1	0,5
Categoría_2		
Elite	0,2	22,7
Premium	0,2	2,3
Q1	0,2	0,2
Q2	0,5	0,0
Q3	0,9	0,0
Q4	1,3	0,0
Q5	2,8	0,0
Prob. supervivencia a 1 año (promedio)	1,79%	99,9%
Prob. supervivencia a 10 años (promedio)	0,00%	99,9%

Tabla 16. Caracterización perfiles de mayor y menor duración

La tabla se interpreta de la siguiente manera: los clientes de mayor duración gastan 18,3 veces lo que gasta el promedio de la cartera en negocios externos, mientras que en el otro extremo los clientes de baja duración gastan 0,1 veces el promedio de la cartera en estos comercios. Se observa que la retención está marcada por dos extremos de clientes. El perfil de mayor duración que corresponde a clientes de alto gasto, de categorías exclusivas y segmentos socioeconómicos altos, tienen una mayor tasa de consumo de créditos, cuentas adicionales y poseen un stock de puntos relativamente alto con el que canjean frecuentemente. En cambio, el perfil de menor duración corresponde a clientes de las categorías Q4 y Q5 (muy bajo gasto), tienen bajo consumo de créditos y no canjean. En conclusión, los clientes muy activos, con alta capacidad de gasto y participación en el club de puntos están fidelizados y poseen alta retención, y los clientes de muy baja actividad, que no consumen productos ni canjean, son clientes muy propensos a la fuga.

6.2.6.3 Variables relevantes

En la siguiente tabla se exhiben las probabilidades de supervivencia promedio al analizar a los clientes según variables del último año relevantes para el modelo y el proyecto.

Variable	% de clientes	Probabilidad de supervivencia al año:	
		5	10
Categoría_2			
Elite	3%	0,94	0,91
Premium	11%	0,93	0,88
Q1	19%	0,91	0,83
Q2	18%	0,83	0,74
Q3	17%	0,73	0,63
Q4	15%	0,60	0,51
Q5	17%	0,18	0,15
Productos financieros			
No tiene	76%	0,62	0,55
Tiene avance	15%	0,90	0,82
Tiene crédito de consumo	16%	0,96	0,92
Canjes			
Sin canje	72%	0,61	0,53
1 canje	17%	0,90	0,84
2 canjes o más	11%	0,94	0,89

Tabla 17. Probabilidades de supervivencia abierto por variables relevantes

Reafirmando lo observado en análisis previos, se observa que la categoría (que se asocia al gasto) está muy relacionada con la retención esperada y a mejor categoría las probabilidades de mantenerse en la empresa son mayores. Por otro lado, se tienen que los productos financieros también están asociados a un mayor compromiso del cliente con la empresa y, por ende, se espera que estos clientes se mantengan durante mayor tiempo tanto por una relación de necesidad como de preferencia. En cuanto a canjes, también se observa una correlación entre la cantidad de canjes y la retención esperada, esto parece razonable ya que es lo deseable de un programa de fidelización.

6.2.7 Resultados CLV

Al combinar ambas componentes, de rentabilidad y retención hasta 10 años, se obtiene la estimación del CLV. En lo siguiente se describen los resultados obtenidos para la cartera de diciembre de 2017, primero de manera general y luego abierto por variables relevantes.

6.2.7.1 CLV

Como primer resultado, se presenta el histograma del valor de los clientes a 10 años para la cartera de diciembre de 2017. A diferencia de la rentabilidad a 3 años, la curva se expande hacia la derecha por la incorporación de la componente de retención.



Ilustración 25. Distribución del Customer Lifetime Value (2018-2028)

Promedio (M\$)	Desv. est. (M\$)	Q1 (M\$)	Mediana (M\$)	Q3 (M\$)
834	1.227	35	348	1.145

Tabla 18. Descriptivos Customer Lifetime Value (2018-2028)

Como estadísticos básicos, se tiene un promedio de 834 M\$ con una desviación típica de 1,45 veces el promedio. El primer y tercer cuartil indican que la mitad de la cartera se encuentra en el rango de 35 M\$ y 1,15 MM\$, con una mediana de 348 M\$ que indica que los valores están más cargados hacia abajo, tal y como se muestra en la distribución.

6.2.7.2 CLV por Categoría/Gasto

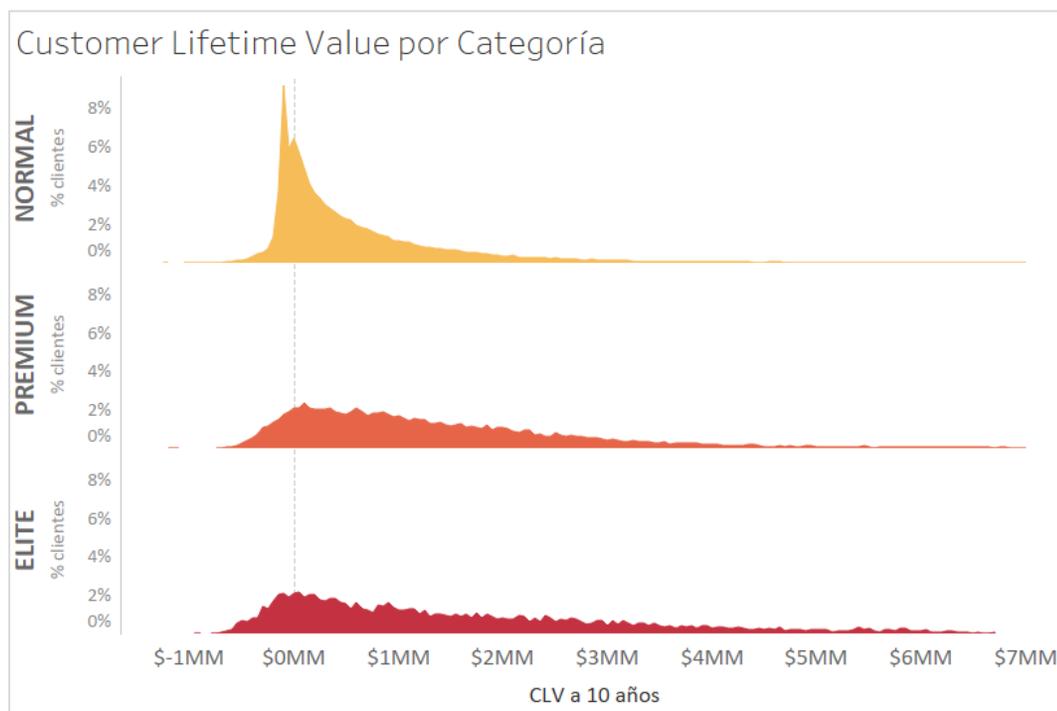


Ilustración 26. Distribución CLV (2018-2028) por Categoría (2017)

Categoría	Promedio (M\$)	Desv. est. (M\$)	Mediana (M\$)
Normal	401	738	134
Premium	1.188	1.223	889
Elite	1.374	1.527	940

Tabla 19. Descriptivos CLV (2018-2028) por Categoría (2017)

Se observa que la distribución de la categoría Normal está cargada hacia la izquierda, mientras que las categorías Premium y Elite poseen distribuciones más planas, con una mayor proporción de valores sobre el millón de pesos. Esto es coherente con los resultados de los sub-modelos de rentabilidad y retención, ya que los clientes de estas dos últimas categorías poseen mayores rentabilidades y mejores tasas de retención, lo que extiende su distribución a valores de rentabilidad más altos. Si bien los Elite poseen mayor dispersión que los Premium, también tienen un promedio de rentabilidad y una mediana mayor, por lo que son la categoría más rentable en el largo plazo.

6.2.7.3 CLV por Canjes

Al abrir los resultados por clientes con y sin canje en el año previo, se observa que las distribuciones difieren bastante. Los clientes con canje se concentran mayormente cerca de cero con un *peak* importante, mientras que los clientes que canjearon poseen una distribución más dispersa y con valores de rentabilidad más altos en promedio.

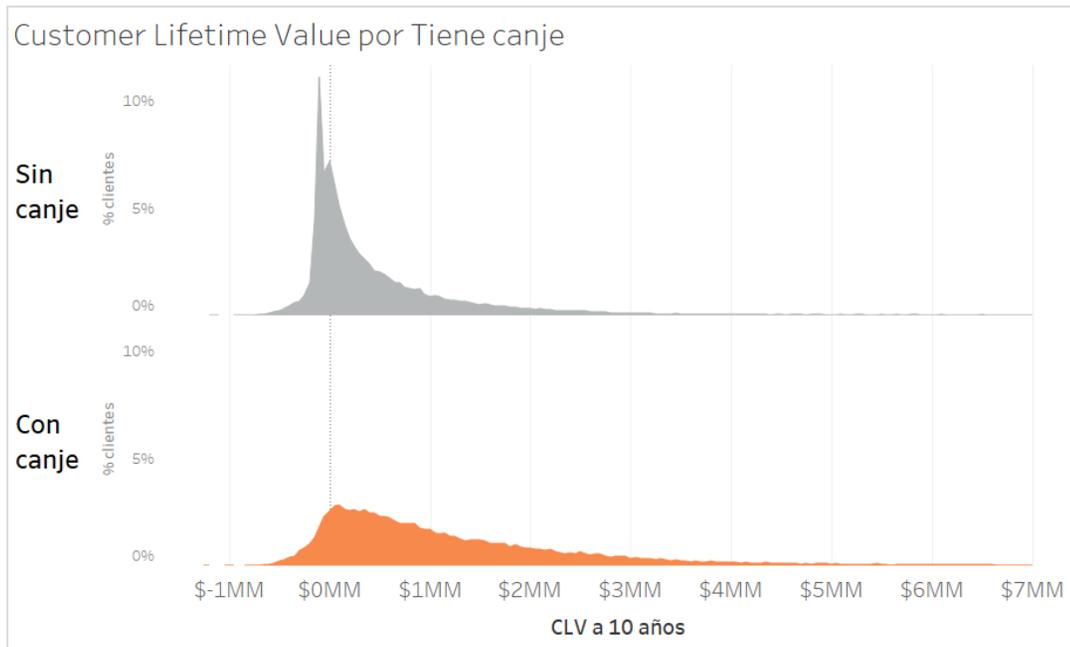


Ilustración 27. Distribución CLV (2018-2028) por Canje (2017)

Tiene canje	Promedio (M\$)	Desv. est. (M\$)	Mediana (M\$)
Sin canje	560	971	179
Con canje	1.562	1.487	1.107

Tabla 20. Descriptivos CLV (2018-2028) por Canje (2017)

Los clientes que no canjearon y si canjearon poseen un valor promedio respectivamente menor y mayor al promedio de la cartera (834 M\$).

Estos resultados indican de forma preliminar de que el canje si tiene un impacto positivo. Sin embargo, para asegurar la existencia de un efecto y conocer su magnitud es necesario utilizar métodos especializados para la inferencia de causalidad. Quizás estas diferencias se dan por casualidad más que causalidad, ya que los clientes pueden pertenecer a un perfil rentable desde antes de canjear o por alguna otra razón correlacionada. Por esto, la siguiente etapa busca abordar este problema y estimar el efecto causal minimizando los sesgos de autoselección.

6.3 Estimación de efecto causal

Para la base de clientes a analizar se consideran clientes cuenta abierta en diciembre de 2014 y con al menos un año de antigüedad, para así identificar los clones en base a las covariables de ese mismo año. Por otro lado, recordar que la marca de actividad se realiza para un periodo de dos años (2015-2016) debido a la duración de los puntos, por lo que también se exige que los clientes sean cuenta abierta en diciembre de 2016. Así no se subestima los resultados del grupo control, ya que por construcción los clientes identificados como activos en el programa serán en su mayoría cuenta abierta, no así los inactivos. Con esto, se obtiene una base de 2,12 millones de clientes para el estudio.

6.3.1 Definición de activos

El criterio de actividad en el club de puntos se construye de forma arbitraria como combinación de 2 métricas de interacción con el club de puntos que buscan: no sesgar por potencial de gasto, es decir, todos pueden participar independiente de los puntos que pueda acumular dada su capacidad de gasto; y que indique en un periodo de tiempo determinado si el cliente participó del club de puntos o no. En base a lo anterior y el análisis de variables disponibles, se seleccionan las métricas: número de canjes y tasa de vencimiento. La primera indica interacciones directas y, la segunda, si el cliente está pendiente de sus puntos en el periodo observado. Como se menciona en Transformación de variables, la tasa de canje utilizada por la empresa posee una limitante al indicar un comportamiento rezagado, por esto es descartada.

Para definir la regla de cliente activo, primero se realiza una limpieza de *outliers* con Mahalanobis al percentil 99,9% de la chi-cuadrado eliminando 1,15% de casos extremos, teniendo en cuenta sólo estas dos variables: número de canjes y tasa de vencimiento. Luego, se utiliza k-Medias para realizar la separación de activos e inactivos, ya que este método permite encontrar el corte óptimo de manera automatizada, minimizando la distancia entre cada cliente hacia el centro de su grupo. Como resultado, se obtiene la regla como combinación lineal de ambas variables (Ilustración 28), en donde, a menor cantidad de canjes se exige una menor tasa de vencimiento y viceversa, y a partir de 7 canjes (en 2 años) no se exige tasa de vencimiento.

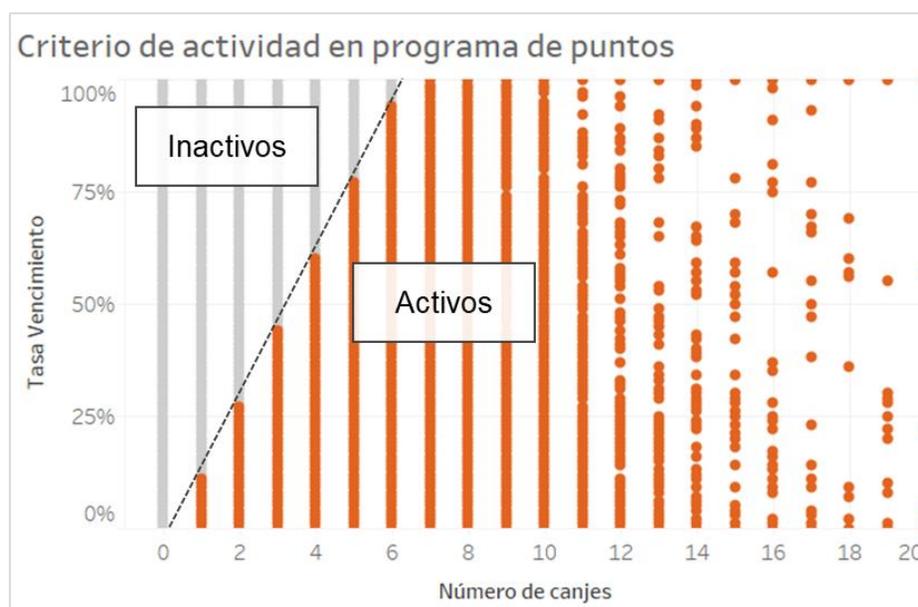


Ilustración 28. Criterio de actividad club de puntos (2015-2016)

De una base total de 2,12 millones de clientes, se tiene un 16% de activos en el programa de fidelización durante el horizonte considerado.

6.3.2 Efecto promedio: estratificación tradicional

Con el CLV y la definición de activo en el club de puntos es posible realizar un contraste estratificado como primer acercamiento al efecto asociado. Este acercamiento consiste en estratificar a los clientes según sus variables del año previo (2014) al periodo de actividad, para así aproximar las distribuciones de los clientes activos e inactivos en estos aspectos del comportamiento. Para este método se deben seleccionar variables categóricas y de manera arbitraria, la cantidad de variables y niveles no debe ser extremadamente grande dado que puede volverse rápidamente infactible el balance entre estratos del grupo activos e inactivos. Como se espera estimar el efecto asociado al club, entonces se seleccionan dimensiones previas que sean relevantes para la participación en este programa. Por lo tanto, se utilizan las variables: “tiene canje”, “stock de puntos”, “gasto” y “monto cupo” que están directamente relacionadas a su actividad previa con los puntos como también de su potencial de canje y acumulación; “tiene crédito de consumo” que representa un compromiso explícito de largo plazo; y “rentabilidad” que representa el comportamiento comercial generalizado del cliente con la empresa. Como resumen del proceso de estratificación, se tienen 1.523 y 1.762 estratos en el grupo activo e inactivo respectivamente, de los cuales cruzan 1.407 estratos con al menos un cliente. A continuación, se exhibe el contraste de resultados sin y con estratificación, en donde sin estratificación corresponde a la base completa de activos e inactivos.

Base completa sin estratificación*	Periodo previo			Periodo posterior			Efecto
	GC	T	Dif.	GC	T	Dif.	
Rentabilidad (M\$)	51,91	134,55	82,64	60,51	157,05	96,54	13,90
CLV (M\$)	418,35	1.208,54	790,19	425,71	1.163,66	737,94	-52,24
Muestra estratificada							
Rentabilidad (M\$)	105,45	103,25	-2,20	120,32	105,43	-14,89	-12,69
CLV (M\$)	842,07	952,32	110,25	771,72	877,69	105,98	-4,27

GC: Promedio grupo control, T: Promedio grupo tratamiento, Dif.: Diferencia promedio

*Tabla 21. Resultados con muestra estratificada
Base completa de clientes, activos e inactivos.

Como primera observación, las diferencias iniciales (2014) entre los clientes activos e inactivos eran de 82,64 M\$ y 790,19 M\$ en rentabilidad y CLV respectivamente, y se reducen a -2,20 M\$ y 110,25 M\$ como consecuencia de la estratificación. Por otro lado, las diferencias de 2017 también se reducen en magnitudes similares. Como resultado, se calculan los efectos como las diferencias en el crecimiento de estas variables. Se observa que la rentabilidad cambia de signo con la estratificación, de 13,90 M\$ a -12,69 M\$ y el valor de los clientes en el largo plazo se reduce desde -52,24 M\$ a sólo -4,27 M\$. Al parecer el efecto del programa de puntos es relativamente pequeño y negativo, quizás porque el mismo costo de los canjes es mayor al retorno en “fidelización” que causa. Sin embargo, durante el proceso de estratificación se observó una gran sensibilidad de los resultados frente a las variables y los niveles seleccionados. Por lo tanto, no se puede concluir en los resultados con fiabilidad y es probable que aún existan sesgos en la estimación. La continua incertidumbre sobre qué variables se deben utilizar para la estratificación es un problema complejo de abordar con este método.

Como solución a lo anterior, surge la alternativa de utilizar un método especializado, el Propensity Score Matching. Este método permite realizar una especie de “estratificación individual” a través de la búsqueda de clones. Este se basa en un modelo de propensión que prioriza las variables según su relevancia con el tratamiento. En los siguientes apartados se procede con este método, primero se modela la propensión a ser activo del club de puntos, se implementa el algoritmo de emparejamiento, para luego finalizar con los resultados respectivos.

6.3.3 Propensity Score

El modelo de propensión a ser activo en el programa de puntos, en 2015-2017, se construye con las variables del cliente en el año anterior. En la siguiente tabla se muestran las 14 variables de mayor relevancia para el modelo de regresión logística, que

significan el 81% de la importancia. Este tiene un *accuracy* de 84,9%⁵, en Anexo 5 se muestra la matriz de confusión y en Anexo 6 se muestra el detalle completo del modelo. Aunque cabe recordar de que el único fin de este es el emparejamiento y no su uso predictivo, por lo que la simpleza del modelo no es prioridad y puede existir correlación entre las variables complicando su interpretación.

Variable	Importancia
Número de canjes	17%
Tasa de canje	16%
Categoría	11%
Porcentaje pagado con la tarjeta en negocios internos	7%
Meses con compra	6%
Gasto 12 meses (\$100 miles)	6%
Distancia al próximo nivel de canje	4%
Compras en negocio interno: tiendas por depto.	4%
Máxima cuota	3%
Con compra en rubro supermercados	2%
Compras en negocio externos	2%
Porcentaje de transacciones con cuotas	2%
Compras en negocio interno: mejoramiento del hogar	1%
Porcentaje del gasto en semestre 2	1%

Tabla 22. Variables más relevantes (81%) para modelo de propensión

Como es de esperar, las variables relacionadas a canjes, categoría y uso de la tarjeta en negocios internos al *holding* son relevantes para la propensión de un cliente a activarse en el club de puntos. Por lo que, al emparejar a través del puntaje de propensión implícitamente se asignarán clones priorizando la igualdad en estas variables. Esto soluciona, en parte, la incertidumbre sobre la selección de variables a la que se incurre con el método de estratificación clásico (6.3.2).

Se selecciona el modelo regresión logística, ya que resulta en el mejor balance de las covariables a pesar de no poseer la mejor capacidad predictiva. En Anexo 7, Anexo 8 y Anexo 9 se presenta gráficamente el balance de las covariables con la diferencia de medias estandarizadas (DME) para el árbol C5, árbol CHAID y regresión logística respectivamente, en donde se muestran las diferencias antes del aplicar *matching* (línea rosa) y al aplicar *matching* con 1 vecino más cercano (línea celeste). A modo de resumen, la siguiente tabla exhibe estadísticos acerca de estas diferencias para cada modelo. Se

⁵ En base de entrenamiento, equivalente al total de la base. No se utiliza partición de testeo dado que el único fin del modelo es calcular los puntajes de propensión para el emparejamiento, no realizar pronósticos.

evidencia que la regresión logística resulta en un balance ideal con ninguna variable sobre las 0,1 desviaciones estándar de diferencia.⁶

Modelo	Prom. DME	Máx. DME	Variables con DME>0,1	Variables con DME>0,25
Árbol C5	0,060	0,154	7	0
Árbol CHAID	0,097	0,457	24	2
Reg. Logística	0,018	0,073	0	0

Tabla 23. Resumen diferencias de medias estandarizadas según modelo de propensión (Muestra 25% con 1 vecino más cercano)

6.3.4 Matching

i. Elección de los parámetros

Calculado el puntaje de propensión es turno de aplicar el algoritmo de *matching*. Para esto primero se definen los siguientes 3 parámetros:

- K vecinos más cercanos: cantidad de clientes inactivos que se van a asignar a cada cliente activo. Mientras mayor 'k', mayor es la representación que tiene el grupo de control y, por ende, la confianza en su resultado. Sin embargo, puede ocurrir que se asignen clientes muy distantes o que se descarten clientes tratados al no encontrar sus 'k' clones. Se realiza un análisis de sensibilidad para encontrar el 'k' adecuado.
- Radio de Caliper: distancia máxima para asignar clones. Funciona como complemento al número de vecinos y soluciona el problema de que se asignen clientes distantes. Todo cliente activo que no posee 'k' inactivos dentro de su radio de propensión, es descartado del análisis. Este radio se establece en 0,2 desviaciones del logit de la propensión, porque minimiza sesgo (Austin, 2011).
- Reposición: es posible reutilizar a los clientes de control para ser emparejado con más de un tratado. La opción de reposición no es recomendada porque puede aumentar el sesgo, a menos que la proporción de candidatos de control sobre tratados sea muy baja. En este caso, se tienen 5,25 candidatos de control por cada tratado, por lo tanto, no es necesaria esta opción.

