



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

**ESTRATEGIA DE PERSONALIZACIÓN Y GESTIÓN DE CAMPAÑAS DE  
PUBLICIDAD WEB EN UN RETAIL FINANCIERO**

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

**SEBASTIÁN ANDRÉS IBARRA ALBORNOZ**

**PROFESOR GUÍA:  
ALEJANDRA PUENTE CHANDÍA**

**MIEMBROS DE LA COMISIÓN:  
TODD LLOYD PEZZUTI  
JOSÉ ANTONIO NALDA REYES**

**SANTIAGO DE CHILE  
2016**

**RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE:** Ingeniero Civil Industrial  
**POR:** Sebastián Andrés Ibarra Albornoz  
**FECHA:** 11/10/2016  
**PROF. GUÍA:** Alejandra Puente Chandía

## **ESTRATEGIA DE PERSONALIZACIÓN Y GESTIÓN DE CAMPAÑAS DE PUBLICIDAD WEB EN UN RETAIL FINANCIERO**

El mercado de las tarjetas de crédito se caracteriza por ser altamente competitivo e importante en transacciones, con montos cercanos al 14% del PIB nacional en conjunto con débito, por lo que entregar una experiencia diferenciadora a los clientes se ha vuelto una necesidad básica. Así, el uso de diferentes canales de comunicación integrados contribuye a generar la diferenciación y, en particular, el sitio web de la empresa presenta un aumento de ingresos de usuarios únicos de un 96% en los últimos 4 años, por lo que surge como una oportunidad para entregar una experiencia personalizada.

De acuerdo a esta necesidad, el objetivo general de este proyecto es diseñar acciones de apoyo a la personalización y gestión de campañas de publicidad web. Así, se quiere responder lo siguiente: ¿quiénes son los clientes que utilizan el sitio web?, ¿quiénes son los clientes que interactúan con campañas de publicidad web? y ¿existen diferencias entre los clientes que interactúan con diferentes formatos de campañas web?

Para responder las preguntas mencionadas, se realizan modelos de clasificación para identificar qué características definen que un cliente tenga mayor probabilidad de interactuar con el sitio web, obteniendo desempeños sobre el 20% de ganancia de información en los principales fenómenos en estudio.

Los principales resultados tienen relación con los perfiles de los clientes que interactúan con el sitio web, siendo los que ingresan al sitio adultos jóvenes, que compran en la web de la multitienda, viajes y utilizan la tarjeta principalmente por el club de puntos. En el caso de campañas, los Shadow Box apuntan a clientes de categoría media o alta, que compran en la web de la multitienda, restaurant, viajes y su tarjeta es Visa. De manera contraria, las Burbujas apuntan a clientes que solicitan avances en efectivos, créditos de consumo, compran en supermercados externos y en cuotas, mientras que Calugas apunta a clientes de categoría media o alta, que tienen transacciones en la web de la multitienda, servicios básicos, educación, restaurant y solicitan avances en efectivo.

Con el fin de generar un aumento de clientes web y aumentar las interacciones con las campañas, se propone gestionar a 280 mil clientes que no utilizan el sitio web mediante email marketing, a cerca de 90 mil clientes con experimentos del formato Shadow Box y a cerca de 110 mil clientes con experimentos del formato Caluga.

En caso de utilizar las estrategias propuestas en el presente proyecto, la compañía generaría ventas incrementales de \$ 9 mil millones en un año, equivalente al 0,3% de las ventas anuales, en el caso de que el 5% de los clientes gestionados cambien su comportamiento.

Finalmente, como trabajos futuros se propone la realización de experimentos que utilicen sesgos cognitivos y herramientas de eyetracking. Además, mejorar el estudio con información detallada de la navegación web que permita aumentar la personalización, tal como duración en cada sección del sitio web y movimientos del puntero en la página web.

## **AGRADECIMIENTOS**

Finalizando esta etapa, quisiera agradecer en primer lugar a mi familia, mi mamá Alicia, mi papá Juan Carlos, mi abuelo Roberto y mi hermana Carolina, por apoyarme, confiar en mis capacidades y decisiones y entregarme cariño durante toda mi vida. Están presentes en todos mis logros.

Agradezco también a mis amigos de la vida y del colegio, con quienes aprendí lo importante de contar con buenas amistades. En especial, a mis mejores amigos, Anto y Javito, quienes me han escuchado, aconsejado y aguantado en cada situación que los he necesitado.

A mis amigos de la universidad que conocí desde los primeros momentos, Pedro, Pablo, Nico, Carlos, Daniel, Silva, Pancho, siempre en todas. A mis amigos que me acompañaron en Industrias, Coni, Vale, Coa, Marti, Raúl, Chicho, gracias por los momentos dentro y fuera de lo académico.

También a mis amigos con los que viví todo el proceso de titulación, Benja, Chicho y Coa, gracias por la empatía y consejos durante esta etapa.

Finalmente, muchas gracias a mi profesora guía, Ale Puente, por haberme dado consejos durante todo el proceso de titulación, pese a la gran cantidad de problemas que tuve, y también a la empresa en la que realicé este proyecto, en especial a José, profesor integrante, por la paciencia y enseñanzas de la vida laboral que me entregó.

## TABLA DE CONTENIDO

1	ANTECEDENTES GENERALES .....	1
1.1	Retail en Chile .....	1
1.2	Tarjetas de Crédito en Chile .....	1
1.3	Retail Financiero .....	2
1.4	Tendencias Web .....	3
1.5	Personalización de Sitios Web .....	3
2	DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN .....	4
3	OBJETIVOS .....	6
3.1	Objetivo General .....	6
3.2	Objetivos Específicos .....	6
4	ALCANCES DEL PROYECTO .....	7
5	MARCO CONCEPTUAL .....	8
5.1	Campañas de Marketing .....	8
5.2	Modelos de Propensión .....	8
6	METODOLOGÍA Y DESARROLLO METODOLÓGICO .....	12
6.1	Selección de Datos y Pre-Procesamiento .....	12
6.2	Descripción de Variables .....	13
6.3	Análisis Exploratorio .....	18
6.4	Modelos de Propensión .....	24
6.5	Conclusión Modelos de Propensión .....	53
6.6	Líneas de Acción .....	55
7	RESULTADOS COMPLEMENTARIOS .....	61
7.1	Clasificación por Comportamiento de Navegación .....	61
7.2	Intensidad Web .....	63
7.3	Score Engagement Web (SEW) .....	64
8	CONCLUSIONES .....	65
8.1	Conclusiones del Proyecto .....	65
8.2	Limitaciones y Trabajos Futuros .....	66
9	BIBLIOGRAFÍA .....	68
10	ANEXOS .....	69
10.1	Anexos de Descripción de Variables .....	69
10.2	Anexos de Análisis Exploratorio .....	70
10.3	Anexos de Modelos de Propensión .....	73

