



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

**IDENTIFICACIÓN DE CLIENTES PROPENSOS AL CANJE NO AÉREO EN EL PROGRAMA DE
FIDELIZACIÓN DE UNA LÍNEA AÉREA**

MEMORIA PARA OPTAR AL TITULO DE INGENIERO CIVIL INDUSTRIAL

LEONARDO CRISTÓBAL FERRARO LÓPEZ

PROFESOR GUÍA:

PABLO MARÍN VICUÑA

MIEMBROS DE LA COMISIÓN:

ALEJANDRA PUENTE CHANDÍA

LUIS ABURTO LAFOURCADE

SANTIAGO DE CHILE

2015

IDENTIFICACIÓN DE CLIENTES PROPENSOS AL CANJE NO AÉREO EN EL PROGRAMA DE FIDELIZACIÓN DE UNA LÍNEA AÉREA

Esta investigación tiene como objetivo identificar a los clientes más propensos al Canje No Aéreo (CNA o canje de catálogo) de LANPASS, uno de los programas de fidelización de LATAM Airline Group. Esto es relevante porque el 80% de los clientes no acredita los suficientes kilómetros (KMS) para canjear pasajes y porque las decisiones comerciales de LANPASS no son complementadas con la información de su base de clientes sino que se enfocan en decisiones de producto.

Para cumplir el objetivo se estudió el comportamiento transaccional de los clientes entre Julio de 2013 y Junio de 2014 en conjunto con sus variables demográficas. Mediante un análisis descriptivo se descubrió que las variables más importantes son el saldo, la acreditación anual de KMS y el N° de CNA realizados durante el último año, en tal orden. Se pudo identificar a los clientes más propensos también a través de la edad, el género, la categoría de socio y la mejor tarjeta de crédito. Se observó que los clientes tienden a canjear en el catálogo sus KMS cercanos a la expiración, llegando a ser hasta 10 veces más propensos al CNA en los últimos 2 meses antes de expirar. Finalmente se vio que los clientes que se frustran más rápido en el proceso de canje aéreo son 5 veces más propensos al CNA que un cliente promedio.

Con estas variables se calibraron tres modelos de propensión; Logit Multinomial con variables transaccionales, Logit Multinomial con variables demográficas y Árbol de Clasificación con método de crecimiento CHAID-Exhaustivo con ambos tipos de variables. El mejor modelo fue el árbol CHAID-E, el cual identificó al 30% de los canjeadores en el primer 1% de clientes. Esta técnica se comparó con criterios sencillos de ordenamiento, el más robusto fue el de los clientes que más han realizado CNA en el último año. Cuantitativamente CHAID-E mostró que genera un incremental de 50% de KMS canjeados por sobre dicho criterio en el primer centil.

Se utilizó un modelo Logit para estudiar la elección entre 3 categorías de canje; Tecnología, GiftCard de Multitienda y "Otras Categorías". Las dos primeras son las categorías más importantes, representando un 40% del CNA histórico. Se encontró que los clientes que canjean en Tecnología tienden a repetir esta preferencia, mientras que los que prefieren Giftcard son los que presentan la menor propensión a canjear en Tecnología

Se recomienda a LANPASS enfocar sus esfuerzos en los clientes que acreditan más de 21.000 KMS al año, tienen un saldo superior a 7.500 KMS y han realizado previamente un canje no aéreo. Este segmento fue el más grande encontrado y con mejor precisión en su clasificación. Se sugiere también atraer a los clientes que cumplan dichas reglas pero que aún no realizan CNA, pues una vez que canjeen, es probable que lo vuelvan a hacer. Estas medidas tienen un incremental esperado de un 50% de KMS en las ventas.

Agradecimientos

A mis padres Juan Cristóbal y María Soledad por su amor infinito y el apoyo que me dieron en cada etapa de este largo camino.

A mis hermanas Carla y Lorena por su cariño y su apoyo.

A Isabel y Marlene, por ayudarme a renacer en la hora más negra.

A Alejandra, Luis y Marcel por el compromiso que pusieron semana a semana en el curso de titulación y por ser fieles a su estilo de enseñanza.

A Pablo por confiar en mí, por su paciencia y por todo lo que aprendí con él.

A la VP de Clientes de LANPASS por la tremenda oportunidad que me ofrecieron.

A Andrea, Bernardita, Camila, Daniel, Felipe, Martin, Matías, Mauricio, Mayara, Natalia, Javier, José Pedro, Tomás y Sergio por su amistad incondicional.

A Eduardo, Franco y Pascual por llenar estos 6 años y medio de alegría y buenos momentos.

A Alejandro, Ariel, Carlos, Gonzalo, Javier, Rafael, Rodrigo-Arturo, Sergio y Patricio por hacer de incluso el día más triste un día mejor.

Al Centro de Estudiantes de Ingeniería Industrial 2014 por llenar mi último año de estudiante.

A Fernanda, por acompañarme en las buenas y en las malas, en especial en los momentos más difíciles de este proceso.

A Ornella y su familia por acompañarme en el último tramo de este camino, y compartir conmigo todas las dificultades que se nos presentaron.

A Nicole y en especial a Daniela por ayudarme en el tramo final.

Un agradecimiento especial a todos lo que conocí durante el último año, que fue muy importante para mí.

A todos los que mencioné y a los que no, gracias por aportar su granito de arena en hacerme quien soy hoy.

Gracias, no podría estar más feliz ni orgulloso.

Tabla de Contenido

1.	Antecedentes Generales e Introducción	1
2.	Descripción y Justificación del Proyecto	3
3.	Objetivos	4
3.1.	Objetivo General	4
3.2.	Objetivos Específicos	4
4.	Metodología	5
5.	Marco Conceptual.....	7
5.1.	Programa de Viajero Frecuente	7
5.2.	Modelos de Clasificación	7
5.2.1.	Regresión Logística.....	7
5.2.2.	Validación de modelos	9
5.2.3.	Arboles de Decisión.....	10
6.	Alcances	12
7.	Análisis Descriptivos	13
7.1.	Análisis Descriptivo de Tipo de Canje	14
7.1.1.	Estudio del canje según variables demográficas	14
7.1.1.1.	Efecto de variables transaccionales sobre variables demográficas	18
7.1.2.	Estudio del canje según Variables transaccionales	21
7.1.3.	Aprendizajes e Hipótesis del Análisis Descriptivo de Tipo de de Canje	30
7.2.	Análisis Descriptivo de Categoría	30
7.2.1.	Estudio de categoría según variables demográficas	30
7.3.	Estudio de categoría según variables transaccionales	34
7.3.1.	Aprendizajes e Hipótesis del Análisis Descriptivo de Categoría.....	38
7.4.	Poder Predictivo de Variables	39
8.	Modelos de Clasificación	41
8.1.	Primera Decisión: Canje Aéreo, Canje No Aéreo o No Canje	41
8.1.1.	Modelos de Regresión Logística	41
8.1.1.1.	Modelo de corto plazo	42
8.1.1.2.	Modelo de Largo Plazo	46
8.1.2.	Arboles de clasificación	50
8.1.2.1.	Curva de Ganancia de información	52
8.1.2.2.	Validación del modelo	53
8.1.3.	Comparación de Modelos.....	54
8.1.3.1.	Criterios Duros	54
8.1.3.2.	Criterios Blandos.....	56
8.1.3.3.	Veredicto	56
8.2.	Segunda Decisión: Elección de categoría	56
8.2.1.	Regresión Logística.....	57

8.2.2.	Valoración de los modelos	62
8.2.2.1.	Primera decisión: Canje Aéreo, Canje No Aéreo o No Canje	63
8.2.2.2.	Segunda Decisión: Elección de categoría	64
9.	Conclusiones	66
9.1.	Conclusiones Generales.....	66
9.2.	Recomendaciones de investigaciones futuras.....	67
9.3.	Recomendaciones Comerciales.....	68
10.	Bibliografía	69
11.	Anexos.....	70

Tabla de Ilustraciones

Ilustración 1. Estrategia de Negocios de LANPASS. Fuente: Elaboración Propia.	1
Ilustración 2. Distribuciones de Acreditación y Canje.....	2
Ilustración 3. Variables seleccionadas.	5
Ilustración 4. Propensión de canje según edad.....	14
Ilustración 5. Propensión de Canje vs Género (% de Clientes.....	15
Ilustración 6. Propensión de Canje vs Mejor Tarjeta de Crédito (% de Clientes).	16
Ilustración 7. Propensión de Canje vs Categoría de Socio (% de Clientes).	17
Ilustración 8. Propensión de Canje vs Región	18
Ilustración 9. Efecto Saldo, Caso Tarjeta de Crédito	19
Ilustración 10. Efecto Saldo, Caso Edad.....	19
Ilustración 11. Propensión Según Mejor TC	20
Ilustración 12. Propensión de Canje Según Edad	20
Ilustración 13. Propensión de Canje vs Acreditación Total (% de Clientes).	21
Ilustración 14. Propensión de Canje según KMS Canjeados en CA.	22
Ilustración 15. Propensión de Canje según KMS Canjeados en CNA.	22
Ilustración 16. Propensión de Canje vs N° de CNA (% de Clientes).	23
Ilustración 17. Propensión de Canje vs N° de CNA (% de Clientes).	23
Ilustración 18. Propensión de Canje vs Existencia de Canjes Recientes (% de Clientes).	24
Ilustración 19. Propensión de Canje vs Tipo de canjeador (% de Clientes).	25
Ilustración 20. Propensión de Canje vs Tipo de Acreditador (Acreditación Promedio).....	25
Ilustración 21. Propensión de Canje según Saldo.	27
Ilustración 22. Propensión de Canje según Saldos mayores a 60k KMS.	27
Ilustración 23. Propensión de Canje vs Caducidad Junio (% de Clientes).	28
Ilustración 24. Propensión De Canje vs Caducidad Diciembre (% de Clientes).....	29
Ilustración 25. Distribución del CNA vs Edad.	31
Ilustración 26. Distribución del CNA Según Género.....	32
Ilustración 27. Distribución del CNA Según Mejor Tarjeta de Crédito.....	33
Ilustración 28. Distribución del CNA Según Categoría de Socio.....	33
Ilustración 29. Distribución del CNA Según Acreditación Promedio Anual.	34
Ilustración 30. Distribución del CNA Según Saldo.	34
Ilustración 31. Distribución del CNA Según Tipo de Canjeador.	35
Ilustración 32. Distribución del CNA Según N° de Canjes Anteriores de Giftcards.....	36
Ilustración 33. Distribución del CNA Según N° de Canjes Anteriores de Tecnología.....	36
Ilustración 34. Distribución del CNA Según N° de Canjes Anteriores de Otras Categorías.	37
Ilustración 35. Poder Predictivo de las Variables para los casos de CNA y CA.....	39
Ilustración 36. Pseudo R-Cuadrado Modelos Corto Plazo.....	43
Ilustración 37. Resultado Test de Ratios de Verosimilitud del Modelo a Corto Plazo.....	43
Ilustración 38. Resultados Regresión Logística del Modelo Corto Plazo.....	44

Ilustración 39. Curva de Ganancia de Información Logit Corto Plazo.....	45
Ilustración 40 Curva de Ganancia de Información Logit Corto Plazo (1º decil).	46
Ilustración 41. Pseudo R-Cuadrado Modelos Largo Plazo.....	47
Ilustración 42. Resultados Test de Ratios de Verosimilitud del Modelo a Largo Plazo.	47
Ilustración 43. Resultados Regresión Logística del Modelo Largo Plazo.....	48
Ilustración 44. Curva de Ganancia de Información Logit Largo y Corto Plazo.	49
Ilustración 45. Curva de Ganancia de Información Logit Largo y Corto Plazo (1º decil).....	49
Ilustración 46. Métodos de Crecimiento: Curva de Ganancia de Información (1º Decil).	51
Ilustración 47. Segmentos según reglas de decisión Arbol CHAID-E.....	51
Ilustración 48. Curva de Ganancia de Información: CHAID-E.....	52
Ilustración 49. Curva de Ganancia de Información: CHAID-E (1º Decil).....	53
Ilustración 50. Validación del Modelo: Curvas de Ganancia de Información	54
Ilustración 51. Resultados Matriz de Confusión.....	54
Ilustración 52. Comparación de Modelos: Curva de Ganancia de Información.....	55
Ilustración 53. Curva de Ganancia de Información: Árboles de Clasificación (1º Decil).	55
Ilustración 54. Tablas de Comparación de Modelos.	57
Ilustración 55. Resultados Árbol: Parámetros Individuales.	58
Ilustración 56. Curva de Ganancia de Información: Regresión Logística - Elección de Categoría.	59
Ilustración 57. Curva de Ganancia de Información: Elección de Categoría.	60
Ilustración 58. Curva de Ganancia de Información: Elección de Categoría (Junio).....	61
Ilustración 59. Curva de Ganancia de Información: Elección de Categoría (Mayo).....	61
Ilustración 60. Valoración de Modelos Según Criterios de Ordenamiento.....	63
Ilustración 61. Valoración de Modelos Según Criterio de Ordenamiento.	64
Ilustración 62. Valoración Modelo de Elección de Categoría.....	64
Ilustración 63. Propensión de Canje según Pertenencia Santander.	70
Ilustración 64 Propensión de Canje vs Categoría de Socio (Saldo Promedio).	71
Ilustración 65. Propensión de Canje vs Categoría de Socio (Acreditación Promedio).	71
Ilustración 66. Propensión de Canje vs Género (Acreditación Promedio).	72
Ilustración 67. Propensión de Canje según Máximo de KMS canjeados en CA.	73
Ilustración 68. Propensión de Canje según Máximo de KMS canjeados en CNA.....	73
Ilustración 69. Propensión de Canje según Tipo de Acreditador.	74
Ilustración 71. Propensión de Canje según Canjes Anteriores.....	75

1. Antecedentes Generales e Introducción

LANPASS es uno de los dos programas de fidelización de LATAM Airline Group. La mayoría de las aerolíneas del mundo tienen programas de fidelización, en particular, FIDELIDADE es el programa de fidelización de TAM, que tras la fusión, también forma parte de LATAM. Los programas de fidelización tienen por objetivo generar una relación estable y duradera con los clientes. Un plan de fidelización busca, a grandes rasgos, captar, convencer y conservar a los clientes¹, bajo el concepto de que es más efectivo económicamente retener a un cliente actual que captar a uno nuevo.

LANPASS funciona como unidad de negocio separada de LAN. El modelo de negocios de LANPASS funciona en base a la venta de Kilómetros (KMS). Los KMS son la moneda del programa de fidelización. Estos se le venden tanto a LAN como a los partners comerciales de la empresa, de esta forma los aliados comerciales pueden otorgarles kilómetros a sus clientes para hacer más atractiva su oferta. Los KMS se valorizan a medida que su acumulación estimule el comportamiento de compra de los clientes de los partners comerciales de LAN.

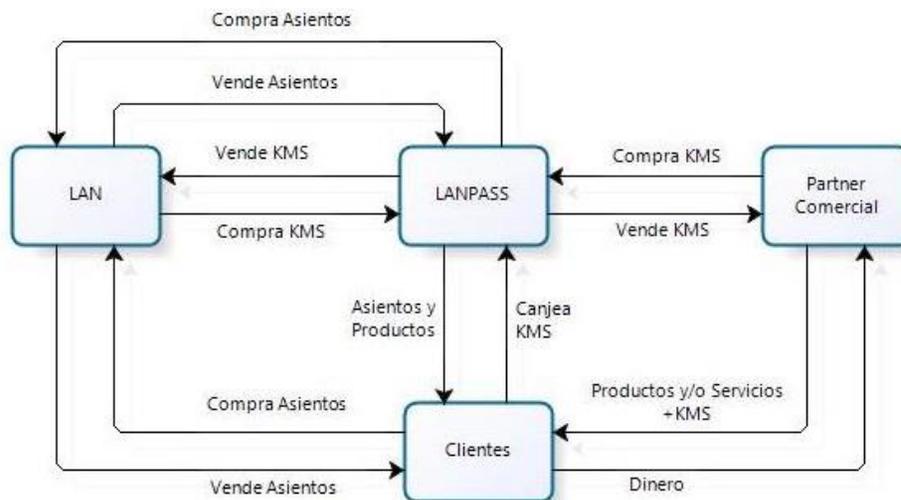


Ilustración 1. Estrategia de Negocios de LANPASS. Fuente: Elaboración Propia.

Los KMS acumulados le permiten al cliente canjear asientos en vuelos a diversos destinos o productos del Catálogo LANPASS, este catálogo es una plataforma virtual que posee un mix de productos similar al de una tienda por departamentos. El catálogo se divide en 9 categorías; Tecnología, Giftcard, Hogar, Belleza, Tiempo Libre, Infantil,

¹ Rodríguez, Santiago (2007). «Capítulo 16». Creatividad en Marketing Directo. Barcelona (España): Ediciones Deusto. p. 272. ISBN 84-234-2498-7. «Creo que un proceso de fidelización es como una navaja suiza que al abrirla muestra tres "C": captar, convencer y conservar»

Donaciones y Promociones. El catálogo funciona exclusivamente con despacho a domicilio.

A partir de ahora nos referiremos al canje de asientos como **Canje Aéreo** o “CA” y al canje en el catálogo LANPASS como **Canje No-Aéreo** o “CNA”.

LAN fideliza al cliente ofreciéndole 4 formas de acreditación (o acumulación) de Kilómetros:

1. Acreditación por vuelo: Los clientes acreditan KMS cada vez que vuelen con LAN, TAM, aerolíneas de la alianza OneWorld o aerolíneas asociadas. El monto de acreditación depende de 3 factores: Distancia entre origen y destino, familia tarifaria del pasaje y categoría de socio.
2. Acreditación por alianza financiera: También se pueden acumular KMS pagando con tarjetas Santander, la acreditación depende del medio de pago, con la siguiente relación:
 - a. Se acreditan 0,5 KMS por cada dólar gastado con tarjeta de débito.
 - b. Se acreditan 1,5 KMS por cada dólar gastado con tarjeta de crédito.
3. Acreditación por alianza comercial: Realizar compras en comercios asociados también permite acumular kilómetros. Los comercios asociados incluyen productos y servicios varios, hogar, entretenimiento, autos, embarcaciones y propiedades entre otros. La acreditación depende de la alianza, en general la acreditación fluctúa entre los 0,5 y 1,5 KMS por dólar gastado
4. Compra de KMS: La última forma de acreditar es comprando KMS por dinero. De esta forma los clientes pueden acercarse al canje invirtiendo una suma que normalmente no sería conveniente.

Para finalizar la contextualización, se presentan las distribuciones agregadas de canje y acreditación para el año 2014:

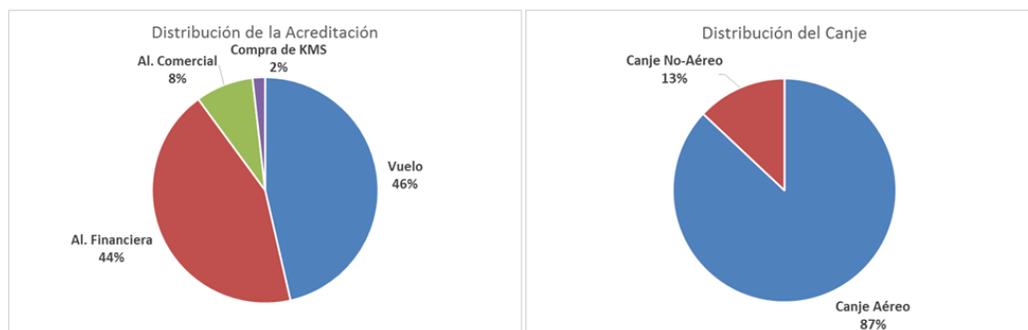


Ilustración 2. Distribuciones de Acreditación y Canje.

2. Descripción y Justificación del Proyecto

Esta investigación tiene como objetivo encontrar, describir y predecir a los clientes más propensos al canje de productos del catálogo de productos de LANPASS (Canje No-Aéreo o CNA). El análisis se complementará con una descripción tangencial del canje de asientos (Canje Aéreo o CA).

Hasta el año 2012, el 80% de los clientes de los programas de viajero frecuente no acumulaban puntos suficientes para canjear asientos.² Estos números denotan la importancia que tienen las alternativas de canje.

Este problema es relevante pues actualmente el área de alianzas comerciales de LANPASS, a cargo de la gestión del CNA, no maneja métricas de propensión de canje. La información que se maneja es a nivel producto, es decir, se sabe que productos y categorías se venden mejor pero no se maneja un perfil de canjeador ni se conocen qué factores influyen en la decisión de canje ni en qué medida estos actúan.

LANPASS posee mucha información de CRM de la base de clientes LAN y datos de las compras realizadas en el catálogo por los clientes, esto presenta una oportunidad para aplicar técnicas de minería de datos y marketing cuantitativo para encontrar información relevante en los grandes volúmenes de datos que la empresa posee y no está aprovechando.

La viabilidad del proyecto se justifica no solo en la disponibilidad actual de los datos, sino también en los bajos costos requeridos para estudiarlos. No es necesario un esfuerzo monetario extra de la empresa para realizar un estudio, pues está al alcance de un analista, sin embargo, hay restricciones de capital humano en el área, lo que motiva la inclusión de un memorista para hacerse cargo del proyecto.

² <http://simpliflying.com/2011/infographic-the-future-of-loyalty-program-will-be-powered-by-social-media/>

3. Objetivos

3.1. Objetivo General

- Identificar a los clientes de LANPASS más propensos al Canje No-aéreo.

3.2. Objetivos Específicos

- Caracterizar a los clientes en base a su comportamiento de canje e información demográfica. .
- Identificar las variables más relevantes que determinan la decisión de canje.
- Generar un modelo de propensión que permita determinar la probabilidad de que un cliente realice un canje.
- Estimar el impacto del modelo desarrollado en las ganancias de la empresa.
- Proponer aplicaciones comerciales en base a los resultados obtenidos.

4. Metodología

Para identificar a los clientes propensos al canje de catálogo, se utilizó la información disponible en las bases de datos de LANPASS. Se consideró adaptar una metodología KDD (Knowledge Discovery in Data) para cumplir los objetivos. Se procedió de la siguiente forma:

1. **Selección de datos:** Se escogieron las bases de datos que contienen la información demográfica y transaccional de los clientes. Estas son:
 - a. Base de datos con la información demográfica y transaccional general de los clientes pertenecientes al programa. De los 2.700.000 clientes se tomó una muestra del 10%.
 - b. Base de datos con el detalle de cada transacción realizada en el catálogo LANPASS. Al cruzar esta base con la anterior se obtuvo una muestra de 500 canjes.

La información extraída de estas bases se resume a continuación:

Demográficas	Transaccionales	Catálogo
Rut	Acreditación Total*	Rut
Género	Acreditación por Tipo de Alianza*	Categoría del producto
Edad	Canje Aéreo*	Precio en KMS
Pertenencia Santander	Canje No Aéreo*	
Mejor Tarjeta de Crédito (Si pertenece al banco)	Saldo*	
Categoría de Socio	Looks en Lan.com	
Región		

*Estas variables se presentan en los formatos: Total de KMS, máximo y mínimo de KMS, y fecha de primera y última transacción.

Ilustración 3. Variables seleccionadas.

2. **Preprocesamiento de datos:** En esta fase se eliminaron los datos inconsistentes y fuera de rango. Por ejemplo, edades negativas o infactibles, regiones pertenecientes a otros países, géneros indefinidos, etc.
3. **Transformación de datos:** Se recodificaron las variables según el uso que se necesitó darle.
 - a. Para observar el efecto de la expiración de KMS en el comportamiento de canje se creó la variable “cercanía a última renovación” a partir de la fecha del último vuelo.
 - b. Para los análisis descriptivos se utilizó un nivel de agregación de intervalos de 10.000 KMS que permitió una visualización adecuada de los datos.
 - c. Para su posterior uso con modelos de regresión logística, fue necesario transformar las variables “Saldo” y “Acreditación total” a sus análogas logarítmicas, pues estas variables distribuyen exponencialmente, lo que es un problema para esta herramienta. Las fases de **modelamiento**,

interpretación y evaluación corresponden al cuerpo de este informe. Se resumirá brevemente en que consiste cada etapa.

4. **Fase de modelamiento:** Esta etapa se divide en dos partes:
 - a. **Análisis descriptivo:** En este se observó el efecto individual de cada variable respecto a la propensión de canje tanto CNA como CA, de esta forma se pudo comprender el comportamiento de cada variable y su posible relación con las otras variables. De esta forma se obtuvo el manejo necesario de los datos para poder utilizar herramientas más sofisticadas de análisis.
 - b. **Modelos de clasificación:** Se utilizaron dos herramientas, ambas permiten obtener la probabilidad de canje de un cliente. Estas son:
 - i. **Regresión Logística Multinomial:** Este método se caracteriza por cuantificar la relación entre cada variable y su propensión de canje.
 - ii. **Árboles de Clasificación:** Por su parte los árboles generan clústeres de clientes y reglas que podrían utilizar los clientes en su decisión de canje.