Con el radio de Caliper fijo y la opción de reposición descartada, se realiza un análisis de sensibilidad variando el número de 'k' vecinos. La siguiente tabla resume los resultados del balance a tener en consideración para seleccionar el número 'k' adecuado.

⁶ El tiempo de procesamiento del algoritmo crece exponencialmente con el tamaño de la base, por lo que se utiliza una muestra del 25% para escoger el modelo de propensión y los parámetros del *matching*. Con esta muestra el algoritmo demora aprox. 1 hora, con el total de la base 13 horas.

K vecinos =	1	2	3	4	5
DME promedio	0,018	0,022	0,025	0,026	0,028
DME máximo	0,073	0,100	0,126	0,142	0,158
Variables con DME>0.1	0	0	1	1	2
Variables con DME>0.25	0	0	0	0	0
Porcentaje de activos emparejados	62%	50%	42%	38%	34%

Tabla 24. Sensibilidad según K vecinos más cercanos
(Muestra 25% con regresión logística)

Se observa que a mayor K empeora el balance de las covariables, aumentando el DME promedio y el DME máximo. En general, el primer indicador se encuentra muy por debajo de 0,1 y el número de variables que superan este valor es muy pequeño en relación con la alta cantidad de variables del modelo de propensión, mientras que ninguna variable supera el límite máximo de 0,25, por lo que existe un balance general muy bueno.

La observación más interesante es el porcentaje de activos que son emparejados y, en consecuencia, utilizados para el análisis final. Empieza en 62% y luego decrece mientras más vecinos se exigen, esto se da porque la razón entre inactivos sobre activos es relativamente baja en los tramos de alta propensión. Esto se puede visualizar gráficamente en el Anexo 10. En conclusión, la opción más adecuada es utilizar K=1, porque a mayor cantidad de vecinos el porcentaje de activos utilizados empieza a decaer disminuyendo la representatividad del grupo tratado sin mejorar el balance.

Cabe recordar que, dado los tiempos de procesamientos, se utiliza una muestra para definir los parámetros, para luego realizar el *matching* final con el total de la base.

ii. Emparejamiento y revisión del balance de las covariables

Se procede a realizar el *matching* con 1 vecino más cercano y un radio de Caliper de 0,2 desviaciones del logit de la propensión. Es decir, a cada cliente activo se le asigna el cliente inactivo de mayor similitud dentro de una distancia máxima. De esta forma, se generan grupos control y tratamiento comparables, lo cual se valida en este sub-apartado.

En la siguiente figura se exhiben las distribuciones de los puntajes de propensión de la base completa sin estratificación (izquierda) y de la muestra emparejada con el algoritmo de *matching* (derecha), separada en ambos casos por clientes activos e inactivos. Se nota que las distribuciones pasan de ser completamente distintas en un principio, a luego tener distribuciones balanceadas como consecuencia del *matching*. En el gráfico del lado izquierdo se tiene la distribución de los activos está relativamente cargada hacia valores de alta propensión (por construcción), mientras que los inactivos poseen una proporción muy alta de clientes concentrados en valores de mínima probabilidad. Luego de aplicar el algoritmo de emparejamiento, se obtiene una muestra conformada por el grupo de control y tratamiento recreados, en donde ambos grupos poseen distribuciones similares, recreando así la aleatorización del experimento.

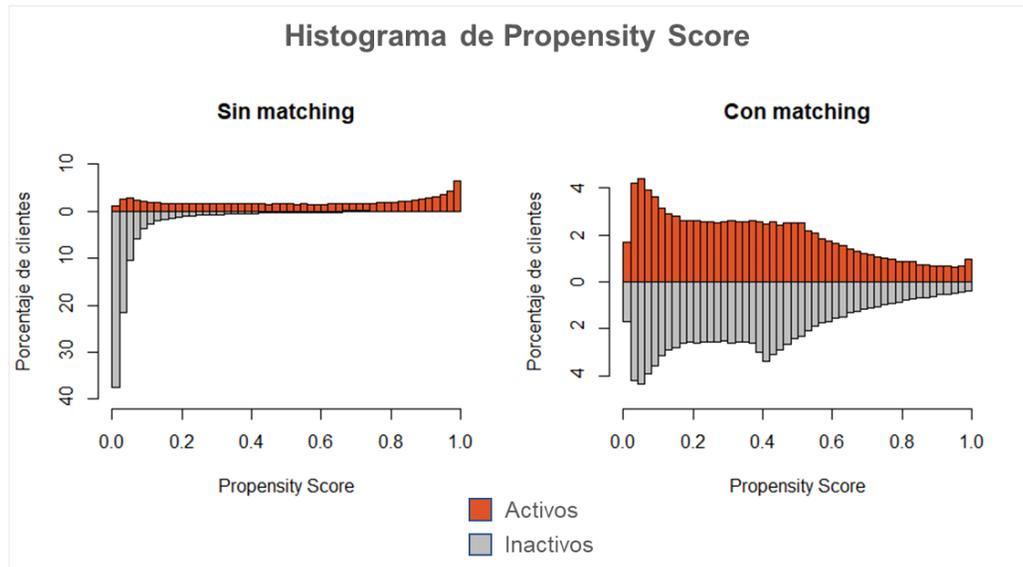


Ilustración 29. Histograma Propensity Score sin y con matching

Como el puntaje de propensión resume las variables previas y relevantes de los clientes, en términos del programa de puntos, estas variables deben quedar balanceadas entre el grupo de activos e inactivos luego del *matching*. Esto se puede comprobar en la siguiente tabla, en donde se presenta el balance de las 14 variables más relevantes para la propensión a ser activo del programa (81% de la importancia total) antes y después del *matching* para así evidenciar las diferencias iniciales y el efecto del emparejamiento.

Variables	Sin matching			Con matching		
	GC	T	DME	GC	T	DME
Clientes (miles)	343	1.780		214	214	
Propensity Score	0,09	0,56	1,887	0,36	0,37	0,037
Canjes	0,24	1,75	1,231	1,07	1,11	0,036
Tasa de canje	0,18	0,77	1,680	0,66	0,66	0,009
Categoría (%)			1,115			0,098
Elite	0,20	4,70		1,20	1,90	
Premium	96,90	55,30		80,50	76,80	
Normal	3,00	40,00		18,30	21,30	
% gasto pagado con tarjeta en negocios internos	0,67	0,91	0,821	0,89	0,88	0,047
Meses con compra	6,74	10,87	1,128	10,40	10,33	0,026
Gasto (M\$)	833	4.107	0,581	2.352	2.605	0,067
Distancia prox. nivel de canje	0,57	0,31	0,967	0,33	0,34	0,035
Compras negocio interno: tiendas p.d.	4,01	11,63	0,860	8,95	9,11	0,019
Máxima cuota	4,51	7,60	0,617	7,25	7,08	0,033
Con compra en rubro: supermercados =1 (%)	33,00	73,00	0,876	65,10	64,60	0,010
Compras negocios externos	10,03	58,24	0,700	31,74	34,85	0,054
% transacciones en cuotas	0,23	0,28	0,226	0,30	0,29	0,031
Compras negocio interno: mej. del hogar	2,35	7,45	0,547	5,28	5,86	0,067
% gasto en semestre 2	0,43	0,50	0,278	0,50	0,50	0,006

GC: Promedio grupo control, T: Promedio grupo tratamiento, DME: Diferencia de medias estandarizada

Tabla 25. Diferencias de covariables pre y post matching (14 variables más relevantes)

En esta tabla se evidencia el efecto del emparejamiento. Se observa que:

- **Tasa de canje:** los promedios de Tasa de canje para los clientes inactivos y activos pasan de 18% y 77% respectivamente, siendo la variable más desbalanceada en un inicio, a quedar con tasas de 66% y 66% para los grupos de control y tratamiento recreados.
- **Categorías:** las proporciones de la Categoría de los clientes pasan de ser completamente distintas a ser relativamente similares luego del emparejamiento, con una diferencia de medias estandarizada inicial de 1,115 desviaciones para luego quedar con un balance adecuado de 0,098 desviaciones.
- **Gasto:** la diferencia de gasto anual en un principio era de más de 3 MM\$ y luego del *matching* se redujo a 250 M\$ aproximadamente.
- **Compras negocios externos:** la diferencia de compras en negocios externos al *holding* en un principio era de casi 50 transacciones, luego del emparejamiento se redujo a sólo 3 transacciones.

En general, el balance de estas variables es excelente, ya que todas resultan en diferencias de medias estandarizadas menores a 0,1. Las diferencias iniciales en características y comportamientos de los clientes activos e inactivos se minimizan de manera transversal como consecuencia de los clones, lo que parece muy complicado de conseguir sin un método como este.

El detalle completo de las variables y sus diferencias de medias sin y con *matching* se encuentran en Anexo 11 y Anexo 12 respectivamente. En Anexo 13 se presenta un resumen gráfico de las diferencias estandarizadas para ambos casos. Se observa que, con *matching* todas las variables quedan perfectamente balanceadas con $DME < 0,1$. A excepción de la variable categórica “Tramo edad” que queda con un DME levemente superior (0,124), pero dentro de lo aceptable ($< 0,25$).

Finalmente, se concluye que los grupos control y tratamiento recreados por el *matching* con Propensity Score poseen un balance en sus covariables apropiado y, por ende, son comparables. Ahora, tal y como se trabaja con experimentos aleatorios controlados, se pueden contrastar los resultados de ambos grupos para inferir causalidad. Por lo tanto, en lo siguiente del informe sólo se procederá a trabajar con el subconjunto emparejado.

6.3.5 Efecto promedio

El siguiente paso es estudiar los resultados asociados al ser activo en el programa de puntos. Para esto, se analizan las diferencias entre los promedios de variables de interés para el negocio y que pueden ser influenciadas por esta activación en el club de puntos. Lo principal es revisar las variables de rentabilidad y CLV, ya que representan el retorno económico asociado. Aunque, con el fin de brindar una mayor comprensión también se analizan otras variables de interés.

La siguiente tabla presenta las diferencias durante los periodos previo y posterior al horizonte definido como “de activación”, 2014 y 2017 respectivamente. Se considera como efecto a la diferencia en el crecimiento de las variables.

Variable	Periodo previo			Periodo posterior			Efecto	Sign.
	GC	T	Dif.	GC	T	Dif.		
Rentabilidad (M\$)	121,4	119,0	-2,5	133,8	146,5	12,7	15,2	***
CLV (M\$)	997,7	1.022,4	24,7	935,8	1.022,1	86,4	61,7	***
Gasto (M\$)	2.352,4	2.605,3	252,9	2.417,9	2.712,1	294,2	41,3	***
en holding	1.568,3	1.636,1	67,8	888,2	1.048,1	159,9	92,1	***
en externos	784,2	969,2	185,1	1.529,7	1.664,0	134,3	-50,8	***
Días con compra	43,2	45,5	2,3	48,7	54,5	5,9	3,6	***
Rubros con compra	6,2	6,2	0,1	6,7	7,3	0,5	0,4	***
Cuota prom. pond.	3,3	3,2	-0,1	3,1	3,1	0,0	0,0	***
% gasto en cuotas	0,58	0,56	-0,02	0,54	0,54	0,00	0,02	***
Tiene crédito (%)	25,5	26,1	0,6	23,9	27,2	3,3	2,7	***
Tiene avance (%)	22,8	22,6	-0,2	19,5	20,9	1,4	1,6	***
Cantidad de canjes	1,1	1,1	0,0	0,8	1,3	0,5	0,5	***
Puntos canjeados	8.474	9.773	1.298	9.826	12.875	3.048	1.750	***

GC: Promedio grupo control, T: Promedio grupo tratamiento, Dif.: Diferencia de promedios

*** significativo al 99%, ** significativo al 95%, * significativo al 90%

Tabla 26. Resultados pre y post matching

En primer lugar, se observa que existe un efecto positivo en la rentabilidad siguiente al periodo de activación y también en el valor del cliente en el largo plazo, de 15 M\$ y 62 M\$ respectivamente (en los tratados). Por lo tanto, al activar un cliente en el programa de fidelización se espera generar un cambio en su comportamiento valorizado en 77 M\$ (rentabilidad + CLV), siempre y cuando el cliente posea un perfil similar al grupo de tratados (Tabla 25).

Este incremental en el valor del cliente se explica por un leve aumento del compromiso con la empresa que se refleja en distintos aspectos de su comportamiento:

- **Transaccionales:** Aumento del uso de la tarjeta en 3,6 días con compra al año y 0,4 rubros nuevos. El gasto anual con la tarjeta crece en 41 M\$ en promedio.
 - Dentro del gasto existe una canibalización de 50,8 M\$ desde negocios externos a internos. Si se analiza el detalle de los promedios en ambos periodos, se evidencia que ambos grupos (control y tratamiento) existe un traspaso del gasto en el sentido inverso y en magnitudes similares. El grupo de control pasa de gastar 1.568 M\$ en el exterior y 784 M\$ en el interior, a 882 M\$ y 1.529 M\$ respectivamente. Mientras que, el grupo tratamiento pasa de gastar 1.636 M\$ en el exterior y 969 M\$ en el interior, a 1.048 M\$ y 1.664 M\$ respectivamente. La variación es similar en ambos grupos contra factuales, por ende, corresponde a una tendencia generalizada e independiente de la participación del tratamiento. Esta observación es importante para análisis posteriores y se explica, según los expertos del negocio, por la estrategia aplicada por la empresa durante estos últimos años que busca incentivar el uso universal de la tarjeta, renovando la percepción que tiene la gente sobre una tarjeta fuertemente asociada al *retail* a una tarjeta de uso general - que puede y debe ser utilizada como medio de pago principal en cualquier comercio del mercado.

Entonces, el efecto causal del programa de puntos en el gasto se interpreta como un gancho de lealtad, ya que, a pesar de esta tendencia generalizada se mantiene una lealtad de los clientes hacia los *retails* del *holding*.

- **Productos financieros:** Respecto a los productos financieros, crece la propensión a consumir créditos de consumo (+2,7%) y avances en efectivo (+1,6%). Este incremento, que puede parecer pequeño, es significativo económicamente dado que son productos muy rentables.
- **Cuotas:** En términos de cuotas, el cliente casi no se ve afectado. No cambia la cuota promedio ponderada, pero si lo hace el porcentaje de gasto en cuotas en un +2%, quizás al aumentar el compromiso con la empresa existe mayor disposición a utilizar el crédito.
- **Puntos:** Por último, se observa que los clientes quieren repetir la experiencia de participar en el club de puntos, ya que aumentan el número de canjes (+0,5) y los puntos canjeados (+1.750) en el periodo posterior. Esto puede estar asociado al aumento del gasto y la preferencia por los negocios del grupo, ya que brindan mayor acumulación de puntos.

Estos cambios son positivos para la relación de la empresa con sus clientes. Aumenta el valor esperado en el largo plazo, explicado por el aumento mismo de la rentabilidad como por un incremento en la retención, ya que por lo visto en 6.2.5 la duración del cliente está relacionada directamente con el gasto, consumo de productos y canje de puntos.

Por último, cabe comparar el efecto inferido anteriormente con la estratificación (en 6.3.2) con el efecto estimado a través de la metodología de Propensity Score Matching. Se puede observar que los resultados pasan a ser positivos y, en particular, el efecto de largo plazo cambia notoriamente en su magnitud, resultado ahora en un valor más razonable que -4,27 M\$.

Variable	Estratificación	Propensity Score Matching
Rentabilidad (M\$)	-12,69	15,2
CLV (M\$)	-4,27	61,7

Tabla 27. Efectos estimados con estratificación versus Propensity Score Matching

Más allá de los resultados, se debe recalcar que el método de estratificación clásico es muy sensible a las variables seleccionadas por el analista, esto porque durante la estimación con este método fueron probadas distintas variables y con distintos niveles categóricos observándose resultados variables tanto en magnitud como en sentido. Además, el hecho de sólo permitir variables categóricas lo vuelve rápidamente infactible cuando se tienen muchas continuas. En cambio, el segundo método permite incorporar mayor cantidad de variables y de todo tipo, y prioriza de manera semi-automática las que son más relevantes para el tratamiento, gracias al modelamiento de la propensión.

6.3.6 Segmentación

Ya calculado el efecto promedio en los tratados, en la presente sección se estudia la posibilidad de encontrar distintos impactos según tipos de clientes, considerando así la heterogeneidad de la cartera. Para esto, se realiza una segmentación en base a la propensión a ser activo del programa de puntos. Independiente de hallarse efectos diferenciados por segmentos o un efecto uniforme para todo tipo de clientes, se genera conocimiento de valor para la gestión del programa de puntos. Esta sección trata la segmentación y caracterización de los clientes.

Utilizando la muestra de clientes ya emparejada en 6.3.4, se prueban métodos de segmentación hasta encontrar el mejor balance dentro de cada segmento. Los métodos a probar son: cuantiles de propensión de los tratados, rangos equidistantes de propensión y k-medias en base a la propensión de los tratados. Además, para cada método se prueban distintas cantidades de segmentos. La siguiente tabla presenta indicadores para la elección del método y el número de segmentos.

Indicador	Método	Cantidad de segmentos							
		2	3	4	5	6	7	8	
Promedio DME*	Cuantiles	0,056	0,055	0,058	0,058	0,058	0,060	0,061	
	Rangos	0,055	0,062	0,064	0,069	0,069	0,070	0,071	
	k-medias	0,057	0,064	0,071	0,077	0,079	0,083	0,086	
Segmentos balanceados**	Cuantiles	1	2	3	4	4	4	5	
	Rangos	2	1	2	3	4	5	6	
	k-medias	2	1	1	2	2	3	2	

* Diferencias de medias estandarizadas para 14 más variables relevantes (81% relevancia de la propensión)

**Segmento tiene todas las variables relevantes con DME bajo 0,25 desviaciones

Tabla 28. Elección de método de segmentación y número de segmentos

El primer indicador señala el balance general de la segmentación y corresponde más específicamente al promedio de las diferencias de medias estandarizadas considerando todas las variables de todos los segmentos generados. Mientras menor este valor, mejor balance. Por otro lado, el segundo indicador señala la cantidad de segmentos balanceados, es decir, con todas sus variables relevantes balanceadas (DME < 0,25). Este último número tiene como tope la cantidad de segmentos generados (columnas) y mientras más cercano este número a la cantidad tope mejor.

Se tiene que con el método de cuantiles se obtienen menores diferencias de medias estandarizadas (Promedio DME), es decir, un mejor balance general. Por otro lado, se observa que, por ejemplo, al generar 5 segmentos, el método de cuantiles es el que tiene mayor efectividad, generando 4 de ellos balanceados, a diferencia de los 3 y 2 con los otros métodos. Continuando con el método por cuantiles, al crear mayor cantidad de

segmentos (6 y 7) continúan existiendo solo 4 de ellos balanceados, por lo que se crean segmentos desbalanceados que no aportan para el análisis. Por lo tanto, se decide utilizar quintiles de propensión, ya que se obtiene un mejor balance y un número de segmentos adecuados para la gestión.

Prosiguiendo con la segmentación basada en los quintiles de propensión de los tratados, se obtienen los siguientes grupos de clientes. En Anexo 14 se presenta una ilustración que muestra de manera gráfica la segmentación en base a la propensión.

Segmento	Mínima propensión (%)	Máxima propensión (%)	N clientes activos	N clientes inactivos
1	0,0	11,4	42.876	42.865
2	11,4	26,3	42.875	42.872
3	26,3	41,9	42.876	45.680
4	41,9	59,5	42.875	44.123
5	59,5	100,0	42.876	38.838

Tabla 29. Quintiles de propensión

Luego, al igual que en 6.3.4, para inferir causalidad a nivel de segmentos es necesario validar que dentro de cada segmento los clientes activos e inactivos son comparables. Por lo tanto, en el siguiente sub-apartado se revisa el balance de las covariables para cada uno de estos grupos. Luego, en el sub-apartado subsiguiente se realiza una caracterización para comprender como se vinculan los perfiles de clientes con las propensiones y los efectos del club de puntos.

i. Balance de covariables por segmento

Al igual que en 6.3.4 Matching, se debe revisar el balance de las covariables entre los clientes activos e inactivos, esta vez para cada segmento. Para revisar este balance, la siguiente tabla resume el balance de las 14 variables más relevantes del modelo de propensión, ordenadas según importancia, que en conjunto significan el 81% de la relevancia del Propensity score.