## ÍNDICE DE TABLAS

Tabla 1: Descriptivos variables cuantitativas de clientes en estudio .....	14
Tabla 2: Correlación entre variables cuantitativas de clientes en estudio .....	14
Tabla 3: Porcentaje de clientes con transacciones en comercios internos .....	17
Tabla 4: Porcentaje de clientes con transacciones en comercios externos .....	17
Tabla 5: Resultados regresión logística modelo de ingreso al sitio web .....	25
Tabla 6: Matriz de confusión de regresión logística de modelo de ingreso al sitio web ..	26
Tabla 7: Matriz de confusión de árbol de decisión de modelo de ingreso al sitio web ...	29
Tabla 8: Resultados regresión logística modelo de publicidad web .....	30
Tabla 9: Matriz de confusión de regresión logística de modelo de publicidad web .....	31
Tabla 10: Matriz de confusión árbol de decisión de modelo de publicidad web .....	33
Tabla 11: Resultados regresión logística modelo de Calugas.....	35
Tabla 12: Matriz de confusión de regresión logística de modelo de Calugas.....	36
Tabla 13: Matriz de confusión de árbol de decisión de modelo de Calugas.....	39
Tabla 14: Resultados regresión logística modelo de Shadow Box.....	40
Tabla 15: Matriz de confusión de regresión logística de modelo Shadow Box.....	41
Tabla 16: Matriz de confusión de árbol de decisión de modelo de Shadow Box.....	43
Tabla 17: Resultados regresión logística de modelo de Burbujas.....	45
Tabla 18: Matriz de confusión de regresión logística de modelo de Burbujas.....	46
Tabla 19: Matriz de confusión de árbol de decisión de modelo de Burbujas.....	47
Tabla 20: Resultados regresión logística modelo campaña de puntos .....	49
Tabla 21: Matriz de confusión regresión logística modelo de Puntos.....	50
Tabla 22: Matriz de confusión de árbol de decisión de modelo de Puntos.....	52
Tabla 23: Variables independientes y sus niveles para migración a canal digital .....	56
Tabla 24: Variables independientes y sus niveles para personalización de campañas ..	58
Tabla 25: Beneficios económicos por la gestión de clientes propuesta .....	61
Tabla 26: Criterios clasificación de clientes web .....	62

# 1 ANTECEDENTES GENERALES

## 1.1 Retail en Chile

A nivel nacional, el desarrollo del comercio al por menor, o retail, ha tenido un crecimiento sostenido y por sobre el resto de las actividades económicas del país. Según cifras del Instituto Nacional de Estadísticas (INE) [1], el Índice de Ventas del Comercio al por Menor, registra al mes de septiembre 2015 un aumento real de 3,1% en los últimos doce meses, siendo los Bienes de Consumo Diversos la cuarta línea de producto que más ha aumentado con un 3,8%.

De acuerdo a la consultora A.T. Kearney, la cual elabora el Índice Global de Desarrollo del Retail (GRDI, por su sigla en inglés Global Retail Development Index) [2], en su reporte para el año 2015 muestra a Chile en la tercera posición, y en el año 2014 en la primera, todo esto considerando que en el atributo de Saturación de Mercado es el peor del Top 10 del ranking. De acuerdo a esta información, se tiene que Chile es un mercado bastante competitivo, por lo que se vuelve necesario el uso de diferentes estrategias de marketing para lograr diferenciación con la competencia.

Asimismo, según cifras del Banco Central de Chile [3], en el año 2015 el sector económico del retail representa un 29% del Producto Interno Bruto (PIB).

## 1.2 Tarjetas de Crédito en Chile

En Chile, el uso de las tarjetas de crédito es uno de los principales medios de pago que existen, representando montos transados que representan cerca de un 14% del Producto Interno Bruto nacional<sup>1</sup> para el año 2013 en conjunto con débito.

Por un lado, esto se debe a la gran cantidad de emisores bancarios y no bancarios (casas comerciales, principalmente) que entregan este tipo de tarjetas, lo que genera un mercado altamente competitivo. Como muestra la **Ilustración 1**, con más de 21 millones de tarjetas vigentes en el mercado, son tres emisores no bancarios quienes representan un mayor porcentaje de este producto.

---

<sup>1</sup> “Actualidad del mercado de las tarjetas y transacciones electrónicas en Chile”, Javier Etcheverry, 18 de marzo de 2015.

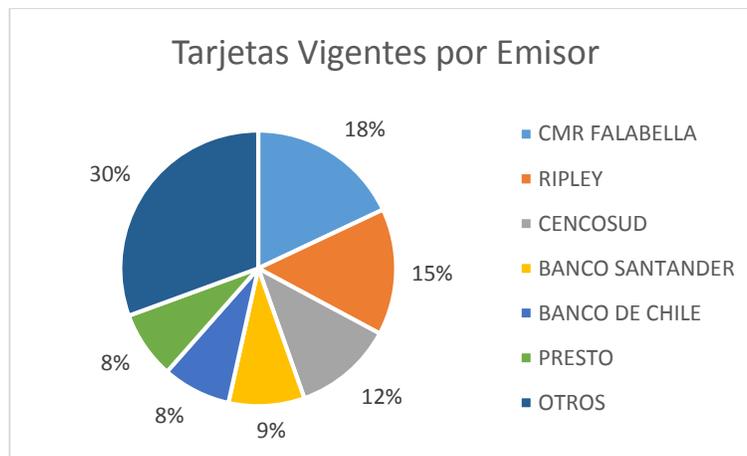


Ilustración 1: Participación de mercado de Tarjetas de Crédito, según Tarjetas Vigentes por Emisor.  
 Fuente: Elaboración propia con datos de la Superintendencia de Bancos e Instituciones Financieras (Junio 2015).

### 1.3 Retail Financiero

La industria del retail financiero hace referencia a la vinculación que tiene una empresa de retail con el negocio financiero, esto a través de tarjetas de crédito principalmente, con el objetivo de que un cliente adquiera mayor capacidad de compra mediante el crédito, que se fidelice al consumidor, conocer las características de cada cliente mediante la identificación de quién realiza la compra y, en general, generar valor a la compañía.

Como se mostró anteriormente, son compañías de esta industria quienes poseen mayor cantidad de tarjetas vigentes, sin embargo, en términos de montos transados, no necesariamente representan los primeros lugares, por lo que existe la necesidad de competir utilizando nuevos canales y mejorando la experiencia del cliente.

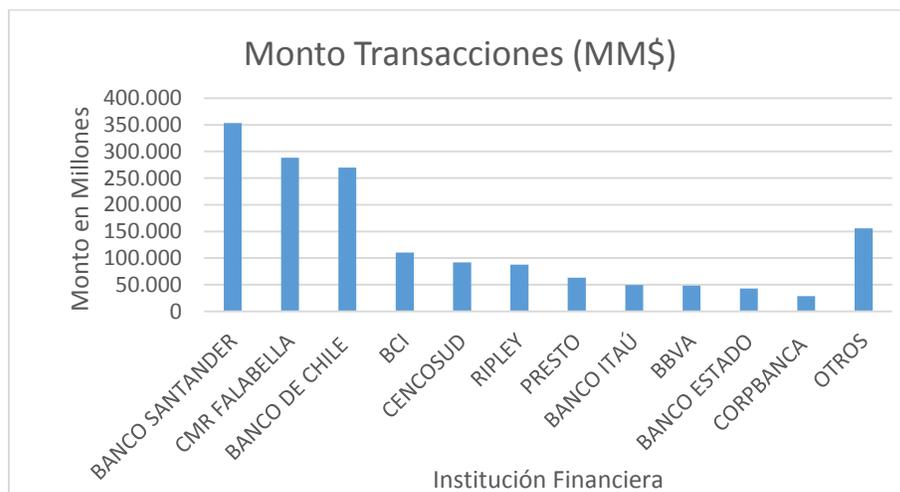


Ilustración 2: Montos transados en MM\$ por Emisor.  
 Fuente: Elaboración propia con datos de la Superintendencia de Bancos e Instituciones Financieras (Junio 2015).