Para su ejecución se utilizó el software IBM SPSS Statistics 21.

5. **Interpretación y evaluación:** La interpretación y evaluación de los resultados se desarrolla en la fase de modelamiento. Las propuestas comerciales y de trabajos futuros se presentan al final de este trabajo.

5. Marco Conceptual

5.1. Programa de Viajero Frecuente

Los programas de viajero frecuente son técnicas de fidelización utilizados por muchas líneas aéreas. Típicamente los clientes acumulan puntos según distancia volada, tramos volados u otras métricas asociadas a un vuelo. Más recientemente las aerolíneas han desarrollado diversas alianzas con empresas de otros rubros como bancos y bencineras, que permiten al cliente tanto acumular como canjear puntos en diversos productos y servicios.

5.2. Modelos de Clasificación

Los modelos de clasificación son técnicas estadísticas que permiten clasificar datos en grupos (comúnmente llamados clases) de acuerdo a diferentes algoritmos.

Estas técnicas se pueden dividir en dos tipos, técnicas de aprendizaje supervisado y no supervisado. La diferencia fundamental entre estos dos tipos es que en las técnicas de aprendizaje supervisado la clase a la que pertenece una entrada de datos es conocida a priori, mientras que en las técnicas no supervisadas esta información se desconoce.

Los datos con los que trabaja esta investigación incluyen la clase a la que pertenece cada cliente, por lo que las técnicas de aprendizaje supervisado se ajustan a las necesidades de este estudio.

Se utilizarán dos técnicas de aprendizaje supervisado en este trabajo, **Regresión Logística** y **Arboles de Clasificación**. De esta forma se tendrán dos formas de clasificar a los clientes, cada una con sus ventajas y debilidades.

5.2.1. Regresión Logística

La regresión logística, popularmente conocida como *Logit*, es un tipo de modelo de clasificación estructural y cuyo supuesto fundamental es que los clientes toman decisiones de manera racional. Se utiliza para predecir una respuesta utilizando variables independientes tanto categóricas como numéricas.

El Logit se basa en el supuesto de que el cliente decide entre las alternativas maximizando su beneficio. Esta utilidad posee dos componentes:

- **Aleatoria:** Define cómo distribuye el error que no se puede observar. En el caso del Logit el error distribuye *Gumbel* o *valor extremo tipo I*.
- **Determinística:** Explica cómo las variables observables afectan la toma decisión del consumidor. Comúnmente es una combinación lineal de las variables independientes.

- Existen dos tipos de Logit: **Binomial**: En donde la variable dependiente puede tomar dos valores. Una decisión que podría modelarse de esta forma sería la decisión de canjear o no canjear.
- **Multinomial**: En donde la variable dependiente puede tomar más de 2 valores. En esta memoria estudiaremos la decisión en donde el cliente decide entre Canje Aéreo, Canje No Aéreo o no canjear.

La Regresión Logística Multinomial genera tantas ecuaciones como alternativas tenga la decisión a modelar. En este caso se tienen tres opciones lo que genera 3 ecuaciones, siendo $Y_i=1$ la categoría de referencia se tiene:

$$\text{Pr ob}(Y_i = 1) = \frac{1}{1 + e^{\alpha_2 + \beta_{12}X_{1i} + \beta_{22}X_{2i}} + e^{\alpha_3 + \beta_{13}X_{1i} + \beta_{23}X_{2i}}}$$

$$\text{Pr ob}(Y_i = 2) = \frac{e^{\alpha_2 + \beta_{12}X_{1i} + \beta_{22}X_{2i}}}{1 + e^{\alpha_2 + \beta_{12}X_{1i} + \beta_{22}X_{2i}} + e^{\alpha_3 + \beta_{13}X_{1i} + \beta_{23}X_{2i}}}$$

$$\text{Pr ob}(Y_i = 3) = \frac{e^{\alpha_3 + \beta_{13}X_{1i} + \beta_{23}X_{2i}}}{1 + e^{\alpha_2 + \beta_{12}X_{1i} + \beta_{22}X_{2i}} + e^{\alpha_3 + \beta_{13}X_{1i} + \beta_{23}X_{2i}}}$$

El modelo estimará tantos vectores de parámetros como opciones existan, menos una. Es decir, en este caso se estimarán 2 vectores β , uno para la opción 2 y otro para la opción 3.

$$\beta_i = (\beta_{0i}, \beta_{1i}, \beta_{2i}, \beta_{3i}, \dots, \beta_{ni})$$

Los parámetros estimados forman la *función logística*.

$$\ln(p_i) = \beta_{0i} + \beta_{1i} * x_{1i} + \beta_{2i} * x_{2i} + \beta_{3i} * x_{3i} + \dots + \beta_{ni} * x_{ni}$$

De esta forma, con los parámetros estimados podremos conocer el efecto que tiene el aumento unitario de una variable independiente X_{ij} en $\ln(p)$, donde p la probabilidad de canje. En términos sencillos, esta técnica asignará $\beta_{ij}=0$ a las variables que no influyan en la probabilidad de canje. Pero lo fundamental es que además de mostrar qué variables influyen o no en la probabilidad de canje, nos dirá en qué magnitud y si estas variables impactan de manera positiva o negativa.

5.2.2. Validación de modelos

La validación es una etapa fundamental para aceptar que el conocimiento que se extrae de los datos es correcto. Esta etapa consiste en verificar que los patrones encontrados en los datos sean robustos a distintos criterios. Se divide en dos partes fundamentales:

Validación dentro de la muestra

Esta parte involucra el uso de test de bondad de ajuste, reducción de verosimilitud y significancia.

Validación fuera de la muestra

También se hace necesario chequear que los patrones encontrados en los datos se repitan a otras muestras independientes. Por ejemplo que un comportamiento se repita en dos muestras de datos que fueron tomados en condiciones similares pero de forma independiente. Esto también se puede realizar separando una muestra en dos submuestras y comprobando que presenten los mismos resultados.

5.2.3. Árboles de Decisión

Similar a la regresión logística, los árboles de decisión permiten predecir una variable dependiente, aunque no se limitan a variables categóricas. Existen dos tipos de árboles; **los árboles de regresión**, que se hacen cargo de trabajar variables dependientes continuas, y **los árboles de clasificación** que trabajan variables discretas, como es el caso de esta investigación, en donde se busca entender cuando el cliente canjea y cuando no.

Árboles de clasificación³

Los árboles de clasificación son útiles para la exploración inicial de datos y apropiados cuando hay un número elevado de datos y existe incertidumbre sobre la manera en que las variables independientes deben introducirse en el modelo.

Una de las ventajas de los árboles de clasificación es que se adaptan fácilmente a datos perdidos sin necesidad de eliminar la observación completa, además, son robustos a errores. A pesar de esto, fallan en no poder cuantificar la importancia de los distintas variables en la clasificación que realizan.

Su mejor aplicación se encuentra en grandes masas de datos, donde las herramientas convencionales de regresión no pueden revelar estructuras complejas. Como por ejemplo, relaciones no lineales entre variables en donde los árboles fácilmente segmentan por tramos dichas relaciones.

Finalmente, los árboles permiten identificar perfiles de clientes de mejor forma por su naturaleza gráfica e intuitiva.

Los árboles de clasificación funcionan de la siguiente forma:

1. Se toma un nodo inicial y se pregunta cómo dividir al conjunto de datos en dos partes homogéneas utilizando una sola variable. Por ejemplo, si una persona acreditó $x > 10.000$ KMS en el último año o si acreditó $x \leq 10.000$ KMS.
2. De la etapa (1) surgen dos nodos, a cada uno de esos nodos se les aplica nuevamente una división usando el mismo procedimiento. El proceso termina una vez que se hayan clasificado todas las observaciones correctamente en su grupo.

Se consideran los siguientes tipos de algoritmos para árboles de decisión⁴:

³ <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema6dm.pdf>

⁴ Rockach L., Maimon O., *Data Mining With Decision Trees: Theory and Applications*, 2nd edition, 2015.

- CRT (Árboles de regresión y clasificación): se caracteriza por entregar clasificación binaria, es decir, cada nodo se divide en dos resultados posibles. Estos resultados pueden ser en forma de clases (clasificación) o en un número real (regresión).
- CHAID (Detección automática de interacción Chi-cuadrado): algoritmo estadístico y multidireccional que crea segmentos y perfiles para alcanzar un resultado deseado, detectando automáticamente las interacciones mediante Chi-cuadrado, eligiendo en cada paso la variable independiente de mayor interacción con la dependiente.
- CHAID Exhaustivo: variante de CHAID que examina todas las divisiones posibles para cada predictor, entregando un resultado más exacto pero con mayor gasto computacional.
- QUEST (Árbol rápido, insesgado, y eficiente): método que sólo puede ser utilizado si la variable dependiente es nominal. Selecciona las variables insesgadas del modelo y crea árboles binarios de forma rápida y eficaz.

6. Alcances

Categorías

Actualmente el catálogo posee 9 categorías. Para el desglose del canje no aéreo solo se utilizaron 3 categorías: GiftCard de Multitienda, Tecnología y Hogar. Estas categorías representan aproximadamente el 80% de los KMS canjeados en el catálogo. Durante la investigación se descubrió que Hogar no representaba diferencias de comportamiento por lo cual se desagregó como “Otras Categorías” y se utilizó como categoría de referencia.

Es importante mencionar que “GiftCard de Multitienda” no es una categoría existente, pero la futura re categorización del catálogo buscará distribuir las Giftcards pertenecientes a la categoría existente “GiftCard” según la categoría que represente la tienda a la que corresponde la tarjeta. De esta forma, se crea una categoría relevante que busca estudiar un comportamiento, más que estudiar las categorías existentes.

Clientes

Los clientes de LANPASS pertenecen a diversas nacionalidades. Se considerarán solo los clientes residentes en Chile, este monto asciende a aproximadamente 2.700.000 personas. La idea detrás de acotar la base de clientes es que la decisión de canje que se busca modelar considere las alternativas de CA y CNA, pues, las limitaciones de despacho no permiten a un cliente en el extranjero canjear en el catálogo.

Horizonte temporal

Para modelar la propensión de canje se considerarán 2 ventanas de tiempo previo al mes de canje:

- Últimos 12 meses: Permite conocer el comportamiento transaccional y de acreditación más reciente.
- Últimos 36 meses: Captura información atinente a la expiración próxima de KMS en los clientes.

El mes de canje de referencia será Junio, el cual es considerado por la empresa como un mes normal de canje, pues no tiene ofertas especiales ni elementos que puedan llevar a la investigación a conclusiones erróneas. Adicionalmente, se utilizó Mayo para comparar y validar los modelos realizados.

7. Análisis Descriptivos

Para obtener una mirada general de la data se realizó un análisis bivariado en el cual se estudian las interacciones de la *propensión de canje* respecto a distintas variables demográficas y transaccionales, una por una.

En primer lugar se estudió la propensión en donde la decisión es canjear pasajes o productos del catálogo, es decir, *Canje Aéreo (CA)* o *Canje No Aéreo (CNA)*.

Posteriormente, dentro de los *Canjes No Aéreos* se estudiaron las preferencias de categoría.

De esta forma se obtiene una primera mirada exhaustiva de los canjeadores no aéreos.

La propensión de canje se modeló como sigue:

$$Pr(\text{Canje } X) = \frac{\text{N}^\circ \text{ de Clientes que canjearon } X}{\text{Total de clientes}}$$

7.1. Análisis Descriptivo de Tipo de Canje

La propensión de canje a estudiar será tanto la propensión de *Canje Aéreo* (CA) como de *Canje No-aéreo* (CNA). El propósito de estudiar la propensión al CA es buscar variables que en su análisis bivariado presenten una amplia diferencia entre sus propensiones. Estas variables serán buenas candidatas a utilizarse en los modelos de clasificación que esta investigación propone.

7.1.1. Estudio del canje según variables demográficas

A continuación se presenta el comportamiento de canje según variables demográficas, estas son:

- Edad
- Género
- Región de residencia
- Categoría de Socio (O Tier)
- Pertenencia a Santander
- Mejor Tarjeta de Crédito

Edad

Como se observa a continuación existe una tendencia al aumento en la *probabilidad de canje aéreo* a medida que aumenta la edad. Esta tendencia no se repite para el *Canje No- Aéreo*, en donde la *probabilidad de canje* aumenta, encontrando su máximo en las personas de **entre 31 y 40 años** y desciende posteriormente a medida que aumenta la edad. Se encuentra un segmento de interés en las personas de **entre 31 y 40 años**, quienes, además de presentar una alta propensión de CNA, también concentran la mayor cantidad de clientes.

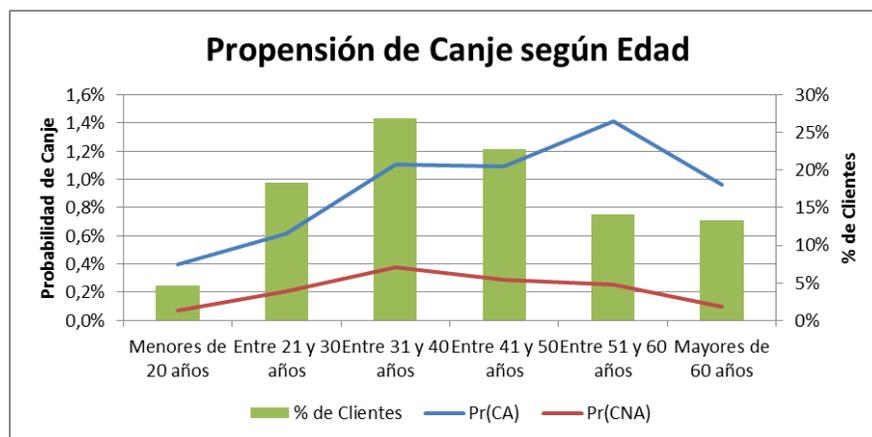


Ilustración 4. Propensión de canje según edad.

Género

En el estudio de género hay dos perspectivas que podemos encontrar, una mirada entre géneros y una mirada dentro de cada género. De estos Análisis se desprende:

- **Análisis Entre Géneros**

- Los hombres son **1,5** veces más propensos al CA que las mujeres.
- Los hombres son **2** veces más propensos al CNA que las mujeres.

Es importante notar que no se puede atribuir directamente causalidad al *género* para concluir respecto a las diferencias de propensión. Estas diferencias se atribuyen a otras diferencias entre los géneros, principalmente transaccionales.

- **Análisis por Género**

- Los hombres son **3,6** veces más propensos al CA que al CNA.
- Las mujeres son **4,6** veces más propensas al CA que al CNA.

A diferencia del Análisis entre Géneros, sí se puede concluir respecto a las preferencias de los clientes dentro del mismo género, pues no existen estas diferencias transaccionales.

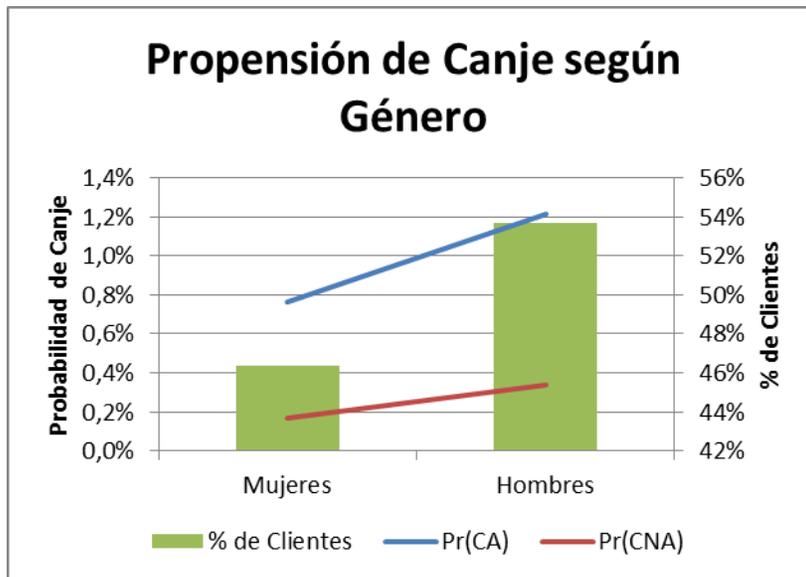


Ilustración 5. Propensión de Canje vs Género (% de Clientes)

Mejor Tarjeta de Crédito

Esta variable categórica nos dice cuál es la mejor tarjeta de crédito que posee un cliente Santander. Los clientes que no tienen tarjeta de crédito son los mismos clientes que no participan de la alianza (Ver **Anexo N°1**). Esto se debe a que los clientes que abren cuenta corriente en Santander poseen todos tarjeta de crédito y automáticamente entran a la base de LANPASS (Esto no incluye a los clientes de cuenta vista).

La mejor tarjeta de crédito es el mejor indicador con que se cuenta sobre el poder adquisitivo de un cliente, a continuación se ordenan de “peor” a “mejor” tarjeta. Se observa que una mejor tarjeta de crédito implica una mayor probabilidad de CA, sin embargo, no parece ser un buen predictor para el CNA. La concavidad de las curvas también es interesante, pues si bien ambas curvas de propensión crecen, la propensión al CA crece cada vez más rápido mientras que la propensión al CNA presenta una pendiente plana.

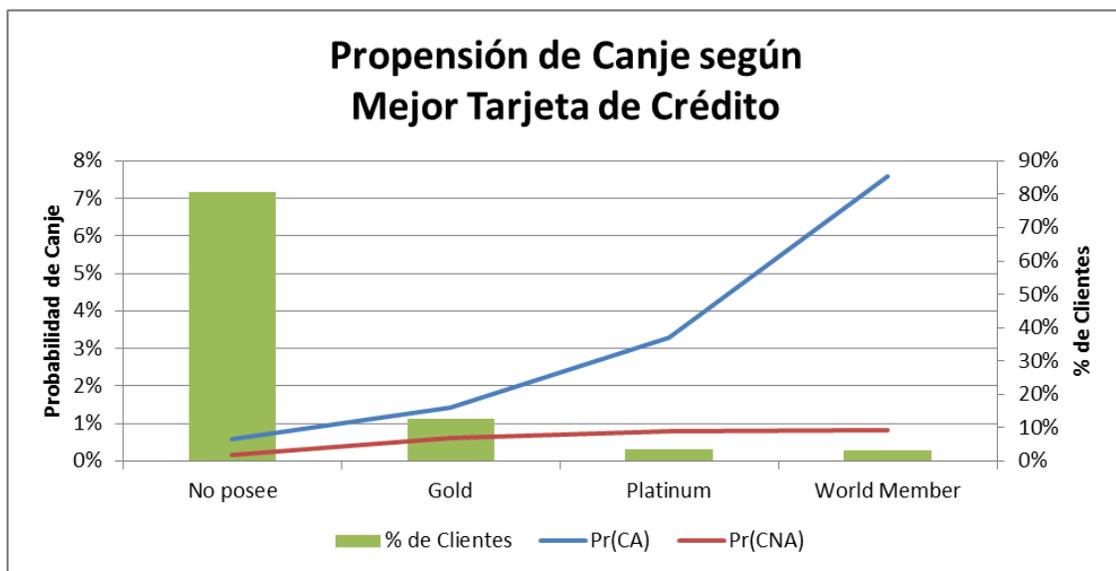


Ilustración 6. Propensión de Canje vs Mejor Tarjeta de Crédito (% de Clientes).

Categoría de socio

La categoría de socio es un indicador de acreditación y/o vuelo. Un cliente que viaja más y/o acredita más tendrá una mejor categoría de socio. Las categorías de socio tienen beneficios para el cliente con los cuales se recompensa su fidelidad.

Las categorías se muestran en orden ascendente, siendo los Comodoro/Black los “mejores”. La propensión de CNA aumenta linealmente a medida que mejora la categoría. Llama la atención la categoría PREMIUM SILVER en donde la diferencia entre propensión de CA y CNA alcanza su máximo.

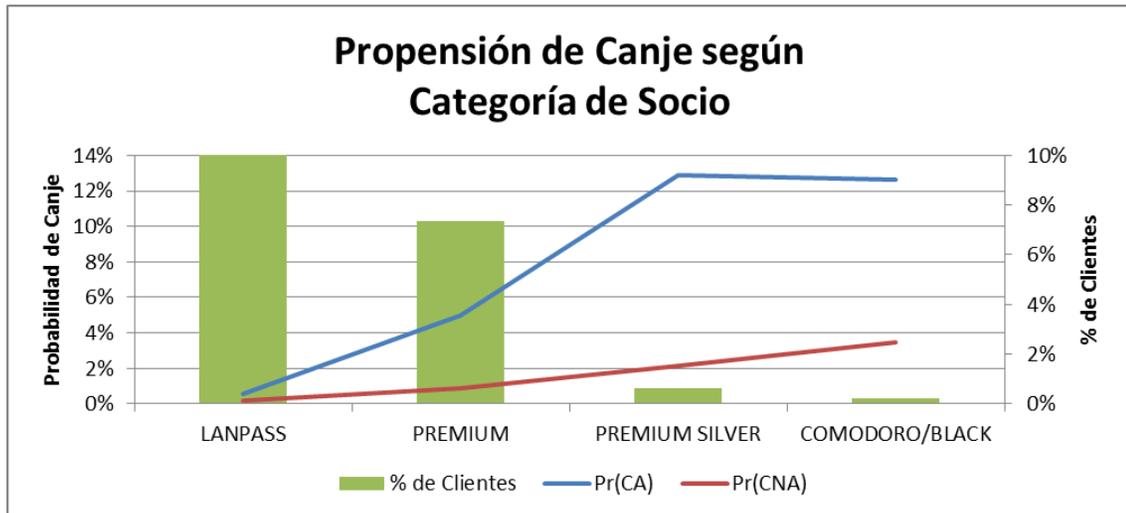


Ilustración 7. Propensión de Canje vs Categoría de Socio (% de Clientes).

Región de residencia

La región de residencia no muestra ser un buen predictor de la probabilidad de canje, sin embargo, dentro de su propia escala se encuentran diferencias de propensión entre las regiones centrales con las norte y sur (con algunas excepciones), estas tienen razones cualitativas, por ejemplo, es razonable pensar que la región de Antofagasta presente mayor propensión de canje dado que los clientes que trabajan en minería suelen volar mucho, y por consiguiente, acreditan bastante.

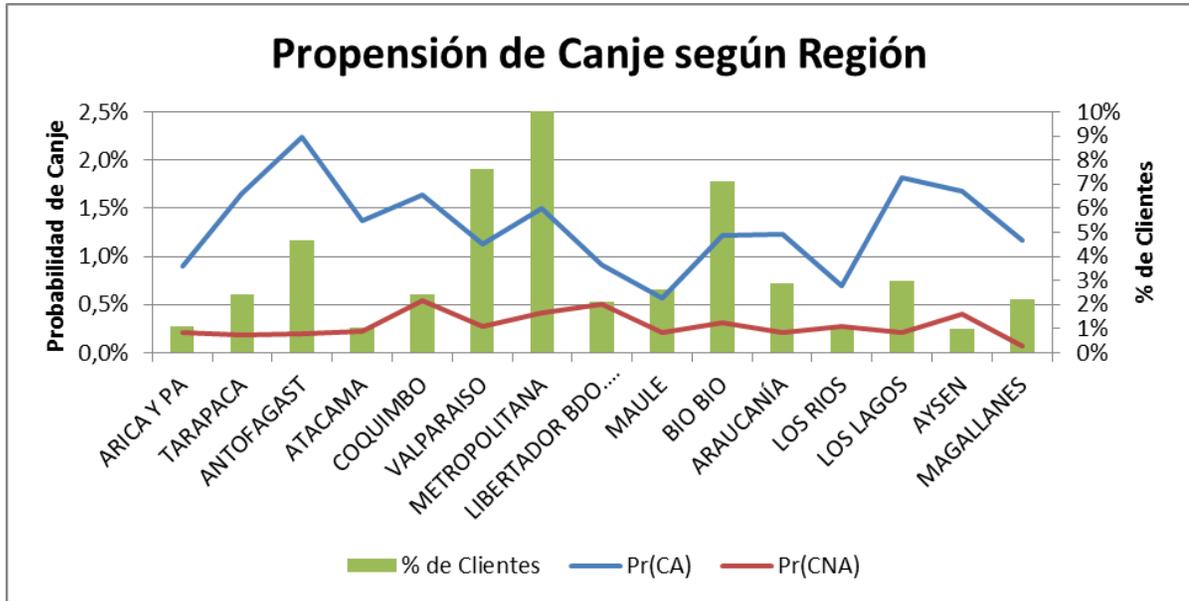


Ilustración 8. Propensión de Canje vs Región

7.1.1.1. Efecto de variables transaccionales sobre variables demográficas

Las variables demográficas no dicen mucho respecto a cómo se comporta un cliente por sí solas, por ejemplo, es probable que un cliente con una muy buena tarjeta de crédito acredite muchos kilómetros por tener un mayor ingreso, o que los hombres acrediten más (y canjeen más) por motivos culturales, en donde la mayoría es el sustento económico del hogar.