Variables relevantes	Diferencia de medias estandarizadas por segmento				
	1	2	3	4	5
Propensity Score	0,001	0,001	0,086	0,092	0,104
Canjes	0,094	0,011	0,003	0,013	0,051
Tasa de canje	0,064	0,180	0,022	0,087	0,231
Categoría	0,069	0,065	0,049	0,069	0,213
% gasto pagado con tarjeta en negocios internos	0,056	0,034	0,060	0,045	0,077
Meses con compra	0,131	0,051	0,010	0,055	0,122
Gasto	0,357	0,234	0,114	0,051	0,046
Distancia prox. nivel de canje	0,027	0,048	0,077	0,075	0,052
Compras negocio interno: tiendas p.d.	0,007	0,036	0,027	0,014	0,009
Máxima cuota	0,024	0,008	0,042	0,071	0,081
Con compra en rubro: supermercados	0,056	0,002	0,020	0,018	0,021
Compras en negocios externos	0,012	0,022	0,013	0,049	0,092
% transacciones en cuotas	0,095	0,001	0,012	0,024	0,004
Compras negocio interno: mej. del hogar	0,064	0,060	0,024	0,034	0,108
% gasto en semestre 2	0,020	0,032	0,013	0,013	0,002

Tabla 30. Diferencias de medias estandarizadas por segmento
(14 variables más relevantes ordenadas por relevancia en la propensión)

Se observa que, en general, el balance por segmentos es peor al balance global (6.3.4 Matching, Tabla 25). El segmento 4 posee un balance excelente, con todas las diferencias de medias estandarizadas (DME) bajo 0,1 desviaciones. Los segmentos 2, 3 y 5 poseen algunas variables por sobre 0,1, aunque ninguna supera el límite de 0,25, por lo que el balance de estos segmentos es aceptable. Por último, el segmento 1 posee una variable desbalanceada, el gasto anual, con una DME de 0,357 y, al ser una variable relevante, esto no puede ser omitido durante el análisis de los resultados. En conclusión, los segmentos 2, 3, 4 y 5 cumplen con un correcto balance y pueden ser comparados para inferir causalidad, no así el segmento 1, con el que no se logró el balance.

ii. Caracterización de segmentos

Para caracterizar a los grupos, la Tabla 31 muestra los promedios (o porcentajes en caso de categóricas) de los tratados, en el periodo previo al horizonte de activación, en algunas de las variables del modelo de propensión en donde se observan mayores diferencias entre segmentos. En Anexo 15 se encuentra el detalle para el total de variables.

Segmento	1	2	3	4	5
Propensión (%)	0,06	0,19	0,34	0,50	0,76
Tasa de canje (%)	0,12	0,54	0,80	0,88	0,94
Número de canjes	0,18	0,75	1,17	1,45	2,01
GSE (%)					
ABC1	9,90	10,20	10,80	12,60	16,70
C2	24,90	26,70	29,70	33,50	43,80
C3	26,30	29,40	33,40	35,10	32,40
D	25,00	24,00	21,60	16,50	6,40
E	3,90	2,20	1,10	0,60	0,10
Sin información	10,10	7,50	3,30	1,80	0,70
Tramo edad (%)					
18-25	3,40	3,10	2,50	2,00	0,50
26-30	10,30	9,40	8,60	7,90	4,00
31-35	11,50	10,90	10,20	10,10	9,20
36-40	11,50	11,00	10,80	11,00	15,00
41-45	11,90	11,70	11,70	12,20	24,30
46-50	11,80	12,30	12,70	12,80	24,80
51-55	11,40	11,80	12,30	12,60	4,30
56-60	10,00	10,60	10,80	11,20	4,90
61-65	7,80	8,20	8,80	9,00	5,10
66+	10,20	11,00	11,50	11,20	7,90
Categoría (%)					
Normal	98,60	94,00	87,60	74,00	29,70
Premium	1,10	5,40	11,50	24,40	64,00
Elite	0,30	0,60	0,90	1,60	6,30
Tramo score de riesgo (%)					
0	12,20	2,90	0,50	0,20	0,10
1-300	1,60	1,40	1,00	1,00	0,70
301-600	28,70	28,60	26,20	23,50	18,60
601-1,000	57,40	67,10	72,20	75,30	80,60
Tiene adicionales (%)	21,30	26,10	29,90	34,00	41,30
Monto cupo (M\$)	991	1.111	1.213	1.359	1.689
Gasto anual (M\$)	1.467	1.928	2.167	2.650	4.816
Porcentaje del gasto en neg. externos (%)	0,20	0,29	0,35	0,38	0,42

Tabla 31. Descriptivos de activos por segmento (variables diferenciadoras modelo de propensión)

Para complementar, se analizan otras variables interesantes que pueden aportar a la caracterización de los segmentos.

Segmento	1	2	3	4	5
Puntos canjeados	1.398	5.166	8.195	11.212	22.893
Stock de puntos	5.651	7.572	8.883	11.241	18.103
Puntos por canje	7.764	6.888	7.004	7.732	11.389
Rubros con compra	3,95	5,52	6,28	6,99	8,44
Compras mensuales en negocios externos	0,87	1,77	2,51	3,36	6,02
Compras mensuales en negocios internos	0,72	1,18	1,46	1,73	2,53
Ticket promedio (M\$)	77	54	46	43	47
Cuota promedio ponderada	3,17	3,33	3,32	3,25	2,90
Porcentaje del gasto pagado en cuotas	0,56	0,59	0,59	0,57	0,51
Tiene avance (%)	15,10	21,90	25,20	26,50	24,50
Tiene crédito de consumo (%)	14,80	22,90	28,00	31,20	33,40
Rentabilidad (M\$)	62	104	129	147	160
CLV (M\$)	531	855	1.053	1.221	1.404

Tabla 32. Promedios de activos por segmento (otras variables)

De las tablas anteriores se desprenden las siguientes observaciones generales:

- Clientes de mayor valor, tanto en rentabilidad y CLV, son más propensos a activarse en el programa de puntos. Esta observación es positiva para el negocio e indica que el club de lealtad está alineado correctamente con el valor de los clientes, alcanzando en mayor medida a los clientes que producen y producirán mayores rentabilidades para la empresa en el futuro.
- A mayor actividad previa en el programa de puntos, existe una mayor propensión a ser activo en el programa de puntos en el siguiente periodo.
- Clientes de estratos socioeconómicos más altos, mejor categoría, mejor puntaje de riesgo y más monto de cupo poseen una tendencia mayor a activarse en el programa. Variables evidentemente relacionadas.
- Clientes que poseen un mayor consumo con la empresa son más propensos a participar del club de fidelización. Este consumo se refleja en una mayor tasa de consumo de créditos, mayor probabilidad de poseer adicionales, mayor cantidad de compras, gasto y, en general, una mayor actividad transaccional con la tarjeta tanto en negocios externos como internos al *holding*.
- El segmento 1 y el segmento 5 representan dos extremos de clientes que participan en el programa de puntos, con un bajo y alto compromiso con la empresa y el club de fidelización respectivamente. Los segmentos intermedios corresponden a distintos grados dentro de esta gama de clientes.

A continuación, se describen las características principales de cada segmento. Se les asigna un apodo para así recordar sus características y referenciarlos en adelante.

Segmento 1: “Oportunistas ocasionales”

- Clientes de valor medio para la empresa, tienen rentabilidad similar al promedio de la cartera y un CLV de 1,2 veces el promedio de esta.⁷
- Clientes de la categoría Normal (98,6%). Tienen mayor riesgo crediticio relativo a los otros segmentos, un 12% tiene puntaje score 0, lo que se explica porque tienen poco historial de pago en la empresa. Por lo mismo, poseen el menor cupo.
- Utilizan con baja frecuencia la tarjeta, 1,5 veces por mes, y tienen un gasto anual de 1,5 MM\$, el 80% de este lo destinan a negocios del *holding*. El *ticket* promedio es el más alto respecto a los demás segmentos, de 77 mil pesos. Al parecer, aprovechan de realizar compras importantes con los descuentos de la tarjeta en los *retails* del grupo.
- Después del segmento 5, son el segmento que menos paga en cuotas. Tienen la menor tasa de consumo de avances y créditos de consumo. Lo que refuerza la idea de que utilizan la tarjeta por sus descuentos.
- Poseen en promedio 5.650 puntos en stock, suficiente para realizar canjes en el primer nivel, en donde existen oportunidades de canje por 1.000 a 5.000 puntos. Sin embargo, el número de canjes por cliente en promedio es de 0,18 y poseen una tasa de canje de 12%, que es muy baja.

En síntesis, corresponden a clientes de gasto medio, que utilizan la tarjeta de manera ocasional, probablemente para aprovechar descuentos en *retail*. Su bajo compromiso con la empresa se traduce en una menor propensión a participar en el programa de puntos y aunque posean un stock de puntos suficiente para realizar canjes, al menos en el primer nivel, estos parecen no estar enterados o interesados en esta oportunidad y sólo utilizan el 12% de sus puntos acumulados. Clientes de valor medio.

Segmento 2: “Cuoterros habituales”

- Clientes de valor medio-alto, tienen rentabilidad de 1,7 veces el promedio de la cartera y un CLV de 1,9 veces el promedio de esta.
- Clientes de la categoría Normal (94,0%) con unos pocos Premium (5,4%). Tienen un monto de cupo promedio de 1,1 MM\$.

⁷ Rentabilidades y CLV calculados para el año 2014, promedios de 63 M\$ y 453 M\$ respectivamente.

- Compran 3 veces al mes. Gastan 1,9 MM\$ al año y el 71% de este gasto es destinado a negocios del *holding*. Tienen la mayor cuota promedio ponderada (3,33) y el mayor porcentaje del gasto pagado en cuotas (59%).
- Canjean 0,75 veces al año, principalmente entre el nivel de 5.000 y 9.000 puntos. Tienen una tasa de canje de 54%, por lo que utilizan la mitad de sus puntos.

Resumiendo, son clientes de categoría normal, que utilizan la tarjeta con frecuencia media-baja y necesitan cuotas. Parecen ser clientes que ven a la tarjeta como medio para aprovechar promociones en los *retails* del *holding*, como también un medio para pagar en crédito. Estos clientes al estar en una relación de necesidad y oportunidad con la tarjeta, también se interesan por los beneficios del programa de puntos, aunque como no están en constante interacción con la empresa, canjean esporádicamente y vencen casi la mitad de sus puntos. Clientes de valor medio-alto.

Segmento 3: “Cuoterios frecuentes”

- Clientes de alto valor, tienen rentabilidad de 2,0 veces el promedio de la cartera y un CLV de 2,7 veces el promedio de esta.
- Clientes categoría Normal (87,6%), con uno Premium de cada 10. Tienen un monto de cupo de 1,2 MM\$ en promedio y un riesgo relativo medio-bajo.
- Compran 4 veces al mes y tienen un gasto anual de 2,2 MM\$, el 65% de este es destinado a negocios del *holding*. El *ticket* promedio es de 46 mil pesos, menor a los segmentos anteriores pero similar a los segmentos 4 y 5.
- Junto con el segmento 2 son los clientes que tienen mayor porcentaje de su gasto pagado en cuotas (59%) y mayor cuota promedio ponderada (3,32). Aunque, a diferencia del segmento anterior, estos clientes utilizan con mayor frecuencia la tarjeta y también consumen más productos financieros, 25% posee avances en efectivo y 28% créditos de consumo.
- En promedio tienen 9 mil puntos en stock y canjean 1,17 veces al año, utilizando el 80% de los puntos que acumulan.

A modo de síntesis, son clientes de categoría normal que utilizan la tarjeta con frecuencia media, incluso en comercios externos y en *tickets* bajos. Tienen una necesidad de financiamiento tanto en cuotas como en avances y/o créditos de consumo. Al tener una interacción con la empresa relativamente constante durante el año, son clientes que están pendientes de sus puntos y son más propensos a participar del club, canjean una vez al año y sólo vencen un 20% de sus puntos acumulados. Clientes de alto valor.

Segmento 4: “Leales”

- Clientes de muy alto valor, tienen rentabilidad de 2,3 veces el promedio de la cartera y un CLV de 2,7 veces el promedio de esta.
- Clientes de categoría Normal (74,0%) más parecidos a los Premium (24,4%), con una pequeña proporción de Elites (1,6%) pero mayor a los segmentos anteriores. Son clientes principalmente de segmentos socioeconómicos C2 y C3, de bajo riesgo y tienen un monto de cupo promedio de 1,4 MM\$.
- Compran 5 veces al mes con la tarjeta. Gastan 2,7 MM\$ al año y el 62% de este gasto es en el *holding*. Tienen el menor *ticket* promedio entre los segmentos, de \$43.000. Al parecer, le dan un uso más general a la tarjeta.
- Compran en cuotas en menor medida que los segmentos 2 y 3, pero en mayor medida que los segmentos 1 y 5. Son el segmento con mayor tasa de consumo de avances (27%) y el segundo segmento en créditos de consumo (31%).
- Tienen en promedio 11.000 puntos en stock y canjean 1,5 veces al año. Tienen tasa de canje de 88%, por lo que rara vez pierden puntos.

En resumen, son clientes de perfil Premium, pertenecen a clase media y media-alta, utilizan la tarjeta con frecuencia y gastan casi 3 MM\$ al año. Son leales tanto a la tarjeta como a los negocios del grupo. Tienden a requerir financiamiento, lo que se evidencia en un consumo de avances y créditos relativamente alto, como también en el uso de cuotas. Su compromiso con la empresa se refleja también en su participación con el programa de puntos, en donde tienen una alta tasa de canje (88%) con más de un canje anual. Clientes de muy alto valor.

Segmento 5: “Punto-fidelizados”

- Clientes de muy alto valor, incluso más que el segmento anterior. Tienen rentabilidad de 2,5 veces el promedio de la cartera y un CLV de 3,1 veces el promedio de esta.
- Más del 70% de los clientes pertenecen a categorías Premium y Elite. Son clientes de muy bajo riesgo, por esto poseen un monto de cupo relativamente mayor.
- Tienen una marcada diferencia en edad respecto a los demás segmentos, concentrándose entre 36 y 50 años. El 40% posee adicionales, por lo que probablemente corresponden a padres de familia. Pertenecen mayormente al segmento C2, aunque poseen una proporción mayor de ABC1 (17%) que los demás segmentos.
- Compran con la tarjeta casi 9 veces al mes y en una alta cantidad de rubros (8,44) respecto los otros segmentos. Tienen un gasto anual muy alto, de 4,8 MM\$, y el 58% de este es destinado a negocios del *holding*.

- Respecto a su necesidad de financiamiento, son el segmento que menos utiliza las cuotas. Pero, por otro lado, tienen la mayor proporción de clientes que consume créditos de consumo.
- En cuanto a puntos, tienen un stock promedio de 18.000 puntos con una utilización casi perfecta, realizando 2 canjes al año.

En síntesis, son principalmente padres de familia, de clase alta y categoría Premium/Elite, utilizan la tarjeta para todo y son muy leales a los *retails* del *holding*. No suelen comprar en cuotas, pero un 30% de ellos consume créditos de consumo. Su compromiso con la empresa probablemente está apalancado por los puntos, ya que canjean 2 veces al año y casi no vencen sus puntos acumulados. Clientes de muy alto valor.

6.3.7 Efecto por segmento

Se tienen 5 segmentos de clientes con distintas características, comportamientos y propensiones a activarse del programa de puntos. Dentro de cada segmento hay clientes activos como inactivos en el club, por lo que se pueden contrastar sus resultados en el periodo posterior al horizonte de activación. El siguiente gráfico resume los efectos estimados a nivel de segmentos, en donde cada grupo es referenciado con su apodo definido en el apartado anterior.

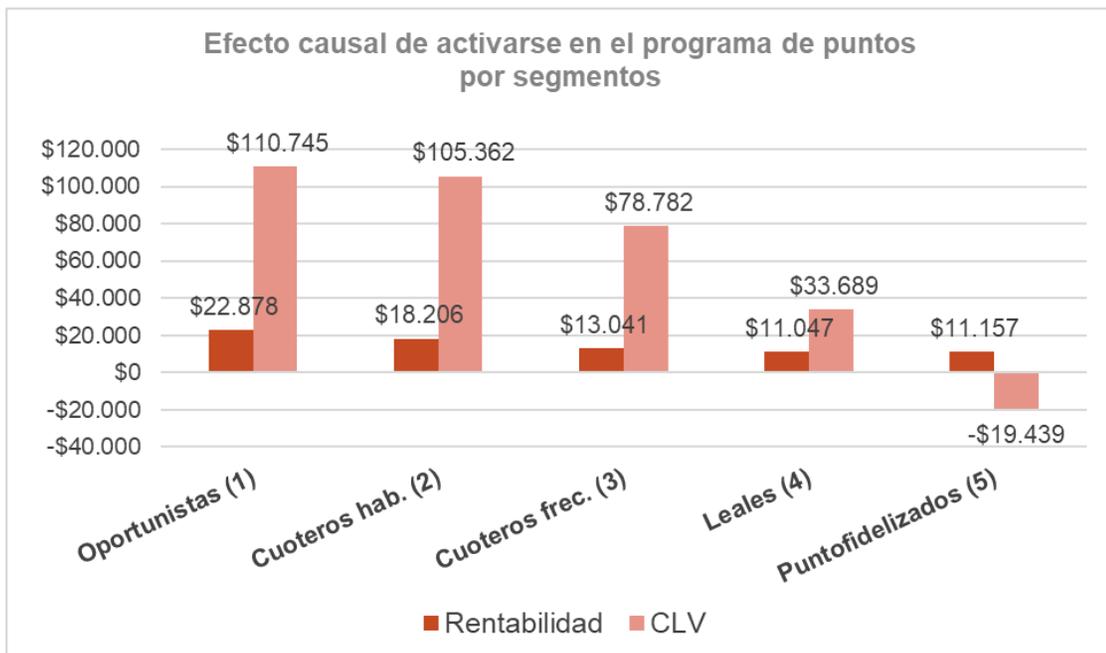


Ilustración 30. Efecto en rentabilidad 2017 y CLV 2018-2028 de activarse en el club de puntos

Como primera observación, existen diferentes efectos según tipos de clientes, lo que valida una de las hipótesis presentes en los objetivos.

Los clientes con mayor potencial al ser activados en el programa de fidelización son los clientes “Oportunistas”, que aumentan su valor en \$134 mil (rentabilidad + CLV). Según el balance logrado en 6.3.6 en este grupo no es adecuado inferir causalidad. No obstante, existe una correlación aparente entre el efecto y la propensión (representada por los segmentos) y, además, la similitud de los efectos observados entre el grupo 1 y el par más cercano, proporcionan pruebas para considerar la inferencia en estos clientes como admisible.

Por otro lado, en el extremo de alta propensión están los “Punto-fidelizados” que al activarse en el programa de puntos disminuyen su valor como clientes en \$-8 mil (rentabilidad + CLV). Llama la atención de que, en este último segmento en particular se genere una pérdida.

Para comprender cómo se explica cada uno de estos efectos, en lo siguiente se analiza cada segmento con mayor detalle.

Efecto segmento 1: “Oportunistas ocasionales”

Variable	Periodo previo			Periodo posterior			Efecto	Sign.
	GC	T	Dif.	GC	T	Dif.		
Rentabilidad (M\$)	58,9	64,3	5,5	77,1	105,4	28,3	22,9	***
CLV (M\$)	503,5	558,4	54,9	554,9	720,6	165,6	110,7	***
Gasto (M\$)	940,3	1.466,6	526,3	967,8	1.495,2	527,4	1,2	
en holding	725,9	1.160,1	434,2	474,0	704,2	230,2	-204,0	***
en externos	214,4	306,4	92,0	493,8	791,0	297,2	205,2	***
Días con compra	19,5	19,2	-0,3	24,8	32,6	7,8	8,1	***
Rubros con compra	4,1	4,0	-0,1	4,5	5,4	0,9	1,0	***
Cuota prom. pond.	3,2	3,2	-0,1	3,1	3,1	0,0	0,0	
% gasto en cuotas	0,59	0,56	-0,03	0,55	0,53	-0,02	0,01	***
Tiene crédito (%)	13,9	14,8	0,9	17,4	22,8	5,4	4,5	***
Tiene avance (%)	16,6	15,1	-1,5	15,4	16,7	1,3	2,8	***
Cantidad de canjes	0,1	0,2	0,0	0,4	1,0	0,6	0,5	***
Puntos canjeados	846	1.398	551	2.584	6.771	4.187	3.635	***

GC: Promedio grupo control, T: Promedio grupo tratamiento, Dif.: Diferencia de promedios

*** significativo al 99%, ** significativo al 95%, * significativo al 90%

Tabla 33. Resultados segmento 1

Los clientes del segmento 1 al activarse en el programa de puntos no aumentan su gasto, pero trasladan gran parte (204 M\$) de este a comercios externos evidenciando una menor importancia por la acumulación de puntos. Compran 8 días más en el año con la tarjeta y la utilizan en un nuevo rubro. La hipótesis detrás de este comportamiento es que este segmento de carácter oportunista se guía por los descuentos que, dada la estrategia de marketing de la empresa durante estos últimos años, ahora son bastante agresivos en alianzas comerciales externas, por ejemplo, 40% de descuento en restaurantes y en

negocios de entretención. Por lo que, puede que el participar del club de puntos potencie este comportamiento característico del cliente, al estar más informados y consciente de los beneficios inmediatos de la tarjeta.

Si bien es un segmento de bajo consumo de productos financieros, tienen un incremental en el consumo de créditos y avances de +4,5% y +2,8% respectivamente. Por último, se observa que la participación en el programa provoca un alza en el número de canjes (+0,5) y los puntos canjeados (+3.665).

Concluyendo, este segmento oportunista, de baja frecuencia y de valor medio, tiene un leve impacto transaccional, aumentando el número de días con compra sin variar su gasto. Más bien, migran su gasto hacia comercios fuera del grupo y le dan un uso más general, parece ser que, en busca de los nuevos descuentos con alianzas externas. Por lo que el programa de puntos parece potenciar este interés de los clientes por los descuentos de la tarjeta. Por otro lado, aumentan su probabilidad de consumir productos financieros, lo que impulsa la rentabilidad del segmento. Antes parecían desconocer los puntos, pero luego de canjear aumentan su propensión a realizarlo nuevamente, ya que ahora conocen el sistema de recompensas. El crecimiento en el consumo de productos financieros y en la actividad con el club también implican un aumento en la retención esperada. Como consecuencia de todo lo anterior, se produce un incremental de 23 M\$ y 111 M\$ en rentabilidad del corto y largo plazo respectivamente.