## **1.4 Tendencias Web**

De acuerdo a información de la Cámara de Comercio, el 70% de los chilenos utilizan Internet, por lo que contar con un canal digital acorde a la realidad se vuelve una necesidad básica para empresas competitivas. Además, en términos de compras en línea, Chile ocupa el primer puesto en el continente, aumentando a un 30% anual, con proyecciones al doble en el año 2020.

El negocio de comercio en línea se ha vuelto cada vez más confiable para el cliente chileno, por lo que además de ser una necesidad es una gran oportunidad de crecer y fidelizar clientes, para lo cual se vuelve fundamental estudiar los comportamientos que se tienen en el canal digital, tanto sitios web como mediante correos electrónicos.

## **1.5 Personalización de Sitios Web<sup>2</sup>**

Actualmente, los clientes se enfrentan a un amplio número de contenido mientras navegan por la web, de lo cual, naturalmente va a relacionarse con el contenido que esté alineado con sus propios intereses y necesidades, mientras que todo el resto no será tomado en cuenta. Dado esto, la personalización juega un rol fundamental, pues permite conectarse directamente con la audiencia objetivo y ofrecer la solución o producto exacto que el cliente busca, lo que aumenta la probabilidad de conversión.

Para realizar esta personalización, es necesario desarrollar estrategias basadas en datos, ya que, de otra forma, se estaría invirtiendo esfuerzos de marketing sin comprender ni valorizar la respuesta del cliente.

Asimismo, una estrategia digital de contenido personalizado cuenta con implicancias en generar una visión omnicanal, ya que ésta sostiene al consumidor como el centro del negocio y apunta a crear una experiencia de compra memorable, lo que requiere que cada cliente encuentre de manera simple lo que busca o necesita.

Lo principales beneficios de generar esta experiencia de compra en todos los puntos de contacto (omnicanal), tienen que ver con la lealtad a la marca, aumentar las tasas de conversión y generar mayor satisfacción y percepción de calidad.

---

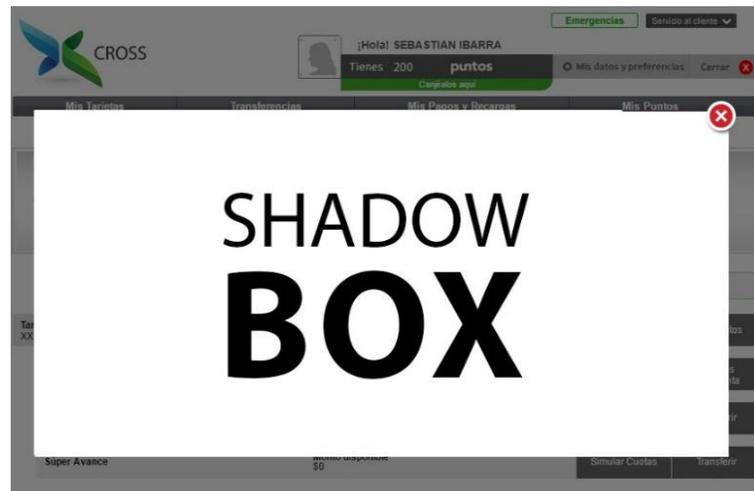
<sup>2</sup> "Why marketers must think 'Personalization'", Daniel Newmann, 25 de marzo de 2015.

## 2 DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN

El proyecto desarrollado en la presente memoria de título consiste en el estudio de la utilización y respuesta de los clientes en el sitio web de un retail financiero. En particular, la respuesta de los clientes se estudiará utilizando modelos de propensión, los cuales incorporan información individual de cada cliente, tales como datos demográficos, transaccionales, contractuales y de navegación web.

El proyecto tiene como objetivo apoyar los procesos de personalización y gestión de campañas de marketing web, lo que apoya el propósito de la empresa de aumentar la interacción de clientes con el canal digital. Esto se define de tal forma, dado que se considera que la primera forma que tiene un cliente para interactuar con el sitio web de la empresa es ingresando e iniciando sesión y luego, es interactuar con la publicidad que se le muestra por medio de dar clic.

Las campañas de marketing a las que se hace referencia se exhiben en el sitio web en diferentes formatos, los cuales corresponden a Shadow Box, Caluga y Burbuja. El Shadow Box, como se muestra en la **Ilustración 3**, es una gráfica que se superpone a la página web y en donde regularmente se oscurece todo lo que la gráfica no cubre con el fin de darle mayor énfasis, y solo se muestra al ingresar al sitio web. Por otro lado, se tiene la Caluga y la Burbuja, como se muestra en la **Ilustración 4**, que corresponden a banners fijos en la página web y que su principal diferencia es que la Caluga incluye una gráfica, mientras que la Burbuja solo utiliza texto.



*Ilustración 3: Representación genérica de Shadow Box.  
Fuente: Elaboración propia.*

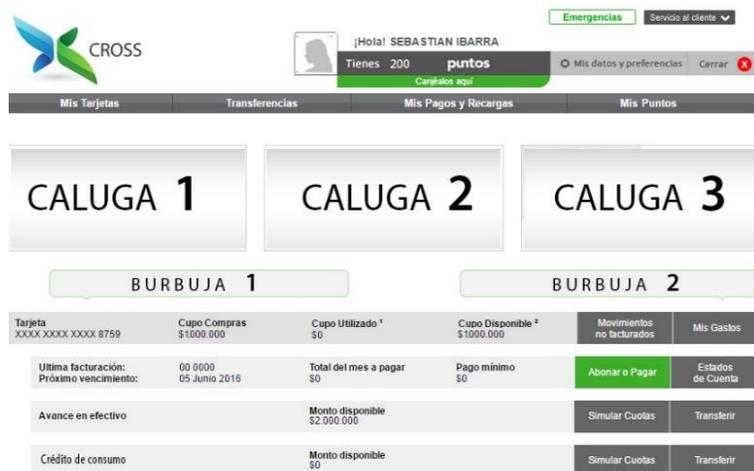


Ilustración 4: Representación genérica de Calugas y Burbujas.  
Fuente: Elaboración propia.

Tal como es la tendencia en los diferentes negocios, el canal digital ha comenzado a adquirir una mayor relevancia en los últimos años, y la compañía en la que se desarrolla el presente proyecto no es la excepción. Como muestra la **Ilustración 5**, la cantidad de usuarios únicos que ingresan mensualmente al sitio presenta un aumento cercano a un 96% en los últimos 4 años. Dado esto, generar proyectos que aporten a entregar una propuesta más personalizada a los clientes es de bastante importancia para la empresa.