Para ilustrar este efecto, veremos cómo varía la propensión de canje cuando la hacemos interactuar con variables transaccionales. En particular, compararemos respecto a las variables más importantes en el momento del canje que son **Saldo** y **Acreditación Anual**.

Efecto Saldo

A continuación se tomaron 2 variables; Mejor Tarjeta de Crédito y Edad, para mostrar como a medida que decrece el saldo disponible, la propensión para ambos tipos de canje disminuye.

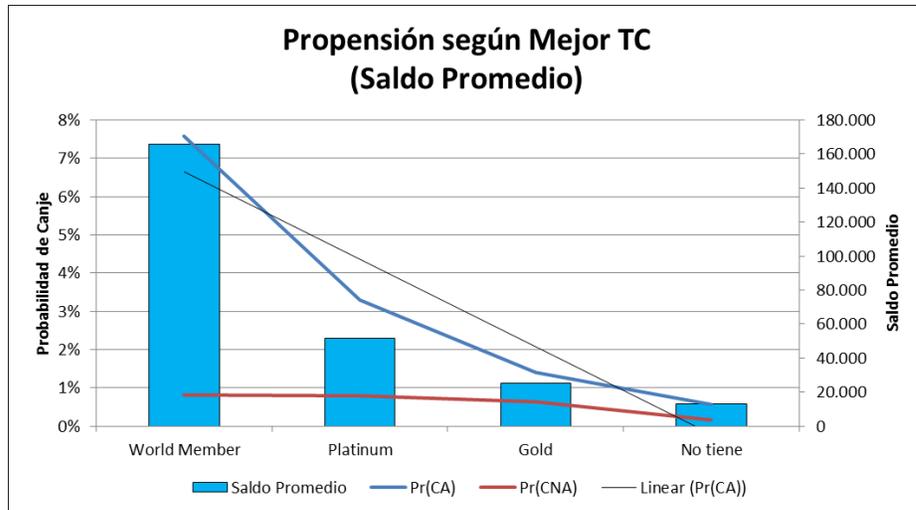


Ilustración 9. Efecto Saldo, Caso Tarjeta de Crédito

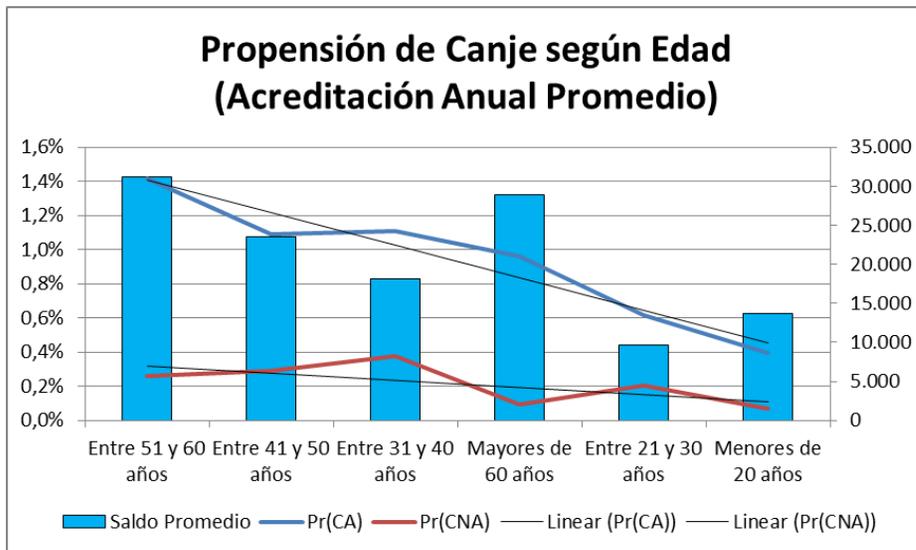


Ilustración 10. Efecto Saldo, Caso Edad

Se observa que existe una tendencia a menores propensiones de canje para menores saldos, sin embargo, este efecto no es perfectamente lineal. Esto se debe a que la propensión de canje no solamente depende del saldo, sino también de otras variables.

En los **Anexos N°2 y N°3** se puede observar que este efecto también se repite en otras variables.

Efecto Acreditación Anual Promedio

Al igual que con el saldo, se observa un efecto similar para la Acreditación Anual Promedio.

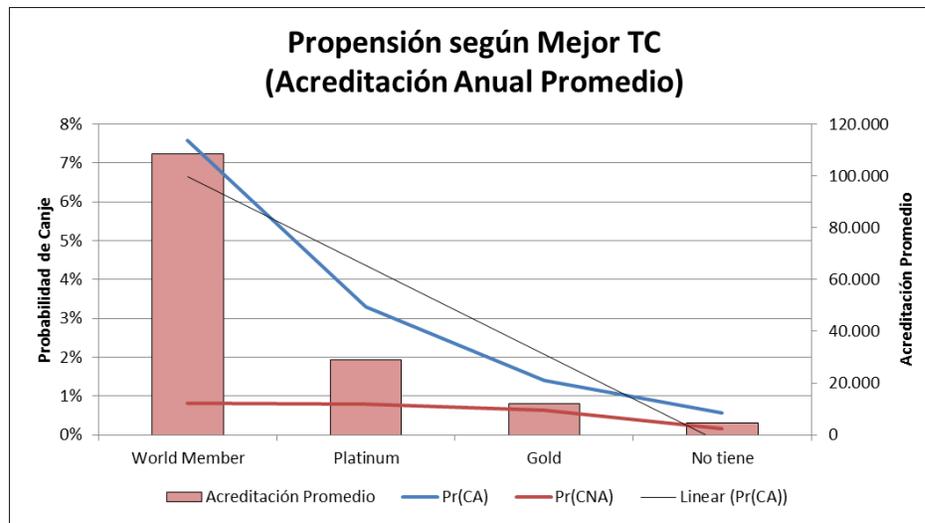


Ilustración 11. Propensión Según Mejor TC

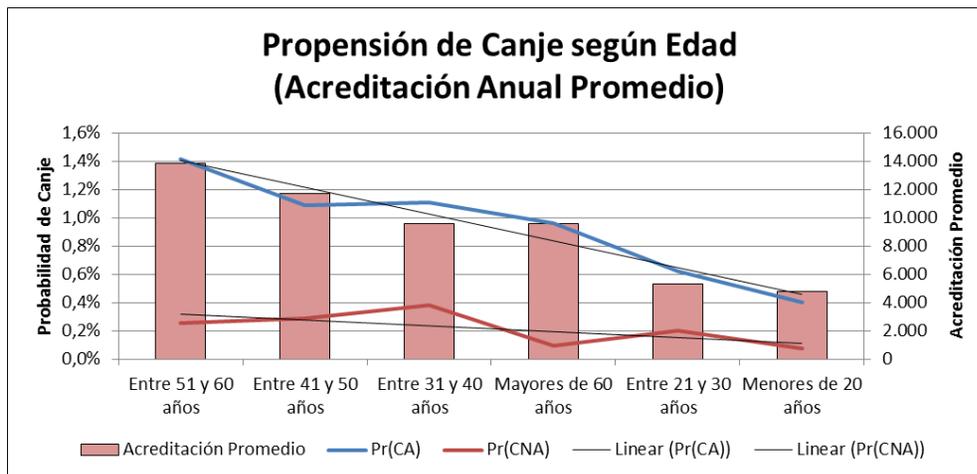


Ilustración 12. Propensión de Canje Según Edad

Podemos aseverar entonces que las variables demográficas tienen un comportamiento transaccional implícito, esto es útil porque nos permitirá predecir un comportamiento transaccional para los clientes nuevos, de quienes naturalmente solo se tendrán datos demográficos.

En los **Anexos N°4 y N°5** se puede observar que este efecto también se repite en otras variables.

7.1.2. Estudio del canje según Variables transaccionales

A continuación veremos el efecto de las variables transaccionales en la propensión de canje. Las variables transaccionales a estudiar son:

- Acreditación
- Canje Aéreo
- Canje No aéreo.
-

Estas variables se trabajaron según 3 indicadores:

- N° de Transacciones
- Total de KMS
- Máximo de KMS

Acreditación Total

La acreditación total parece ser un buen predictor de CA, sin embargo, para CNA esta variable por sí sola no muestra tener un efecto igual de fuerte. A mayor acreditación mayor propensión al CNA, pero la diferencia entre la propensión de CA y CNA también aumenta. Destaca también un quiebre en la curva de CA un aumento en la pendiente para acreditaciones mayores a 40.000 KMS.

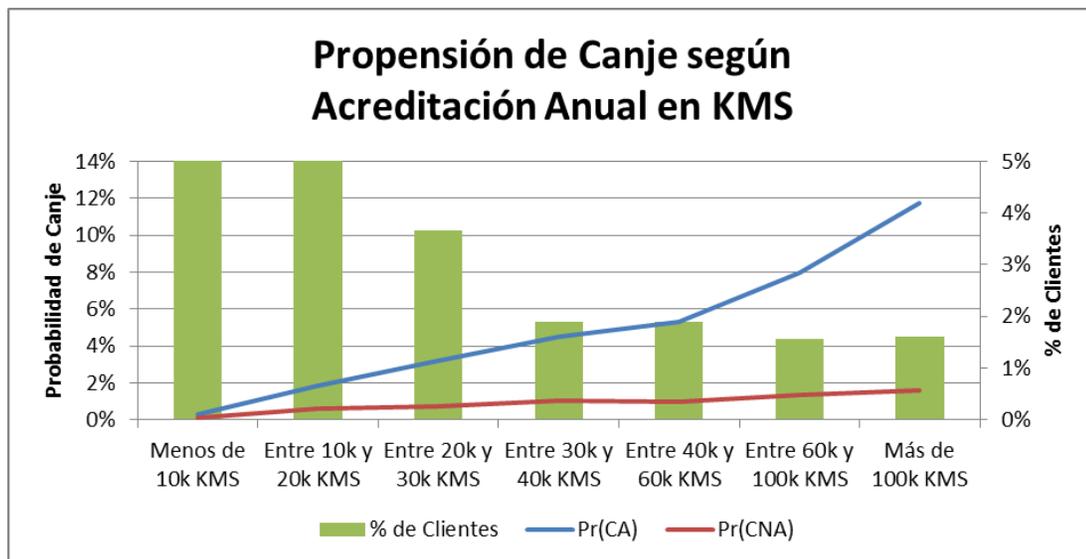


Ilustración 13. Propensión de Canje vs Acreditación Total (% de Clientes).

Canje Aéreo y No-Aéreo en KMS

El Canje Aéreo es un interesante predictor del propio Canje Aéreo, los clientes que canjean más de 100.000 KMS en el último año tienen una probabilidad de 9% de canjear al mes siguiente de pasado dicho tiempo. Este número es interesante,

pues, si asumimos que esta probabilidad se mantiene constante a lo largo del año, nos encontraríamos con un cliente que canjea asientos todos los años. Otra observación relevante es que la variable “KMS canjeados en CA” discrimina más fuertemente a los clientes propensos a CA de los de CNA que en el caso de la variable “KMS canjeados en CNA” que si bien, muestra que un cliente que canjea más en CNA, será más propenso a canjear CNA que CA, esta diferencia de propensión es menor que en su similar “KMS canjeados en CA”. Además, los clientes que más canjean CA muestran ser más propensos al CA que su análogo de CNA con CNA.

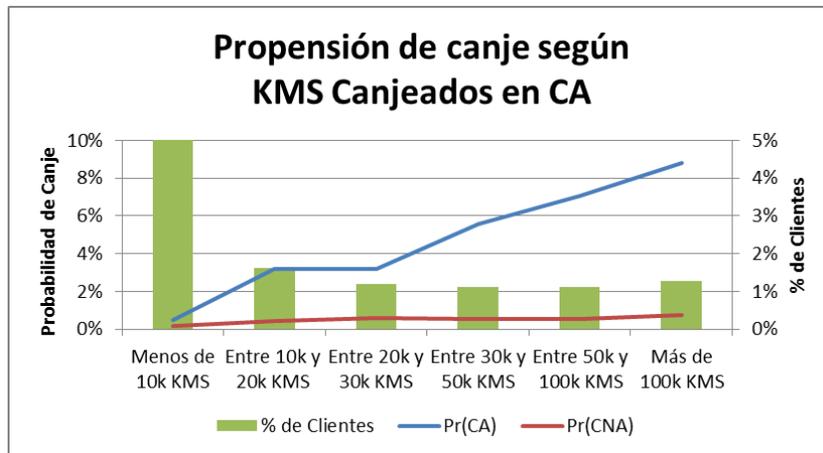


Ilustración 14. Propensión de Canje según KMS Canjeados en CA.

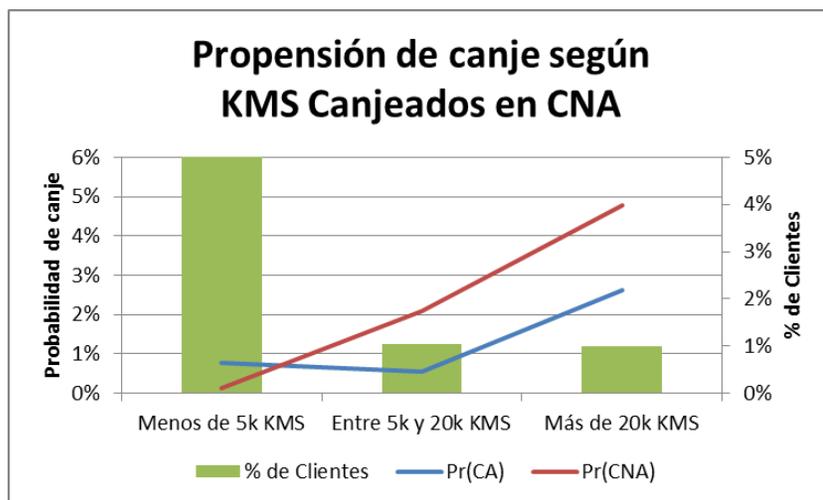


Ilustración 15. Propensión de Canje según KMS Canjeados en CNA.

N° de CA y CNA

En el N° de canjes encontramos segmentos de clientes con alta propensión de CA, comparable con “Acreditación Total en KMS” Y “N° de Acreditaciones”. A su vez encontramos el segmento con la máxima probabilidad de CNA de entre todas las

variables transaccionales que es el de los clientes que canjearon 2 veces o más en CNA el último año.

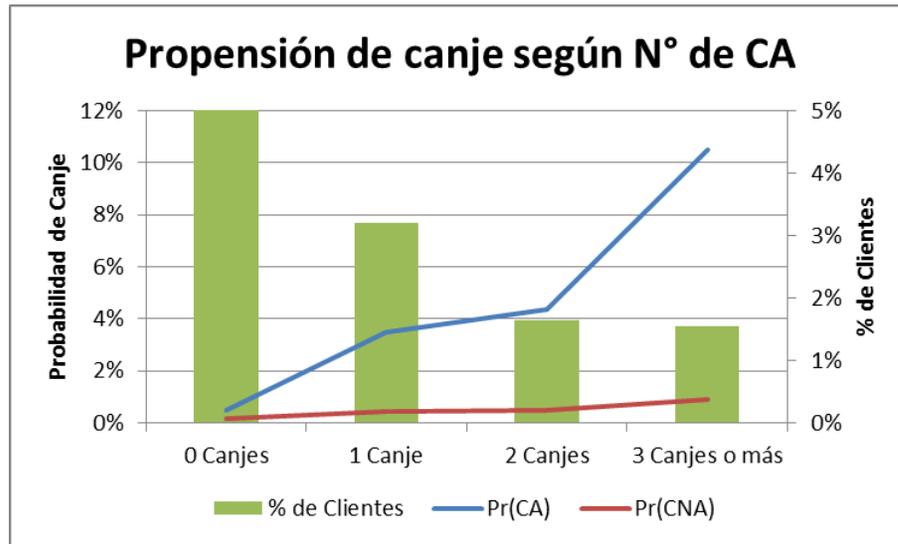


Ilustración 16. Propensión de Canje vs N° de CNA (% de Clientes).

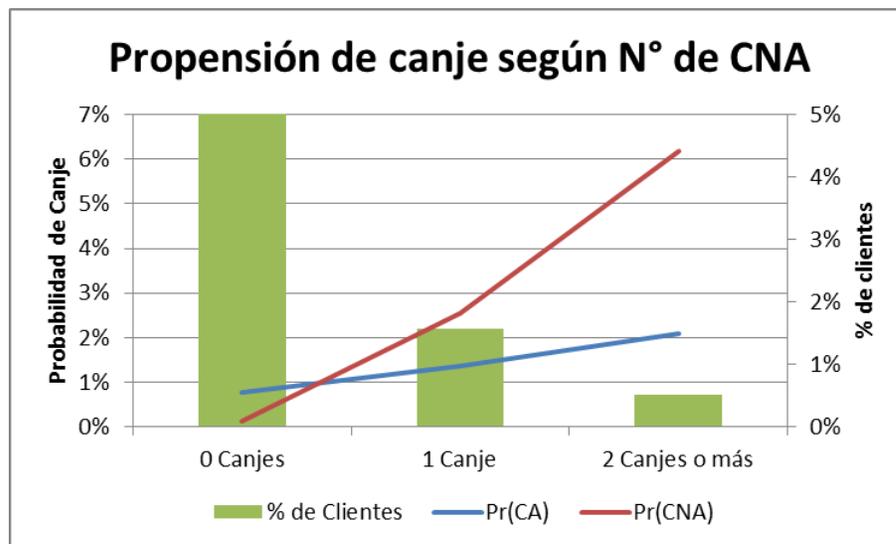


Ilustración 17. Propensión de Canje vs N° de CNA (% de Clientes).

Existencia de Canjes Recientes

El análisis anterior motiva la búsqueda de una variable general que ilustre la propensión de canje ante la existencia de cualquier tipo de canje previo al momento del análisis, esto será útil para saber cómo influye el canje a modo general en las propensiones de CA y CNA. Además, da luces de la satisfacción de los clientes, pues si están satisfechos, deberían tender a repetir el comportamiento.

Se observa que los clientes que han canjeado durante el último año son 7 veces más propensos a ambos tipos de canje que los clientes que no. Las diferencias de escala hacen que la propensión de CA alcance valores mayores pero el efecto es el mismo para ambos tipos de canje. Se confirma que los clientes alcanzan cierto nivel de satisfacción y repiten el comportamiento de canje.

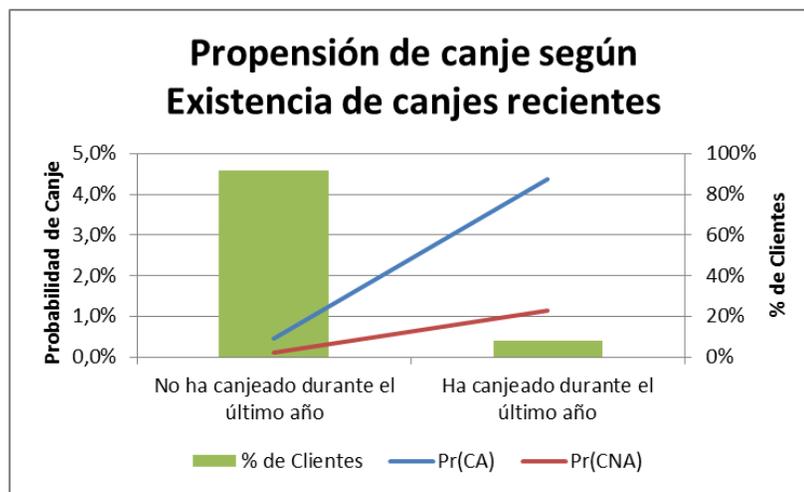


Ilustración 18. Propensión de Canje vs Existencia de Canjes Recientes (% de Clientes).

Tipo de Canjeador

Similar a “Tipo de Acreditador”, se define un canjeador “Aéreo” o “Catálogo” si es que canjean el 80% de sus KMS en un determinado tipo de canje. Esta definición excluye el mes en el que se calcula la propensión, por lo que es una forma más de utilizar el comportamiento anterior para comprender el comportamiento futuro.

Como se observa existen muy pocos clientes tipo Mix, pero son los más propensos a ambos tipos de canje. Se identifica también un perfil de canjeador de catálogo, pues se verifica que los canjeadores “Catálogo” no solo son muy propensos al CNA, sino también lo son en comparación al CA.

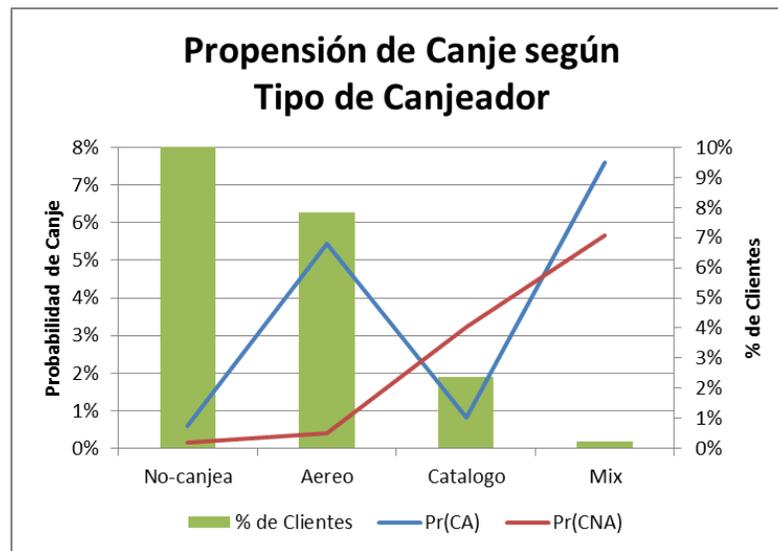


Ilustración 19. Propensión de Canje vs Tipo de canjeador (% de Clientes).

Respecto a la acreditación promedio de los segmentos definidos, los clientes mix acreditan en promedio más que los clientes “Catálogo” y “Aéreo” en conjunto lo que se podría interpretar como un perfil de cliente que canjea el remanente de KMS que no utiliza en viajes, en el catálogo.

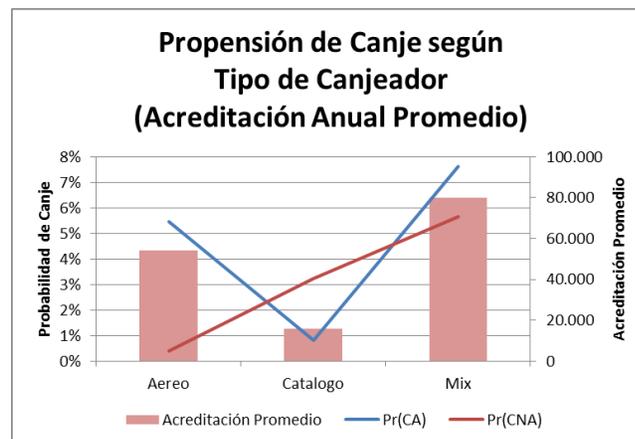


Ilustración 20. Propensión de Canje vs Tipo de Acreditador (Acreditación Promedio).

Vista de Pasajes en Lan.com

El propósito de este análisis es observar la propensión de canje de los clientes que ingresan a canjear o cotizar pasajes a la web. El proceso se divide en 3 etapas; cotización de pasajes, despliegue de precio y la venta del pasaje. Estas etapas por simplicidad se nombrarán “Looks”, “Paso 3” y “Venta”. Este proceso genera 3 segmentos de “Lookers” los cuales se compararán con los clientes que no cotizan pasajes (No-Lookers). Para evaluar, se toma el comportamiento del cliente en el mes anterior al canje.

El único resultado sorprendente respecto al CA es la alta propensión al CA para los clientes que canjearon pasajes en el mes anterior, observado en el segmento “Paso 3-Venta” del gráfico .

Por el lado del CNA se observa que los clientes que entran a canjear pasajes son en general más propensos al CNA que el promedio de los que no.