Efecto segmento 2: “Cuoterros habituales”

Variable	Periodo previo			Periodo posterior			Efecto	Sign.
	GC	T	Dif.	GC	T	Dif.		
Rentabilidad (M\$)	107,4	100,3	-7,1	121,1	132,2	11,1	18,2	***
CLV (M\$)	854,0	855,8	1,8	828,3	935,4	107,2	105,4	***
Gasto (M\$)	1.512,9	1.927,9	415,0	1.573,8	1.984,0	410,1	-4,9	
en holding	1.053,7	1.368,7	315,0	688,3	857,3	169,0	-146,0	***
en externos	459,2	559,3	100,0	885,5	1.126,6	241,2	141,1	***
Días con compra	32,7	33,5	0,8	37,2	44,1	6,9	6,2	***
Rubros con compra	5,5	5,5	0,0	5,9	6,6	0,6	0,6	***
Cuota prom. pond.	3,4	3,3	-0,1	3,2	3,2	0,0	0,1	***
% gasto en cuotas	0,60	0,59	-0,01	0,57	0,57	0,00	0,01	***
Tiene crédito (%)	23,5	22,9	-0,6	23,1	26,8	3,7	4,3	***
Tiene avance (%)	22,3	21,9	-0,4	19,2	20,4	1,2	1,6	***
Cantidad de canjes	0,8	0,8	0,0	0,7	1,2	0,5	0,5	***
Puntos canjeados	4.638	5.166	528	5.462	9.305	3.843	3.315	***

GC: Promedio grupo control, T: Promedio grupo tratamiento, Dif.: Diferencia de promedios

*** significativo al 99%, ** significativo al 95%, * significativo al 90%

Tabla 34. Resultados segmento 2

Los clientes del segmento 2 tienen un impacto similar al grupo anterior, trasladan su gasto desde los *retails* del *holding* hacia comercios externos sin variar su gasto anual. Tienen

un crecimiento de 6 días con compra al año y compran en 0,6 nuevos rubros. Los resultados muestran un incremental estadísticamente significativo en el uso de cuotas, aunque de una magnitud muy pequeña. La tasa de consumo de créditos y avances aumentan en +4,3% y +1,6% respectivamente. En cuanto a puntos, se observa un efecto similar al grupo anterior, crece el número canjes (+0,5) y puntos canjeados (+3.315).

Son clientes que necesitan cuotas, compran con frecuencia media baja y, al igual que el segmento 1, también poseen cierto afán por los descuentos, ya que pasaron parte de su gasto a comercios externos, probablemente motivado por la estrategia de descuentos con alianzas comerciales. La hipótesis que relaciona el club de puntos y este último incremento es que los clientes, al participar del programa de puntos, están más informados sobre los otros beneficios que ofrece la empresa.

Tienen un efecto positivo en el consumo de productos financieros que implica mejores rentabilidades y retención, y un aumento en la frecuencia de canje que, por su lado, impacta principalmente la retención. Como resultado, se produce un incremental de 18 M\$ y 105 M\$ en rentabilidad del corto y largo plazo respectivamente.

Efecto segmento 3: “Cuoterios frecuentes”

Variable	Periodo previo			Periodo posterior			Efecto	Sign.
	GC	T	Dif.	GC	T	Dif.		
Rentabilidad (M\$)	133,2	124,6	-8,7	144,5	148,9	4,4	13,0	***
CLV (M\$)	1.066,4	1.039,0	-27,3	969,7	1.021,2	51,4	78,8	***
Gasto (M\$)	1.933,5	2.166,8	233,3	1.978,4	2.218,0	239,6	6,3	
en holding	1.263,6	1.400,1	136,5	814,8	944,9	130,1	-6,5	
en externos	669,9	766,7	96,8	1.163,6	1.273,1	109,6	12,8	
Días con compra	42,6	43,3	0,7	45,5	50,5	4,9	4,2	***
Rubros con compra	6,3	6,3	0,0	6,7	7,1	0,4	0,4	***
Cuota prom. pond.	3,4	3,3	-0,1	3,2	3,2	0,0	0,0	***
% gasto en cuotas	0,59	0,59	0,00	0,57	0,57	0,00	0,00	***
Tiene crédito (%)	28,7	28,0	-0,7	25,7	28,1	2,4	3,1	***
Tiene avance (%)	25,7	25,2	-0,5	21,5	22,3	0,8	1,3	***
Cantidad de canjes	1,2	1,2	0,0	0,9	1,3	0,5	0,5	***
Puntos canjeados	7.732	8.195	462	7.744	10.806	3.061	2.599	***

GC: Promedio grupo control, T: Promedio grupo tratamiento, Dif.: Diferencia de promedios

*** significativo al 99%, ** significativo al 95%, * significativo al 90%

Tabla 35. Resultados segmento 3

Este segmento de clientes no presenta efectos significativos a nivel de gasto. Aumentan 4 días con compra al año y en promedio compran en 0,4 nuevos rubros. Respecto a productos financieros, crecen su tasa de consumo de créditos (+3,1%) y de avances en efectivo (+1,3%). En cuanto a puntos, el participar del programa aumenta la propensión de los clientes a canjear nuevamente, aumentando en 0,5 el número de canjes y en 2.599 los puntos canjeados al año.

En síntesis, estos clientes de frecuencia media, que ya utilizaban la tarjeta en comercios externos, necesitan de financiamiento a través de cuotas y productos financieros. Luego de activarse en el club de puntos, casi no presentan impacto a nivel transaccional, pero sí un incremento en la propensión a productos financieros de entre 1 y 3 puntos porcentuales, esto aumenta la rentabilidad del segmento en 13 M\$ al siguiente periodo. También crece su interés por repetir esta experiencia con el programa y, a pesar de que acumulan los mismos puntos que antes, canjean más veces y más puntos. Estos incrementales de rentabilidad y de actividad en el club, en conjunto, significa un mayor valor del cliente en el largo plazo (+79 M\$).

Efecto segmento 4: “Leales”

Variable	Periodo previo			Periodo posterior			Efecto	Sign.
	GC	T	Dif.	GC	T	Dif.		
Rentabilidad (M\$)	150,6	144,0	-6,6	162,7	167,1	4,5	11,0	***
CLV (M\$)	1.226,9	1.214,9	-12,0	1.100,0	1.121,7	21,7	33,7	***
Gasto (M\$)	2.518,1	2.649,8	131,8	2.565,2	2.748,9	183,7	51,9	***
en holding	1.621,2	1.600,7	-20,5	980,0	1.087,8	107,8	128,4	***
en externos	896,8	1.049,1	152,3	1.585,2	1.661,1	75,8	-76,5	***
Días con compra	51,1	53,2	2,1	55,2	59,7	4,4	2,3	***
Rubros con compra	7,0	7,0	0,0	7,5	7,8	0,3	0,3	***
Cuota prom. pond.	3,3	3,3	0,0	3,1	3,1	0,0	0,0	
% gasto en cuotas	0,58	0,57	-0,01	0,55	0,55	0,00	0,01	***
Tiene crédito (%)	31,0	31,2	0,2	27,3	29,6	2,3	2,1	***
Tiene avance (%)	26,3	26,5	0,2	21,5	23,2	1,7	1,5	***
Cantidad de canjes	1,4	1,5	0,0	1,0	1,4	0,5	0,4	***
Puntos canjeados	10.416	11.212	796	10.814	13.362	2.548	1.752	***

GC: Promedio grupo control, T: Promedio grupo tratamiento, Dif.: Diferencia de promedios

*** significativo al 99%, ** significativo al 95%, * significativo al 90%

Tabla 36. Resultados segmento 4

El segmento 4 corresponde a clientes de un perfil tipo Premium, leales a la tarjeta y el *holding*, tienen un alto gasto, utilizan cuotas y tienen un alto consumo de productos financieros. Al activarse en el club de puntos, aumentan su gasto en 52 M\$ y mantienen una lealtad hacia los *retails* del grupo, traspasando 77 M\$ desde negocios externos a internos. Esta preferencia puede deberse a que han desarrollado una mayor afinidad con las marcas internas o por el simple hecho de que comprar en estos *retails* les brinda mayor cantidad de puntos. Si bien ya son un segmento de alto consumo de productos financieros y puntos, luego de estar activos en el club los clientes son levemente más propensos a consumir créditos (+2,1%) y avances (+1,5%) como también a canjear más veces (0,4) y más puntos (1.752). Entonces, estos clientes que ya son leales y muy valiosos para la empresa poseen un impacto positivo que aumenta aún más su compromiso y su lealtad hacia los negocios internos, desarrollando también un leve incremento en el consumo de productos financieros. Resultando en un crecimiento de la rentabilidad siguiente de +11 M\$ y de +34 M\$ en el largo plazo.

Efecto segmento 5: “Punto-fidelizados”

Variable	Periodo previo			Periodo posterior			Efecto	Sign.
	GC	T	Dif.	GC	T	Dif.		
Rentabilidad (M\$)	159,0	161,8	2,7	164,9	178,8	13,9	11,2	***
CLV (M\$)	1.360,7	1.443,8	83,1	1.248,3	1.311,9	63,6	-19,4	**
Gasto (M\$)	5.142,2	4.815,5	-326,6	5.299,5	5.114,5	-185,0	141,6	***
en holding	3.364,1	2.650,8	-713,3	1.547,7	1.646,2	98,5	811,8	***
en externos	1.778,1	2.164,7	386,6	3.751,8	3.468,3	-283,5	-670,1	***
Días con compra	72,8	78,5	5,7	83,9	85,9	2,0	-3,7	***
Rubros con compra	8,3	8,4	0,2	9,2	9,4	0,2	0,0	
Cuota prom. pond.	2,9	2,9	0,0	2,8	2,8	0,0	0,1	***
% gasto en cuotas	0,50	0,51	0,01	0,48	0,49	0,01	0,00	***
Tiene crédito (%)	30,7	33,4	2,7	26,0	28,9	2,9	0,2	*
Tiene avance (%)	23,0	24,5	1,5	19,4	21,8	2,4	0,9	***
Cantidad de canjes	1,9	2,0	0,1	1,1	1,6	0,5	0,4	***
Puntos canjeados	19.795	22.893	3.098	23.965	24.131	166	-2.932	***

GC: Promedio grupo control, T: Promedio grupo tratamiento, Dif.: Diferencia de promedios

*** significativo al 99%, ** significativo al 95%, * significativo al 90%

Tabla 37. Resultados segmento 5

El segmento de fidelizados por puntos posee una alta capacidad de gasto, utilizando la tarjeta para todo. Al participar en el programa de puntos incrementan su gasto en 142 M\$, con una fuerte lealtad hacia los negocios del *holding*, traspasando 670 M\$ del gasto desde comercios externos a internos. Este efecto es coherente con la descripción realizada del segmento, ya que tienen una preferencia por los *retails* asociados que parece estar muy apalancada por los puntos. A pesar de este incremental, disminuye el número de días con compra en una magnitud insignificante respecto a la cantidad total de días que utilizan la tarjeta.

En cuanto a productos financieros, la tasa de adquisición de créditos no se ve afectada por la participación en el club de puntos, mientras que la de avances en efectivo tiene un alza menor al 1%. Estos clientes al estar en un nivel de tan alta actividad con la empresa parecen ya haber alcanzado su máximo potencial de consumo para estos productos y el participar en el programa no les afecta su propensión a contratar estos.

Por otro lado, crece la cantidad de canjes (+0,4) y disminuye la cantidad de puntos canjeados (-2.932). Es decir, aumenta la frecuencia de canje, pero gastan menos puntos. Esto es contra intuitivo, ya que el incremento en el gasto les brinda mayor acumulación.

El aumento del gasto, especialmente en negocios del *holding* y el mínimo crecimiento en el consumo de avances producen un impacto positivo en la rentabilidad siguiente valorizado en +11 M\$. Estos clientes presentan un efecto en su rentabilidad de largo plazo relativamente pequeño con respecto a su valor inicial (1.443 M\$), pero en sentido

negativo (-19 M\$). Esto llama bastante la atención y probablemente está relacionado a la baja del gasto en comercios externos y/o a la leve caída de puntos canjeados por año, ya que son las únicas variables observadas que presentan cambios significantes y diferentes respecto a los otros segmentos. Esto no debe alarmar, dado que aún siguen siendo clientes muy rentables y esta pequeña pérdida de valor a 10 años, más bien puede interpretarse como un subsidio hacia el resto de los negocios del grupo, ya que la lealtad y el gasto incremental en estos *retails* es muy potente.

Análisis general

Como observación general, se nota que luego de haber participado activamente en el programa de puntos los segmentos de menor actividad y menor compromiso previo con la empresa no alteran su monto gastado con la tarjeta, pero si demuestran un mayor interés por utilizarla en comercios fuera del *holding*. Esto puede estar ligado a las alianzas comerciales, ya que anteriormente estos clientes se guiaban por los descuentos en negocios internos, y luego de participar en el club desarrollan un mayor conocimiento acerca de los beneficios y las nuevas ofertas de la tarjeta en comercios externos, en donde existe un gran catálogo de descuentos en restaurantes, salones de belleza, cines, entre otros. Por último, estos clientes tienen una base de consumo relativamente menor a los otros segmentos y al activarse en el club de puntos incrementan su nivel de compromiso, desde un alza en el consumo de productos financieros a una mayor cantidad de canjes y puntos utilizados. Esto implica impactos positivos relativamente altos en las rentabilidades y la retención esperada.

En el otro extremo, se tiene a segmentos de mayor gasto, consumo de productos financieros, uso de puntos y, en general, clientes de muy alto valor. Al activarse en el programa de fidelización, estos presentan un alza importante en el gasto demostrando su preferencia a los negocios del *holding*, la que es probable que esté apalancada por los incentivos cruzados del club de puntos. Mientras que, en términos de productos financieros y canje de puntos el impacto es casi insignificante, quizás porque ya han alcanzado el tope de su potencial. Por esto, el impacto rentable en estos clientes es bajo en comparación a los segmentos del párrafo anterior.

Similar a lo observado en la caracterización de segmentos, estos segmentos representan dos extremos de clientes, de bajo y alto compromiso con la empresa y su club de fidelización, y los segmentos intermedios corresponden a distintos grados dentro de esta gama de clientes. Este compromiso y el valor rentable de los clientes está directamente correlacionado. Por lo que, los clientes de menor valor tienen un mayor potencial de crecimiento que los clientes que ya, se puede decir, están fidelizados. Esto se refleja en los efectos estimados de la activación en el programa de puntos, ya que es mayor en los clientes de menor valor y viceversa.

6.4 Recomendaciones

En esta sección se estima el potencial económico de gestionar a los clientes al activarlos en el programa de puntos, según el criterio de actividad definido, los segmentos generados y sus efectos estimados en las secciones anteriores. Para posteriormente realizar recomendaciones tácticas acerca de la gestión de la cartera y el club de puntos con el fin de potenciar el valor de los clientes.

6.4.1 Cálculo del potencial económico

Para estimar el potencial económico de gestionar a la cartera actual⁸ a través del programa de fidelización, se desarrollan los siguientes pasos:

1. Se calcula la propensión de los clientes a ser activos del programa de puntos según su comportamiento del último año, utilizando el modelo desarrollado en 6.3.3 Propensity Score.
2. En base a la propensión calculada, se clasifica a los clientes dentro de los 5 segmentos estudiados anteriormente en 6.3.5.
3. En el segmento 1 (de baja propensión) se descartan los clientes con menos de 1.000 puntos porque no tienen suficientes puntos para canjear. Esto no se realiza para los otros segmentos, ya que sus propensiones consideran implícitamente que tienen el stock de puntos necesario o, si los clientes ya han canjeado, que pueden volver a acumular los puntos.
4. Luego, para cada segmento se asume como efecto potencial la suma de los incrementales estimados en 6.3.7 Efecto por segmento. Es decir, el impacto en la rentabilidad siguiente al periodo de activación más el efecto de largo plazo (CLV). Dado este supuesto, se calcula el valor incremental de la cartera al activar distinto porcentaje de clientes en cada segmento.

Como resultado de este proceso, se obtiene la siguiente estimación.

⁸ A diciembre de 2017.

Segmento	N clientes (miles (%))	Rent. promedio (M\$)	CLV promedio (M\$)	Efecto segmento (M\$)	Beneficio* (MM\$) de activar al		
					1%	5%	10%
1	1.290 (56%)	67,9	407,2	133,6	1.724	8.621	17.242
2	326 (14%)	128,9	792,3	123,6	403	2.013	4.025
3	195 (8%)	151,3	955,2	91,8	179	894	1.788
4	143 (6%)	175,3	1121,1	44,7	64	319	638
5	368 (16%)	183,5	1377,4	-8,3	-30	-152	-305
TOTAL	2.321 (100%)				2.339	11.694	23.388

Tabla 38. Potencial económico de activar a los clientes en el club de fidelización
*Beneficio: Valor incremental de la cartera a 11 años (efecto en rentabilidad y CLV)

La gran mayoría de los clientes (56%) pertenece al segmento 1, los “oportunistas ocasionales”. Esto concuerda con la hipótesis (mencionada en 6.2.3) de que la tarjeta está fuertemente asociada al *retail* en la mente de los consumidores y, por ello, gran parte la utiliza como tarjeta secundaria para aprovechar sus ofertas exclusivas. De todas formas, dado el efecto estimado y la gran cantidad de clientes potenciales en este grupo, si se logra activar al 10% de estos clientes, se espera un retorno de aproximadamente 17.242 MM\$ a 11 años.

Luego, los segmentos 2, 3 y 4 poseen una cantidad relativamente menor de clientes y también un potencial económico menor, al activar el 10% de los clientes de cada segmento en el programa de puntos se genera un incremental monetario de aproximadamente 4.025 MM\$, 1.788 MM\$ y 638 MM\$ respectivamente.

Por último, el segmento 5 que corresponde a aproximadamente 370 mil clientes “Punto-fidelizados”, genera pérdidas para la empresa al activarse en el club de lealtad. Esto porque el efecto de superar este umbral definido como de activación en el programa para estos clientes es negativo, pero casi insignificante considerando su valor promedio inicial (-0,6% del CLV).

6.4.2 Acerca de la estructura del programa

Como se menciona anteriormente en 6.3.5, la propensión de los clientes y su valor rentable ya sea en la rentabilidad anual o en el CLV estimado, están correlacionadas positivamente. Es decir, los clientes de mayor valor para la empresa (segmento 5) son más propensos a participar del club de puntos y viceversa. Además, en el mismo análisis se observa que la propensión está positivamente correlacionada con el stock de puntos y la cantidad de puntos canjeados. Considerando de que cada punto es un costo para la empresa y de que a mayores niveles de canje el costo marginal es creciente, entonces los recursos del programa están correctamente enfocados en los clientes de mayor valor, al menos desde el punto de vista agregado a nivel de segmentos.

Respecto al efecto estimado, en general, se observa que los clientes de menor valor tienen mayor potencial de crecimiento y un impacto relativamente mayor. Este efecto, además del valor monetario, resulta en un mayor compromiso con la empresa y su programa de fidelización. En cambio, en los clientes de mayor valor el impacto es relativamente menor. Así mismo, los segmentos también pueden entenderse como estados de valor, en donde, el club de puntos desempeña una labor importante incrementando las probabilidades de transición a estados de mayor valor para el negocio, siempre y cuando el cliente tenga las características y capacidades económicas para avanzar. Desde una perspectiva similar, se puede interpretar al club de fidelización como un aporte o “puerta de entrada” para que los clientes de menor valor desarrollen su afinidad con la empresa y, por ende, su potencial como clientes – valga la redundancia. Mientras que en los clientes de mayor el valor, que ya se encuentran en un grado de fidelización mayor, el programa aporta en el objetivo de mantener una relación satisfactoria para el cliente, motivada por recompensas hacia su lealtad con la empresa.

Por lo presentado en los párrafos anteriores y a modo de evaluación general, se puede afirmar que la estructura del programa de puntos está correctamente alineada con el valor de los clientes y cumple la función de desarrollar el potencial de estos. La recompensación de los clientes brinda consecuencias positivas para el negocio, principalmente, incrementando el gasto promedio de los clientes con la tarjeta, el consumo de productos financieros y el interés de estos por continuar canjeando. Por lo tanto, no es necesario realizar cambios estructurales en el club de puntos.

6.4.3 Acerca de la gestión por segmentos

Segmento	Propensión promedio (%)
1	4
2	18
3	33
4	50
5	83

Tabla 39. Propensión a activarse en el club de puntos por segmento (2017)

Se debe destacar que las propensiones de los segmentos 1 y 2 son relativamente pequeñas, en promedio de 4% y 18% respectivamente. Sin embargo, por lo estimado en la Tabla 38 (6.4.1) estos clientes poseen el mayor potencial económico, por lo que se recomienda como parte principal de la táctica aumentar la propensión de estos clientes a participar del club de puntos. Esto se puede implementar a través de distintas formas, por ejemplo:

- Ampliar la oferta de canjes en niveles bajos.
- Personalizar la comunicación según líneas de canje y nivel de puntos más cercano al cliente, para así mejorar la tasa de respuesta y a la vez acelerar el gasto.

- SMS informativo cuando alcance el nivel de 5.000 puntos.
- Premio de puntos por primer canje.
- Premio de puntos por contratación de créditos o avances, para así incentivar el consumo de productos cruzados.
- Regalos de 500 puntos por compras superiores a 100.000 M\$ en *retails* internos, ya que se guían bastante por los descuentos de la tarjeta en estos negocios.

La implementación de la táctica para estos segmentos puede ser llevada a cabo de varias formas y se presentan algunas ideas a modo de ejemplo. El diseño de las operaciones que más se adecuan a las necesidades y que son más efectivas para cada segmento queda fuera del alcance del trabajo. Lo importante es que la táctica se debe enfocar en aumentar la propensión de activación en el club de estos dos grupos, educando al cliente de manera continua con informativos, recordatorios e incentivos.

Continuando con el resto de los segmentos, se tiene que el grupo 3 corresponde a clientes de compra frecuente, alta tasa de canje (80%), canjean más de una vez al año y son de alto valor según lo observado en su caracterización (6.3.5). Tienen un efecto potencial de + 92 M\$ en su rentabilidad de largo plazo y están próximos a alcanzar un nivel de muy alto valor, los clientes “Leales”. A pesar de ser clientes que han experimentado el canje de puntos, poseen una propensión del 33%. Es decir, uno de cada tres clientes va a activarse por sí solo y los otros dos requieren de un “empujón”. Para este segmento entonces, se recomiendan utilizar tácticas más oportunas, que les brinde la ayuda necesaria en el momento preciso. Por ejemplo:

- Enviar un aviso anticipado de que en 30 días van a vencer parte de sus puntos.
- Regalos de puntos como factor sorpresa para acercar al cliente al próximo nivel de canje.
- Premio de puntos por contratación de créditos o avances, para así incentivar el consumo de productos cruzados.
- La mayoría de este segmento pertenece a la categoría Normal y puede que esto sea un factor importante del vencimiento de sus puntos, por lo que ofrecerle el traspaso a la tarjeta Premium (a los pre-aprobados) con especial énfasis en la mayor duración de los puntos puede ser beneficioso para la relación cliente-empresa. De paso, puede desarrollar un sentimiento de exclusividad en el cliente que aumente aún más su valor.