Ilustración 5: Evolución ingresos al sitio web durante período de estudio.  
Fuente: Elaboración Propia.

Los resultados de este proyecto entrega información cuantitativa sobre qué características de un cliente lo hacen ser más propensos a interactuar con el contenido del sitio web, por lo que, de no realizarse este estudio, se mantendría el método en que

se dirigen las campañas de marketing actualmente, el cual no incorporan estudios analíticos sobre las características que tienen mayor relación con la interacción. En este sentido, la utilización de los resultados del proyecto tiene un impacto en la experiencia del cliente.

La realización del proyecto contempla una metodología que incluye un levantamiento de información que posee la empresa, consolidar las distintas fuentes de datos en una base analítica adecuada para posteriores análisis, utilización de herramientas de minería de datos y modelamiento econométrico y, por último, adaptación de los resultados de forma que sea accionable por la empresa, por lo que el proyecto cuenta con la complejidad propia de un proyecto profesional y de un estudio analítico.

El desarrollo del proyecto presenta una innovación en términos de la información que utiliza y el fenómeno que busca explicar, ya que se mezclan datos del comportamiento transaccional del cliente con el comportamiento online, apuntando a incorporar una nueva dimensión de cómo entender al cliente, y así entregarle una propuesta de valor más personalizada.

Finalmente, este proyecto se diferencia de trabajos anteriores en base a su componente de innovación, y de considerar el negocio web de forma diferente al negocio presencial, al menos en términos de cómo el cliente lo enfrenta e interactúa ante el contenido e información que se le entrega.

## **3 OBJETIVOS**

### **3.1 Objetivo General**

Diseñar acciones de apoyo a la personalización y gestión de campañas de publicidad web utilizando información demográfica, de navegación y transaccional de los clientes.

### **3.2 Objetivos Específicos**

- ✓ Determinar el conjunto de variables que mejor describan la interacción con el sitio web.
- ✓ Diseñar e implementar modelos econométricos y de minería de datos para describir los perfiles de clientes más propensos a ingresar al sitio web y clicar la publicidad web.
- ✓ Analizar los perfiles de clientes entregados por los modelos de propensión.

- ✓ Diseñar conceptualmente una campaña dirigida a clientes más propensos.
- ✓ Formular líneas de acción de acuerdo a los resultados obtenidos.

## 4 ALCANCES DEL PROYECTO

El estudio cuenta con los siguientes alcances:

- ✓ En un principio, se comienza con una base de datos que incluyen a todos los clientes que cuenten con la tarjeta de crédito sin bloqueos por morosidad, y que hayan realizado la apertura de su cuenta antes de julio de 2015.
- ✓ En una segunda etapa, los clientes estudiados corresponden a quienes cuentan con la tarjeta de crédito sin bloqueos por morosidad, que hayan realizado la apertura de su cuenta antes de julio de 2015 y hayan ingresado al sitio web entre los meses de julio de 2015 hasta marzo de 2016, por limitaciones propias de la recolección de datos de la empresa.
- ✓ En términos de la presente memoria de título, se considera como variables dependientes (en estudio) a los eventos de ingresar al sitio web e iniciar sesión en un caso y al evento en el cual un cliente dentro del sitio web, tras haber iniciado sesión, hace clic sobre el contenido publicitario que se exhibe.
- ✓ Para estudiar la propensión de un cliente a interactuar con el contenido publicitario que se exhibe se agruparán campañas de marketing, con el fin de contar con suficientes casos positivos para poder obtener resultados estadísticamente significativos, teniendo diferentes niveles de modelos: agrupación de todas las campañas (Publicidad Web), agrupación por formato (Shadow Box, Calugas y Burbujas) y una agrupación a nivel de un tipo de campañas (Campañas de Puntos).
- ✓ El estudio no considera la forma, en términos gráficos, de las campañas de marketing, ya que no se cuenta con la información necesaria para obtener los resultados esperados. Dado esto, este estudio no incorpora efectos de sesgos cognitivos o comportamiento del consumidor relacionados a cómo se muestra el mensaje.
- ✓ El proyecto no incluirá una fase de testeo mediante la realización de experimentos, ya que no es factible por parte de la empresa, por lo que se limitará a crear el diseño conceptual de una propuesta de piloto.
- ✓ En la sección de clasificación de clientes, los usos del sitio web en los que se basa el estudio se determinarán en base a las dimensiones que la compañía considera más relevante.

## 5 MARCO CONCEPTUAL

### 5.1 Campañas de Marketing

Por campaña de marketing se hace referencia a actividades específicas que se diseñan para promocionar un producto, servicio o negocio. Una campaña de marketing es una serie de etapas coordinadas que incluyen el uso de diferentes medios (televisión, radio, internet, entre otros) usando diferentes tipos de publicidad.

En el caso particular de este estudio, por campañas de marketing se hace referencia a la publicidad realizada por los tres formatos que ofrece el sitio web: Shadow Box, Calugas y Burbujas.

### 5.2 Modelos de Propensión

Son modelos diseñados para estimar la respuesta que tiene un individuo ante una intervención o estímulo al se le expone. Este tipo de modelo entrega la respuesta más probable del individuo en términos de diferentes variables, y ayuda a evaluar cuál tienen mayor impacto en la respuesta.

Existen diferentes métodos para modelar la propensión de un individuo, a continuación, se describen las que se utilizarán en el presente estudio:

#### 5.2.1 Regresión Logística Binaria

Es un modelo de regresión en donde la variable dependiente es categórica. La regresión logística binaria [4] es usada para estimar la probabilidad de una respuesta binaria basada en una o más variables independientes.

La probabilidad que el agente  $n$  elija la alternativa  $i$  puede ser descrita mediante la siguiente ecuación:

$$P_{ni} = \frac{\exp(v_{ni})}{1 + \exp(v_{ni})} = \frac{\exp(\beta_{ni} \cdot X_n)}{1 + \exp(\beta_{ni} \cdot X_n)}$$

En donde  $v_{ni}$  corresponde a la componente determinística de la utilidad,  $X_n$  es un vector que caracteriza al individuo  $n$  y el vector de parámetros  $\beta_{ni}$  que representa la

relación entre estas variables y la utilidad de escoger la alternativa  $i$  mientras que la componente aleatoria de la utilidad  $\varepsilon_{ni}$  se asume que distribuye i.i.d. valor extremo tipo I.

Con el fin de evaluar modelos de este tipo, se tiene la función de verosimilitud, la cual se puede describir de la siguiente forma:

$$L = \prod_i^N f(x_i|\theta)$$

En donde se tiene que  $(x_1, \dots, x_N)$  es una muestra de datos de una variable aleatoria  $X$ , cuya función de probabilidad es  $f(X|\theta)$ . El estimador de máxima verosimilitud corresponde al parámetro  $\hat{\theta}$  tal que maximiza la función de verosimilitud  $L$ .