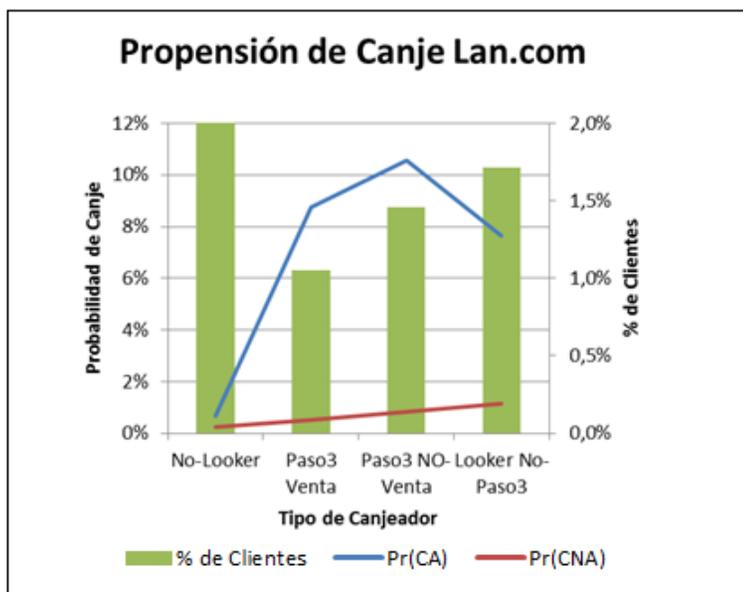


Ilustración 21. Propensión de Canje vs Looks en Lan.com.

En el gráfico se observa que los clientes que no miran pasajes en la web son poco propensos a ambos tipos de canje.

Lo más importante de este análisis es que los clientes son más propensos mientras antes se frustran en el proceso de canje. Esto se ilustra en la curva roja, en donde el último segmento es el cliente que se frustró más rápido, mientras que el segundo es el que se no se frustró.

En efecto, la propensión al CNA es 5 veces más que el promedio para los más frustrados (Looker No-Paso3) y 3 veces mayor para los de frustración media (Paso3 No-Venta).

Saldo

El saldo es la variable más interesante de este análisis. Por sí sola puede que no diga mucho, sin embargo esta variable es fundamental para la decisión de canje pues un cliente sin saldo no puede canjear. En la base de datos utilizada la gran mayoría de los clientes tienen saldos insuficientes para cualquier tipo de canje.

Las curvas de propensión tienen una forma similar a las de la variable “Acreditación Total en KMS”, esto se explica en que el saldo es un output de canje y acreditación, y que el canje en general es bajo.

Se observa que a partir de los 60.000 KMS de saldo la propensión de CNA baja hasta los 90.000 KMS, esto se puede interpretar como un segmento en donde la cercanía al canje internacional hace menos atractivo el CNA pues este aleja al cliente de su meta de vuelo. Posteriormente la propensión vuelve a subir, probablemente porque estos clientes tienen saldo de sobra para volar y canjear en el catálogo. Esto se ve complementado también por el quiebre de la curva de propensión de CA a partir de los 90.000 KMS.

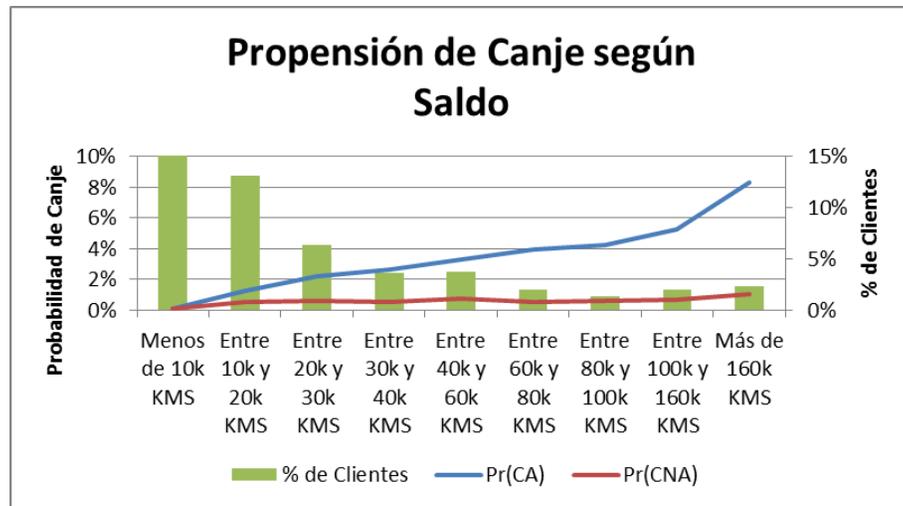


Ilustración 21. Propensión de Canje según Saldo.

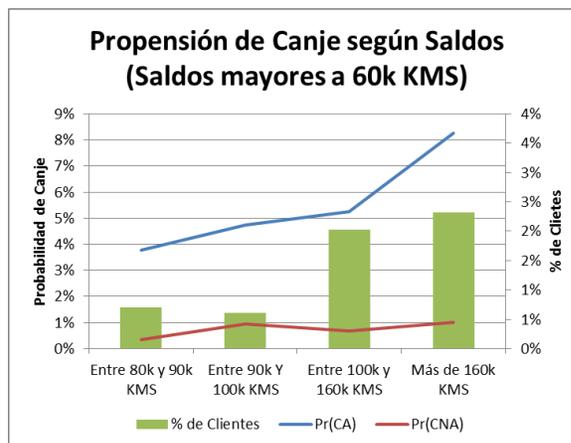


Ilustración 22. Propensión de Canje según Saldos mayores a 60k KMS.

Tiempo desde la última renovación de KMS

La expiración de KMS es una variable interesante, pues revela el comportamiento de los clientes cuando están a punto de perder los KMS que acreditaron recientemente. Los KMS acreditados por vuelo expiran a 3 años exactos de ser acreditados, además, este tipo de acreditación renueva el saldo total del cliente por los mismos 3 años. Los KMS acreditados por otros medios caducan el 31 de diciembre del segundo año desde que fueron acreditados.

Como se ve a continuación, los clientes se comportan de manera muy distinta respecto al CA y al CNA en función de cuánto tiempo ha pasado desde la última vez que renovaron su saldo de KMS. Los clientes que renovaron hace poco tienden a tener una propensión al CA muy alta y al CNA muy baja. Por otro lado, cuando los KMS están a punto de expirar, la propensión al CNA muestra un aumento interesante, lo cual se puede interpretar como una forma de sacarle el mayor valor a los KMS que de otra forma no se usarían pues el CA involucra otros recursos como tiempo, dinero los cuales le quitan atractivo al CA como decisión bajo apuro. Se concluye que una arista importante del CNA es ser una alternativa que entregue valor a los clientes cuyos KMS están a punto de expirar.

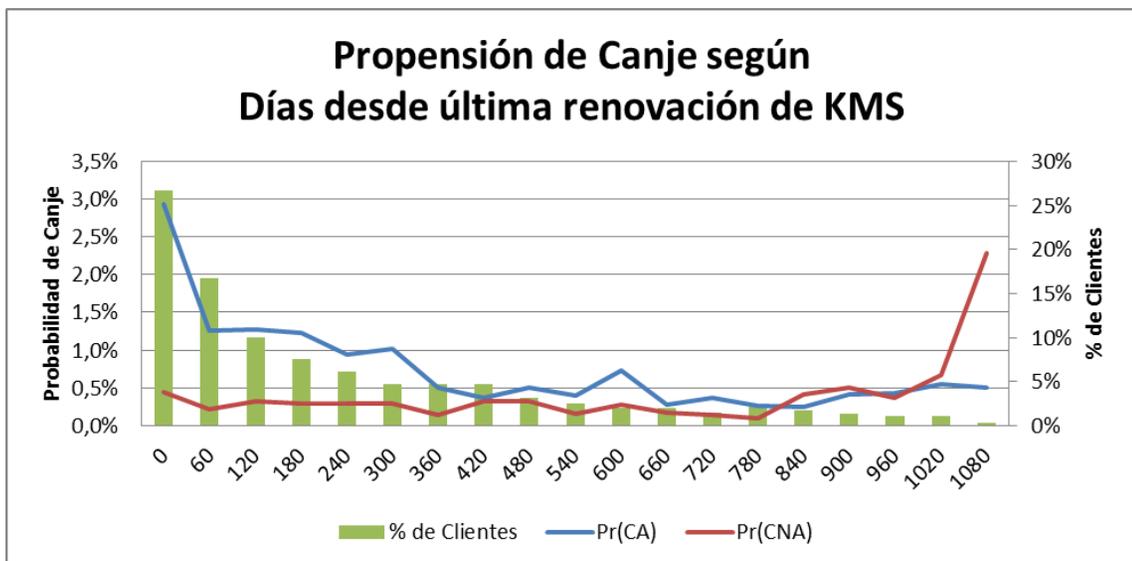


Ilustración 23. Propensión de Canje vs Caducidad Junio (% de Clientes).

Es de interés de la empresa ver cómo se comportan los clientes cercanos a la fecha de expiración relativa a los KMS que no fueron acumulados por vuelo. Como se observa, la curva de CNA es plana a lo largo del tiempo y esta supera en menos de un año al CA. Esto no tiene que ver solo con la caducidad de los KMS, sino también con el efecto de la navidad, en donde el catálogo aumenta sus ventas por la naturaleza de esta festividad.

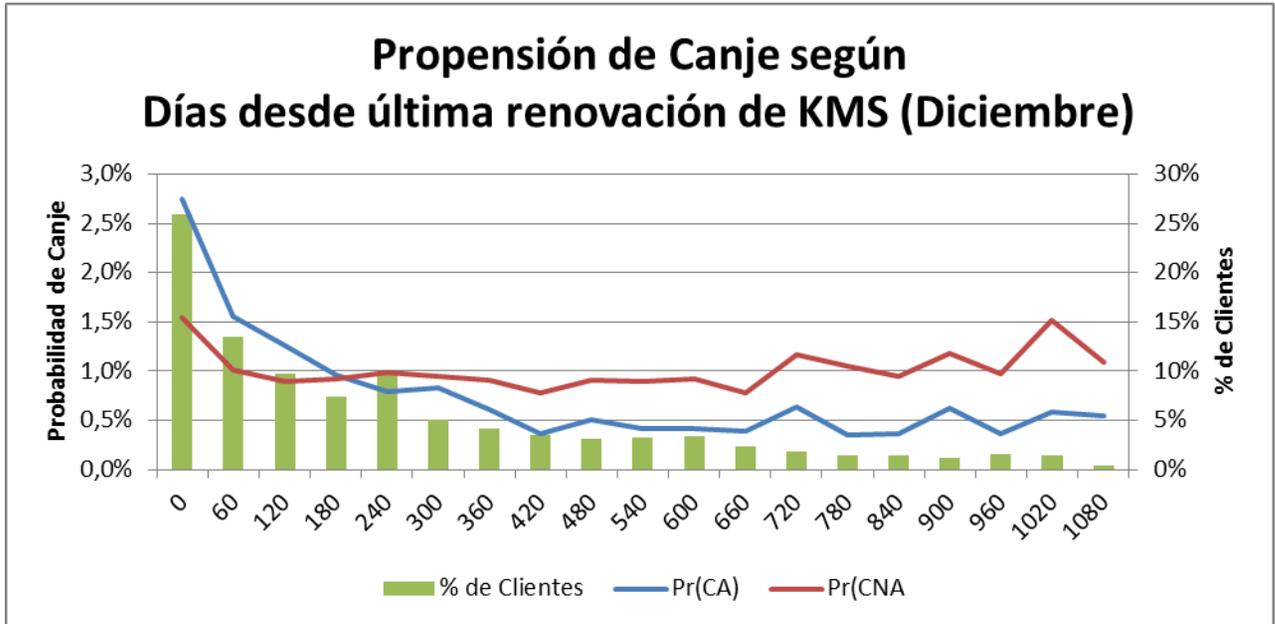


Ilustración 24. Propensión De Canje vs Caducidad Diciembre (% de Clientes).

Entre los **Anexos N°6 y N°x** se puede observar el estudio de las siguientes variables:

- Máximo de CA en KMS.
- Máximo de CNA en KMS.
- Tipo de Acreditador.
- Tipo de Canjes Anteriores

7.1.3. Aprendizajes e Hipótesis del Análisis Descriptivo de Tipo de de Canje

Aprendizajes

Se dedujo que en general el canje aéreo aumenta a medida que un cliente presenta mayor actividad transaccional, ya sea canje o acreditación. Además el canje no-aéreo no presenta diferencias importantes a nivel de transacciones, excepto en el propio canje no-aéreo en donde una mayor actividad transaccional previa sí implica un aumento en la propensión de canje.

Se encontró que los clientes utilizan el CNA como una forma de liquidar sus KMS que están cercanos a expirar.

Otro descubrimiento importante es que los clientes mientras más rápido se frustran en el proceso online de CA, más propensos son a realizar un CNA.

Finalmente observamos las variables codificadas como “Total de KMS” suelen ser más informativas que sus similares “N° de Transacciones” y “Máximo de KMS por transacción”.

Hipótesis

La primera hipótesis es que los modelos en general tenderán a clasificar a los clientes como canjeadores aéreos o no canjeadores, dado esto, será necesario utilizar criterios de ordenamiento por sobre criterios de clasificación al estudiar el poder predictivo de los modelos. Así podremos cumplir el objetivo de esta investigación, identificar a los clientes propensos al Canje No Aéreo.

La segunda hipótesis que se desprende de este análisis es que las variables demográficas son “generadoras” de las variables transaccionales, es decir, que los clientes que cumplen con cierto perfil tendrán un comportamiento transaccional similar al de sus pares en el largo plazo. Esto es relevante pues se podrá predecir el comportamiento de los clientes nuevos en base a su perfil demográfico.

7.2. Análisis Descriptivo de Categoría

Dentro de los canjeadores No-Aéreos nos interesará saber qué motiva la elección de de las distintas categorías disponibles. A continuación mostraremos las variables más relevantes que influyen en la decisión de categoría.

7.2.1. Estudio de categoría según variables demográficas

A continuación se mostrará en qué proporción distribuyen los clientes sus preferencias de categoría de acuerdo a las variables demográficas vistas anteriormente

Edad

La preferencia por canjear GiftCards encuentra su mayor nivel en el rango más joven, decreciendo a medida que los segmentos envejecen. Este efecto es inverso para la elección de “Otras Categorías”, es decir, los clientes de mayor edad son los que más canjean y los jóvenes los que menos.

Se observa que la preferencia por Tecnología se concentra en los rangos etarios medios. Esto se interpreta como una incapacidad de los clientes más jóvenes de canjear estos productos debido a su alto valor y a un desinterés de los clientes de mayor edad por esta categoría de producto.

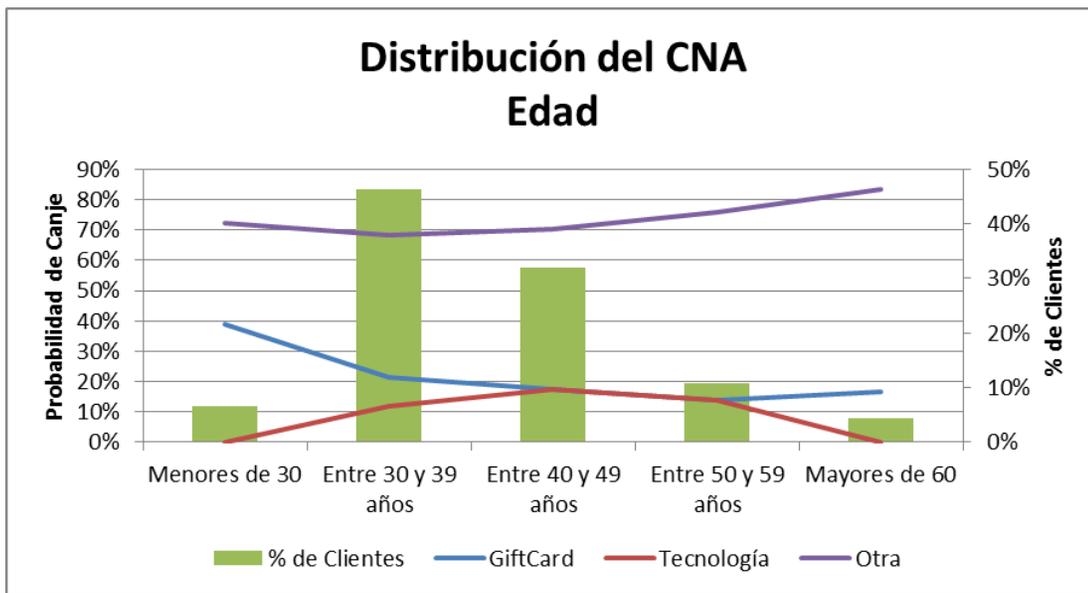


Ilustración 25. Distribución del CNA vs Edad.

Género

En el caso del género se graficó de forma diferente la preferencia. El objetivo es ver como se comportan ambos géneros respecto a la media de la muestra.

Se obtienen dos insights de esta variable:

- Los hombres prefieren canjear en Tecnología por sobre las demás categorías.
- Las mujeres son menos propensas al canje en Tecnología, en cambio, prefieren canjear en Giftcards o en Otras Categorías.

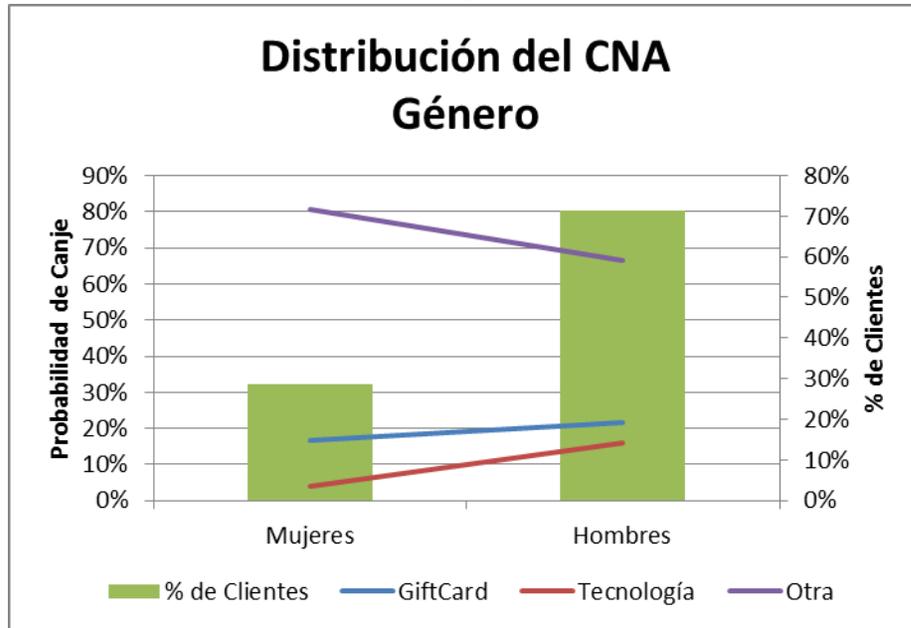


Ilustración 26. Distribución del CNA Según Género.

Mejor Tarjeta de Crédito

En el caso de la mejor Tarjeta de Crédito solo hay un efecto lineal entre una mejoría en la Tarjeta de Crédito y la disminución de la preferencia por el canje de “Otras Categorías”.

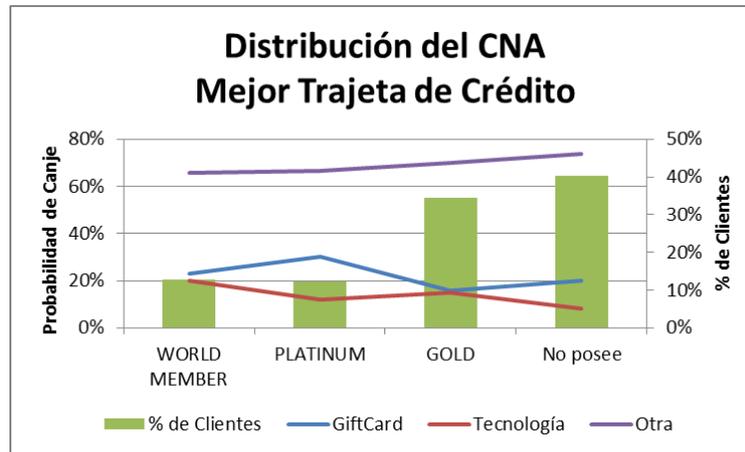


Ilustración 27. Distribución del CNA Según Mejor Tarjeta de Crédito.

Categoría de Socio

Para la Categoría de Socio se observa que la categoría más exclusiva, COMODORO/BLACK, no muestra interés por canjear GiftCards pero muestra el interés más alto por productos de Tecnología dentro de este análisis. Es interesante también ver que los clientes de la categoría más baja, LANPASS, tienen una baja preferencia por Giftcards. Esto se debe a que dado que estos clientes acreditan poco, prefieren ahorrar para un premio más grande, que gastar sus ahorros en una Giftcard, que es más asequible.

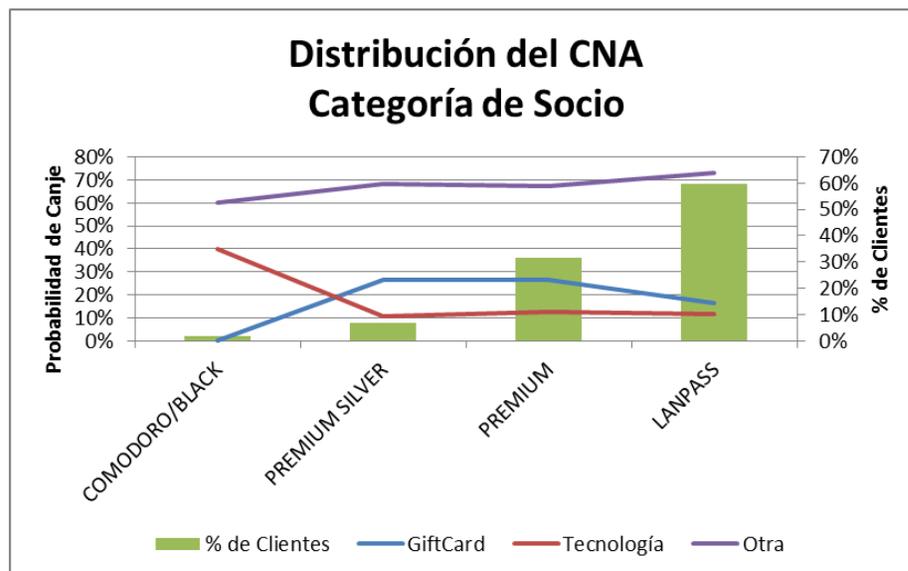


Ilustración 28. Distribución del CNA Según Categoría de Socio.

7.3. Estudio de categoría según variables transaccionales

Acreditación Promedio Anual

Esta variable no muestra un comportamiento lineal en sus preferencias a través de los rangos de acreditación. Sin embargo, se observa que las preferencias por GiftCard son más altas en los rangos bajos y altos de acreditación, mientras que “Otras Categorías” tiene una alta preferencia en los rangos muy bajos y medios de acreditación. Para el caso de Tecnología se observa un ligero aumento en la preferencia a medida que los clientes acreditan más.

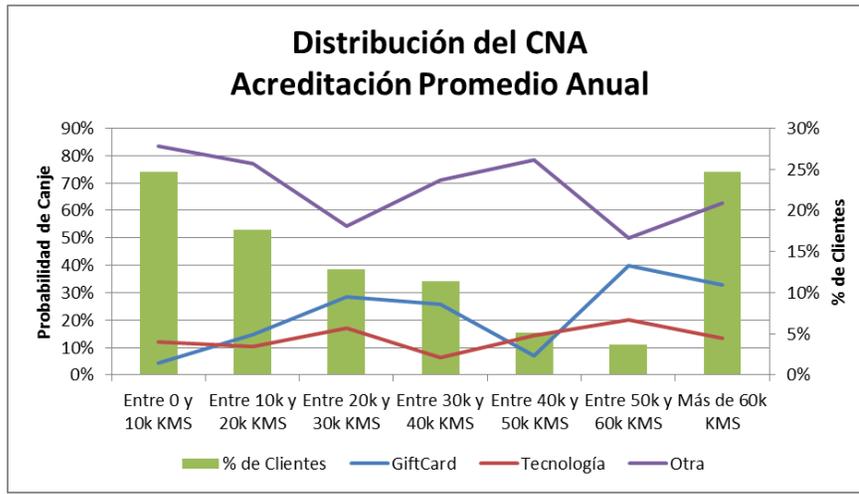


Ilustración 29. Distribución del CNA Según Acreditación Promedio Anual.