Respecto al segmento 4, estos también poseen un efecto positivo, pero relativamente menor. Como son clientes ya muy comprometidos con la empresa, de alta experiencia en el club de puntos y con una propensión a activarse en el programa de 50%, basta con

mantener las gestiones que se realizan actualmente y enviarles avisos anticipados frente a vencimientos importantes de puntos.

Finalmente, el segmento 5 de los “Punto-fidelizados”, son clientes muy valiosos para la empresa y conocen perfectamente el sistema de recompensas. Poseen un impacto negativo al activarse en el programa de puntos, es decir, se genera una pérdida de valor cuando se activan, aunque esta es casi insignificante en relación su CLV. Como se menciona en 6.3.7, esta pérdida debe interpretarse como un subsidio hacia los negocios del *holding*, ya que se genera un incremental de + 812 M\$ en el gasto anual dedicado a estos *retails*. Además, la propensión de este segmento a activarse en el club de puntos es relativamente alta (88%), por lo que no es necesario gestionarlos para promover su activación. En cambio, ya que estos clientes son fanáticos de los puntos, se pueden utilizar premios de puntos como ganchos para promover el cruce de productos.

Resumiendo lo dicho en este sub-apartado, la táctica debe considerar:

- Aumentar la propensión de los segmentos 1 y 2 a través de educar e incentivar al cliente con comunicaciones, regalos y/o premios de bajo costo, pero con constancia para que el cliente aprenda y se interese por las recompensas. También es importante mantener una oferta amplia en niveles de canje bajos.
- Para el segmento 3, que ya tiene experiencia con el club de puntos, hay que darle ese “empujón” o factor sorpresa oportuno para que canjee.
- Al segmento 4 basta con enviarles avisos anticipados para que no caduquen sus puntos y mantenerlos al tanto de las nuevas oportunidades de canje, pero ya son experimentados en el programa.
- Al segmento 5 no es necesario regalarle puntos y/o gestionarlos para que se activen en el club. Más bien, se deben utilizar los puntos como atractivo para que consuman otros productos.

Para llevar a cabo esta táctica e implementar la gestión a nivel de segmentos, se debe considerar que el beneficio asociado a la gestión equivale al monto máximo que puede disponer la empresa para invertir en la activación de cada cliente. Para esto, se considera

$$\begin{aligned} ValorPorGestionar_s &= ValorPotencial_s - ValorEsperado_s \\ &= (1 - Propension_s) * Efecto_s \end{aligned}$$

En donde, $ValorPotencial_s$ corresponde al incremental asociado a la activación del cliente del segmento s y $ValorEsperado_s$ corresponde al incremental que se genera de natural en el segmento s por la estructura misma del club, es decir, dado el efecto y la propensión bruta de los clientes. Por lo tanto, el $ValorPorGestionar_s$ corresponde al beneficio a buscar con la gestión y, por ende, corresponde a la máxima disposición a invertir en estas actividades.

Segmento	Propensión (%)	Efecto (M\$)	Beneficio marginal por gestionar (M\$)
1	4	133,6	129,6
2	18	123,6	101,3
3	33	91,8	61,5
4	50	44,7	22,4
5	83	-8,3	0

Tabla 40. Beneficio marginal por gestionar por segmento (Equivalente a la máxima inversión a realizar por cliente)

En base a lo calculado en la tabla, se tiene que la máxima disposición a invertir por un cliente del segmento 1 debiese ser de 130 M\$ en la gestión para promover su activación en el club de puntos. Pero invertir este monto por cliente puede parecer una cantidad imprudente y arriesgada, por lo que se recomienda destinar una parte del presupuesto de fidelización a estas actividades y utilizar este dato como una razón para asignar los costos de gestión por segmento. Por ejemplo, se pueden destinar 10 M\$, 7,5 M\$ y 5 M\$ para cada cliente del segmento 1, 2 y 3 respectivamente, lo que en equivalencia a puntos correspondería a 6.250, 5.000 y 3.000 puntos aproximadamente a regalar por cliente. Por último, se aconseja utilizar métodos para priorizar la gestión de los clientes y obtener mejores tasas de respuesta, ya que invertir en todos los clientes de los primeros 3 segmentos (1,81 millones de clientes) puede ser una actividad poco eficiente. Por ejemplo: se puede priorizar a los clientes por su propensión a activarse; se pueden pilotear acciones de marketing y a través de un modelamiento por Uplift identificar a los clientes más influenciables; también puede resultar interesante priorizar a los clientes de mayor propensión a productos financieros o a comprar en cuotas, ya que estos tienen mayor potencial rentable para el negocio y es a quienes hay que fidelizar (6.2.3).

VII. CONCLUSIONES

Existe bastante investigación acerca del efecto de los clubes de lealtad en el comportamiento de los consumidores, sobre su aceleración en la frecuencia de compra, el aumento del gasto y el cruce de productos. Generalmente, esto se ha estudiado en el corto plazo, previo y posterior al canje, y no existe mucha literatura acerca del efecto en la relación cliente-empresa en el largo plazo. Este trabajo aborda esta falta y, además, cuantifica el valor de esta relación en términos monetarios. De esta forma, se mide el impacto del programa de fidelización de un *retail*/financiero en coherencia con su objetivo, fidelizar a los clientes para rentabilizarlos en el largo plazo.

Para cumplir el objetivo, se implementó una interesante metodología basada en el modelamiento del Customer Lifetime Value (CLV) y el método Propensity Score Matching (PSM). El primero, además de entregar un conocimiento importante a la empresa y servir como base para trabajos futuros sobre el valor de la cartera, se utiliza como *input* para estudiar el efecto causal de participar en el programa de puntos para así medir el impacto en la rentabilidad de los clientes en el largo plazo. Para el análisis de causalidad se utiliza el PSM que permite recrear los grupos control y tratamiento en experimentos no controlados, obteniéndose satisfactoriamente grupos similares en cuanto a sus comportamientos y características en el periodo observado, esto garantiza de que las muestras son comparables tal y como se tratase de un experimento aleatorizado, para así inferir causalidad del tratamiento - activación en el programa de puntos.

Acerca de la metodología propuesta, es posible extender su uso a otras áreas del negocio para estudiar y cuantificar los efectos en el desarrollo del valor de los clientes. Por ejemplo, en hitos importantes para el cliente en donde existe autoselección como:

- contratar pago automático de cuentas. Existe la hipótesis de que un cobro frecuente al cliente puede provocar un aumento en su compromiso con la tarjeta, por lo que sería interesante de estudiar.
- traspasarse de tarjeta, ya sea desde una Visa normal a una tarjeta Visa Platinum o una Signature.
- agregar una cuenta adicional, entre otros.

En lo siguiente se describen las conclusiones particulares obtenidas con el modelo del CLV y posteriormente con el método de *matching* para la estimación del efecto del programa de puntos.

7.1 Sobre el Customer Lifetime Value

Se elaboró el modelo de CLV a 10 años como combinación de dos sub-modelos, uno de rentabilidad a 3 años y un modelo de retención anual a 10 años. Las principales conclusiones son las siguientes:

- El modelo de rentabilidad que obtuvo mejor desempeño es una red neuronal Perceptrón multicapa, superando a una regresión lineal y un árbol CHAID. Esta red posee un WAPE de 65% implicando este porcentaje como error porcentual promedio respecto al resultado real a nivel de clientes, lo cual está dentro de lo esperado considerando trabajos anteriores y el plazo pronosticado.
- Como resultado general, se espera que la rentabilidad promedio a 3 años para la cartera de diciembre de 2017 sea de aproximadamente 185 M\$ (en valor presente), lo cual significa un crecimiento de 19 M\$ sobre lo observado para la cartera de diciembre de 2014 en sus 3 siguientes años. Este pronóstico posee una distribución muy cargada la izquierda, con una mediana de 85 M\$ y el cuartil inferior de -13 M\$. La hipótesis que explica esta concentración corresponde a que, como tarjeta fuertemente asociada al *retail*, existe un alto porcentaje de la cartera (48%) que utiliza la tarjeta casi únicamente en los negocios del *holding* (80% del gasto o más) por sus promociones exclusivas, más no la usan como una tarjeta principal.
- Los ingresos que produce actualmente el cliente están fuertemente correlacionados con sus rentabilidades futuras. Por lo que, los productos financieros y el uso de cuotas surgen como principales indicadores de rentabilidad. En particular, se tiene que en el 2017:
 - un 16% de los clientes contrató avances en efectivo y, en promedio, se espera que generen 576 M\$ más en rentabilidad a 3 años que quienes no lo hicieron;
 - un 15% de los clientes contrató créditos de consumo y, en promedio, se espera que generen 425 M\$ más en rentabilidad a 3 años que quienes no lo hicieron;
 - un 48% de los clientes compró al menos una vez en 6 cuotas o más y, en promedio, se espera que generen 281 M\$ más en rentabilidad a 3 años que quienes compraron en el periodo, pero en cuotas menores.
- Respecto a la componente de retención, se utiliza el criterio de fuga de la empresa que corresponde a 12 meses sin compras ni pagos. En base a esto, se modela la probabilidad de supervivencia para cada año extra hasta un plazo de 10 años. El modelo que obtuvo mejor desempeño es el modelo Weibull, dado el criterio de Akaike que permite comparar la calidad de ajuste relativa a otros modelos.
- A diciembre de 2017 existen 307 mil clientes fugados. Del resto, se espera que el 6,6% se fugue en el siguiente año y el 71,8% se mantenga por 11 periodos o más. Este pronóstico sujeto al supuesto de que el cliente se fuga cuando posee menos de 50% de probabilidades de sobrevivir en el periodo, lo que se conoce como tiempo de vida medio.
- La retención del cliente está directamente relacionada con su nivel de actividad en la empresa. Clientes de mayor gasto, mejor categoría, que compran fuera del *holding*, que consumen otros productos, que llevan años activos, tienen bajo *recency* y canjean puntos, son los clientes que tienen mejores tasas de retención. En particular, para el pronóstico de la cartera a diciembre de 2017 se tiene que:

- las categorías Elite y Premium tiene la mejor retención esperada a 10 años, de 91% y 88% respectivamente. En cambio, clientes del quintil 4 y 5 poseen las más bajas, de 51% y 15% respectivamente;
- quienes contrataron créditos de consumo poseen una retención esperada de 91% a 10 años, quienes contrataron avances de 82% en promedio y quienes no consumieron estos productos poseen una retención esperada de 55%;
- respecto al programa de puntos, las probabilidades de supervivencia a 10 años corresponden a 89%, 84% y 53% para clientes que realizaron 2 canjes, 1 canje y quienes no realizaron canjes en el 2017 respectivamente.
- Luego de modeladas ambas componentes, rentabilidad y retención, se consolidan para crear la esperanza de las rentabilidades futuras de cada cliente. El valor promedio de la cartera a 10 años es de 834 M\$ con una mediana de 348 M\$.
 - Respecto a las categorías, se tiene que los clientes Elite, Premium y Normal tienen un valor promedio de 1.527 M\$, 1.223 M\$ y 738 M\$ respectivamente. Por otro lado, los clientes que canjearon en el último año se espera que renten 1.562 M\$ versus los 560 M\$ de quienes no lo hicieron. Según lo observado en cada sub-modelo y lo recién mencionado, es de esperar que exista causalidad respecto al programa de puntos, pero con un efecto de magnitud menor a estas diferencias, ya que existe un sesgo de autoselección latente.
- Al igual que en los resultados del modelo de retención, se tiene una distribución del CLV de los clientes muy cargada hacia la izquierda. Para trabajar el valor de la cartera entonces, existe un gran porcentaje que utiliza la tarjeta con baja frecuencia y tiene un alto potencial de crecimiento. Lo que se recomienda es identificar a clientes propensos a los créditos de consumo, a avances y a pagar en cuotas, ya que son los de mayor potencial en términos rentables. Luego, se deben enfocar los esfuerzos en desarrollar la retención de estos clientes, por ejemplo, a través de:
 - ofertas agresivas y personalizadas en créditos de consumo, ya que promueven un compromiso de largo plazo;
 - gestión continua con el programa de puntos para incentivar y educarlos respecto al sistema de recompensas;
 - metas de descuentos y metas de compra en nuevos rubros para que empiecen a darle un uso general a la tarjeta fuera del *holding*,
 - ofrecer traspasos a tarjetas mejores con mantención gratuita durante un plazo de tiempo, para que así asciendan de categoría, tengan mejores beneficios y sientan una mayor exclusividad.

7.2 Sobre el efecto del programa de fidelización

Se estimó el impacto del club de puntos en el valor rentable de los clientes en el largo plazo, analizando también el detalle de qué ocurre en sus comportamientos. Para esto se definió un criterio de actividad en el programa como una combinación lineal entre métricas de interacción con el club de puntos, el número de canjes y la tasa de vencimiento. En base a esto, se utilizó PSM y se obtuvo las siguientes conclusiones.

- El porcentaje activo en el club de puntos durante 2015-2016 corresponde a un 16%, en donde los clientes presentaron un incremental de 15 M\$ en la rentabilidad siguiente y de 62 M\$ en las rentabilidades a 10 años, como efectos de corto y largo plazo respectivamente. Por lo tanto, queda demostrado que el programa de puntos sí tiene un efecto positivo en la fidelización y el desarrollo del valor del cliente, agregando un valor de 77 M\$ a 11 años promedio en los clientes tratados. Este se explica por un leve aumento del compromiso del cliente con la empresa. Más concretamente porque:
 - aumenta en 41 M\$ el gasto en negocios del *holding*, sumado a una preferencia por *retails* internos, posiblemente porque brindan mayor cantidad de puntos;
 - aumenta un 2% el monto consumido en cuotas, ya que clientes propensos a las cuotas sienten mayor comodidad a la hora de endeudarse con la tarjeta (luego de haber experimentado el club de puntos) porque saben que si le dan un uso continuo serán recompensados;
 - aumenta la propensión a consumir productos financieros en 2,7% en créditos y en 1,6% en avances, si bien son incrementales relativamente pequeños, el beneficio en rentabilidad es muy importante para el negocio;
 - el cliente quiere repetir la experiencia, aumentando el número de canjes (+0,5) y los puntos canjeados (+1.750) por año. Esto genera un impacto positivo en la retención.

Se profundizó y estudió el efecto a nivel de 5 segmentos comprobando que los clientes presentan distintos efectos. Estos grupos fueron construidos en base a los quintiles de la propensión a participar del programa y se obtuvo lo siguiente:

- Segmentos de baja propensión corresponden a clientes de menor consumo, menor actividad en el club de puntos y menor valor para la empresa. En cambio, segmentos de muy alta propensión corresponden a clientes de alto consumo con la tarjeta y de productos financieros, clientes muy activos, que canjean y son de muy alto valor para el negocio. Los perfiles intermedios, a su vez, corresponden a distintos grados dentro de esta gama de clientes. Esto evidencia una correlación positiva entre el valor rentable del cliente y su propensión a participar del club puntos, por ende, sus recursos se están destinando a los clientes más rentables. Entonces, se concluye que la estructura del club está correctamente planteada y no requiere de cambios importantes.

- En los segmentos de menor valor existe un mayor impacto, dado que son clientes que tienen un mayor potencial de crecimiento. En el segmento 1, 2 y 3 (de menor propensión) los incrementales promedio en el valor del cliente causado por la activación en el club corresponden a 134 M\$, 123 M\$ y 92 M\$ respectivamente. Este efecto se da principalmente por un aumento en la propensión a créditos de consumo y avances en efectivo, como también por un aumento en el interés por volver a canjear en los próximos años (fidelización).
- En cambio, los segmentos 4 y 5 son clientes ya fidelizados, que utilizan la tarjeta como medio de pago principal, tienen una alta tasa de consumo de productos y tienen experiencia previa con los puntos. Tienen menor potencial de crecimiento y esto se refleja en un menor impacto, esto porque ya alcanzaron una base de valor muy alta respecto a los otros segmentos. Ambos segmentos presentan un incremental positivo en el gasto, con especial preferencia por los *retails* del *holding*, ya que les brinda mayor acumulación de puntos. El impacto en el consumo de productos financieros es muy pequeño e incluso nulo para el segmento 5 en créditos de consumo. Respecto al club de puntos, el segmento 4 aumenta en 1.800 sus puntos canjeados al año, mientras que el segmento 5 disminuye en casi 3.000 puntos. Como resultado, los segmentos 4 y 5 tienen efectos de +45 M\$ y -8 M\$ respectivamente. Si bien, este último presenta una pérdida, esta es casi insignificante (<1%) en relación con su valor inicial y debe entenderse como un subsidio hacia los negocios del *holding*, ya que este segmento en particular aumenta en 812 M\$ su gasto anual en estos *retails*.

Dado el conocimiento generado, se cuantificó el potencial económico de llevar a cabo una gestión para activar a los distintos segmentos. Al activar al 10% de los clientes del segmento 1, 2, 3 y 4 se genera un beneficio incremental en el valor de la cartera de 17.242 MM\$, 4.025 MM\$, 1.788 MM\$ y 638 MM\$ respectivamente. Por lo que se recomienda lo siguiente:

- La táctica se debe enfocar en activar a los clientes del segmento 1 y 2 por su mayor potencial y efecto esperado. Dado que no poseen mucha experiencia ni interés por el programa de puntos, tienen propensiones muy bajas, 4% y 18% respectivamente. Para este perfil de clientes se recomienda ampliar la oferta de canjes en niveles bajos, enviarles ofertas de canjes personalizadas y aplicar campañas de educación e incentivos constantes para que se introduzcan en el sistema de las recompensas. En estos clientes el sistema de recompensas puede interpretarse como una puerta de entrada al ecosistema de beneficios que brinda la empresa a sus clientes.
- En segundo lugar, se debe gestionar a los clientes del segmento 3. Si bien ya conocen el programa de puntos, su propensión promedio es de 33%. Por lo que se recomienda una táctica más oportuna, brindarles en el momento preciso el apoyo que necesitan para activarse. Esto puede ser a través de recordatorios de próximos puntos a vencer, ofertas de canje, regalos sorpresa que los acerquen al próximo nivel de canje, etc.
- El segmento 4 posee bastante experiencia con el programa de puntos y una propensión del 50%, por lo tanto, basta con enviarles recordatorios de productos que se pueden canjear o avisos de vencimiento para mantenerlos al tanto. El segmento 5,

por su parte, no es necesario gestionarlos ya que tienen una propensión del 88%. En ambos segmentos el club de lealtad cumple la función de mantenerlos fidelizados. No es necesario destinar gran cantidad de recursos adicionales para incentivar su actividad en el club, más bien se deben utilizar los puntos como atractivo para potenciar el cruce de productos.

- Respecto a la inversión que se debe destinar para estas actividades, se recomienda utilizar como guía la razón del beneficio que puede generar la gestión para cada segmento. Por ejemplo, se puede disponer de 10 M\$, 7,5 M\$ y 5 M\$ por cliente del segmento 1, 2 y 3 respectivamente, lo que en equivalencia a puntos correspondería a 6.250, 5.000 y 3.000 puntos a regalar por cada cliente. Invertir esto en todos los clientes no es eficiente, más bien se recomienda enfocarse en los clientes de mayor propensión a participar del club y con mayor propensión a productos financieros o las cuotas, ya que esto mejora las tasas de respuesta y tiene un retorno potencial mayor.

7.3 Limitaciones y trabajos futuros

Respecto al CLV, se propone perfeccionar el modelo incorporando distintos horizontes de tiempo en el entrenamiento de este. Para este trabajo solo se contaba con rentabilidades anuales para años completos, desde enero a diciembre. Entonces, el modelo fue entrenado con información del cliente para años fijos y se recomienda que sea utilizado de esta manera, por ejemplo, con información de enero a diciembre 2017 para pronosticar los años futuros. De no ser utilizado así, por ejemplo, con información de julio a junio, puede resultar en estimaciones sesgadas, dado que se utilizaron variables del tipo tendencias o gasto semestral que capturan estacionalidades implícitas. Por lo tanto, si se desea tener un modelo más robusto y que permita calcular el CLV de manera frecuente durante distintos periodos del año, se recomienda realizar el cálculo de rentabilidad para todos los periodos y reentrenar el modelo.

En la misma línea del CLV, este modelo fue elaborado solo para los clientes que poseen al menos un año de cuenta abierta. Por lo que no es posible estimar el valor de los clientes nuevos hasta que posean este periodo de tiempo, por temas de la información en sus covariables. Se recomienda realizar un CLV especial para estos clientes, por ejemplo, que considere sus primeros tres meses en la empresa y de esta forma empezar la gestión de estos de manera temprana y personalizada según las necesidades de cada uno.

Respecto al programa de puntos, se propuso una métrica de interacción con el club del tipo continua. Sin embargo, el criterio fue definido de forma binaria. Esto implica que el umbral de activación corresponde a una línea, arbitraria, en dos dimensiones. Entonces, es muy probable que clientes que cercanos al corte fueron emparejados (*match*) con clientes que están del otro lado del corte, pero a una distancia mínima. Es decir, para efectos prácticos, clientes con grados de participación en el programa de puntos iguales pero separados por décimas. Esto puede haber reducido o subestimado el verdadero valor del club de puntos. Por lo tanto y a modo de análisis de sensibilidad, se recomienda

evaluar el tratamiento a nivel de dosis, es decir, con grados de actividad o derechamente con métricas categóricas de menor ambigüedad, como lo es el nivel de canje.

Respecto al conocimiento generado sobre el efecto del programa de puntos, se propone aterrizar las tácticas a través del diseño de campañas que se adecuen a las necesidades de cada segmento, optimizando la comunicación y el uso de recursos para obtener mejores tasas de respuesta. Dado que estos segmentos son relativamente grandes, también se requiere de una lógica para priorizar a los clientes. Una idea para esto último, puede ser el diseño de tácticas que funcionen en conjunto. Por ejemplo, potenciar el consumo de cuotas a través de incentivos con puntos y viceversa.