### 5.2.2 Árboles de Decisión

Los árboles de decisión [5] corresponden a un modelo predictivo utilizado en estadística, minería de datos y machine learning. Esta metodología permite predecir variables tanto categóricas como continuas en función de diferentes variables explicativas a través de la sucesiva división de la base de datos en grupos disjuntos de datos. Cada nodo interior corresponde a una de las variables explicativas y cada rama representa un valor de la variable objetivo dados los valores de las variables explicativas representadas por el camino desde la raíz hasta la rama. Por ejemplo, como se muestra en la **Ilustración 6**, la decisión de comprar un celular se explica con la variable Edad, entregando que quienes tengan menos de 30 años son más propensos que los mayores a 30 años.

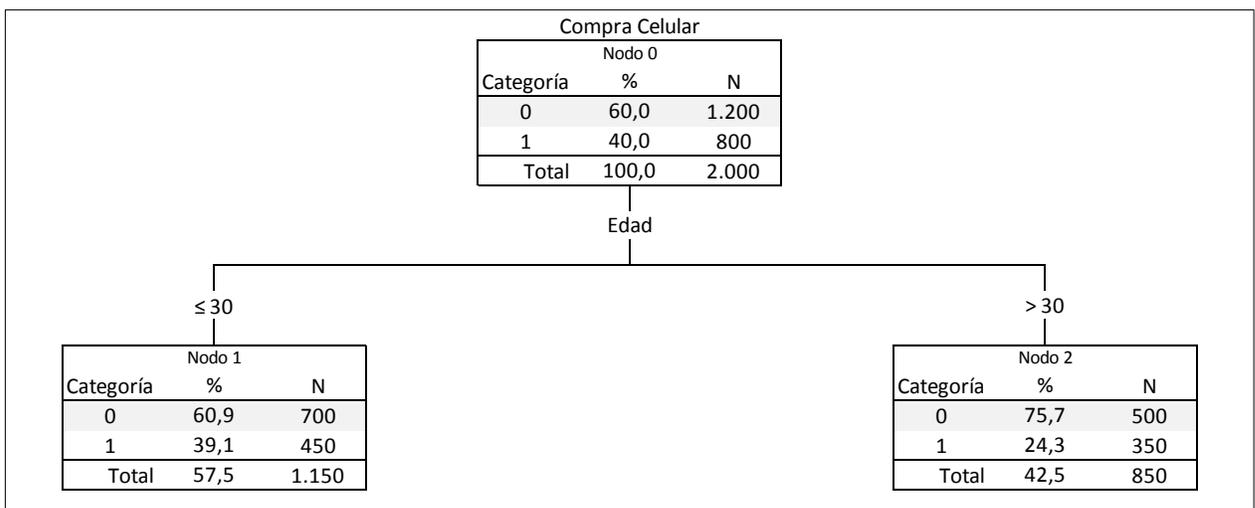


Ilustración 6: Ejemplo gráfico de un algoritmo de árbol de decisión.  
Fuente: Elaboración propia.

Este método cuenta con diferentes tipos de algoritmos, los cuales representan distintos criterios de generación de nodos o cortes; a continuación, se presentan los más utilizados:

- ✓ **CART:** Árbol de clasificación y regresión *Classification and Regression Tree*

Este tipo de árbol permite utilizar una variable dependiente categórica o continua, de la cual se dividen los casos en sólo dos nodos hijos.

El criterio de división para generar nodos se realiza a través del criterio de la impureza de Gini, el cual es una medida de la frecuencia en que un elemento seleccionado aleatoriamente del conjunto sería etiquetado incorrectamente si se etiquetó aleatoriamente de acuerdo a la distribución de las etiquetas del subconjunto.

La ecuación de esta métrica se calcula de la siguiente forma, siendo  $f_i$  la probabilidad de acertar y  $(1 - f_i)$  la probabilidad de equivocarse para un conjunto de  $m$  elementos:

$$I_G(f) = \sum_{i=1}^m f_i(1 - f_i) = \sum_{i \neq k} f_i f_k$$

Este algoritmo utiliza una poda previa mediante un algoritmo de un paso (single-pass algorithm), y permite modelar el problema con variables tanto categóricas como continuas.

- ✓ **CHAID:** Detector Automático de Chi-Cuadrado de Interacción *Chi-Squared Automatic Interaction Detector*

Este tipo de árbol crea dos o más nodos a partir de un nodo padre únicamente en a través de variables categóricas, en caso de tener variables nominales, se generan distintas categorías, y para realizar la división lo hace en base a pruebas de significancia Chi-Cuadrado, el cual tiene la ventaja de no ser paramétrico.

Para realizar la prueba de significancia Chi-Cuadrado se utiliza el siguiente estadístico:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \sim \chi_k^2$$

Donde  $\chi^2$  corresponde al estadístico de prueba acumulativa de Pearson,  $O_i$  al número de observaciones tipo I,  $E_i$  a la frecuencia teórica esperada de tipo I.

Este algoritmo utiliza una poda previa mediante el test Chi-Cuadrado de independencia, y permite modelar el problema con variables tanto categóricas y continuas.

- ✓ **QUEST:** Árbol Estadístico Rápido, Insesgado y Eficiente *Quick, Unbiased, Efficient, Statistical Tree*

Este algoritmo requiere de una variable dependiente categórica, de la cual puede generar solamente dos nodos hijos, y permite utilizar variables categóricas o continuas para explicar el problema.

El método de división de este algoritmo depende de la naturaleza de la variable: utiliza un ANOVA de dos vías para examinar la diferencia de dos variables independientes continuas u ordinales y una prueba de significancia Chi-Cuadrado en el caso de las variables categóricas, al igual que el árbol CHAID. A diferencia del resto de los algoritmos explicados anteriormente, éste utiliza una poda posterior a obtener los nodos.

### 5.2.3 Análisis COR/ROC (Característica Operativa Relativa/Relative Operating Characteristic)

Este análisis se utiliza para la comparación de modelos de clasificación y predicción, y se basa en los diferentes indicadores de la matriz de confusión. Con el fin de obtener una representación gráfica, es que se utiliza la curva COR, la cual enfrenta a la sensibilidad del modelo (probabilidad de detección o tasa de verdades positivas, mide proporción de los casos positivos que se clasifican correctamente) y la especificidad (tasa de verdaderos negativos, mide la proporción de casos negativos que se clasifican correctamente como negativos). Utilizando esta curva, se calcula el área bajo la curva (AUC), la cual al ser más cercana a 1 o 100% se tiene un mejor desempeño.

Este análisis es bastante útil para bases de datos desbalanceadas, dado que es independiente de la tasa de respuestas positivas que se tengan, por lo que en este estudio se utilizará como complemento para seleccionar modelos.

## 6 METODOLOGÍA Y DESARROLLO METODOLÓGICO

Con el fin de obtener los objetivos propuestos anteriormente, se diseñó una metodología dividida en etapas que constituyen el proyecto. A continuación, se explican estas etapas y los principales resultados que se obtienen de cada una:

### 6.1 Selección de Datos y Pre-Procesamiento

La empresa en la que se desarrolla el proyecto cuenta con almacenamiento de datos (Data Warehouse) con distintas fuentes, tales como el sitio web, el uso de la tarjeta de crédito, información contractual, entre otros, el cual contiene un gran número de datos, por lo que es de suma importancia definir cuáles serán relevantes para poder accionar sobre ellos y permitan modelar el fenómeno en estudio.