Saldo

En el caso del saldo se puede observar que las preferencias por Giftcard y Tecnología aumentan a medida que los clientes tienen más saldo, en desmedro de su preferencia por “Otras Categorías”.

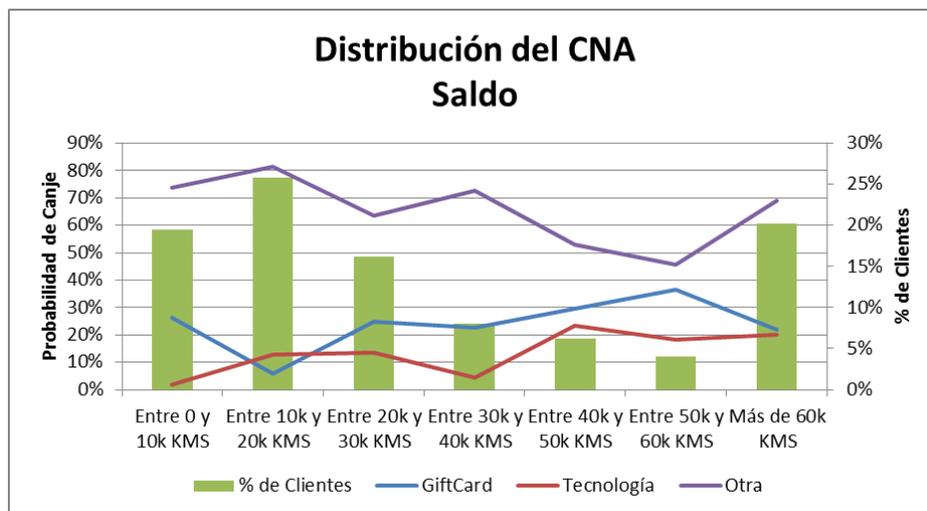


Ilustración 30. Distribución del CNA Según Saldo.

Tipo de Canjeador

Los distintos tipos de canjeadores muestran una preferencia similar por Tecnología. Es interesante ver que los canjeadores tipo “Aéreo” tienen la mayor preferencia por “Otras Categorías”, esto nos dice que tienen las preferencias más variadas de canje dentro del CNA.

Los clientes tipo Mix presentan la mayor preferencia al canje de GiftCard.

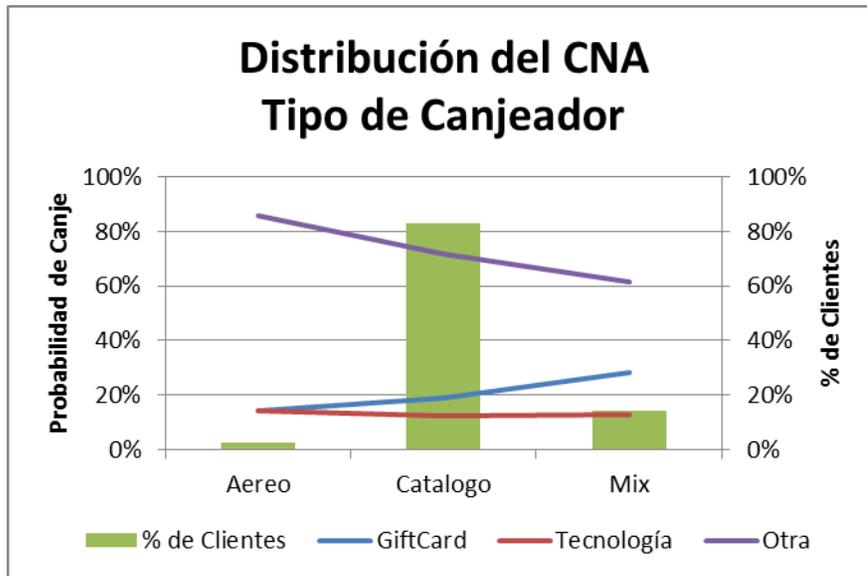


Ilustración 31. Distribución del CNA Según Tipo de Canjeador.

Canjes Anteriores

Cuando vemos las preferencias de los clientes, según las categorías que eligieron en compras anteriores, podemos observar que a medida que más canjearon los clientes en Tecnología y GiftCard, es más probable que vuelvan a repetir su elección.

Esto también es cierto para los clientes que canjearon en otra categoría, pero la tendencia no es lineal. No es razonable concluir respecto a esta “anomalía”, pues la categoría es muy surtida.

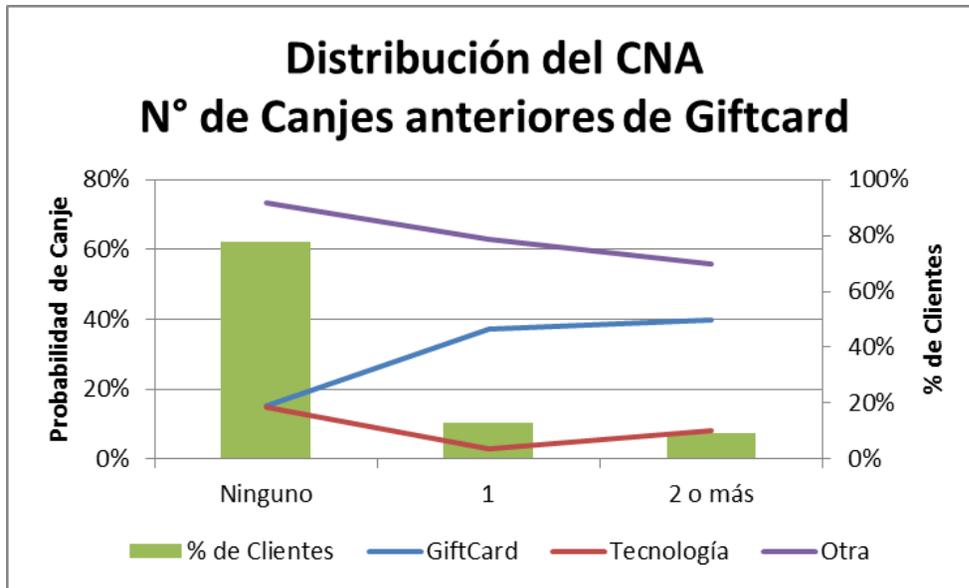


Ilustración 32. Distribución del CNA Según Nº de Canjes Anteriores de Giftcards.

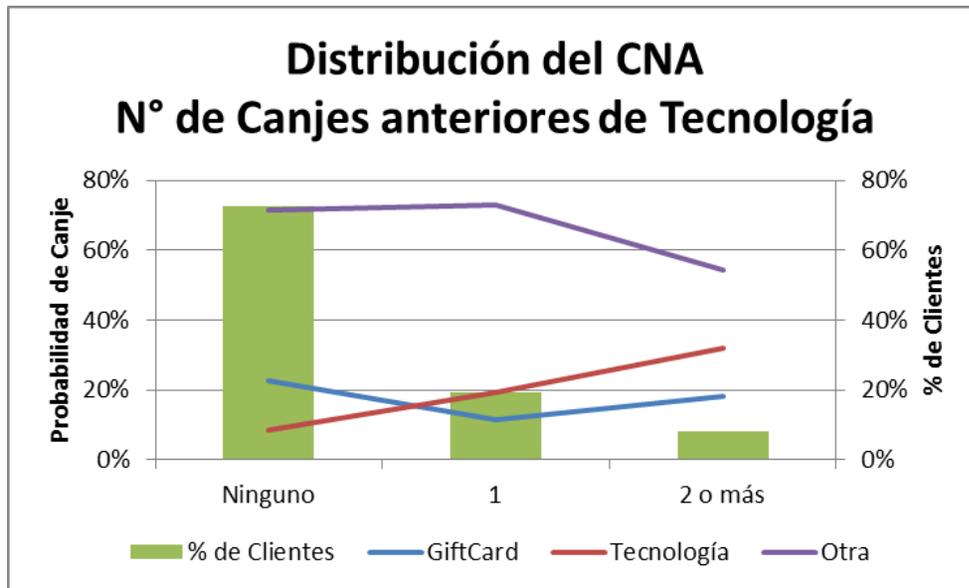


Ilustración 33. Distribución del CNA Según Nº de Canjes Anteriores de Tecnología.

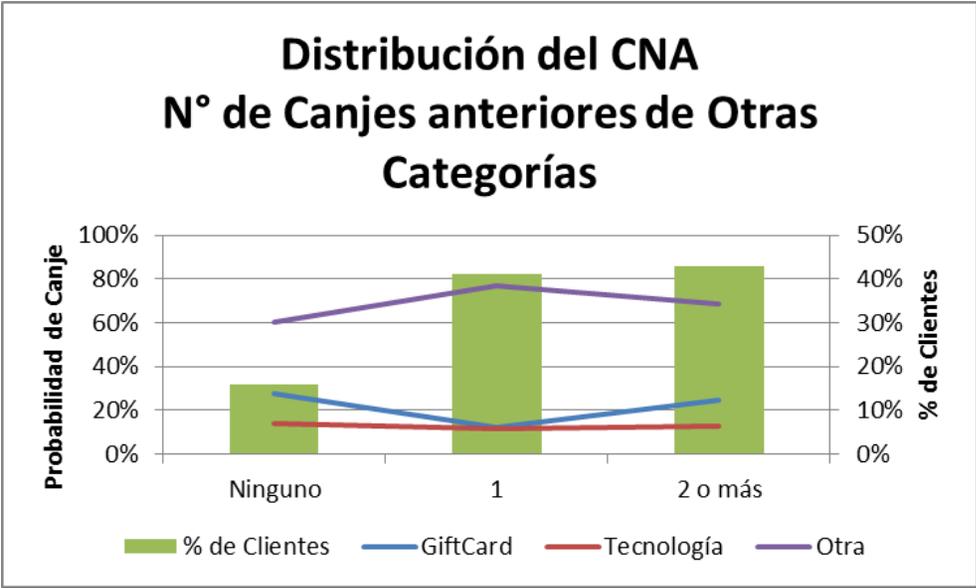


Ilustración 34. Distribución del CNA Según N° de Canjes Anteriores de Otras Categorías.

7.3.1. Aprendizajes e Hipótesis del Análisis Descriptivo de Categoría

Aprendizajes

La preferencia por las categorías depende mucho del nivel de acreditación, en general, los clientes con mayor poder adquisitivo tendrán una mayor preferencia por Tecnología, dado que es una categoría más cara, mientras que los con menos poder preferirán GiftCard u “Otras Categorías”.

No es sensato concluir respecto a “Otras Categorías” pues esta clasificación es muy ambigua, sin embargo, su presencia nos permite observar con mayor precisión lo que pasa con las otras dos categorías, que son de mayor interés.

Hipótesis

La hipótesis principal de este análisis es que los canjes anteriores en cada categoría sean la variable más influyente en la elección de categoría para la construcción de un modelo de propensión, siendo posible que estas sean las únicas variables significativas.

Otra hipótesis es que existe una canibalización por parte de las Giftcards de multitienda hacia las otras dos categorías, dado que los clientes podrán comprar productos de cualquier categoría. Lamentablemente no existe información respecto a lo que hacen los clientes con las Giftcards que adquieren mediante CNA, por lo cual solo podremos ver si es que la canibalización existe, pero no sabremos en qué medida.

Una tercera hipótesis es que la elección de canje depende de la oferta y el valor en KMS de cada producto en el momento de la compra, sin embargo, esto sale de los alcances de este trabajo.

7.4. Poder Predictivo de Variables

Para finalizar el análisis descriptivo veremos el poder predictivo de las variables. Esta herramienta nos dirá que tan informativa es una variable respecto a la propensión de canje respecto a que no utilizáramos ninguna variable para predecir el canje. El poder predictivo corresponde a la suma del valor absoluto de las diferencias de *odds* que cada categoría presenta en una variable respecto a la propensión promedio del total de la muestra.

A continuación se muestra el poder predictivo de las variables tanto para el Canje No Aéreo como para el Canje Aéreo.

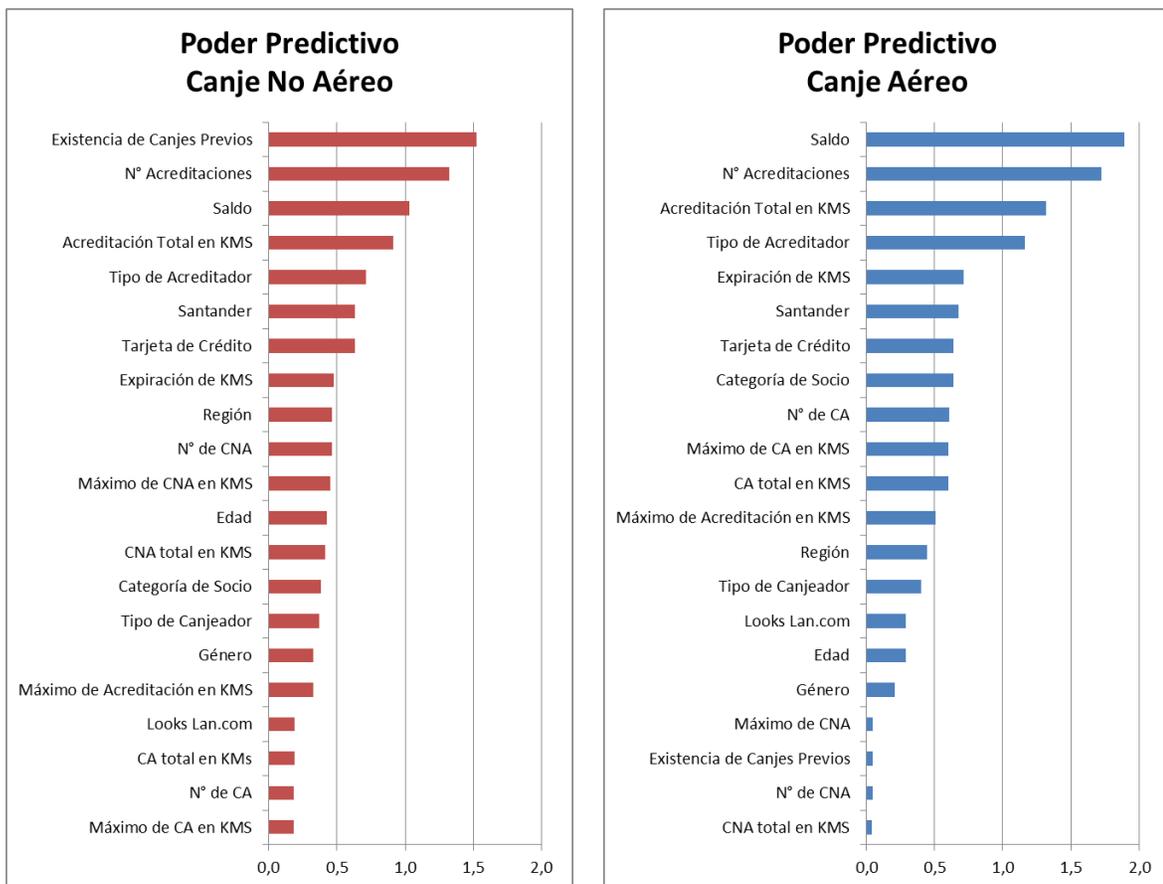


Ilustración 35. Poder Predictivo de las Variables para los casos de CNA y CA.

Como era de esperarse, las variables que mejor predicen el canje son el Saldo y la Acreditación, las cuales, como vimos previamente, están fuertemente relacionadas.

Además se repite lo observado en el análisis descriptivo bivariado; el CA es un buen predictor de CA, el CNA es un buen predictor de CNA y que el máximo de una variable es en general un peor predictor de canje que sus complementos “*Total*” y “*N° de Transacciones*”.

El resultado más interesante es que la existencia de canjes previos, de cualquier tipo, son el mejor predictor de CNA. Esto quiere decir que el CNA tiene una alta preferencia entre los clientes que ya están familiarizados con la oferta del catálogo LANPASS.

8. Modelos de Clasificación

Esta fase de la investigación tiene como objetivo utilizar las variables estudiadas en el análisis descriptivo, de manera conjunta, para obtener la probabilidad de que un cliente realice un canje. De esta forma se podrán conocer a los clientes más propensos a realizar Canje No aéreo y de qué forma son caracterizados por las variables utilizadas.

Se utilizarán dos herramientas para obtener las probabilidades de canje, estas son la **Regresión Logística Multinomial** y **los Árboles de clasificación**.

8.1. Primera Decisión: Canje Aéreo, Canje No Aéreo o No Canje

Para representar esta decisión usaremos tres modelos, dos con regresiones logísticas multinomiales y uno con árboles de clasificación.

Al primer Logit lo llamaremos **modelo de corto plazo** porque utilizará la información transaccional reciente de los clientes para realizar las predicciones.

Por su parte, al segundo Logit lo denominaremos **modelo de largo plazo** pues incluirá a clientes que aún no tienen historial de transacciones, pero que comparten características demográficas con los que sí tienen. De esta forma esperamos que los clientes nuevos se vayan asimilando a sus pares demográficos a medida que vayan creando un historial transaccional.

La elección de dos modelos se justifica en que las variables transaccionales excluyen a las demográficas a la hora de realizar una regresión. Para verificar si existe alguna interacción importante entre estas variables utilizaremos árboles de clasificación.

8.1.1. Modelos de Regresión Logística

La regresión logística nos dirá que variables y en qué magnitud influyen en la probabilidad de canje. Mediante la experimentación con distintos conjuntos de variables se buscará que los modelos cumplan 3 condiciones fundamentales.

Condición 1: Ajusten lo mejor posible y sus parámetros sean significativos.

Condición 2: Que los parámetros tengan una interpretación con sentido comercial.

Condición 3: Entreguen la mejor curva de ganancia de información.

8.1.1.1. Modelo de corto plazo

Recordemos que el modelo de corto plazo considera variables únicamente transaccionales

Se experimentó con distintas variables, de los modelos que cumplieron los dos primeros criterios, se eligió el que entregaba la mejor curva de información.

Las variables utilizadas en este primer modelo son:

- **Acreditación Total (Logaritmo):** Como sugiere la intuición, y confirman los análisis descriptivos, es más probable que un cliente realice un canje si acredita más. Se utilizó logaritmo para normalizar la distribución exponencial de los montos de acreditación.
- **Saldo (Logaritmo):** Es la variable crucial en el canje, un saldo insuficiente se traduce en imposibilidad de canje, por lo que es razonable utilizarla en este modelo. Como se confirmó en la sección anterior, es efectivamente un buen predictor de ambos tipos de canje. Por los mismos motivos que la Acreditación Total se trabaja en logaritmo y el ajuste del modelo también mejora.
- **Número de CNA previos:** Se observó en la sección de poder predictivo que esta variable influye positivamente en la propensión de CNA y negativamente en la de CA. En particular tiene de los poderes predictivo más altos de CNA, que es el núcleo de este estudio.

A continuación veremos que el modelo construido con estas variables cumple los 3 criterios:

Condición 1

Pseudo R-Square	
Cox and Snell	,027
Nagelkerke	,195
McFadden	,184

Ilustración 36. Pseudo R-Cuadrado Modelos Corto Plazo.

Dentro de los modelos probados el modelo elegido presenta el mayor Pseudo R-cuadrado de McFadden. Un modelo excelente tendrá valores entre 0,2 y 0,4 por lo cual calificaremos este modelo como muy bueno.

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	44285,975	18747,249	2	0,000
Acreditación Anual (Log)	26225,915	687,188	2	,000
Saldo(Log)	26043,823	505,097	2	,000
N° de CNA	25854,878	316,152	2	,000

Ilustración 37. Resultado Test de Ratios de Verosimilitud del Modelo a Corto Plazo.

Por su lado, los parámetros muestran ser todos significativos, poseen un p-valor menor a 0.5. Veremos más detalles de los parámetros a continuación.

Condición 2

Como se observa en la tabla siguiente, para ambas decisiones las variables son significativas a nivel de elección, excepto el **N° de CNA** para la decisión de Canje Aéreo.

Los parámetros tienen sentido, pues:

- El saldo y la acreditación anual son positivos y con valores elevados, esto concuerda con que es necesario acreditar un saldo suficiente para poder realizar cualquier canje.
- El N° de CNA anteriores es positivo para la elección de CNA.
- El N° de CNA muestra ser negativo, sin embargo, su valor es muy bajo y no es significativo, por lo que asumiremos que no influye en la elección de CA.

Elección		B	Std. Error	Sig.
CNA	Intercept	-11,016	,273	0,000
	Acreditación Anual (Log)	2,651	,430	,000
	Saldo(Log)	5,970	,533	,000
	N° de CNA	,642	,034	,000
CA	Intercept	-12,452	,162	0,000
	Acreditación Anual (Log)	6,872	,335	,000
	Saldo(Log)	6,709	,338	,000
	N° de CNA	-,007	,056	,898

Ilustración 38. Resultados Regresión Logística del Modelo Corto Plazo.

La categoría de referencia es No Canjear.

En términos de magnitud observamos:

- El saldo es la variable más importante en la decisión de CNA, teniendo más del doble de importancia que la acreditación anual. Esto a priori resulta trivial, sin embargo si comparamos con el caso de la elección de CA, vemos que para esta decisión el saldo y la acreditación son prácticamente iguales.
- El N° de CNA no tiene influencia alguna en la decisión de CA, pero una leve importancia en la elección de CNA, dando luces de que los canjeadores no aéreos tienden a repetir su comportamiento.

Condición 3

La curva de ganancia de información tiene una forma sinusoidal. Esto se interpreta como una buena capacidad de identificar a los clientes más propensos al CNA pero mala para identificar al último 30%, siendo incluso peor que no utilizando ningún método de clasificación.

De hecho, este modelo puede identificar fácilmente al 50% de los canjeadores en el primer 10% de los clientes más propensos que arrojó el modelo.

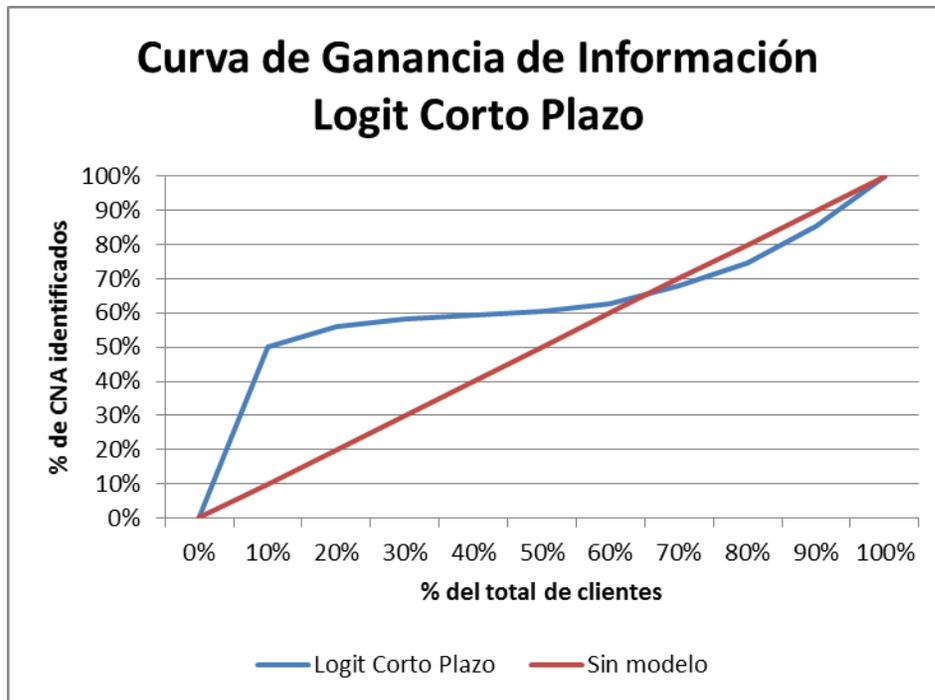


Ilustración 39. Curva de Ganancia de Información Logit Corto Plazo.

El poder predictivo muestra ser bueno en el primer decir. Veamos cómo se comporta el modelo en este segmento.

La curva a este nivel de agregación muestra ser cóncava, esto quiere decir que el modelo predice muy bien en el primer centil y empieza a perder pendiente a medida que los clientes más propensos son identificados en los centiles anteriores.