En particular, en los resultados del segmento 5, los “Punto-fidelizados”, se observó un efecto negativo en su valor como clientes para la empresa, pero un efecto relativamente grande en términos de gasto en negocios del *holding* que puede significar un subsidio por parte de la empresa que desarrolla la lealtad de los clientes hacia estos comercios. Esto puede ocurrir en ambas direcciones y de distintas maneras, por lo tanto, sería interesante conocer cómo interactúan los valores de los clientes (CLV) de cada una de estas empresas asociadas. Quizás, con una metodología similar a la del presente trabajo se podría estudiar cómo afectan acciones de los otros negocios al valor del cliente de esta empresa y viceversa, de modo de evaluar y desarrollar estrategias que optimicen el bienestar común.

VIII. BIBLIOGRAFÍA

- Ares, F. (2008). Valorización marginal del programa de pasajero frecuente en una línea aérea. *Memoria para optar al título de Ingeniero Civil Industrial*. Universidad de Chile.
- Austin, P. (2011). Optimal Caliper Widths for Propensity-Score Matching When Estimating Differences in Means and Differences in Proportions in Observational Studies. *Pharmaceutical Statistics* 10.2, 150-161.
- Carboni, A. (2012). Estudio de las respuestas de los consumidores ante un programa de lealtad. *Tesis para optar al título de Magister en gestión de operaciones*. Universidad de Chile.
- Fader, P. H. (2005). RFM and CLV: Using Iso-CLV Customer for Customer Base Analysis. *Journal of Marketing Research*, Vol. 42, 415-30.
- Fader, P. H. (2015). Probability Models for Customer-Base Analysis. 26th Anual Advanced Research Techniques Forum.
- Fayyad, U. P.-S. (1996). The KDD Process for extracting useful from volumes of data. *Communications of the ACM*, Vol. 39, No. 11.
- Gallardo, C. (2016). 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*. Universidad de Chile.
- Gu, X. R. (1993). Comparison of multivariate matching methods: structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2: 405-420.
- Gupta, S. D. (2006). Modeling Customer Lifetime Value. *Journal of Service Research*, 9: 139.
- Gupta, S. L. (2003). Customers as Assets. *Journal of Interactive Marketing*, 17 (1), 9-24.
- Hadi, A. (1992). Identifying Multiple Outliers in Multivariate Data. *Journal of the Royal Statistical Society. Series B: Methodological*, 54(3).
- Klein, J. M. (2003). Survival analysis techniques for censored and truncated data. Chapter 12.
- Kumar, V. (2006). *CLV: A Path to Higher Profitability*. University of Connecticut, Storrs.
- Morrison, D. S. (1980). *Jobs, Strikes, and War: Probability Models for Duration*. Columbia University.
- Pfeifer, P. C. (2000). Modeling Customers Relationship as Markov Chains. *Journal of Interactive Marketing*, 14 (2), 43-55.
- Pickering, M. (1999). Not all Customers are Created Equal, Looking at Lifetime Value. *Charter magazine, the Journal of the Australian Institute of Chartered Accountants*, (Vol. 70 No. 9. Pages 38 to 40).
- Reuse, F. (2011). Estimación del valor de clientes fidelizados en una cadena de supermercados utilizando modelos Jerárquicos Bayesianos. *Memoria para optar al título de Ingeniero Civil Industrial*. Universidad de Chile.
- Ríos, C. (2013). Análisis del efecto de un club de fidelización en el comportamiento de clientes bajo régimen contractual. *Memoria para optar al título de Ingeniero Civil Industrial*. Universidad de Chile.
- Rodríguez, V. (2013). Estimación del Customer Lifetime Value a nivel de clientes, de un banco usando variables transaccionales y sociodemográficas. *Memoria para optar al título de Ingeniero Civil Industrial*. Universidad de Chile.
- Rosenbaum, R. R. (1984). Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc.*, 79;516-24.

- Rubin, D. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74: 318-328.
- Rubin, D. R. (1983). The Central Role of the Propensity Score in Observational Studies for Causal. *Biometrika*, Vol. 70, No. 1, pp. 41-55.
- Schmittlein, D. M. (1987). Counting Your Customers: Who Are They and What Will They Do Next? *Management Science*, Vol. 33, No. 1, 1-24.
- Sekhon, J. (2011). Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. *Journal of Statistics Software*, Vol. 42 (7).
- Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25 (1), 1-21.
- Stuart, E. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, Volume 66, Issue 8, Supplement, 84-90.
- Therneau, T. (2015). A Package for Survival Analysis in S.
- Troncoso, I. (2016). Estudio del efecto de un programa de fidelización en el comportamiento de compra de sus clientes. *Memoria para optar el título de Ingeniera Civil Industrial*. Universidad de Chile.
- Uncles, M. D. (2003). Customer loyalty and customer loyalty programs. *Journal of Consumer Marketing*, Vol. 20 Issue: 4, pp.294-316.
- Urzúa, P. (2007). Estimación del Customer Lifetime Value mediante técnicas supervisadas de data mining en una empresa de retail. *Memoria para optar al título de Ingeniero Civil Industrial*. Universidad de Chile.
- Yang, D. D. (2012). A unified approach to measuring the effect size between two groups using SAS. *Departments of Quantitative Health Sciences and Outcomes Research, Cleveland Clinic*.
- Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirement, matching metrics, and monte carlo evidence. *Review of Economics and Statistics*, 86 (1), 91-107.

IX. ANEXOS

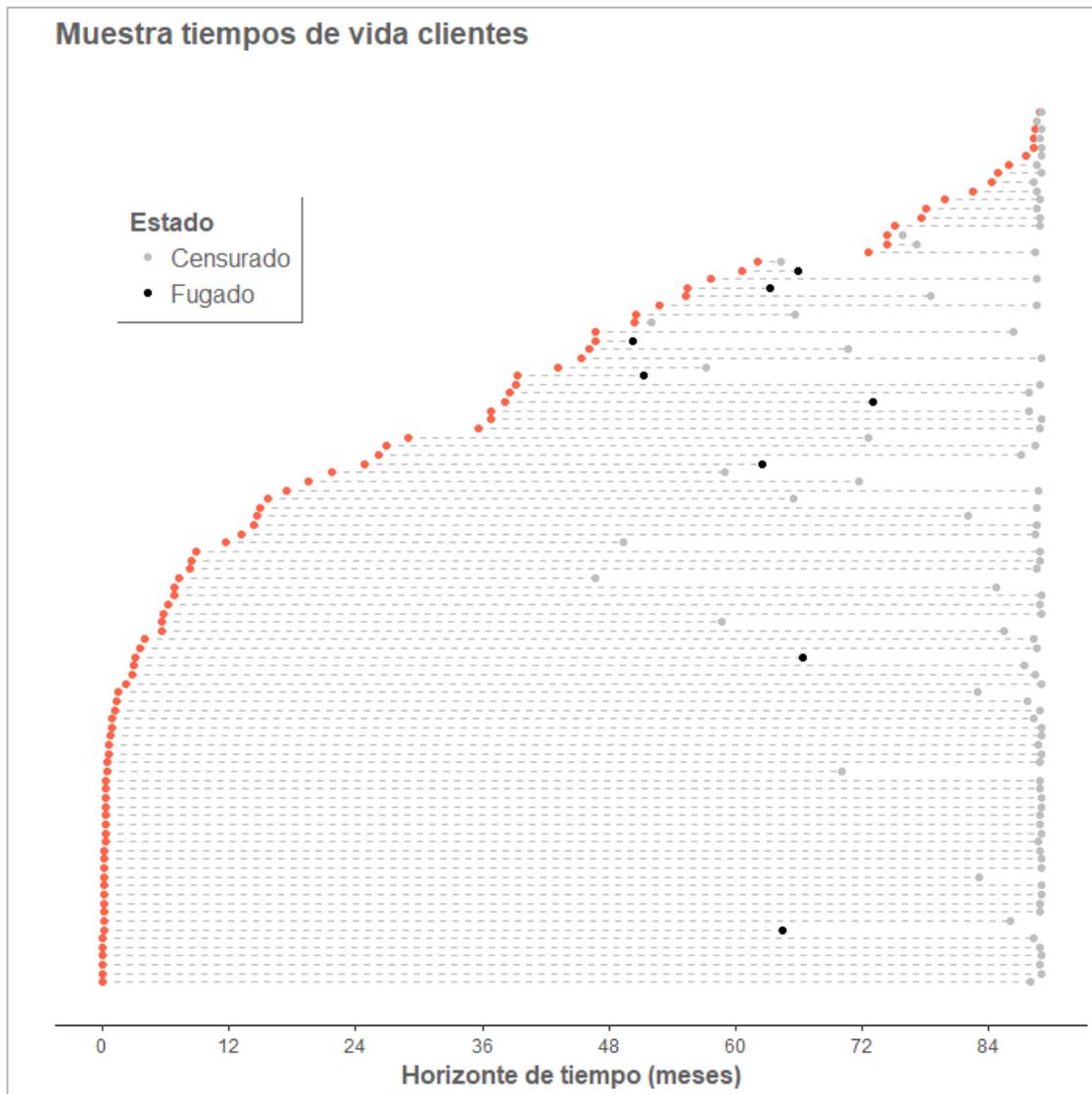
9.1 Anexos A: Variables

Nombre y tipo de variable (c: categórica, d: dummy, #: numérica)	
Identificador del cliente	Transaccional
Sociodemográficas	# Recency en meses
C Sexo	C Tramo de recency
C Estado civil	# Frecuencia de facturación (emisión de estados de cuenta)
# Edad	# Cuota promedio ponderada por montos
C Tramo edad	# Máxima cuota
C GSE	# Máximo monto por transacción
C Zona geográfica	# Porcentaje de transacciones con cuotas
Contrato	# Porcentaje de gasto en cuotas
# Antigüedad	# Porcentaje de intereses sobre gasto
D Se traspasó de tarjeta en últimos 12 meses	# Porcentaje pagado con la tarjeta en negocios internos
D Se traspasó de tarjeta en últimos 6 meses	# Días con compra
# Mono de cupo	# Meses con compra
C Tramo monto de cupo	# Gasto 1er semestre
D Tiene adicionales	# Gasto 2do semestre
# Número de adicionales	# Gasto
Productos	C Trimestre de mayor gasto
C Categoría	# Variación gasto semestral
C Categoría_2 (Normal: con quintiles de gasto)	# Variación gasto trimestres 1-2
D Tiene seguro desgravamen	# Variación gasto trimestres 3-4
D Tiene seguro cesantía	# Porcentaje de gasto trimestre 1
D Tiene seguro otros	# Porcentaje de gasto trimestre 2
D Tiene créditos de consumo	# Porcentaje de gasto trimestre 3
# Monto en créditos de consumo últ. año	# Porcentaje de gasto trimestre 4
# Número de créditos de consumo últ. año	# Porcentaje del gasto en semestre 2
D Tiene avance	D Con compra en negocio interno: tiendas por depto.
# Monto en avances últ. año	D Con compra en negocio interno: mejoramiento del hogar
# Número de avances últ. año	D Con compra en negocio interno: supermercado
D Tiene pago automático	D Con compra en negocios externos
# Monto en pago automático últ. año	# Número de negocios (internos) con compra
# Número de transacciones automáticas últ. año	D Compras en negocio interno: tiendas por depto.
D Tiene BIP post-pago	# Compras en negocio interno: mejoramiento del hogar
Riesgo	# Compras en negocio interno: supermercado
C Tramo score de riesgo	# Compras en negocio interno: otros
# Monto disponible	# Compras en negocio externos
# Porcentaje cupo disponible	# Gasto en negocio interno: tiendas por depto.
C Tramo porcentaje disponible	# Gasto en negocio interno: mejoramiento del hogar
C Tramo porcentaje disponible últ. año	# Gasto en negocio interno: supermercado
# Meses en mora últ. año	# Gasto en negocio interno: otros
# Porcentaje de estados de cuenta con atrasos	# Gasto en negocios externos
# Porcentaje de estados de cuenta prepagados	# Porcentaje del gasto en negocio interno: tiendas p.d.
Puntos	# Porcentaje del gasto en negocio interno: mej. hogar
# Tasa de canje	# Porcentaje del gasto en negocio interno: supermercado
# Stock de puntos	# Porcentaje del gasto en negocios externos
# Puntos acumulados últ. año	# Número de rubros semestre 1
D Con canje últ. año	# Número de rubros semestre 2
# Número de canjes últ. año	# Número de rubros con compra
# Puntos canjeados últ. año	# Variación semestral de rubros con compra
# Distancia al próximo nivel de canje	# Con compra en rubro transporte
Canales	D Con compra en rubro entretenimiento
# Ingresos web últimos 6 meses	D Con compra en rubro educación
D Activo web últimos 6 meses	D Con compra en rubro salud
# Variación semestral de ingresos web	D Con compra en rubro supermercados
D Activo sucursal en últimos 6 meses	D Con compra en rubro recaudación
# Interacciones sucursales en últimos 6 meses	D Con compra en rubro tiendas por departamento
D Activo tótems en últimos 6 meses	D Con compra en rubro seguros
# Días con ingreso tótems en últimos 6 meses	D Con compra en rubro viajes
Rentabilidad	D Con compra en rubro hogar

# Rentabilidad últ. año	D Gasto en rubro transporte
# Ingresos últ. año	# Gasto en rubro entretenimiento
# Costos últ. año	# Gasto en rubro educación
# Porcentaje de ingresos por productos financieros	# Gasto en rubro salud
# Porcentaje de ingresos por intereses	# Gasto en rubro supermercados
# Porcentaje de costos por riesgo	# Gasto en rubro recaudación
# Porcentaje de costos por fidelización	# Gasto en rubro tiendas por departamento
# Costo canjes cliente	# Gasto en rubro seguros
	# Gasto en rubro viajes
	# Gasto en rubro hogar

Anexo 1. Tabla de variables

9.2 Anexos B: Modelo de retención

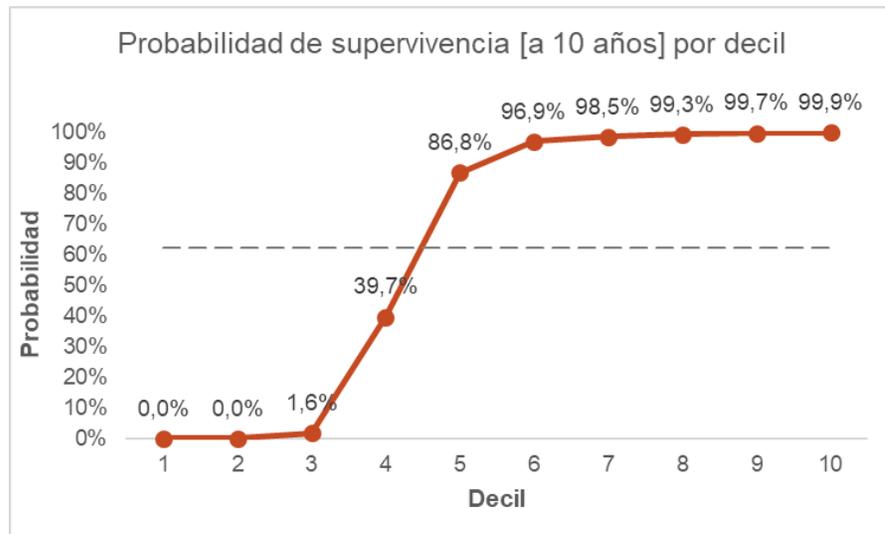


Anexo 2. Muestra de tiempos de vida de los clientes, fugados y censurados (se muestran 100 clientes aleatorios, el horizonte de observación inicia en ene-2011)

Variable	Coeff.	ETR	LB	UB
Parámetro de escala	0,49			
Intercepto	1,74			
Gasto negocio interno: supermercado (\$100 miles)	0,08	1,08	1,07	1,09
Gasto negocios externos (\$100 miles)	0,03	1,03	1,03	1,03
Tiempo activo (años)	0,68	1,97	1,96	1,98
Tiene seguro desgravamen	-0,08	0,92	0,91	0,93
Tiene seguro cesantía	0,16	1,18	1,17	1,19
Ratio deuda / cupo	0,92	2,51	2,45	2,57
Tiene crédito de consumo	-0,32	0,73	0,70	0,75
Tiene avances	0,07	1,07	1,06	1,09
Meses con mora último año	-0,09	0,92	0,91	0,92
Tiene adicionales	-0,18	0,84	0,82	0,85
Cantidad de canjes último año	-0,10	0,91	0,90	0,91
Puntos en stock (miles)	0,01	1,01	1,01	1,01
Tramo score riesgo a) 0	-	-	-	-

Tramo score riesgo b) 1-300	0,40	1,49	1,46	1,52
Tramo score riesgo c) 301-600	0,40	1,49	1,47	1,51
Tramo score riesgo d) 601-1.000	0,33	1,40	1,38	1,41
Categoría a) Q5	-	-	-	-
Categoría b) Q4	-0,00	1,00	0,98	1,01
Categoría c) Q3	0,09	1,10	1,08	1,11
Categoría d) Q2	0,16	1,18	1,15	1,20
Categoría e) Q1	0,16	1,17	1,13	1,21
Categoría f) PREMIUM	0,91	2,48	2,33	2,64
Categoría g) ELITE	0,91	2,48	2,23	2,76
Tramo edad a) 18-25	-	-	-	-
Tramo edad b) 26-30	-0,03	0,97	0,95	0,99
Tramo edad c) 31-35	-0,18	0,83	0,82	0,85
Tramo edad d) 41-45	-0,19	0,83	0,81	0,85
Tramo edad e) 46-50	-0,19	0,83	0,82	0,85
Tramo edad f) 51-55	-0,18	0,84	0,82	0,86
Tramo edad g) 56-60	-0,18	0,82	0,80	0,84
Tramo edad h) 61-65	-0,20	0,83	0,81	0,85
Tramo edad i) 66+	-0,20	0,77	0,75	0,78
GSE a) Sin información	-	-	-	-
GSE b) E	-0,24	0,79	0,77	0,81
GSE c) D	0,35	1,41	1,40	1,43
GSE d) C3	0,18	1,19	1,18	1,21
GSE e) C2	0,05	1,05	1,03	1,06
GSE f) ABC1	0,08	1,08	1,06	1,10
Tramo cupo a) 0 – 0,1 MM\$	-	-	-	-
Tramo cupo b) 0,1 – 0,2 MM\$	0,04	1,04	1,02	1,06
Tramo cupo c) 0,2 – 0,3 MM\$	-0,17	0,84	0,83	0,85
Tramo cupo d) 0,3 – 0,5 MM\$	-0,12	0,88	0,87	0,90
Tramo cupo e) 0,5 – 0,75 MM\$	-0,31	0,74	0,72	0,75
Tramo cupo f) 0,75 – 1,0 MM\$	-0,27	0,76	0,75	0,77
Tramo cupo g) 1,0 – 1,25 MM\$	-0,29	0,75	0,73	0,76
Tramo cupo h) 1,25 – 1,5 MM\$	-0,23	0,80	0,78	0,82
Tramo cupo i) 1,5 – 2,0 MM\$	-0,32	0,72	0,71	0,74
Tramo cupo j) 2,0 MM\$ o más	-0,32	0,72	0,71	0,74
Tramo porcentaje disponible a) 0-20%	-	-	-	-
Tramo porcentaje disponible b) 21-40%	0,29	1,34	1,30	1,38
Tramo porcentaje disponible c) 41-60%	0,40	1,49	1,45	1,54
Tramo porcentaje disponible d) 61-80%	0,45	1,56	1,52	1,60
Tramo porcentaje disponible e) 81-100%	0,36	1,44	1,40	1,47
Categoría Q5 * Con canje	-0,14	0,87	0,84	0,90
Categoría Q4 * Con canje	-0,14	0,87	0,85	0,89
Categoría Q3 * Con canje	-0,19	0,83	0,81	0,85
Categoría Q2 * Con canje	-0,10	0,91	0,88	0,93
Categoría Q1 * Con canje	0,03	1,03	0,99	1,07
Categoría PREMIUM * Con canje	-0,50	0,61	0,56	0,66
Categoría ELITE * Con canje	-0,03	0,97	0,80	1,17

Anexo 3. Modelo de retención Weibull: variables y coeficientes



Anexo 4. Lift por decil de supervivencia a 10 años (2028)