Para lograr trabajar con la alta cantidad de datos con los que cuenta la empresa, se hace uso de lenguaje SQL para comenzar a extraer la información necesaria de las distintas bases de datos y la consolidación en una base analítica completa para realizar los análisis posteriores.

A continuación, se explican las bases de datos utilizadas y la principal información que se obtuvo de cada una:

- ✓ Demográfica: Muestra información individual por cliente, como edad, género, zona geográfica, entre otros.
- ✓ Contrato: Muestra información del contrato que un cliente tiene con la empresa, como qué tipo de tarjeta tiene, el cupo que tiene en sus tarjetas, fecha de apertura de la cuenta, entre otros.
- ✓ Transacciones: Muestra información acerca de las compras que realiza el cliente, tanto dentro como fuera de las empresas pertenecientes al holding, incluyendo categoría o rubro de la transacción.
- ✓ Productos Financieros: Muestra información acerca de los avances en efectivo y créditos de consumo realizados por el cliente.
- ✓ Web: Muestra información acerca de los clics que realiza un cliente en las diferentes secciones del sitio web de la empresa, incluyendo la cantidad y la fecha en que sucede.
- ✓ Campañas: Muestra información acerca de las campañas de marketing realizadas con los diferentes atributos que las representa, como objetivo, formato, público objetivo, entre otros.

Utilizando la información contenida en las bases de datos mencionadas, se consolidan tres bases analíticas para lograr entender diferentes comportamientos de los clientes web.

Primero, con la finalidad de estudiar a todos los clientes de la empresa, se genera una base analítica que incluye tanto a clientes que ingresan al sitio web como los que no ingresan, en donde el fenómeno que se quiere estudiar es el del ingreso al sitio web, el cual se modela como variable dependiente, en donde se tiene un universo de 2.033.667 clientes, y se espera comprender las características de quienes ingresan para luego realizar acciones de marketing que permitan migrar clientes hacia el canal digital.

Luego, una segunda base analítica, corresponde a un subconjunto de la anterior, la cual incluye a los clientes que ingresaron durante los meses de Febrero y Marzo 2016, en donde se tiene un universo de 627.713 clientes, la cual contiene las interacciones con todos los tipos de publicidad web, es decir, los tres formatos disponibles: Shadow Box, Caluga y Burbuja. Con esta base se trabajará para estudiar la interacción con la publicidad web en general y también a nivel de formato.

Finalmente, se genera una base analítica correspondiente al estudio particular de una campaña publicitaria, la cual está relacionada con el club de puntos de la empresa, con el fin de entender si existen diferencias significativas con el estudio por formato. En este caso particular, la base contiene un universo de 57.802 clientes.

## **6.2 Descripción de Variables**

Para comprender el universo de clientes sobre el cual se está trabajando y la composición de los datos que los representan, se genera un análisis descriptivo de las diferentes variables consideradas, diferenciando entre variables de acuerdo a su naturaleza, categóricas o continuas.

### **6.2.1 Variables cuantitativas**

En primer lugar, en la **Tabla 1**, se muestran los descriptivos de las variables cuantitativas que se incluyen en el análisis, mostrando el valor promedio, la desviación estándar, el máximo y el mínimo valor.

Variable	Cliente No Web				Cliente Web			
	Media	Desviación estándar	Mínimo	Máximo	Media	Desviación estándar	Mínimo	Máximo
Edad	53	14	19	99	42	13	19	99
Cupo (M\$)	1.122	1.054	1	80.000	1.480	1.265	1	50.000
Antigüedad	12	8	1	35	11	8	1	35
Ticket Promedio	47.391	70.700	-	19.014.071	46.666	63.312	-	10.273.923

Tabla 1: Descriptivos variables cuantitativas de clientes en estudio.  
Fuente: Elaboración propia.

Asimismo, como se puede ver en la **Tabla 2**, entre estas variables existe la mayor correlación entre la variable Antigüedad y las variables Edad y Cupo, sin embargo, se considera una correlación positiva moderada, por lo que, a priori, se mantienen ambas variables en el análisis.

Variables	Edad	Cupo	Antigüedad	Ticket Promedio
Edad	1	,106	,418	-,004
Cupo	-	1	,479	,148
Antigüedad	-	-	1	,031
Ticket Promedio	-	-	-	1

Tabla 2: Correlación<sup>3</sup> entre variables cuantitativas de clientes en estudio.  
Fuente: Elaboración propia.

La variable Edad para los clientes web se concentra cerca de 42 años, como se puede ver en el **Anexo I**, por lo que corresponden en su mayoría a adultos jóvenes, mientras que, para los clientes no web, se concentran cerca de 53 años, por lo que estos clientes son mayores que los que ingresan.

La variable de Cupo muestra el monto en dinero que dispone cada cliente para gastar y se tiene que un 47,4% para clientes web y un 61,5% para clientes no web tiene un monto de cupo menor o igual a \$1.000.000, un 28,1% para clientes web y un 22,7% para clientes no web entre \$1.000.000 a \$2.000.000 y un 24,5% para clientes web y un 15,8% para clientes no web un cupo mayor a \$2.000.000, por lo que la mayoría de los clientes cuenta con un bajo cupo.

## 6.2.2 Variables cualitativas

Las variables cualitativas que se incluyen en el estudio corresponden a las variables demográficas Género y Zona Geográfica, variables contractuales como Tipo de Cliente, Tipo de Tarjeta y Tenencia Adicional, variables transaccionales como Contrato de PAT, Transacciones en Comercio Interno (empresas pertenecientes al holding) y

<sup>3</sup> La correlación es significativa en el nivel 0,01 (bilateral) para cada combinación.

Transacciones en Comercio Externo, variables de productos financieras como Solicita Avance y Solicita Crédito de Consumo y variables que corresponden a métricas de la empresa sobre las preferencias de un cliente por ciertos usos de la tarjeta de crédito.

El universo de clientes se divide en 50,9% de género masculino y 49,1% de género femenino. En cuanto a la variable Zona Geográfica, indica en qué zona de Chile reside o declara el cliente para la apertura de su cuenta de tarjeta de crédito, la cual se detalla a continuación:

- Norte Grande: I, II y XV regiones de Chile.
- Norte Chico: III y IV regiones de Chile.
- Central: V y VI regiones de Chile.
- Sur: VII, VIII, IX y XIV regiones de Chile.
- Austral: X, XI y XII regiones de Chile.

Para el caso de la Región Metropolitana, debido a la su alta población, se subdivide en 6 zonas:

- Norte RM: Considera las comunas del norte de Santiago, como Recoleta, Independencia, Conchalí, entre otras.
- Poniente RM: Considera las comunas de la zona poniente de Santiago, como Maipú, Pudahuel, Quinta Normal, entre otras.
- Sur RM: Considera las comunas de la zona sur de Santiago, como Lo Espejo, Puente Alto, San Ramón, entre otras.
- Oriente RM: Considera comunas de la zona oriente de Santiago, como Las Condes, Lo Barnechea y Providencia, entre otras.
- Centro RM: Considera la comuna de Santiago.
- Exterior RM: Considera las comunas en las afueras de Santiago que pertenecen a la Región Metropolitana.