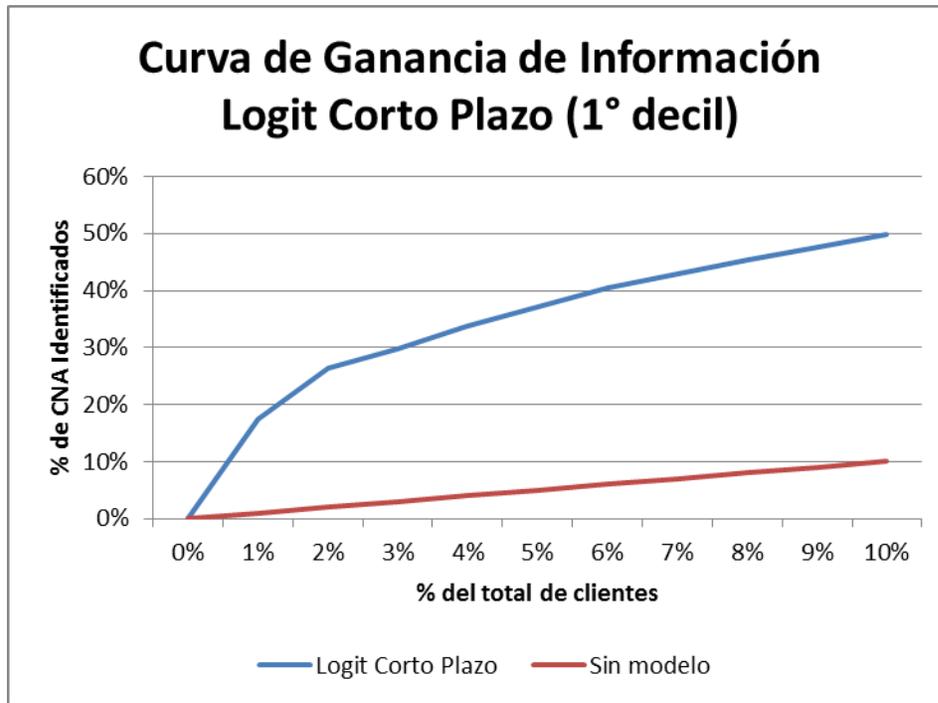


Ilustración 40 Curva de Ganancia de Información Logit Corto Plazo (1º decil).

8.1.1.2. Modelo de Largo Plazo

Recordemos que en este modelo usaremos las variables que fueron rechazadas por el modelo de corto plazo y veremos cuan buenas son identificando a los canjeadores no aéreos.

Las variables utilizadas en este modelo son:

- Mejor Tarjeta de crédito
- Género
- Categoría de Socio
- Edad

Veamos que este modelo también cumple con las 3 condiciones

Condición 1

En el ajuste vemos la primera divergencia entre este modelo y el previamente estudiado. El R-cuadrado de McFadden es de 0,1 lo cual no es ideal pues se aleja de los parámetros ideales de entre 0,2 y 0,4.

Pseudo R-Square	
Cox and Snell	,015
Nagelkerke	,111
McFadden	,104

Ilustración 41. Pseudo R-Cuadrado Modelos Largo Plazo.

Por otro lado, vemos que las variables son significativas a nivel individual y que en su test de reducción de verosimilitud, como era de esperar, las variables que tienen fuerte relación con el comportamiento transaccional presentan mejoras más poderosas en el modelo que la edad y el género.

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	4359,188 ^a	0,000	0	,000
Edad	4381,823	22,636	2	,000
Mejor Tarjeta	4821,268	462,080	6	,000
Género	4388,412	29,224	4	,000
Categoría de Socio	5340,710	981,522	6	,000

Ilustración 42. Resultados Test de Ratios de Verosimilitud del Modelo a Largo Plazo.

Condición 2

En la siguiente tabla se muestran las variables que a nivel de elección fueron significativas.

Respecto a la magnitud y sentido de estas variables se observa que:

- Los clientes más jóvenes y más viejos son los que preferirán el CNA en comparación al CA, como vimos en el análisis descriptivo la brecha de propensión favorecía el CA para los rangos etarios medios.

- Los que pertenecen a las dos peores categorías de socio tienden a ser más propensos al CNA. De hecho, los clientes de la peor categoría (LANPASS) tienen un beta 3 veces mayor a la categoría que le sigue (PREMIUM).
- La tarjeta de crédito presenta resultados interesantes pues todos los parámetros son negativos. Sin embargo, se puede destacar que el efecto es menos desfavorable para los clientes con Tarjeta Platinum, quienes tienen un poder adquisitivo medio. El efecto es más desfavorable para los clientes Gold, seguido de cerca por los clientes World Member. Esto tiene sentido pues los clientes Gold son el grueso de clientes comunes del banco, mientras que los World Member tienen el poder adquisitivo para adquirir los productos del catálogo en otros lugares y sacarle un mayor valor a sus puntos.

Elección		B	Std. Error	Sig.
CNA	Intercept	2,390	,104	,000
	Edad	-,352	,155	,024
	Mejor Tarjeta: World Member	-,776	,082	,000
	Mejor Tarjeta: Platinum	-,304	,095	,001
	Mejor Tarjeta: Gold	-,954	,061	,000
	Categoría: LANPASS	2,999	,099	,000
	Categoría: Premium	1,120	,096	,000
CA	Intercept	-1,107	,245	,000
	Edad	-1,673	,358	,000
	Mejor Tarjeta: World Member	-,838	,204	,000
	Mejor Tarjeta: Gold	,498	,116	,000
	Género: Femenino	-,388	,107	,000

La categoría de referencia es No Canjear.

Ilustración 43. Resultados Regresión Logística del Modelo Largo Plazo.

Condición 3

Respecto a la capacidad de identificar a los canjeadores, este modelo identifica a un 35% de los clientes en el primer decil, lo cual no es tan malo en comparación al 50% del modelo de corto plazo.

Una observación interesante es que el poder predictivo se comporta de manera más regular que en el modelo anterior, teniendo un poder predictivo aceptable incluso en el 50% de clientes menos propensos.

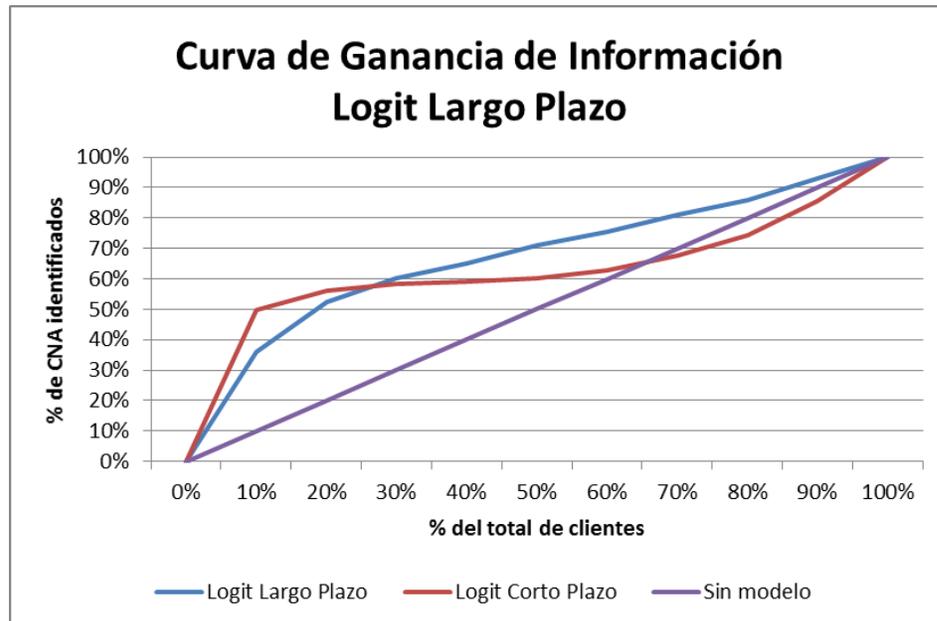


Ilustración 44. Curva de Ganancia de Información Logit Largo y Corto Plazo.

Explorando lo que pasa en el primer decil vemos que el poder predictivo nuevamente es menor que en el modelo previo.

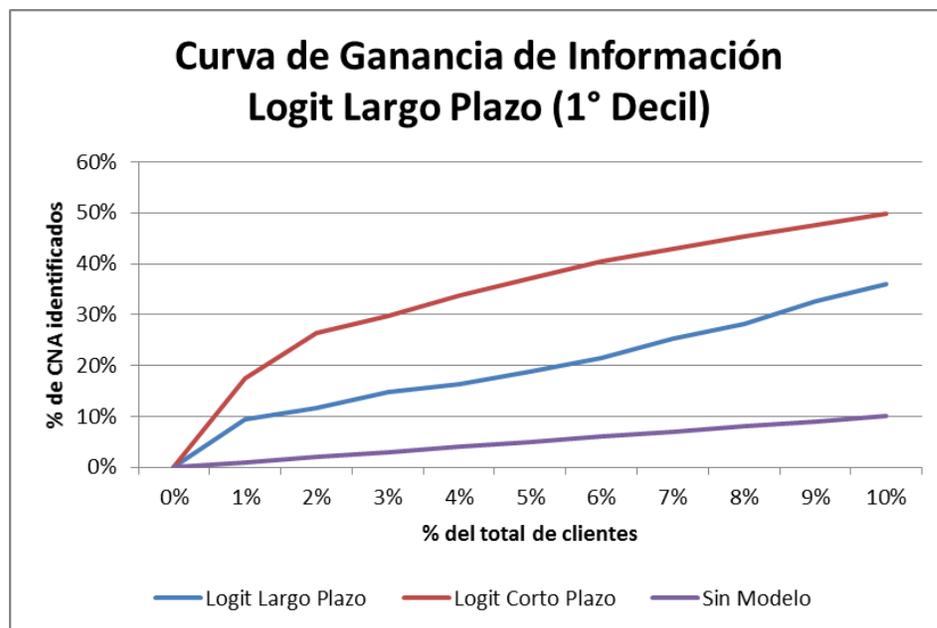


Ilustración 45. Curva de Ganancia de Información Logit Largo y Corto Plazo (1º decil).

Podemos concluir que el modelo de corto plazo tiene la virtud de predecir muy bien al 25% de los clientes más propensos pero que empieza a fallar después, siendo superado

por el modelo de largo plazo. No es recomendable utilizar este modelo si se utilizará para trabajar sobre más del 60% de los clientes.

Debido a la forma recta de la curva de ganancia del modelo de largo plazo, es posible que el modelo solo funcione en el primer 25% de clientes para luego clasificar a los clientes al azar, como sería el caso de no tener ningún modelo predictivo.

8.1.2. Árboles de clasificación

A diferencia de Logit, los árboles de clasificación nos permitirán utilizar todas las variables disponibles y la propia herramienta elegirá las que son relevantes, las clasificará y asignará probabilidades de canje a los datos. Esto no solo nos dará otra forma de identificar variables relevantes en la decisión de canje, sino que también nos dará reglas de decisión que serán útiles para entender la decisión de canje.

Las variables utilizadas en este árbol fueron:

- Edad.
- Tiempo desde última renovación.
- Categoría de Socio.
- Mejor Tarjeta de Crédito.
- Saldo.
- Acreditación durante el último año.
- KMS gastados en CA durante el último año
- KMS gastados en CNA durante el último año

El método de crecimiento escogido es CHAID exhaustivo, esto se debe a que el tiempo de cómputo adicional respecto a CHAID no presentó problema con el conjunto de datos utilizado. De esta forma tendremos una herramienta más confiable sin ningún costo relevante.

Además como se observa a continuación, los métodos de crecimiento CHAID son mejores identificando a los canjeadores que los métodos CRT y QUEST.

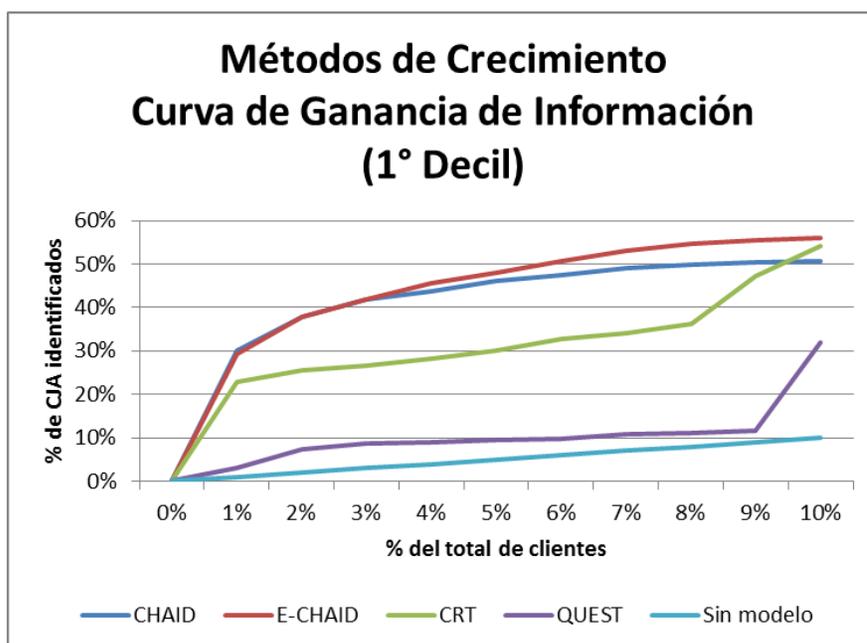


Ilustración 46. Métodos de Crecimiento: Curva de Ganancia de Información (1º Decil).

El árbol muestra resultados interesantes, pues los nodos con mayor concentración de canjeadores no aéreos nos entregan las reglas de decisión asociadas a esos clientes. A continuación se presentan las reglas más importantes de decisión encontradas, con las cuales se identifica al 72,4% de los clientes que realizaron CNA.

Regla	Porcentaje de Canjeadores	Precisión de la clasificación
Acreditaron más de 21.000 KMS el último año, realizaron CNA el último año y poseen saldo superior a 7.500 KMS	20,34%	9,0%
Acreditaron más de 21.000 KMS el último año, no han realizado CNA el último año y pertenecen a las categorías LANPASS o PREMIUM	16,04%	0,50%
Acreditaron entre 4.200 y 9.100 KMS el último año y poseen saldo superior a 7.500 KMS	14,74%	0,50%
Acreditaron entre 9.100 y 21.000 KMS el último año pero no canjearon nada durante ese año	12,31%	0,40%
Acreditaron entre 9.100 y 21.000 KMS el último año y realizaron algún CNA durante ese año	8,96%	3,80%
Otros Segmentos	27,61%	0,07%
Total	100,00%	0,03%

Ilustración 47. Segmentos según reglas de decisión Arbol CHAID-E.

Los segmentos con mejor precisión coincidieron ser los segmentos de clientes que realizaron CNA durante el último año. Los segmentos menos precisos aglutinan más del 40% de los canjeadores.

8.1.2.1. Curva de Ganancia de información

Respecto al total de la muestra, la ganancia de información de los árboles de clasificación llama la atención por su forma, en el primer 25% de clientes identifica la misma cantidad de clientes que el modelo de regresión logística de corto plazo, mientras que el 75% restante es identificado de la misma forma que el modelo de largo plazo.

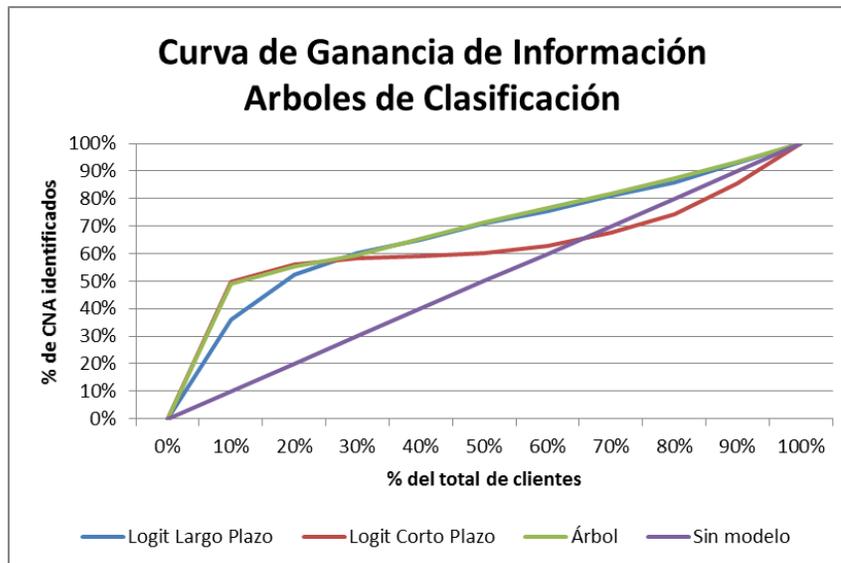


Ilustración 48. Curva de Ganancia de Información: CHAID-E.

Ya vimos que lo árboles de clasificación rinden de igual forma que el modelo de corto plazo en el primer cuarto de clientes. Para saber qué técnica es mejor para identificar a los clientes más propensos al CNA comparemos ambas técnicas en el primer decil.

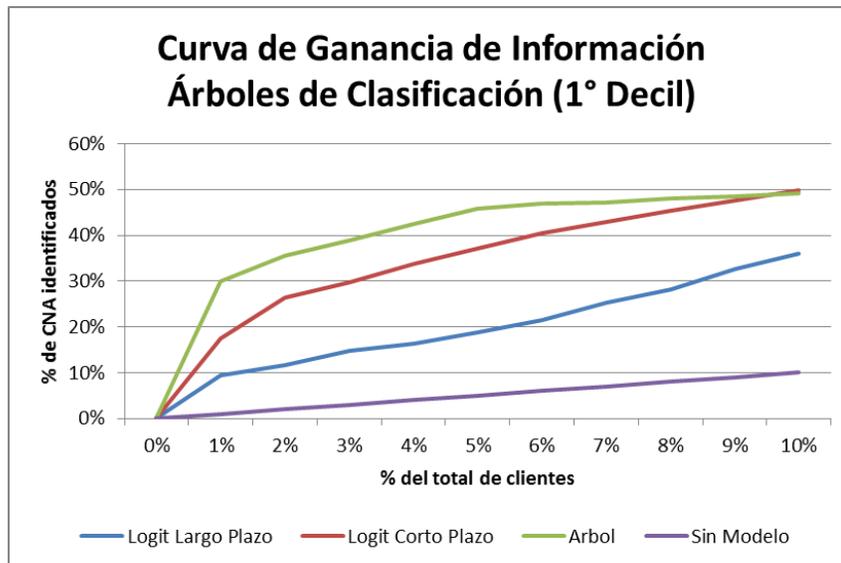


Ilustración 49. Curva de Ganancia de Información: CHAID-E (1º Decil).

Como se observa, en el primer centil, el árbol identificó el doble de canjeadores que el Logit de corto plazo. Esta diferencia es fundamental, pues en general, la capacidad predictiva en los primeros centiles es la más útil comercialmente pues es una garantía de que los clientes que pertenecen a dicho centil serán propensos al canje.

Concluimos que el árbol de clasificación es la mejor técnica de las utilizadas para identificar a los canjeadores de catálogo.

8.1.2.2. Validación del modelo

Para calibrar este modelo en particular, usaremos la técnica del 70-30. En términos simples, dividiremos la muestra en dos, una muestra con el 30% del total de los datos y otra con el 70%. Con esta última se calibrará un modelo y se evaluará si coincide con el modelo de la submuestra restante.

Como se observa, las curvas de ganancia de información son estructuralmente iguales, y más importante aún se comportan igual en el primer decil. Dado esto, consideraremos que el modelo es válido.

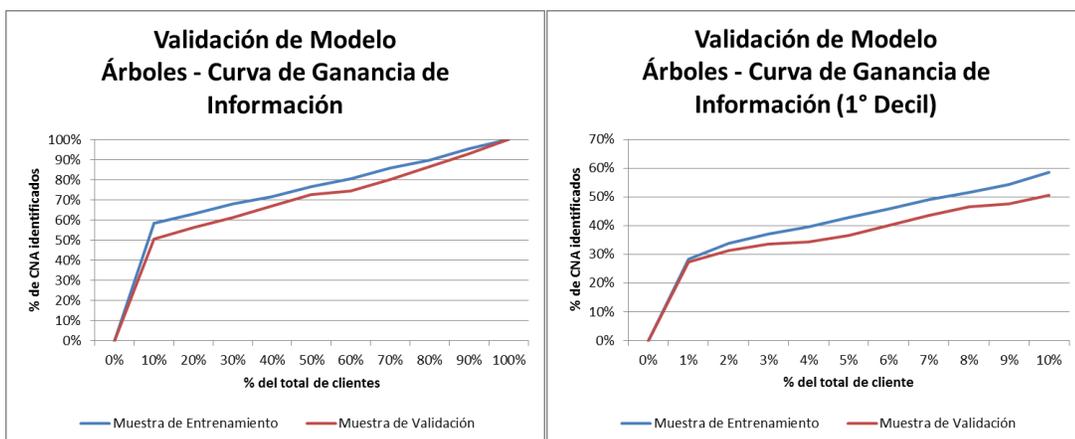


Ilustración 50. Validación del Modelo: Curvas de Ganancia de Información

8.1.3. Comparación de Modelos

A continuación compararemos los resultados obtenidos con las distintas herramientas de forma más integral. Para esto utilizaremos dos tipos de criterios; criterios duros basados en la perspectiva matemática de los modelos y criterios blandos basados en la perspectiva comercial del problema.

8.1.3.1. Criterios Duros

Matriz de Confusión

Las matrices de confusión no aportan mucho en este caso. Se observa que los modelos no son capaces de predecir la decisión de canje, sin embargo, esto se debe a que los modelos algorítmicamente buscan reducir el error, lo que decanta en clasificaciones triviales. Esto no presenta un problema pues ya vimos que los modelos tienen capacidad predictiva, pero deberemos utilizar otros criterios para comparar.

Clasificación: Logit Corto Plazo					Clasificación: Logit Largo Plazo					Clasificación: Arbol CHAID-Exhaustivo							
Observado	Predicho				Porcentaje Correcto	Observed	Predicho				Porcentaje Correcto	Observado	Predicho				Porcentaje Correcto
	No Canjea	CNA	CA				No Canjea	CNA	CA				No Canjea	CNA	CA		
No canjea	210173	42	0		100,0%	No canjea	210215	0	0		100,0%	No canjea	210215	0	0		100,0%
CNA	520	16	0		3,0%	CNA	536	0	0		0,0%	CNA	536	0	0		0,0%
CA	2130	1	0		0,0%	CA	2131	0	0		0,0%	CA	2131	0	0		0,0%
Porcentaje Total	100,0%	,0%	0,0%		98,7%	Porcentaje Total	100,0%	0,0%	0,0%		98,7%	Porcentaje Total	100,0%	0,0%	0,0%		98,7%

Ilustración 51. Resultados Matriz de Confusión.

Curva de Ganancia de Información

Según este criterio, el mejor modelo es el árbol de clasificación. Esto se debe a tres razones:

1. El árbol CHAID-E predice de mejor manera en el primer centil de clientes.
2. El árbol CHAID-E predice de mejor manera en el 25% inicial de clientes.

3. El árbol CHAID-E cubre las debilidades del Logit de corto plazo en el último 75% de clientes.

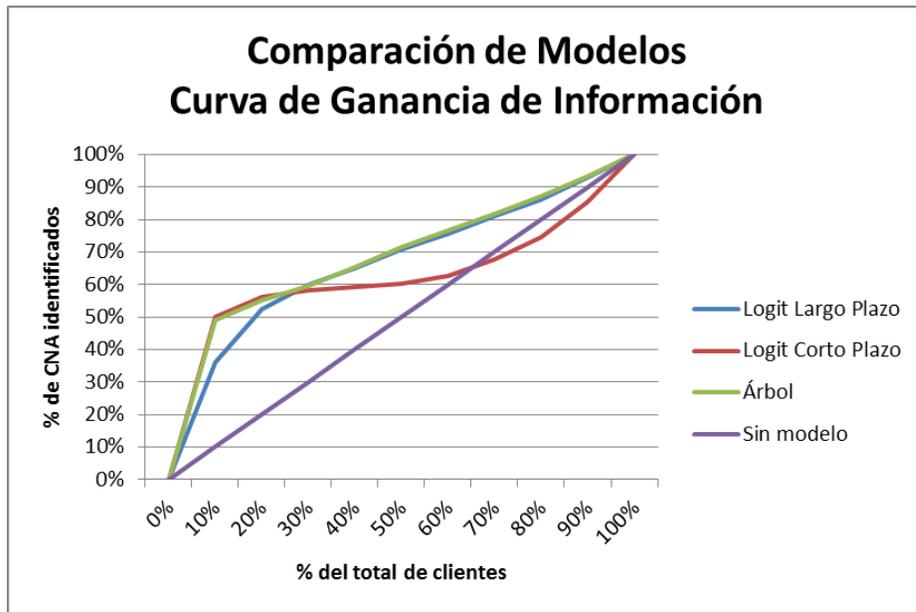


Ilustración 52. Comparación de Modelos: Curva de Ganancia de Información.