9.3 Anexos C: Modelo de propensión

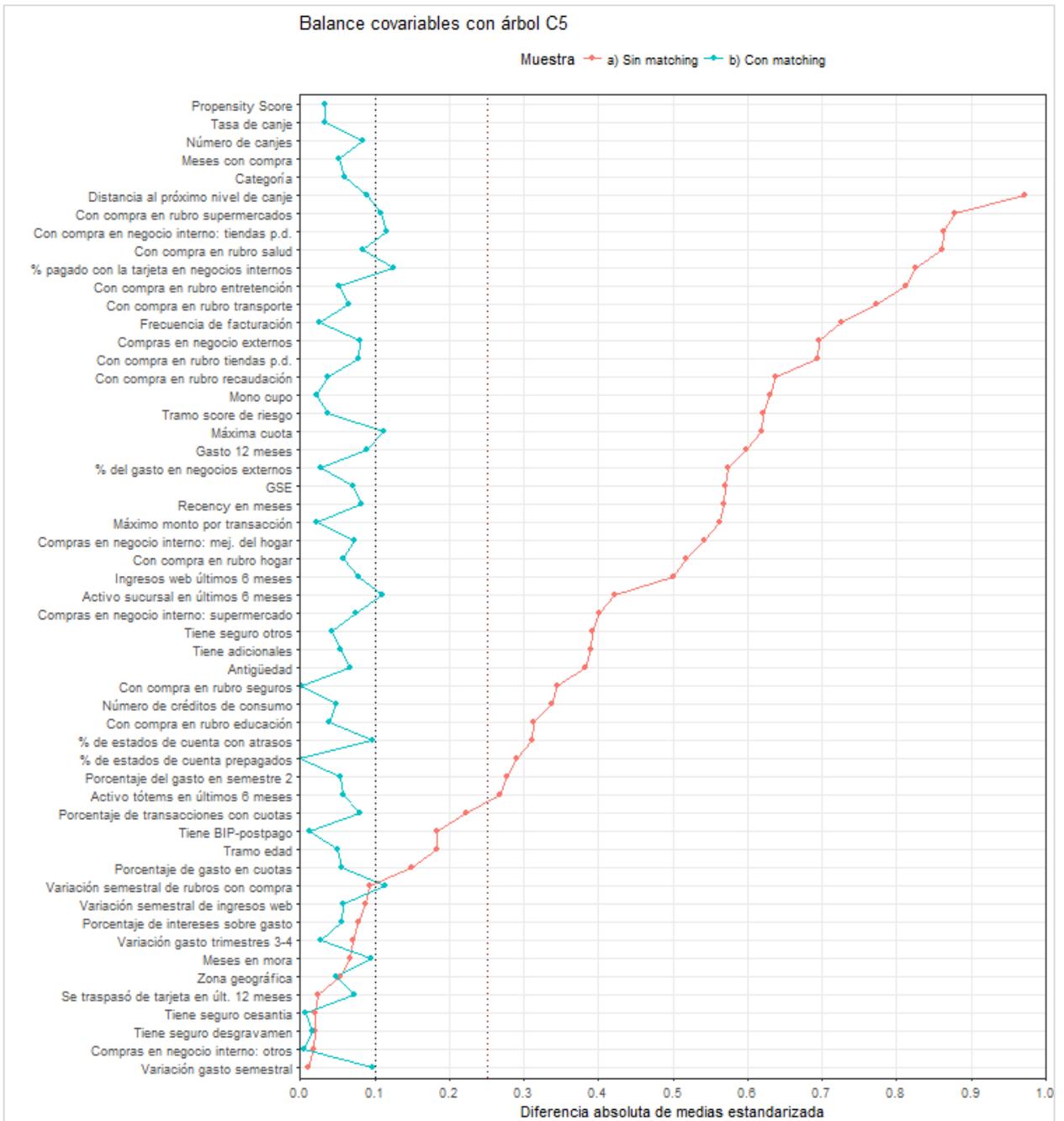
Pronóstico Regresión Logística			Accuracy
Activo	Activo 13,4%	Inactivo 2,7%	84,9%
Inactivo	12,4%	71,5%	
Pronóstico Árbol C5			Accuracy
Activo	Activo 13,8%	Inactivo 2,4%	85,1%
Inactivo	12,6%	71,3%	
Pronóstico Árbol CHAID			Accuracy
Activo	Activo 14,1%	Inactivo 2,0%	81,7%
Inactivo	16,3%	67,5%	

Anexo 5. Matriz de confusión y desempeño de modelos de propensión

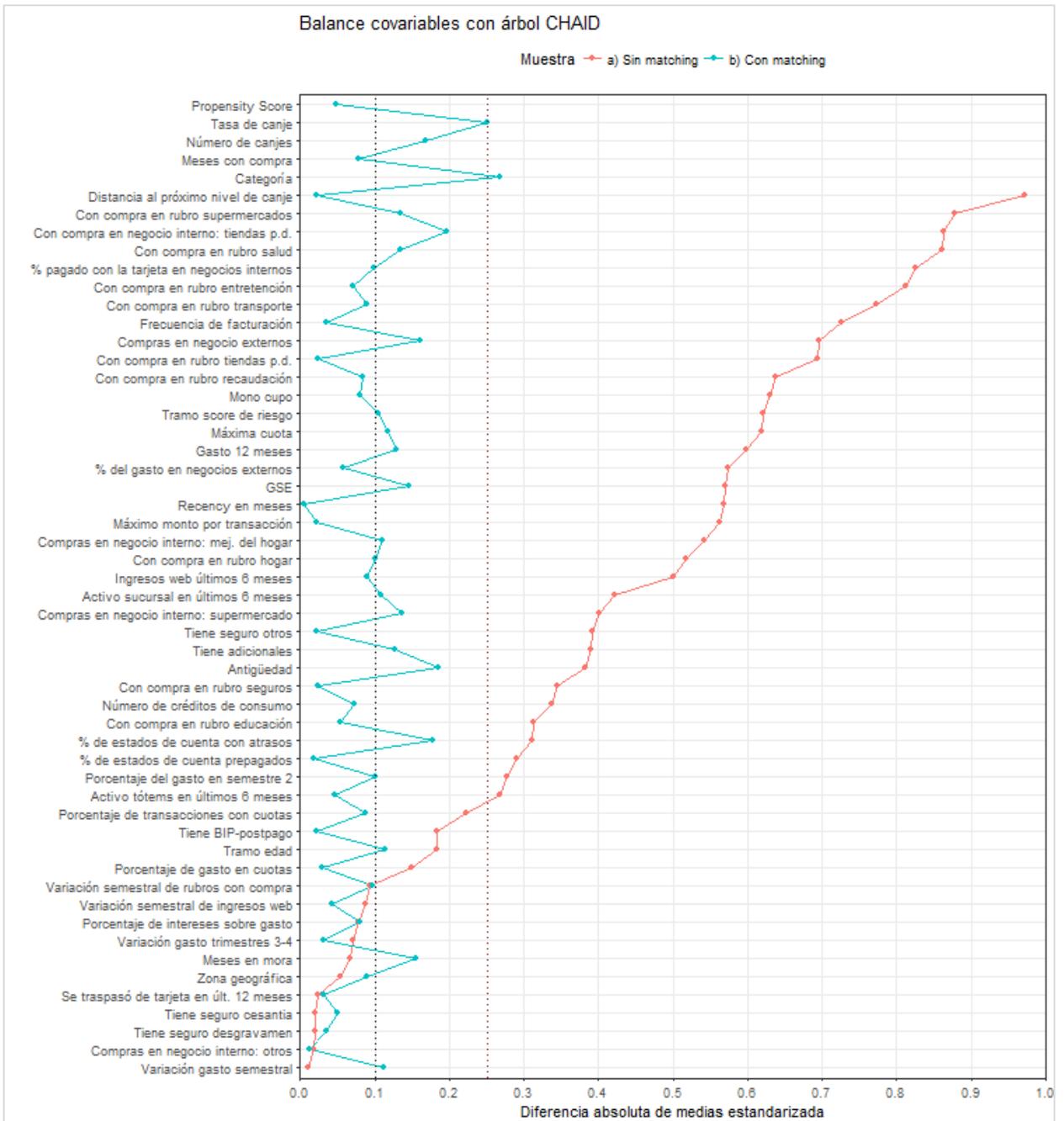
Variable	Importancia	B	Sig.	Exp(B)
Intercepto		-2,172	0	0,114
Número de canjes	17,2%	0,593	0	1,810
Tasa de canje	15,8%	1,228	0	3,415
Categoría_2 Premium	10,6%			
Categoría_2 Elite	10,6%	0,042	0,184	1,043
Categoría_2 Normal	10,6%	-1,429	0	0,240
Porcentaje pagado con la tarjeta en neg. internos	6,5%	1,064	0	2,898
Meses con compra	5,8%	0,054	0	1,055
Gasto 12 meses	5,5%	0,010	0	1,010
Distancia al próximo nivel de canje	4,3%	-0,588	0	0,556
Compras en negocio interno: tiendas por depto.	3,5%	0,021	0	1,022
Máxima cuota	3,3%	0,036	0	1,037
Con compra en rubro supermercados (=0)	2,1%	-0,200	0	0,818
Compras en negocio externos	1,8%	0,002	0	1,002
Porcentaje de transacciones con cuotas	1,6%	0,761	0	2,141
Compras en negocio interno: mej. del hogar	1,5%	0,014	0	1,014
Porcentaje del gasto en semestre 2	1,3%	0,712	0	2,037
Con compra en rubro tiendas p. d. (=0)	1,3%	-0,154	0	0,857

Máximo monto por transacción	1,2%	0,022	0	1,022
Con compra en rubro salud	1,2%	-0,122	0	0,886
Antigüedad	1,1%	0,026	0	1,026
Activo sucursal en últimos 6 meses (=0)	1,1%	-0,176	0	0,839
Con compra en rubro transporte	1,0%	-0,119	0	0,888
Compras en negocio interno: supermercado	1,0%	0,010	0	1,010
Con compra en rubro entretenimiento (=0)	0,9%	-0,080	0	0,923
Ingresos web últimos 6 meses	0,9%	0,004	0	1,004
Variación semestral de rubros con compra	0,6%	0,284	0	1,329
Con compra en rubro hogar (=0)	0,6%	-0,113	0	0,893
Con compra en rubro recaudación (=0)	0,6%	-0,060	0	0,942
Recency en meses	0,6%	-0,004	0,002	0,996
Tiene adicionales (=0)	0,6%	-0,098	0	0,907
Monto cupo	0,5%	0,003	0	1,003
Porcentaje de estados de cuenta prepagados	0,5%	-0,237	0	0,789
Porcentaje de estados de cuenta con atrasos	0,5%	-0,600	0	0,549
Activo tótems en últimos 6 meses (=0)	0,5%	-0,088	0	0,916
GSE Sin información	0,5%			
GSE ABC1	0,5%	-0,244	0	0,784
GSE C2	0,5%	-0,055	0	0,947
GSE C3	0,5%	0,016	0,218	1,016
GSE D	0,5%	-0,195	0	0,822
GSE E	0,5%	-0,412	0	0,663
Con compra en rubro seguros (=0)	0,4%	-0,055	0	0,946
Tramo edad 66+	0,3%			
Tramo edad 18-25	0,3%	0,060	0,002	1,061
Tramo edad 26-30	0,3%	0,191	0	1,210
Tramo edad 31-35	0,3%	0,160	0	1,173
Tramo edad 36-40	0,3%	0,156	0	1,169
Tramo edad 41-45	0,3%	0,162	0	1,176
Tramo edad 46-50	0,3%	0,158	0	1,171
Tramo edad 51-55	0,3%	0,146	0	1,157
Tramo edad 56-60	0,3%	0,142	0	1,153
Tramo edad 61-65	0,3%	0,172	0	1,188
Porcentaje de intereses sobre gasto	0,3%	-1,087	0	0,337
Variación gasto semestral	0,3%	-0,411	0	0,663
Variación semestral de ingresos web	0,3%	0,126	0	1,134
Con compra en rubro educación (=0)	0,3%	-0,073	0	0,930
Se traspasó de tarjeta en últimos 12 meses (=0)	0,3%	-0,192	0	0,825
Número de créditos de consumo	0,2%	0,010	0,007	1,010
Porcentaje de gasto en cuotas	0,2%	-0,327	0	0,721
Variación gasto trimestres 3-4	0,2%	-0,031	0	0,969
Compras en negocio interno: otros	0,2%	-0,019	0,009	0,981
Tiene seguro desgravamen	0,1%	-0,097	0	0,908
Tiene seguro cesantía (=0)	0,1%	0,033	0	1,033
Zona Sur	0,1%			
Zona Austral	0,1%	-0,011	0,515	0,989
Zona Centro	0,1%	-0,001	0,944	0,999
Zona Norte Chico	0,1%	-0,010	0,579	0,990
Zona Norte Grande	0,1%	-0,073	0	0,929
Zona RM	0,1%	-0,103	0	0,903
Tiene BIP-postpago (=0)	0,0%	0,116	0	1,123
Meses en mora	0,0%	-0,024	0	0,976
Tramo score de riesgo 601-1.000	0,0%			
Tramo score de riesgo 0	0,0%	0,106	0	1,112
Tramo score de riesgo 1-300	0,0%	0,089	0	1,093
Tramo score de riesgo 301-600	0,0%	-0,018	0,015	0,983
Tiene seguro otros	0,0%	0,098	0	1,103
Frecuencia de facturación	0,0%	-0,807	0	0,446
Porcentaje del gasto en negocios externos	0,0%	-0,492	0	0,611

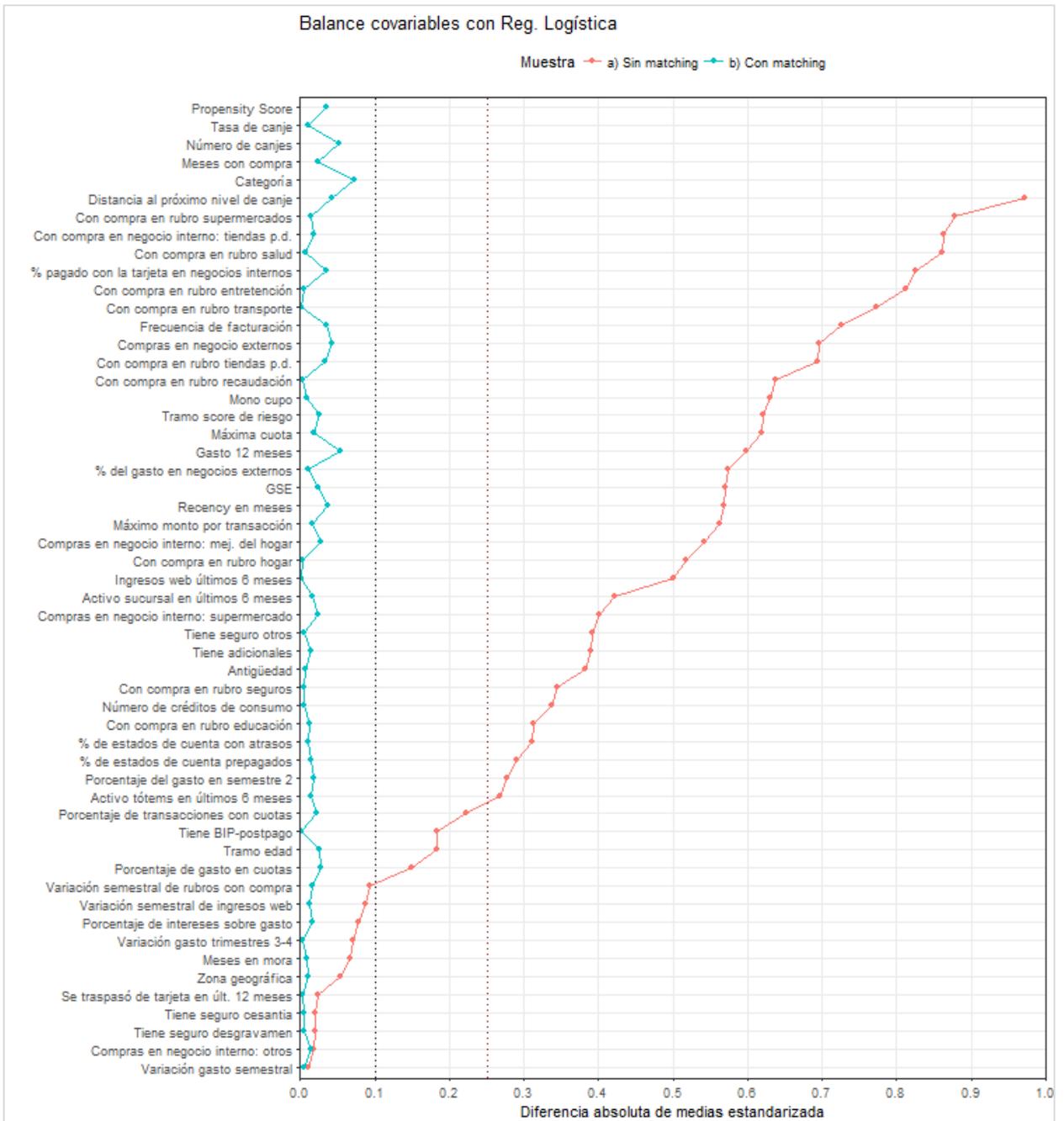
Anexo 6. Modelo de propensión a activarse en programa de puntos (Regresión logística)



Anexo 7. Diferencias de medias estandarizadas antes y después del matching con modelo de propensión árbol C5 (1 vecino más cercano).
Diferencias mayores a 1.0 no se muestran porque exceden el límite del gráfico.

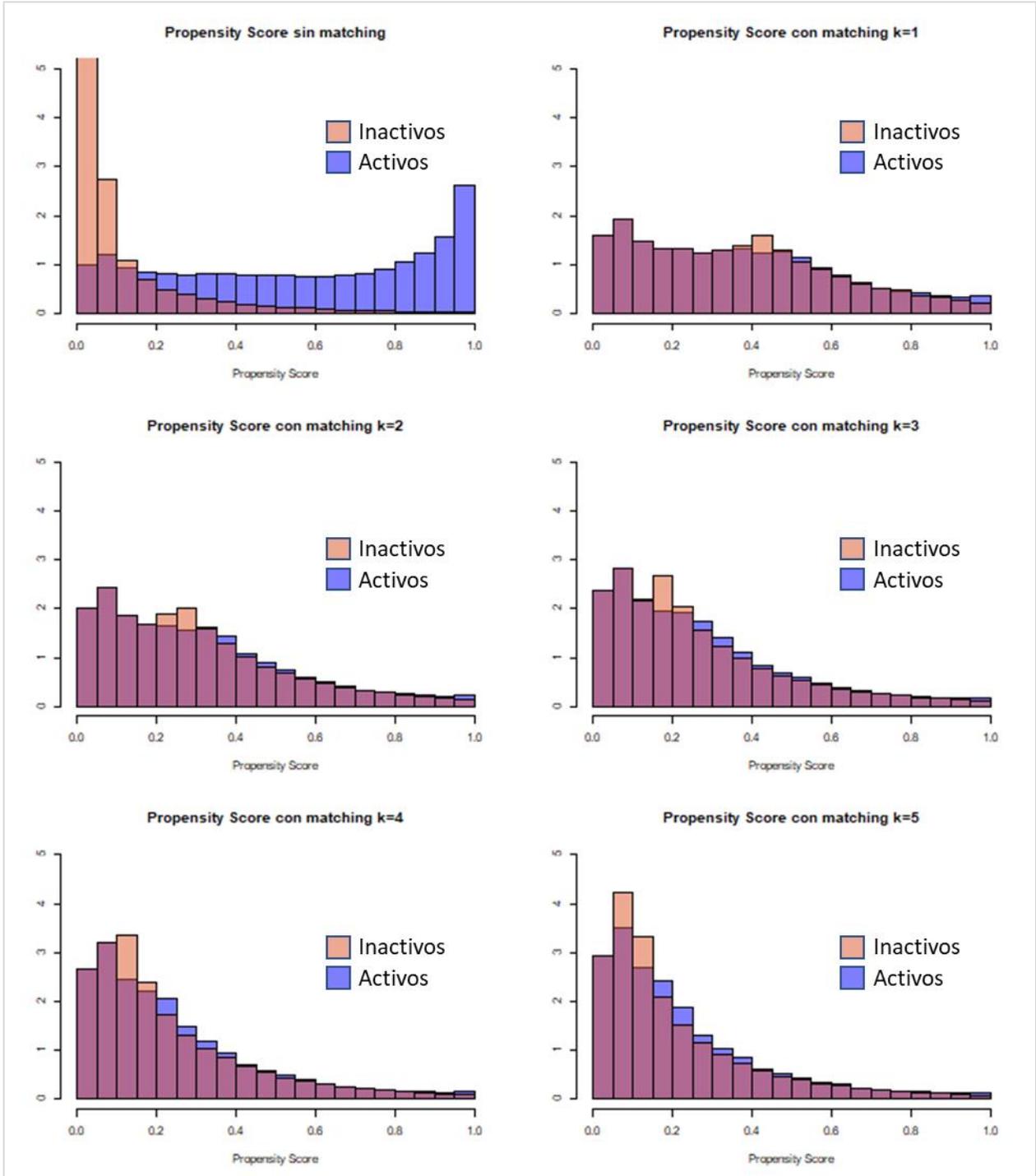


Anexo 8. Diferencias de medias estandarizadas antes y después del matching con modelo de propensión árbol CHAID (1 vecino más cercano).
 Diferencias mayores a 1.0 no se muestran porque exceden el límite del gráfico.



Anexo 9. Diferencias de medias estandarizadas antes y después del matching con modelo de propensión regresión logística (1 vecino más cercano).
 Diferencias mayores a 1.0 no se muestran porque exceden el límite del gráfico.