Como se muestra en la **Ilustración 7**, los clientes del universo de estudio se concentran en su mayoría en la Región Metropolitana (50,7% web y 42,9% no web), mientras que los clientes pertenecientes al Norte Chico representan la menor proporción (5,1% clientes web y no web).

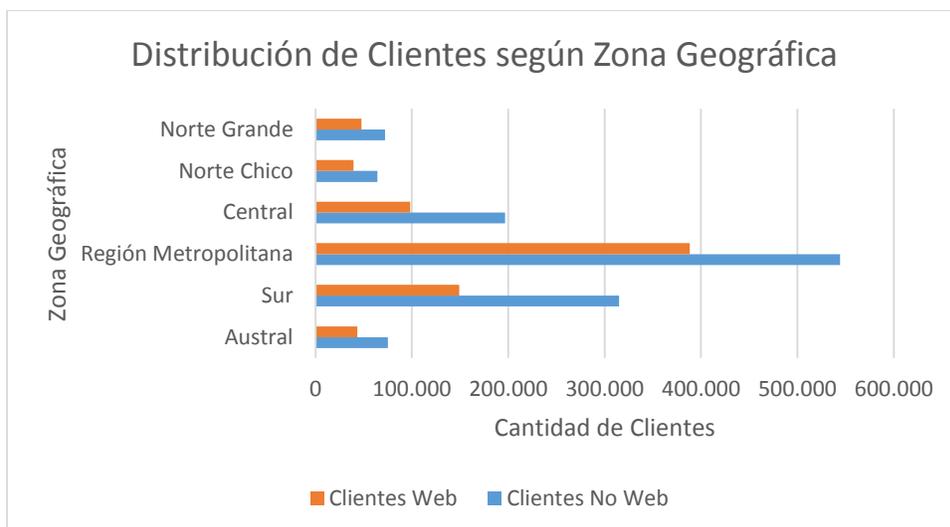


Ilustración 7: Distribución de clientes según zona geográfica.  
Fuente: Elaboración propia.

Luego, como se muestra en la **Ilustración 8**, para los clientes de la Región Metropolitana, se concentran en las comunas del Sur de la región (32,4% clientes web y 29,5% no web), mientras que las comunas del Centro de la región representan la menor proporción (4,4% clientes web y 6,9% no web).

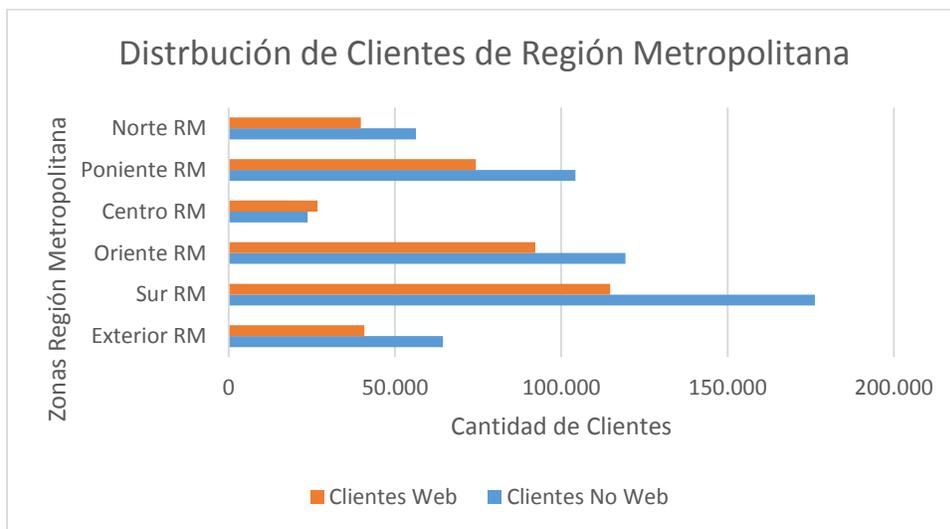


Ilustración 8: Distribución de clientes de la región metropolitana según zona geográfica.  
Fuente: Elaboración propia.

Con respecto a las variables contractuales, se tiene que la Categoría del Cliente, la cual se define en función de cómo interactúa un cliente con el club de lealtad de la empresa, se divide en 78,2% web y 93% web que corresponden al nivel bajo, un 17,8% web y 6,3% no web a un nivel medio y un 4,0% web y 0,7% no web que son de nivel alto.

En cuanto al Tipo de Tarjeta, la base está compuesta por un 82,7% de clientes web y un 71,0% no web con tarjeta Visa. Finalmente, en la Tenencia de Adicional se observa que un 21,3% de clientes web y un 24,0% no web efectivamente cuenta con una tarjeta adicional.

Con respecto a las variables transaccionales, se tiene que el Contrato de PAT se compone por un 8,2% de clientes web y 7,6% no web que efectivamente contratan este servicio. En cuanto a las transacciones por comercio, se explican en la **Tabla 3** para el Comercio Interno<sup>4</sup> y en la **Tabla 4** para el Comercio Externo<sup>5</sup>.

Comercio Interno	% Clientes Total	% Clientes Sitio Web
<b>Retail</b>	85,6%	30,7%
<b>Mejoramiento del Hogar</b>	68,1%	7,2%
<b>Supermercados</b>	35,4%	32,0%

*Tabla 3: Porcentaje de clientes que han realizado transacciones en comercios internos.*

*Fuente: Elaboración propia.*

Rubro Comercio Externo	% Clientes No Web	% Cliente Web
<b>Auto - Accesorios</b>	7,2%	15,5%
<b>Clínica</b>	8,4%	20,6%
<b>Combustible</b>	13,3%	29,9%
<b>Farmacia</b>	22,2%	39,8%
<b>Mejoramiento del Hogar</b>	9,1%	16,4%
<b>Tienda por Departamentos</b>	34,9%	55,8%
<b>Educación</b>	24,7%	41,7%
<b>Restaurant</b>	10,2%	30,2%
<b>Supermercados</b>	22,1%	43,1%
<b>Telefonía</b>	6,1%	15,8%
<b>Viajes</b>	8,7%	27,9%
<b>E-Commerce</b>	1,9%	6,8%

*Tabla 4: Porcentaje de clientes que han realizado transacción en comercios externos.*

*Fuente: Elaboración propia.*

Con respecto a las variables que corresponden al consumo de productos financieros, se tiene que un 30,6% de clientes web y un 23,4% de clientes no web ha solicitado un Avance, mientras que un 17,2% web y un 9,8% no web ha solicitado un Crédito de Consumo.

<sup>4</sup> En el contexto del proyecto, por Retail se hará referencia a Tienda por Departamentos.

<sup>5</sup> Se hará la diferencia entre negocios pertenecientes al holding como Internos y negocios pertenecientes a la competencia como Externos.

Finalmente, con respecto a las métricas de la empresa, se tiene que un 20,8% de clientes web y un 23,6% de los no web se consideran enganchados con el Financiamiento, un 24,9% web y un 18,1% no web enganchados con Descuentos y finalmente un 38,1% web y un 18,7% no web enganchados con los Puntos del club de lealtad [6].