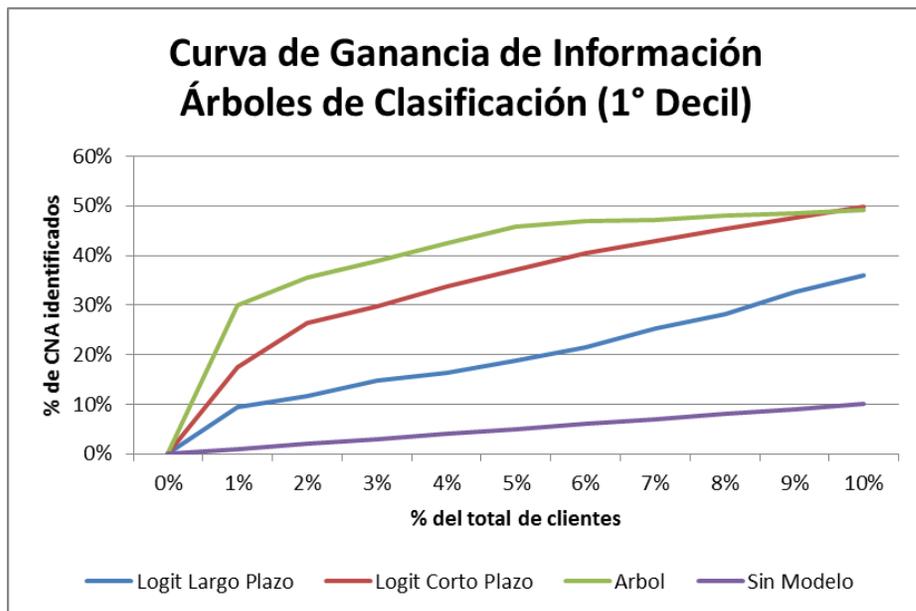


Ilustración 53. Curva de Ganancia de Información: Árboles de Clasificación (1º Decil).

8.1.3.2. Criterios Blandos

Los criterios blandos se basan en la utilidad que le podemos dar a los modelos utilizados.

Modelo de Regresión Logística:

- Permiten conocer en qué magnitud las variables influyen la decisión de canje

Modelo Árbol CHAID-Exhaustivo:

- Permite identificar clusters de clientes
- Muestra reglas de decisión útiles y aplicables
- Su uso es simple y ayudará a identificar nuevos patrones de decisión a medida que varía la oferta y la demanda del catálogo con el tiempo

8.1.3.3. Veredicto

Las dos técnicas utilizadas para modelar la elección de canje tienen fortalezas y debilidades. Por esta razón no es sensato preferir una, sino aprovechar el conocimiento que ambas entregan.

Dicho esto se concluye lo siguiente:

- Para obtener la mejor predicción de canje no aéreo, es recomendable utilizar árboles de clasificación.
- Para entender las reglas de decisión que siguen los clientes es útil utilizar árboles de clasificación.
- Para conocer en qué medida las variables afectan la decisión de canje, es recomendable utilizar regresiones logísticas con variables transaccionales.

8.2. Segunda Decisión: Elección de categoría

Ya habiendo identificado a los clientes más propensos al CNA, veremos cómo distribuyen su elección a través de las 3 categorías que enmarca el alcance de esta investigación: Tecnología, GiftCard de Multitienda y “Otras Categorías”.

Debido a la cantidad de reducida de datos, no es posible utilizar arboles de clasificación ni técnicas complejas de regresión logística, como el Logit anidado.

Utilizaremos entonces la regresión logística multinomial como lo hicimos con los modelos de corto y largo plazo.

8.2.1. Regresión Logística

Siguiendo la metodología del proyecto, se experimentó con todas las variables disponibles previamente más las variables asociadas al historial transaccional de los clientes dentro del catálogo, estas son:

- N° de canjes previos en cada una de las categorías.
- KMS canjeados en cada una de las categorías

No fue posible calibrar modelos significativos con las variables previamente vistas, ni con mezclas de estas con las variables asociadas al canje de catálogo. Sin embargo, se pudo calibrar un modelo de regresión logística multinomial utilizando las variables de catálogo.

El mejor modelo encontrado utiliza 3 variables:

- N° de canjes previos en Tecnología.
- N° de canjes previos en GiftCard.
- N° de canjes previos en Otras categorías.

Condición 1

En cuanto al ajuste se observa que el modelo tiene un Pseudo R-cuadrado de McFadden de 0,074. Esto está muy por debajo de lo deseado, pero es lo más alto que se encontró dado el contexto.

Pseudo R-cuadrado		Likelihood Ratio Tests				
		Effect	Model Fitting Criteria	Likelihood Ratio Tests		
			-2 Log of Likelihood of Reduced Model	Chi-square	df	Sig.
Cox y Snell	,123	Intercept	558,914	196,740	2	,000
Nagelkerke	,148	N° de Canjes previos en "Otras"	444,150	81,976	2	,000
		N° de Canjes previos en GC	392,269	30,095	2	,000
McFadden	,074	N° de Canjes previos en Tecnología	369,378	7,204	2	,027

Ilustración 54. Tablas de Comparación de Modelos.

En cuanto a la significancia individual de las variables todas cumplen con tener un p-valor menor a 0,05 lo que las hace significativas. Todas inducen una mejora al modelo, esto tiene sentido y concuerda con lo visto en la decisión de tipo de canje, donde vimos que el CNA pasado era un predictor significativo del CNA futuro.

Condición 2

En cuanto a los parámetros individuales podemos ver que la elección previa de una categoría predice bien la elección futura de la misma categoría.

Además, es posible observar que las GiftCards de Multitienda y la Tecnología son preferencias “opuestas” de canje, pues para cada categoría, un canje anterior de la otra categoría reduce la probabilidad de canjear en la categoría en cuestión. Por ejemplo, si alguien canjea una GiftCard es menos probable que en el futuro canjee en Tecnología.

Elección		B	Std. Error	Sig.
GC	Intersección	-1,274	,113	,000
	N° de Canjes previos en "Otras"	-,952	,929	,306
	N° de Canjes previos en Tecnología	-1,805	,919	,050
	N° de Canjes previos en GC	4,516	,720	,000
Tecnología	Intersección	-1,393	,137	,000
	N° de Canjes previos en "Otras"	-3,638	1,532	,018
	N° de Canjes previos en Tecnología	3,256	,695	,000
	N° de Canjes previos en GC	-11,204	3,180	,000

Ilustración 55. Resultados Árbol: Parámetros Individuales.

Condición 3

Finalmente, en las curvas de ganancia de información vemos que el modelo no es muy bueno para predecir la elección de GiftCards o de “Otra Categoría”. Esto tiene sentido pues en la práctica el cliente puede utilizar la Giftcard para canjear algún sustituto de algún producto del catálogo que pertenezca a “Otras categorías”.

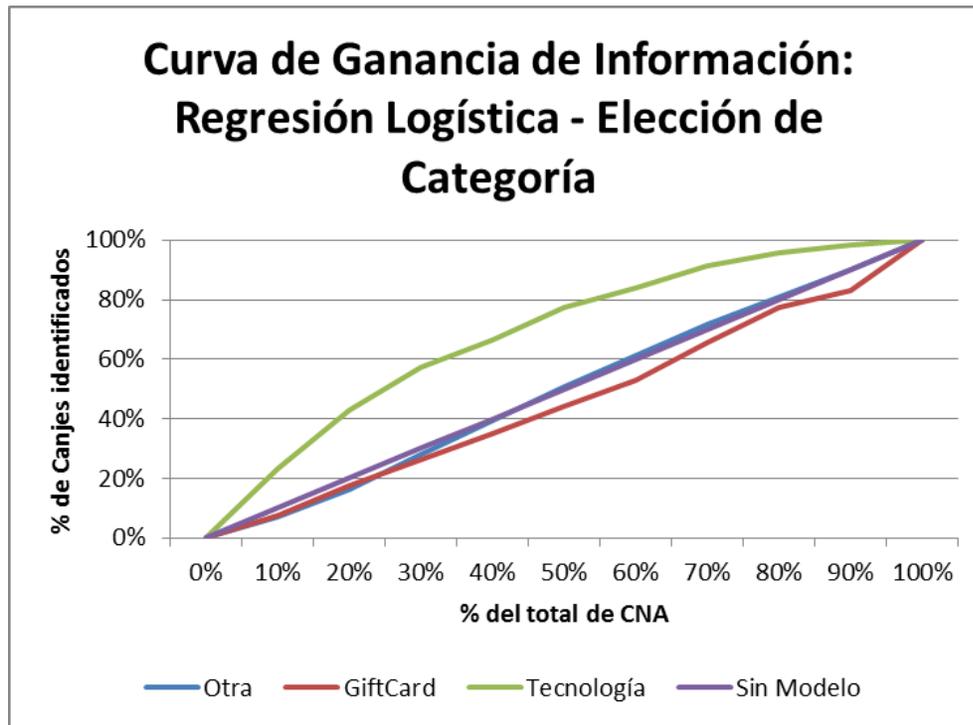


Ilustración 56. Curva de Ganancia de Información: Regresión Logística - Elección de Categoría.

En el caso de tecnología es comprensible que la “fidelidad” a la categoría muestre ser un buen predictor, esto se explica en que Tecnología presenta un canje mayoritario de hombres, los cuales presentan en general una mayor fidelidad de compra. Sin embargo, esta tendencia no se repite para las categorías restantes.

Si cambiamos el criterio de ordenamiento de “Propensión al canje de GC/Otra de manera decreciente” a “Propensión al canje de Tecnología de manera creciente” y dibujamos nuevamente las curvas, nos damos cuenta que se gana información respecto al canje de GC, mientras que “Otras Categorías” sigue sin entregar información alguna.

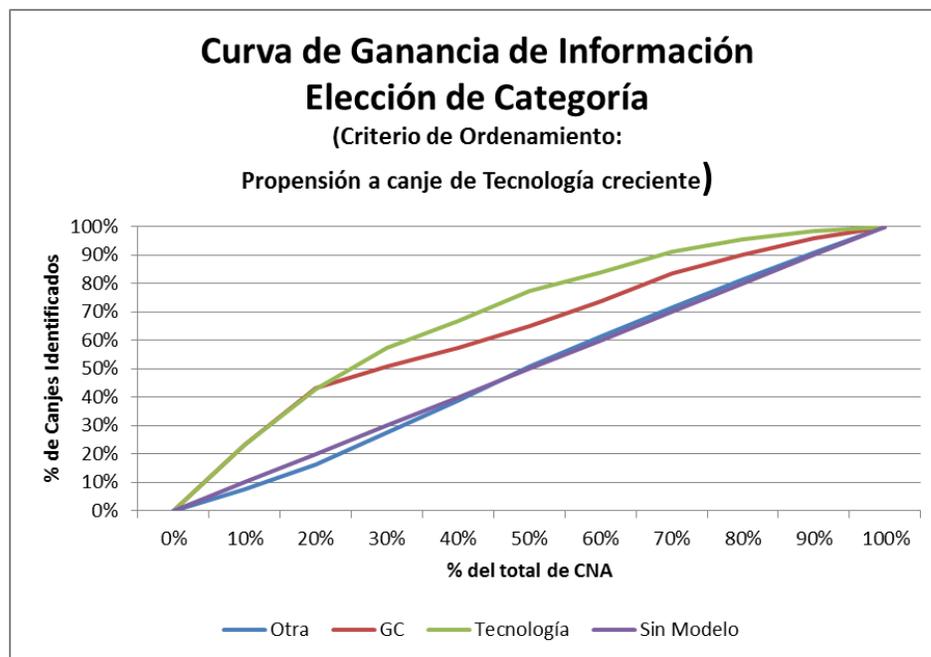


Ilustración 57. Curva de Ganancia de Información: Elección de Categoría.

De esta forma podemos concluir que los clientes que canjean GiftCards y Tecnología son distintos. El cliente que canjea GiftCard es averso al canje de Tecnología, mientras que el cliente que canjea Tecnología persiste en su preferencia a dicha categoría.

Validación

Como ya comentamos, el conjunto de datos es pequeño, por lo que no es sensato validar la información utilizando submuestras del conjunto de dato. Una alternativa razonable es comparar las curvas con las que obtenemos al calibrar el modelo con una de un mes adyacente al mes con el cual se realizó el modelo.

El mes de elección fue el mes de mayo, julio se descartó debido a que la oferta del catálogo fue especial durante ese mes, lo que podría sesgar los datos y llevarnos a conclusiones erróneas.

Podemos ver claramente que las curvas poseen la misma forma para las 3 categorías. La categoría Tecnología muestra el mismo poder predictivo en los primeros deciles; 20% en el primero y 40% en el segundo.

Concluimos que el modelo es consistente.

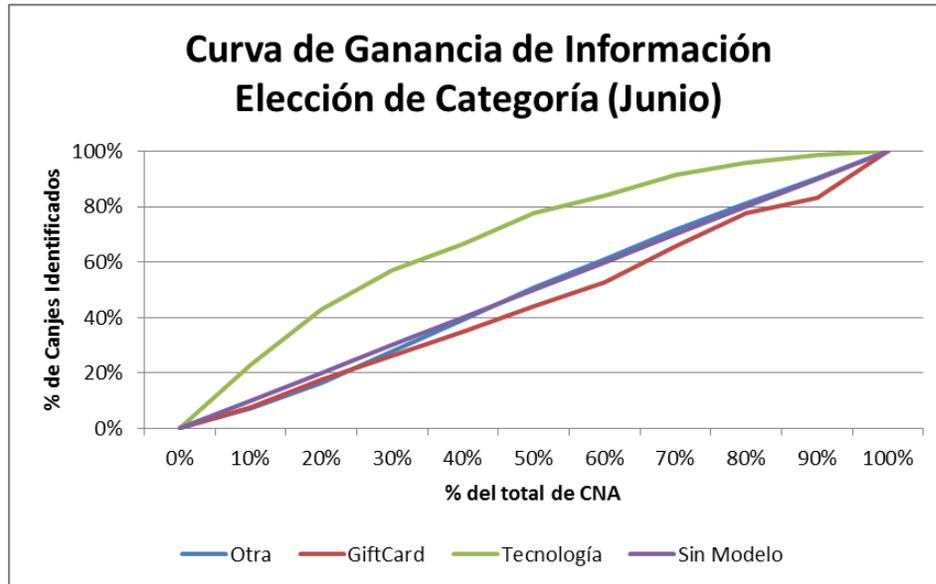


Ilustración 58. Curva de Ganancia de Información: Elección de Categoría (Junio).

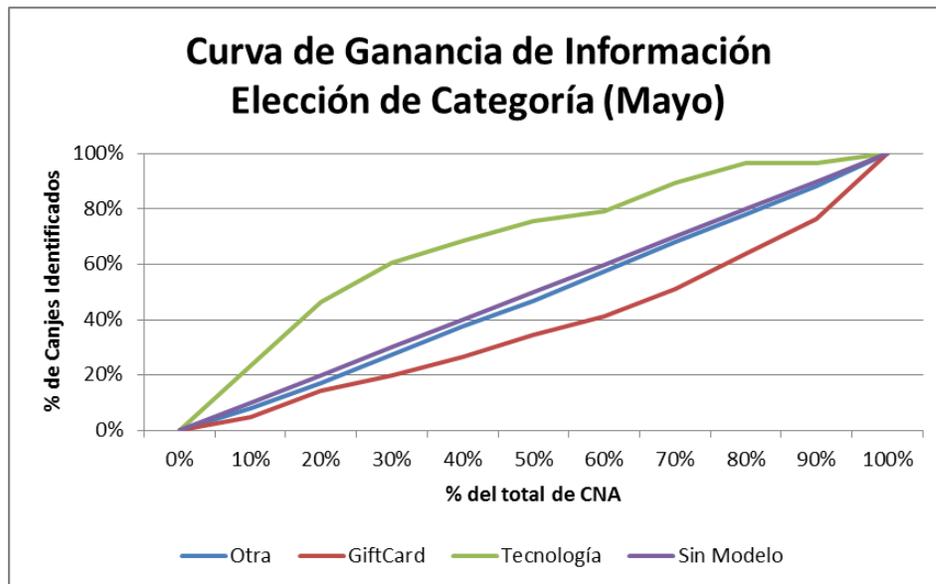


Ilustración 59. Curva de Ganancia de Información: Elección de Categoría (Mayo).

Veredicto

Según vimos, los clientes más propensos a canjear en Tecnología son los clientes que ya canjearon previamente en esa categoría.

Por otro lado los clientes que canjean Giftcard no persisten en su preferencia, pero podemos afirmar que tienen aversión a canjear en Tecnología.

En el caso de “Otras Categorías” era esperable que el modelo no entregara ganancia alguna de información debido a que aglutina 6 subcategorías.

8.2.2. Valoración de los modelos

Para saber cuánto pueden aportar los modelos construidos en esta memoria se simulará la respuesta de los clientes a una eventual campaña de e-mail marketing.

Se pondrán a prueba los modelos a prueba de acuerdo a 3 niveles de agregación; primer centil, quinto centil y primer decil. Es decir el 1%, 5% y 10% del total de clientes respectivamente.

En cada uno de estos niveles se verá el porcentaje incremental de KMS que tendría dicha campaña si es que se compara con algunos criterios básicos de ordenamiento.

8.2.2.1. Primera decisión: Canje Aéreo, Canje No Aéreo o No Canje

Para ambos modelos de regresión logística y el de árboles de clasificación se comparó con los siguientes criterios de ordenamiento:

1. Saldo, de mayor a menor.
2. Acreditación Total en el último año, de mayor a menor.
3. Acreditación en alianzas financieras en el último año, de mayor a menor.
4. Acreditación en alianzas comerciales en el último año, de mayor a menor.
5. Acreditación por vuelo en el último año, de mayor a menor.
6. Total de Canje Aéreo en el último año, de mayor a menor.
7. Total de Canje No Aéreo en el último año, de mayor a menor.
8. Distancia a la última renovación de KMS, de mayor a menor.
9. Orden al azar.

Bajo los criterios más débiles se observa que todos los modelos calibrados presentan una mejoría importante. Es de particular interés ver que los 3 modelos vencen al criterio de los clientes con mayor saldo, pues es un criterio que puede parecer poderoso pero que se verifica que no lo es. De hecho, la cercanía a caducidad de KMS y la acreditación por vuelo y alianza comercial tiene poderes predictivos bastante cercanos a los modelos.

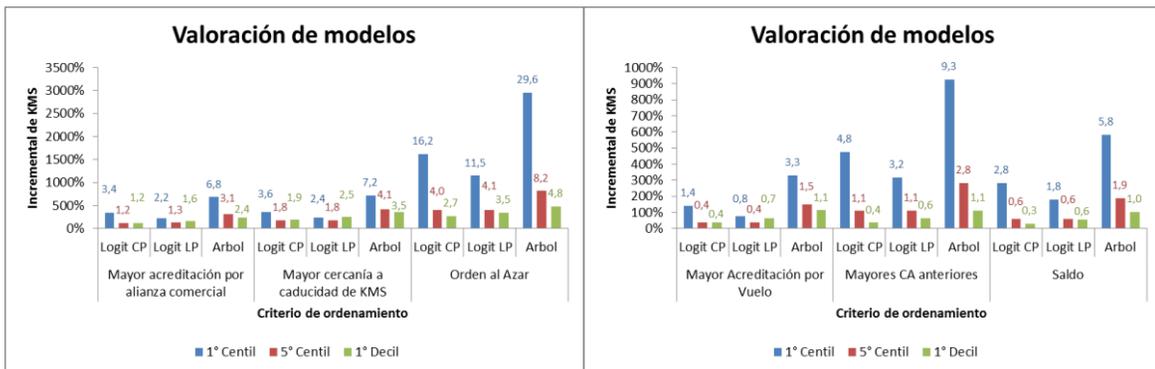


Ilustración 60. Valoración de Modelos Según Criterios de Ordenamiento.

Sin embargo, lo fundamental para probar la utilidad de los modelos es ver cómo se comportan contra otros criterios de ordenamiento poco sofisticados pero poderosos, como lo son la acreditación, la acreditación por vuelo y el CNA anterior.

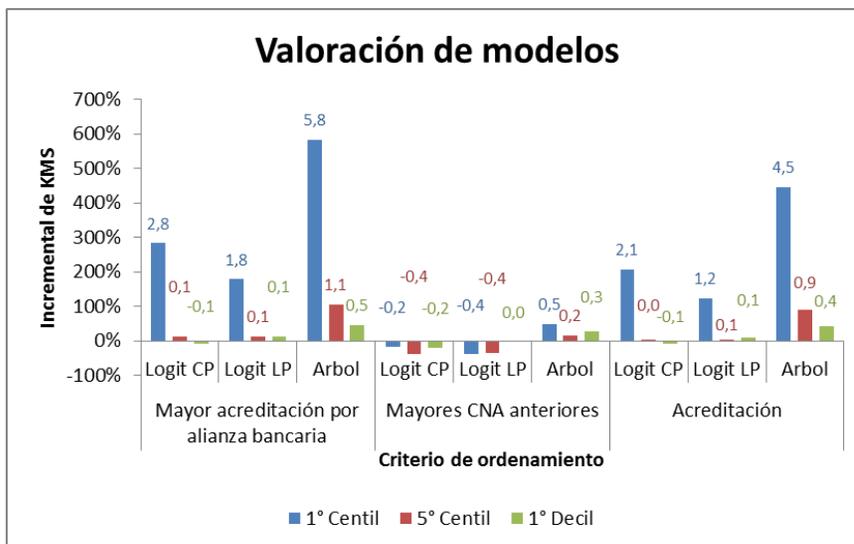


Ilustración 61. Valoración de Modelos Según Criterio de Ordenamiento.

En esta comparación vemos que solo el modelo de árboles de clasificación sobrevive al criterio más poderoso, “Mayor cantidad de KMS canjeados en CNA previamente”. Encontramos entonces el valor de esta investigación, saber que en el 1% de clientes más propensos al canje No Aéreo, se obtiene un incremental de 50% de KMS. Esta ventaja se mantiene a través del primer decil, ganando por un 20% en el primer 5% y por un 30% en el primer 10%.

8.2.2.2. Segunda Decisión: Elección de categoría

La decisión de categoría se estudió condicional en que los clientes ya habían realizado un CNA. Por esto no será posible valorar de la misma forma esta decisión.

Esta evaluación se realizó midiendo la precisión de la regresión en los dos primeros deciles.

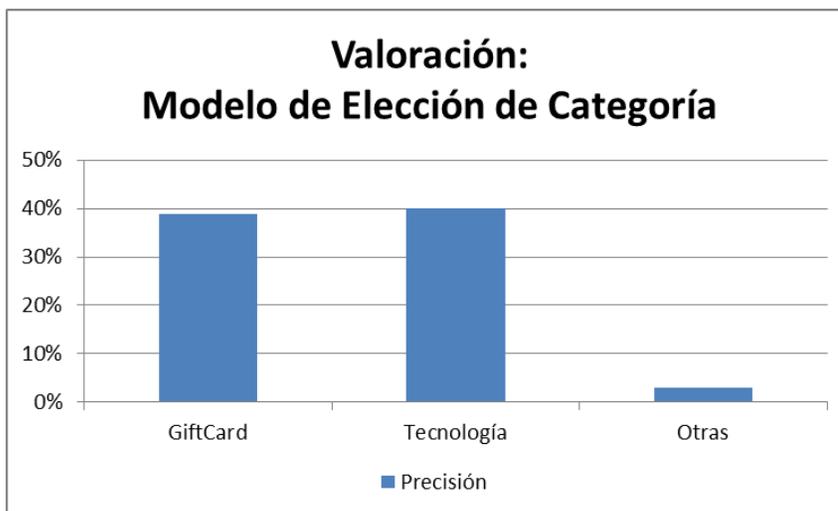


Ilustración 62. Valoración Modelo de Elección de Categoría.

Se Observa que la precisión se acerca al 40% en las categorías sobre las que se tiene poder predictivo, mientras que solo se tiene un 3% para el caso de "Otras" Categorías. Esto es un buen indicador de que se logra identificar a los clientes en los primeros deciles.

9. Conclusiones

9.1. Conclusiones Generales

La investigación realizada se basó en una metodología que podríamos describir como “bottom-up”, es decir, se partió estudiando la data de manera individual y superficial con un análisis descriptivo y avanzamos hacia modelos más sofisticados utilizando técnicas de clasificación. En cada una de estas etapas se encontró información que permitió ir conociendo paulatinamente al canjeador de catálogo.