9.4 Anexos D: Matching



Anexo 10. Distribuciones de Propensity Score para grupo control y tratamiento según k-vecinos

Diferencias sin matching (base total)	Inactivos*	Activos*	DME
N clientes	1.779.552	342.505	
Propensity Score	0,09 (0,15)	0,56 (0,32)	1,887
Número de canjes	0,24 (0,59)	1,75 (1,63)	1,231
Tasa de canje	0,18 (0,34)	0,77 (0,36)	1,68
Categoría_2 Premium			1,115
Elite	3259 (0,2)	16114 (4,7)	
Normal	1723683 (96,9)	189549 (55,3)	
Premium	52610 (3,0)	136842 (40,0)	
% pagado con la tarjeta en neg. internos	0,67 (0,37)	0,91 (0,15)	0,821
Meses con compra	6,74 (4,56)	10,87 (2,44)	1,128
Gasto 12 meses	833,12 (1.693,53)	4.106,80 (7.786,66)	0,581
Distancia al próx. nivel de canje	0,57 (0,31)	0,31 (0,23)	0,967
Compras en negocio interno: tiendas p.d.	4,01 (5,54)	11,63 (11,25)	0,860
Máxima cuota	4,51 (4,52)	7,60 (5,44)	0,617
Con compra en rubro supermercados	587338 (33,0)	250147 (73,0)	0,876
Compras en negocios externos	10,03 (27,05)	58,24 (93,50)	0,700
Porcentaje de transacciones con cuotas	0,23 (0,25)	0,28 (0,23)	0,226
Compras en negocio interno: mej. hogar	2,35 (4,56)	7,45 (12,36)	0,547
Porcentaje del gasto en semestre 2	0,43 (0,31)	0,50 (0,17)	0,278
Con compra en rubro tiendas p.d.	1317484 (74,0)	332177 (97,0)	0,690
Máximo monto por transacción	172,31 (265,23)	453,40 (684,00)	0,542
Con compra en rubro salud	402936 (22,6)	211039 (61,6)	0,859
Antigüedad	11,03 (7,67)	13,96 (7,56)	0,386
Activo sucursal en últimos 6 meses	636569 (35,8)	193176 (56,4)	0,423
Con compra en rubro transporte	303539 (17,1)	175694 (51,3)	0,774
Compras en negocio interno: supermercado	1,06 (4,36)	5,41 (14,52)	0,405
Con compra en rubro entretenimiento	382908 (21,5)	199204 (58,2)	0,807
Ingresos web últimos 6 meses	2,57 (7,34)	8,75 (15,98)	0,497
Con compra en rubro hogar	90517 (5,1)	77464 (22,6)	0,525
Variación semestral de rubros con compra	0,07 (0,51)	0,11 (0,37)	0,091
Con compra en rubro recaudación	345533 (19,4)	165083 (48,2)	0,639
Tiene adicionales	346475 (19,5)	126352 (36,9)	0,395
Recency en meses	3,78 (8,45)	0,29 (2,22)	0,565
Monto cupo	875,17 (804,02)	1.466,11 (1.023,59)	0,642
Porcentaje de estados de cuenta prepagados	0,05 (0,16)	0,02 (0,08)	0,294
Activo tótems en últimos 6 meses	216027 (12,1)	76066 (22,2)	0,269
Porcentaje de EECC con atrasos	0,07 (0,14)	0,03 (0,09)	0,305
GSE			0,568
ABC1	165851 (9,3)	50997 (14,9)	
C2	407319 (22,9)	121591 (35,5)	
C3	441461 (24,8)	105214 (30,7)	
D	505152 (28,4)	49608 (14,5)	
E	109760 (6,2)	3676 (1,1)	
Sin información	150009 (8,4)	11419 (3,3)	
Con compra en rubro seguros	1193792 (67,1)	279395 (81,6)	0,336
Tramo edad			0,186
18-25	59853 (3,4)	7091 (2,1)	
26-30	147535 (8,3)	27683 (8,1)	
31-35	176652 (9,9)	35245 (10,3)	
36-40	183965 (10,3)	38899 (11,4)	
41-45	196036 (11,0)	42515 (12,4)	
46-50	206798 (11,6)	44749 (13,1)	
51-55	205692 (11,6)	43700 (12,8)	
56-60	182699 (10,3)	38227 (11,2)	
61-65	147503 (8,3)	29780 (8,7)	
66+	272819 (15,3)	34616 (10,1)	
Porcentaje de intereses sobre gasto	0,02 (0,07)	0,01 (0,04)	0,079
Variación gasto semestral	0,11 (0,64)	0,10 (0,54)	0,012
Variación semestral de ingresos web	0,04 (0,37)	0,07 (0,46)	0,091
Con compra en rubro educación	19707 (1,1)	25190 (7,4)	0,314
Se traspasó de tarjeta en últ. 12 meses	150389 (8,5)	26291 (7,7)	0,028

Número de créditos de consumo	0,13 (0,50)	0,38 (0,89)	0,340
Porcentaje de gasto en cuotas	0,48 (0,38)	0,54 (0,31)	0,155
Variación gasto últimos 3 meses	0,14 (0,65)	0,18 (0,60)	0,071
Compras en negocio interno: otros	0,03 (3,02)	0,11 (0,55)	0,033
Tiene seguro desgravamen	1101330 (61,9)	208090 (60,8)	0,023
Tiene seguro cesantía	339938 (19,1)	62596 (18,3)	0,021
Zona geográfica			0,053
Austral	63384 (3,6)	11862 (3,5)	
Centro	566773 (31,8)	102881 (30,0)	
Norte chico	55434 (3,1)	10526 (3,1)	
Norte grande	133331 (7,5)	26073 (7,6)	
RM	819827 (46,1)	166119 (48,5)	
Sur	140803 (7,9)	25044 (7,3)	
Tiene BIP-postpago	41287 (2,3)	20855 (6,1)	0,189
Meses en mora	2,88 (3,51)	2,66 (3,30)	0,066
Tramo score de riesgo			0,620
0	328958 (18,5)	7029 (2,1)	
1-300	27496 (1,5)	3415 (1,0)	
301-600	484507 (27,2)	76170 (22,2)	
601-1.000	938591 (52,7)	255891 (74,7)	
Tiene seguro otros	443058 (24,9)	147308 (43,0)	0,390
Porcentaje del gasto en neg. externos	0,19 (0,31)	0,37 (0,33)	0,575
Frecuencia de facturación	0,77 (0,36)	0,97 (0,14)	0,724

Anexo 11. Balance de covariables sin matching.

**Para variables numéricas: Promedio (Desviación estándar),*

Para variables categóricas: Cantidad de clientes (porcentaje).

***Variables ordenadas de mayor a menor relevancia para la propensión.*

Diferencias con matching	Inactivos*	Activos*	DME
N clientes	214.378	214.378	
Propensity Score	0,36 (0,24)	0,37 (0,25)	0,037
Número de canjes	1,07 (1,02)	1,11 (1,06)	0,036
Tasa de canje	0,66 (0,39)	0,66 (0,41)	0,009
Categoría_2 Premium			0,098
Elite	2584 (1,2)	4122 (1,9)	
Normal	172570 (80,5)	164592 (76,8)	
Premium	39224 (18,3)	45664 (21,3)	
% pagado con la tarjeta en neg. internos	0,89 (0,16)	0,88 (0,17)	0,047
Meses con compra	10,40 (2,74)	10,33 (2,88)	0,026
Gasto 12 meses	2.352,43 (3.849,82)	2.605,34 (3.718,61)	0,067
Distancia al próx. nivel de canje	0,33 (0,24)	0,34 (0,25)	0,035
Compras en negocio interno: tiendas p.d.	8,95 (8,80)	9,11 (8,63)	0,019
Máxima cuota	7,25 (5,55)	7,08 (4,97)	0,033
Con compra en rubro supermercados	139507 (65,1)	138443 (64,6)	0,010
Compras en negocios externos	31,74 (53,15)	34,85 (60,86)	0,054
Porcentaje de transacciones con cuotas	0,30 (0,24)	0,29 (0,23)	0,031
Compras en negocio interno: mej. hogar	5,28 (8,58)	5,86 (8,76)	0,067
Porcentaje del gasto en semestre 2	0,50 (0,20)	0,50 (0,18)	0,006
Con compra en rubro tiendas p.d.	206000 (96,1)	204586 (95,4)	0,033
Máximo monto por transacción	346,20 (495,80)	356,90 (430,57)	0,023
Con compra en rubro salud	110188 (51,4)	109758 (51,2)	0,004
Antigüedad	13,13 (7,97)	13,19 (7,28)	0,008
Activo sucursal en últimos 6 meses	113712 (53,0)	111794 (52,1)	0,018
Con compra en rubro transporte	86921 (40,5)	88676 (41,4)	0,017
Compras en negocio interno: supermercado	3,08 (9,16)	3,26 (8,94)	0,020
Con compra en rubro entretenimiento	103862 (48,4)	103552 (48,3)	0,003
Ingresos web últimos 6 meses	6,52 (12,51)	6,46 (12,62)	0,005
Con compra en rubro hogar	32199 (15,0)	33310 (15,5)	0,014
Variación semestral de rubros con compra	0,13 (0,42)	0,12 (0,40)	0,018

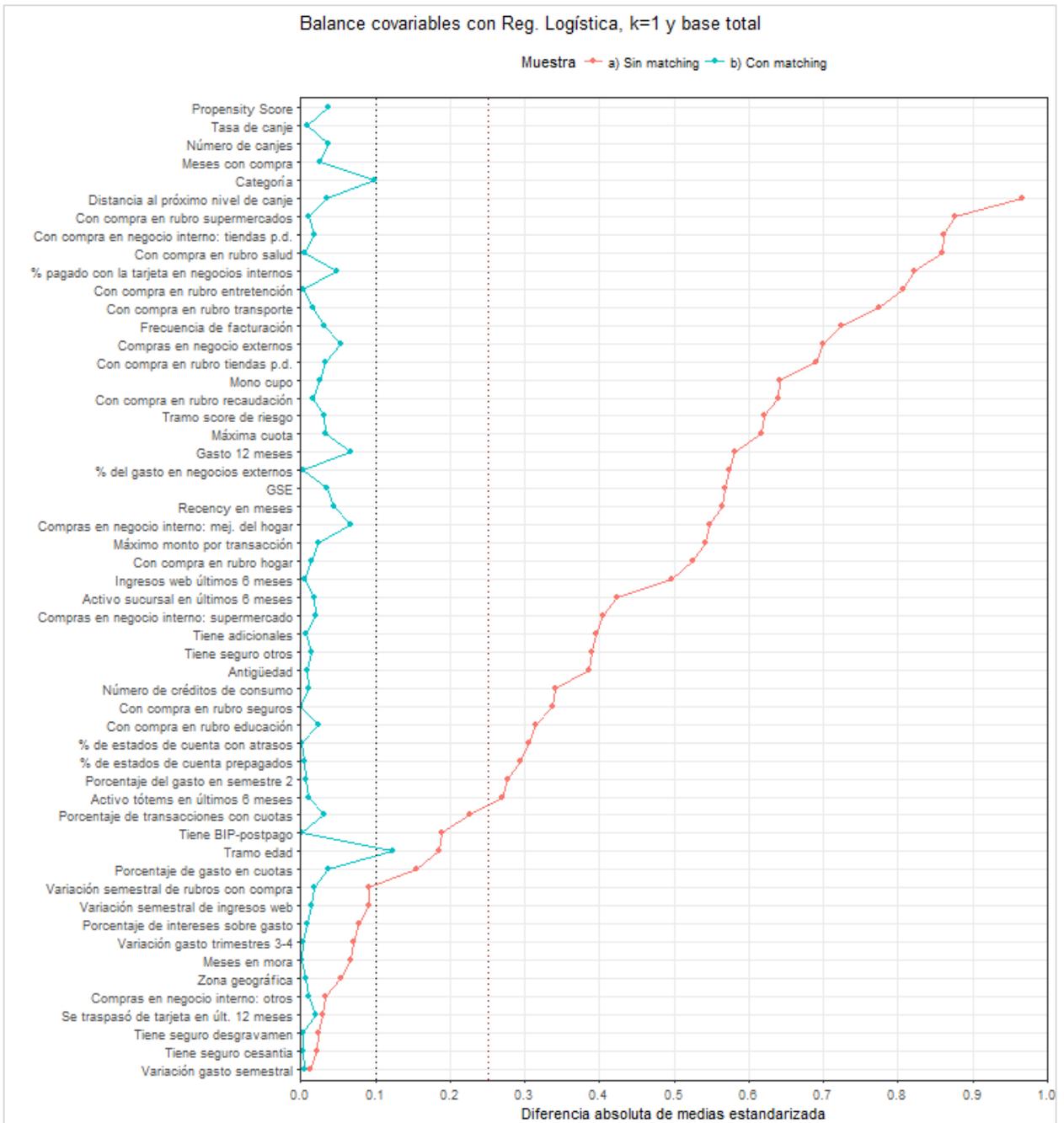
Con compra en rubro recaudación	84126 (39,2)	85842 (40,0)	0,016
Tiene adicionales	64738 (30,2)	65430 (30,5)	0,007
Recency en meses	0,34 (2,14)	0,45 (2,77)	0,043
Monto cupo	1.249,97 (883,17)	1.272,56 (893,91)	0,025
Porcentaje de estados de cuenta prepagados	0,02 (0,08)	0,02 (0,09)	0,006
Activo tótems en últimos 6 meses	45004 (21,0)	44121 (20,6)	0,010
Porcentaje de EECC con atrasos	0,04 (0,10)	0,04 (0,10)	<0,001
GSE			0,035
ABC1	24948 (11,6)	25806 (12,0)	
C2	65828 (30,7)	68006 (31,7)	
C3	68070 (31,8)	67096 (31,3)	
D	41768 (19,5)	40057 (18,7)	
E	3054 (1,4)	3369 (1,6)	
Sin información	10710 (5,0)	10044 (4,7)	
Con compra en rubro seguros	170830 (79,7)	170844 (79,7)	<0,001
Tramo edad			0,122
18-25	5616 (2,6)	4976 (2,3)	
26-30	19807 (9,2)	17221 (8,0)	
31-35	23040 (10,7)	22232 (10,4)	
36-40	24306 (11,3)	25457 (11,9)	
41-45	25878 (12,1)	30828 (14,4)	
46-50	26900 (12,5)	31933 (14,9)	
51-55	26025 (12,1)	22446 (10,5)	
56-60	22925 (10,7)	20325 (9,5)	
61-65	18052 (8,4)	16732 (7,8)	
66+	21829 (10,2)	22228 (10,4)	
Porcentaje de intereses sobre gasto	0,01 (0,04)	0,01 (0,05)	0,010
Variación gasto semestral	0,10 (0,60)	0,10 (0,57)	0,006
Variación semestral de ingresos web	0,08 (0,45)	0,08 (0,46)	0,014
Con compra en rubro educación	8497 (4,0)	9549 (4,5)	0,024
Se traspasó de tarjeta en últ. 12 meses	20043 (9,3)	18773 (8,8)	0,021
Número de créditos de consumo	0,32 (0,81)	0,31 (0,77)	0,010
Porcentaje de gasto en cuotas	0,58 (0,32)	0,56 (0,31)	0,037
Variación gasto últimos 3 meses	0,19 (0,65)	0,19 (0,63)	0,002
Compras en negocio interno: otros	0,07 (0,39)	0,07 (0,46)	0,010
Tiene seguro desgravamen	131425 (61,3)	131739 (61,5)	0,003
Tiene seguro cesantía	40708 (19,0)	40894 (19,1)	0,002
Zona geográfica			0,007
Austral	7780 (3,6)	7779 (3,6)	
Centro	66109 (30,8)	65894 (30,7)	
Norte chico	6509 (3,0)	6556 (3,1)	
Norte grande	15748 (7,3)	16103 (7,5)	
RM	101932 (47,5)	101591 (47,4)	
Sur	16300 (7,6)	16455 (7,7)	
Tiene BIP-postpago	10391 (4,8)	10316 (4,8)	0,002
Meses en mora	2,90 (3,35)	2,91 (3,41)	0,001
Tramo score de riesgo			0,031
0	5888 (2,7)	6846 (3,2)	
1-300	2700 (1,3)	2494 (1,2)	
301-600	55340 (25,8)	53866 (25,1)	
601-1.000	150450 (70,2)	151172 (70,5)	
Tiene seguro otros	81219 (37,9)	82631 (38,5)	0,014
Porcentaje del gasto en neg. externos	0,33 (0,33)	0,33 (0,33)	0,004
Frecuencia de facturación	0,96 (0,16)	0,95 (0,17)	0,031

Anexo 12. Balance de covariables con matching reg. log y k=1.

*Para variables numéricas: Promedio (Desviación estándar),

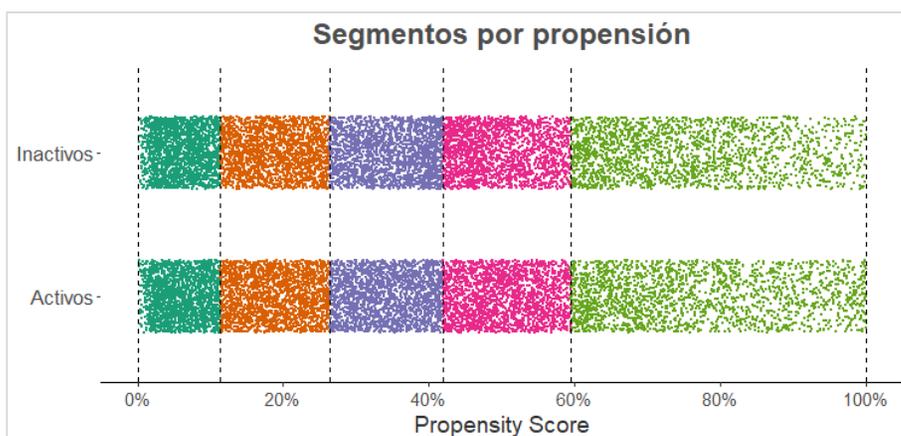
Para variables categóricas: Cantidad de clientes (porcentaje).

**Variables ordenadas de mayor a menor relevancia para la propensión.



Anexo 13. Diferencias de medias estandarizadas antes y después del matching final. Modelo de propensión regresión logística, 1 vecino más cercano, radio de caliper y base total. Diferencias mayores a 1.0 no se muestran porque exceden el límite del gráfico.

9.5 Anexos E: Segmentos



Anexo 14. Visualización de quintiles en base a propensión

Promedios (o proporciones %)	Segmentos				
	1	2	3	4	5
Propensity Score (%)	0.06	0.19	0.34	0.50	0.76
Número de canjes últ. año	0.18	0.75	1.17	1.45	2.01
Tasa de canje (%)	0.12	0.54	0.80	0.88	0.94
Categoría (%)					
Elite	0.30	0.60	0.90	1.60	6.30
Normal	98.60	94.00	87.60	74.00	29.70
Premium	1.10	5.40	11.50	24.40	64.00
Porcentaje pagado con la tarj. en neg. internos (%)	0.82	0.87	0.89	0.91	0.92
Meses con compra	7.58	10.08	10.98	11.35	11.68
Gasto (M\$)	1466.55	1927.91	2166.85	2649.84	4815.53
Distancia al próximo nivel de canje	0.46	0.35	0.32	0.30	0.26
Compras en negocio interno: tiendas por depto.	4.58	7.53	9.25	10.62	13.58
Máxima cuota	5.89	6.92	7.18	7.61	7.80
Con compra en rubro supermercados	39.20	59.20	67.70	74.40	82.50
Compras en negocio externos	10.48	21.21	30.08	40.29	72.21
Porcentaje de transacciones con cuotas (%)	0.27	0.30	0.31	0.30	0.27
Compras en negocio interno: mej. del hogar	2.81	4.43	5.24	6.35	10.46
Porcentaje del gasto en semestre 2 (%)	0.48	0.50	0.50	0.50	0.50
Con compra en rubro tiendas por departamento (%)	83.40	96.90	98.60	98.90	99.40
Máximo monto por transacción (M\$)	290.20	310.41	322.91	358.61	502.34
Con compra en rubro salud (%)	25.10	44.20	54.00	61.60	71.00
Antigüedad	11.89	12.40	12.94	13.77	14.94
Activo sucursal en últimos 6 meses (%)	42.10	51.50	53.50	55.90	57.70
Con compra en rubro transporte (%)	19.40	33.80	41.40	49.00	63.20
Compras en negocio interno: supermercado	1.24	2.18	2.93	3.73	6.21
Con compra en rubro entretenimiento (%)	25.90	42.00	49.50	56.30	67.90
Ingresos web últimos 6 meses	3.42	5.41	6.34	7.24	9.90
Con compra en rubro hogar (%)	5.30	10.70	15.30	19.00	27.30
Variación semestral de rubros con compra	0.16	0.16	0.11	0.10	0.08
Con compra en rubro recaudación (%)	21.70	33.80	40.40	46.10	58.30
Tiene adicionales	9134.00	11181.00	12825.00	14593.00	17697.00
Recency en meses	1.88	0.20	0.09	0.06	0.03
Monto de cupo (M\$)	991.17	1110.59	1213.06	1359.21	1688.80
Porcentaje de estados de cuenta prepagados	0.05	0.02	0.02	0.01	0.01
Activo tótems en últimos 6 meses (%)	15.50	20.00	21.00	22.10	24.20
Porcentaje de estados de cuenta con atrasos	0.07	0.05	0.04	0.03	0.03
GSE (%)					
ABC1	9.90	10.20	10.80	12.60	16.70

C2	24.90	26.70	29.70	33.50	43.80
C3	26.30	29.40	33.40	35.10	32.40
D	25.00	24.00	21.60	16.50	6.40
E	3.90	2.20	1.10	0.60	0.10
Sin información	10.10	7.50	3.30	1.80	0.70
Con compra en rubro seguros (%)	71.00	79.20	81.40	82.30	84.60
Tramo edad (%)					
18-25	3.40	3.10	2.50	2.00	0.50
26-30	10.30	9.40	8.60	7.90	4.00
31-35	11.50	10.90	10.20	10.10	9.20
36-40	11.50	11.00	10.80	11.00	15.00
41-45	11.90	11.70	11.70	12.20	24.30
46-50	11.80	12.30	12.70	12.80	24.80
51-55	11.40	11.80	12.30	12.60	4.30
56-60	10.00	10.60	10.80	11.20	4.90
61-65	7.80	8.20	8.80	9.00	5.10
66+	10.20	11.00	11.50	11.20	7.90
Porcentaje de intereses sobre gasto	0.02	0.01	0.01	0.01	0.01
Variación gasto semestral	0.10	0.11	0.11	0.10	0.10
Variación semestral de ingresos web	0.07	0.10	0.08	0.07	0.06
Con compra en rubro educación (%)	1.40	2.60	3.70	4.90	9.60
Se traspasó de tarjeta en últimos 12 meses (%)	11.80	12.00	8.30	6.80	5.00
Número de créditos de consumo últ. año	0.14	0.25	0.33	0.39	0.43
Porcentaje de gasto en cuotas	0.56	0.59	0.59	0.57	0.51
Variación gasto trimestres 3-4	0.16	0.19	0.19	0.19	0.20
Compras en negocio interno: otros (M\$)	0.04	0.05	0.06	0.07	0.13
Tiene seguro desgravamen (%)	59.90	61.30	62.80	62.50	60.80
Tiene seguro cesantía (%)	19.60	19.10	18.80	18.50	19.50
Zona geográfica (%)					
Sur austral	3.60	3.80	3.70	3.60	3.40
Centro	31.10	30.80	31.60	31.00	29.10
Norte chico	3.30	3.20	2.80	2.80	3.20
Norte grande	7.60	7.30	6.90	7.10	8.70
Región Metropolitana	46.80	47.20	47.00	47.80	48.10
Sur	7.60	7.80	7.90	7.70	7.40
Tiene BIP post-pago (%)	2.20	3.90	5.20	6.00	6.70
Meses en mora últ. año	3.03	3.06	2.97	2.79	2.68
Tramo score de riesgo (%)					
0	12.20	2.90	0.50	0.20	0.10
1-300	1.60	1.40	1.00	1.00	0.70
301-600	28.70	28.60	26.20	23.50	18.60
601-1.000	57.40	67.10	72.20	75.30	80.60
Tiene seguro otros (%)	28.00	35.70	38.70	41.10	49.20
Porcentaje del gasto en negocios externos (%)	0.20	0.29	0.35	0.38	0.42
Frecuencia de facturación (% meses con EECC)	0.83	0.95	0.99	0.99	1.00

*Anexo 15. Caracterización de segmentos (promedios y proporciones)
(Todas las variables del modelo de propensión ordenadas según relevancia para el modelo)*