De la decisión de canje, en donde los clientes deciden si realizar un canje aéreo, canje no aéreo o no canjear concluimos que los clientes son más propensos a canjear a medida que presentan una mayor actividad transaccional previa, ya sea de acreditación o canje. Este efecto es mucho más poderoso para la decisión de canje aéreo que para canje no aéreo. Además, en general se observa que los clientes son más propensos al CA que al CNA, existiendo una excepción relevante que es la alta actividad de CNA realizada previamente por los clientes. Las variables demográficas mostraron ser menos poderosas en general para predecir canje, pero se confirmó que estas tienen un comportamiento transaccional implícito.

Se descubrió que los clientes con KMS cercanos a expirar tienden a liquidar su saldo en CNA por sobre CA. Además, los clientes son más propensos al CNA mientras menos avanzan en el proceso de canje de un pasaje aéreo.

Los modelos de regresión logística nos permitieron encontrar que las variables más importantes que determinan el CNA son el saldo, la acreditación total en el último año y el N° de Canjes No aéreos realizados en el último año, en ese orden. Además, confirmaron que las variables demográficas, en conjunto, generan un buen predictor de CNA. Sin embargo, la mejor técnica para detectar propensión mostró ser el árbol de clasificación CHAID exhaustivo. Esta técnica es capaz de identificar al 30% de los canjeadores de catálogo en el primer centil de clientes, mejor que el 18% que identifica la mejor regresión logística.

El árbol de clasificación también nos permitió identificar algunas reglas de decisión que median la elección de CNA, estas explican el 72,4% del CNA y son los clientes que:

- Acreditaron más de 21.000 KMS el último año, realizaron CNA el último año y poseen saldo superior a 7.500 KMS.
- Acreditaron más de 21.000 KMS el último año, no han realizado CNA el último año y pertenecen a las categorías LANPASS o PREMIUM
- Acreditaron entre 4.200 y 9.100 KMS el último año y poseen saldo superior a 7.500 KMS
- Acreditaron entre 9.100 y 21.000 KMS el último año pero no canjearon nada durante ese año

- Acreditaron entre 9.100 y 21.000 KMS el último año y realizaron algún CNA durante ese año

En el estudio descriptivo de categorías observamos que los clientes con mayor acreditación tienden a preferir canjear en tecnología, esto se ve tanto en las variables transaccionales como en las demográficas con información transaccional implícita. Además esta categoría es preferida por clientes de entre 40 y 50 años, mientras que las demás categorías son de preferencia de los clientes más adultos y más jóvenes que los que prefieren Tecnología. En términos demográficos encontramos que solo “Otras Categorías” presenta mayor preferencia femenina.

En cuanto a los modelos de clasificación, se utilizó una regresión logística multinomial y se encontró significancia solo en las variables relacionadas con el historial de compra en las categorías en cuestión. Con esto se descubrió que el canjeador de Tecnología tiende a repetir su preferencia de canje, mientras que el canjeador de Giftcards es adverso al canje de Tecnología.

En la valoración económica de la investigación se encontró que existen criterios básicos de ordenamiento que son efectivos en la identificación de clientes propensos al CNA. El más poderoso es la “*Mayor cantidad de KMS canjeados en CNA previamente*”. Este criterio es tan efectivo que muestra ser un mejor predictor de canje que las regresiones logísticas utilizadas. El árbol de clasificación justifica el uso de técnicas de marketing cuantitativo pues es capaz de generar un 50% más de KMS en el primer decil que el criterio básico más fuerte.

En cuanto al cumplimiento del objetivo general, este se considera logrado. Usando la información disponible de la empresa se realizó una mirada integral del proceso de canje en el cual las variables demográficas y transaccionales, de manera individual y conjunta se convirtieron en herramientas que además de poder identificar a los clientes más propensos al CNA, revelaron reglas de decisión, importancia de variables y otros datos útiles en el conocimiento del cliente.

9.2. Recomendaciones de investigaciones futuras

En términos de fondo, al momento de realizarse esta memoria, LANPASS ya contaba con un conocimiento íntegro del fenómeno de acreditación de sus clientes. Esta investigación complementa dicho conocimiento a nivel de tipo de canje de forma satisfactoria.

Sin embargo, el nivel de agregación definido en los alcances del proyecto no permite un conocimiento profundo a nivel de categoría y producto.

Estas dos razones hacen sugerir una investigación enfocada en el producto, pues ya se conoce al cliente.

Respecto a la forma de la investigación, esta tiene limitaciones pues no considera el cambio de las preferencias de los clientes a través del tiempo ni el surtido ni el precio de la oferta en cada momento del tiempo. Esto motiva el estudio de la sensibilidad del cliente en cuanto a la oferta y el precio.

En cuanto a las reglas de decisión obtenidas, es probable que los puntos de corte varíen con el tiempo pero que estas reglas no cambien estructuralmente. Por este motivo no se considera necesario un estudio dinámico del proceso de canje.

Es importante destacar que durante el año 2015 han surgido nuevas formas de acumulación mediante la alianza Santander, lo que podría hacer necesario reevaluar el impacto de este método de acumulación en el futuro.

9.3. Recomendaciones Comerciales

Se recomienda enfocar los esfuerzos de marketing al segmento de los clientes que acreditaron más de 21.000 KMS el último año, que realizaron CNA el último año y que poseen saldo superior a 7.500 KMS. Este segmento destaca no solo poseer saldo suficiente para realizar canje sino además mostró ser el segmento más grande y con mejor precisión en su identificación.

Este segmento tiene más de 5.000 clientes, lo que permite a la empresa realizar campañas de e-mail marketing y obtener un incremental potencial de 50% de KMS respecto a lo que obtendría ofreciéndole canje a los clientes que más han realizado CNA previamente. Además se recomienda atraer a los clientes que no han canjeado previamente en el catálogo y que tienen el poder adquisitivo para hacerlo, pues según se vió, una vez que los clientes realizan un canje, son propensos a seguir haciéndolo.

En cuanto a las categorías de canje, se encontró evidencia que los clientes tienden a repetir su preferencia en tecnología, por lo cual es recomendable insistir en la oferta de esta categoría. A los clientes menos propensos a canjear en tecnología se sugiere ofrecerles Giftcards.

10. Bibliografía

1. JUDITH BARRERA RETAMALES. 2014 Diseño de un producto puntos más dinero para la compra de un pasaje aéreo. Memoria (Ingeniería Civil Industrial).Chile, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
2. NICOLE PEREIRA BIZAMA. 2014 Identificación de clientes con patrones de consumo eléctrico fraudulento. Memoria (Ingeniería Civil Industrial).Chile, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
3. KAREM PADILLA VERGARA. 2014 Identificación de clientes de alto valor para el desarrollo de alianzas de una empresa. Memoria (Ingeniería Civil Industrial).Chile, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
4. JUAN MIGUEL MARIN Análisis de Clúster y árboles de clasificación [Consulta: 24 de Marzo de 2013]
<http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema6dm.pdf>
5. GOIC MARCEL. 2013."Logit". En: CURSO IN5602, Marketing II, Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
6. JAMES LATTIN, PAUL GREEN, DOUGLAS CARROLL: Analyzing Multivariate Data. Belmont, Brooks/Cole, 2003
7. ROCKACH L., MAIMON O., Data Mining With Decision Trees: Theory and Applications, 2nd edition, 2015.
8. ALARCON ROSA, BRAVO NIKOLAS, SAFFIE RODRIGO. 2015 REST Web Service Description for Graph-Based Service Discovery. Lecture Notes in Computer Science pp 461-478

11. Anexos

Anexo 1: Pertenencia Santander

El banco Santander es la alianza más poderosa de LANPASS, pues el 17% de los clientes LANPASS participan de esta alianza. Santander tiene la virtud de que permite a los clientes acumular kilómetros utilizando sus medios de pago, acumulando kilómetros en todas las transacciones.

Similar al estudio de género, veremos el efecto de la pertenencia al banco Santander entre grupos y dentro de los grupos.

- **Análisis entre grupos:** Los clientes Santander son **6** veces más propensos al Canje Aéreo que los clientes que no pertenecen al banco. A su vez, los **clientes Santander son 5,5 veces** más propensos al canje no-aéreo que los No-Santander.
- **Análisis dentro de los grupos:** Los clientes Santander son **4** veces más propensos al CA que al CNA, mientras que los No-Santander son **3,8** veces más propensos al CA.

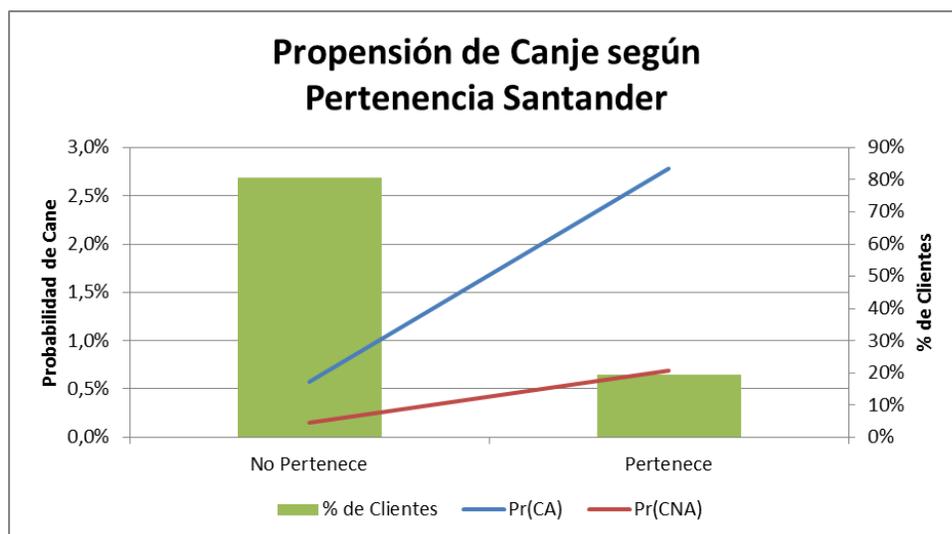


Ilustración 63. Propensión de Canje según Pertenencia Santander.

Anexo 2: Propensión de Canje según Categoría de Socio

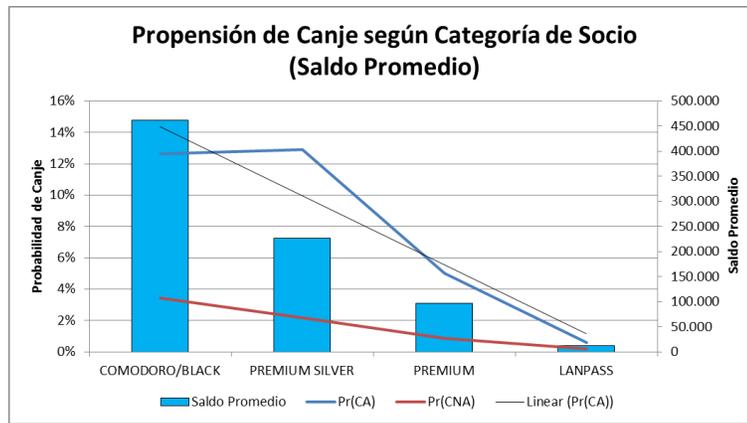


Ilustración 64 Propensión de Canje vs Categoría de Socio (Saldo Promedio).

Anexo 3: Propensión de Canje según Género

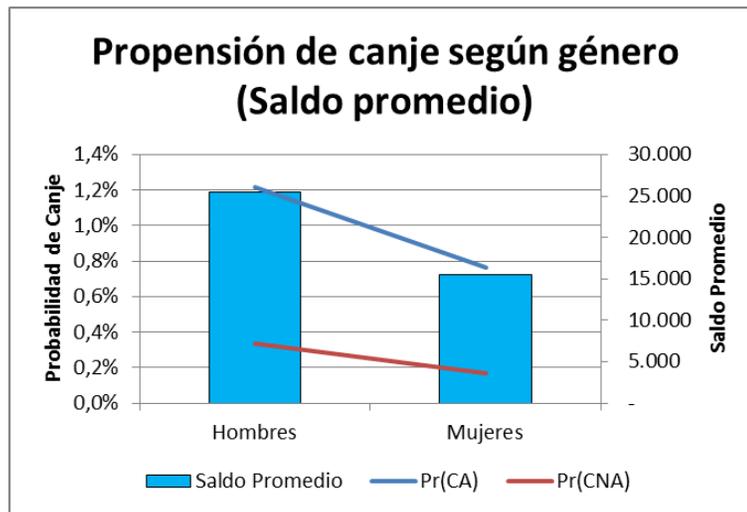


Ilustración 65. Propensión de Canje vs Género (Saldo Promedio).

Anexo 4: Propensión de Canje según Categoría de Socio

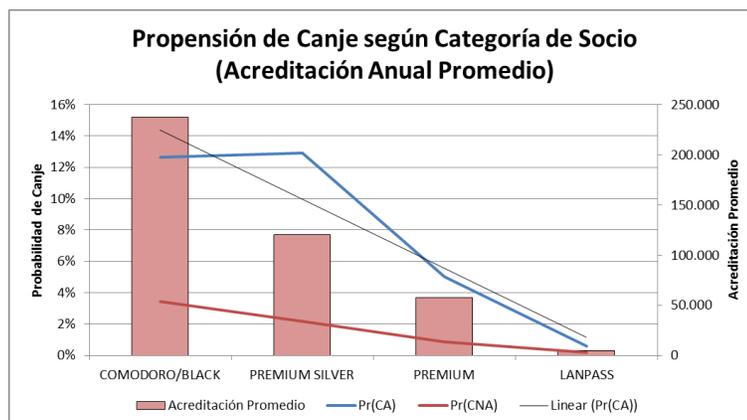


Ilustración 65. Propensión de Canje vs Categoría de Socio (Acreditación Promedio).

Anexo 5: Propensión de Canje según Género

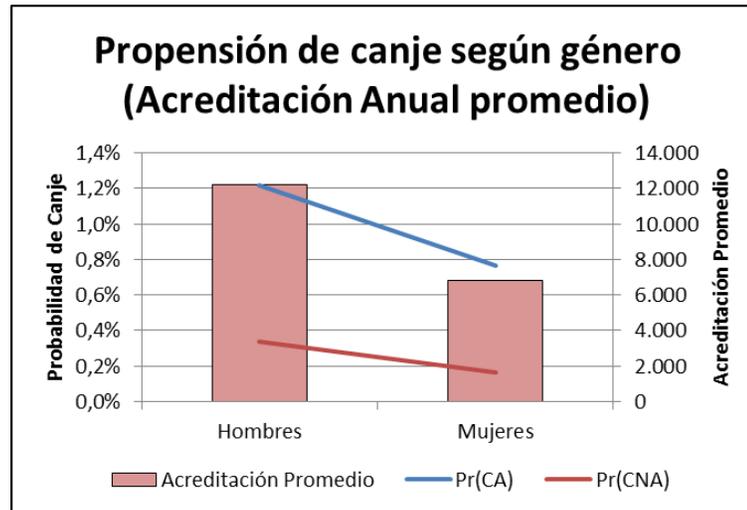


Ilustración 66. Propensión de Canje vs Género (Acreditación Promedio).

Anexo 6: Máximo de CA y CNA en KMS

En acreditación total vimos que los máximos de acreditación no aportaban mucho respecto al Número de Acreditaciones o a los KMS Acreditados. En el caso de CA y CNA se presenta un fenómeno interesante, las curvas de CNA y CA presentan convexidades complementarias, es decir, mientras “*máximo de CA*” como predictor de CA y CNA genera curvas cóncavas de propensión, “*máximo de CNA*” genera curvas convexas. A pesar de este fenómeno, la propensión de CA alcanza un máximo de 7% en los clientes que canjearon un vuelo más caro, mientras que el análogo de CNA alcanza un máximo de 5% para los clientes que canjearon un producto más caro.

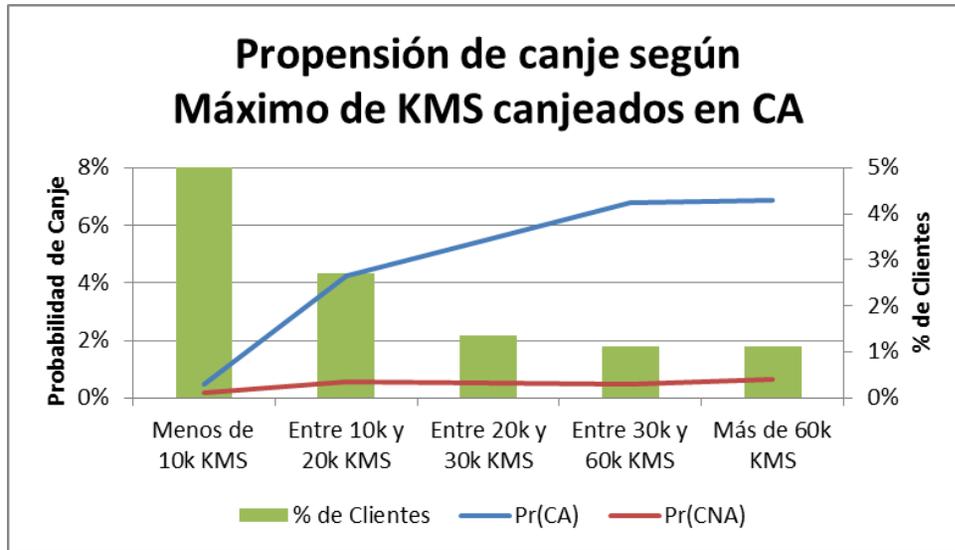


Ilustración 67. Propensión de Canje según Máximo de KMS canjeados en CA.

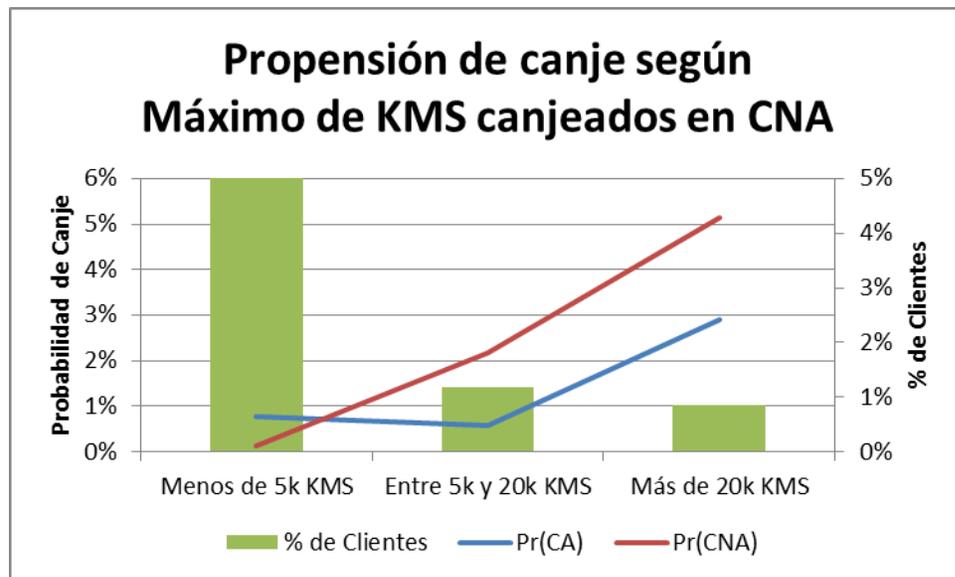


Ilustración 68. Propensión de Canje según Máximo de KMS canjeados en CNA.

Anexo 7: Tipo de Acreditador

Según los tres tipos principales de acreditación se define el tipo de acreditador según como cada cliente distribuya su acreditación. Si un cliente acredita el 80% de sus KMS mediante vuelo, entonces su tipo de acreditación se definirá como “Vuelo”, el mismo criterio se aplica para alianza comercial y bancaria. Si un cliente acredita menos de un 80% en los tipos de acreditación antes mencionados, el cliente será etiquetado como “Mix”, finalmente, si no tiene acreditación, se llamará “Ninguno”.

Se observa que los clientes “Mix” son los más propensos al CA y en particular, los que muestran una mayor diferencia respecto a su propensión al CNA, esto habla de un cliente que utiliza todos los medios de acreditación para alcanzar un premio de vuelo. El segundo lugar lo ocupan los acreditadores por vuelo, esto tiene sentido pues en general los vuelos implican una gran acreditación de KMS, por lo que los clientes estarán constantemente cercanos al canje.

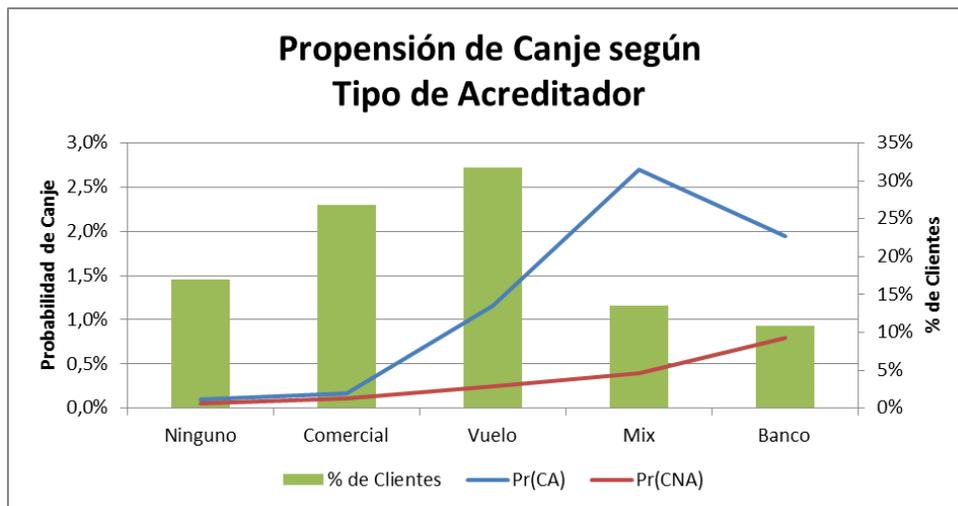


Ilustración 69. Propensión de Canje según Tipo de Acreditador.

Es interesante también conocer la acreditación promedio que presenta cada tipo de acreditador. Los clientes Banco y Mix son los que más acreditan, lo que más sorprende es la gran diferencia que presentan respecto a los acreditadores por vuelo, esto se puede deber a que las transacciones bancarias son un comportamiento cotidiano no comparable con un viaje en avión, lo que puede verse reflejado en un alto promedio de acreditación. De la misma forma, las alianzas comerciales también requieren esfuerzos adicionales para lograr la acreditación, lo que explica el bajo promedio.

Anexo 8: Tipo de Canjes Anteriores

Esta variable presenta diferencias interesantes en las propensiones. Se observa que los clientes que han realizado alguna vez CNA son casi 18 veces más propensos al CNA que los clientes que nunca han canjeado, mientras que los que adicionalmente también canjearon alguna vez CA son 27 veces más propensos al CNA que los clientes que nunca han canjeado.

El grupo de los clientes que solo han realizado CNA es de particular interés pues, además de tener una alta propensión al CNA, tienen una propensión mayor al CNA que al CA.

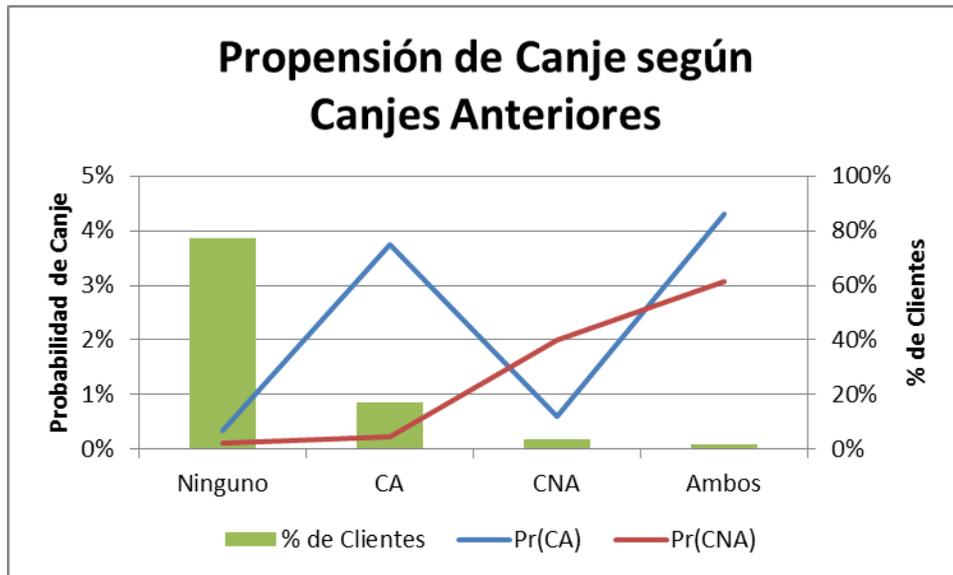


Ilustración 70. Propensión de Canje según Canjes Anteriores.