### 6.2.3 Variables dependientes

Se tiene que una variable dependiente o en estudio corresponde al ingreso al sitio web, en particular, al inicio de sesión, con el fin de poder identificar al cliente. En este caso, se tiene un 37,7% de los clientes del universo como caso positivo.

Para el caso de la interacción con los diferentes formatos de las campañas web se tiene que un 8,3% de los clientes ha clicado en Shadow Box, un 7,3% de los clientes ha clicado en Calugas y un 5,5% de los clientes en Burbujas.

Finalmente, para el caso de la campaña de puntos en estudio, se tiene que un 10,5% de los clientes ha clicado.

## 6.3 Análisis Exploratorio

Con la finalidad de comprender cómo se relacionan las variables antes descritas con las variables dependientes, se realizan análisis univariados entre las características de los clientes y el fenómeno en estudio. A continuación, se muestran los más relevantes:

### 6.3.1 Ingresos al sitio web<sup>6</sup>

Para el caso del ingreso al sitio web, se estudia en primera instancia cómo se relaciona la proporción de clientes que efectivamente ingresan al sitio e inician sesión de acuerdo a características demográficas.

Realizando un análisis univariado para cada característica, se puede apreciar en la **Ilustración 9**, que no existen muchas diferencias de ingresos entre mujeres y hombres al sitio web de la empresa, por lo que este rasgo no entrega mayor información acerca de quiénes son los que ingresan al sitio web.

---

<sup>6</sup> Anexos III al VI muestran relación otras variables relevantes.

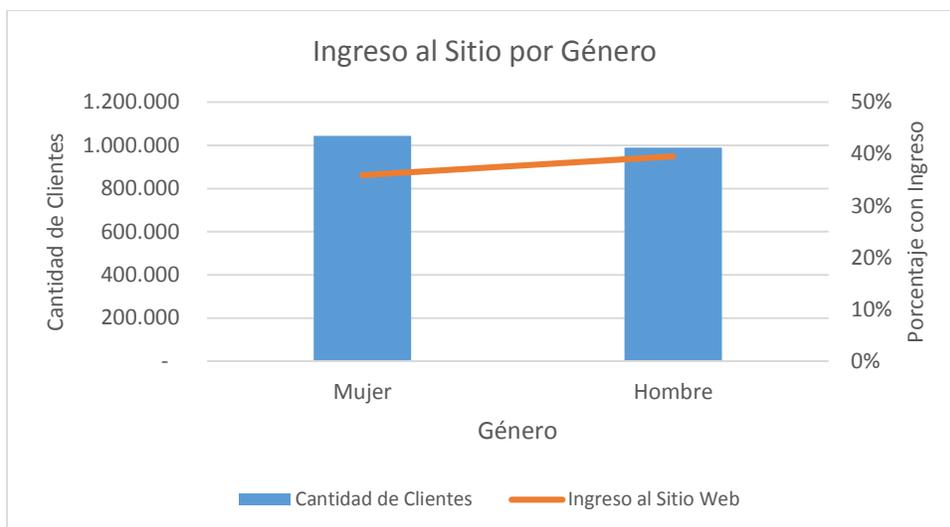


Ilustración 9: Relación entre ingresos al sitio web y el género del cliente.  
Fuente: Elaboración propia.

Por otro lado, se puede ver en la **Ilustración 10**, que la edad es un rasgo bastante relevante al momento de ingresar al sitio web, teniendo que a medida que aumenta la edad del cliente, hay una menor proporción que ingresa al sitio web, lo que hace sentido con la intuición que, al ser un estudio del canal digital, se tenga mayor penetración en rangos etarios menores.

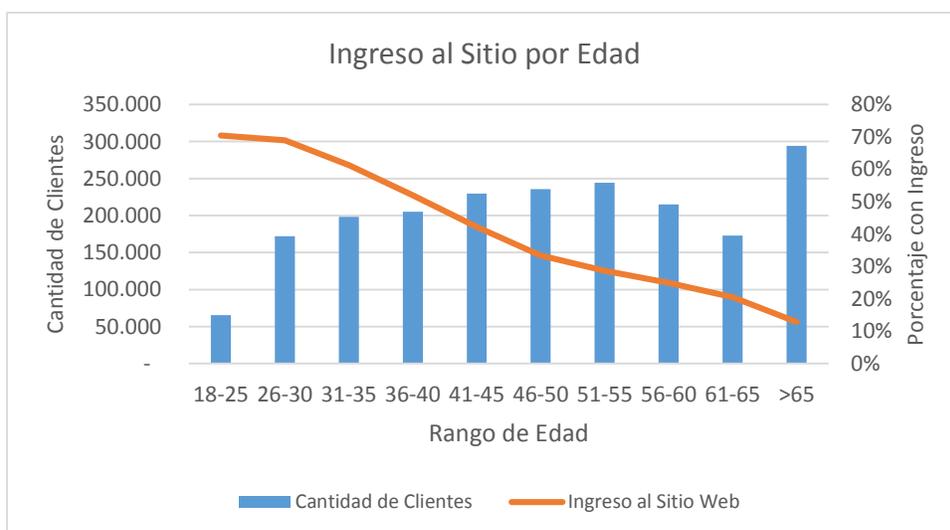
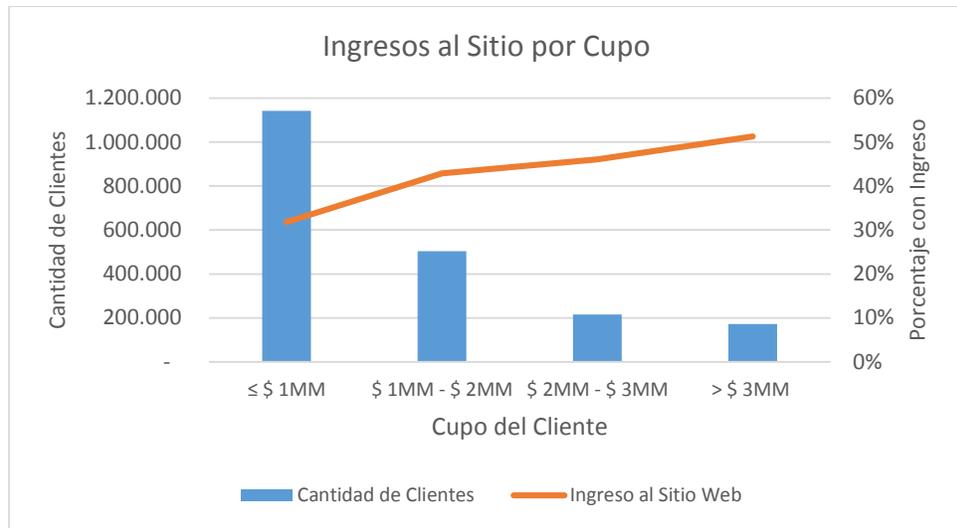


Ilustración 10: Relación entre ingresos al sitio web y la edad del cliente.  
Fuente: Elaboración propia.

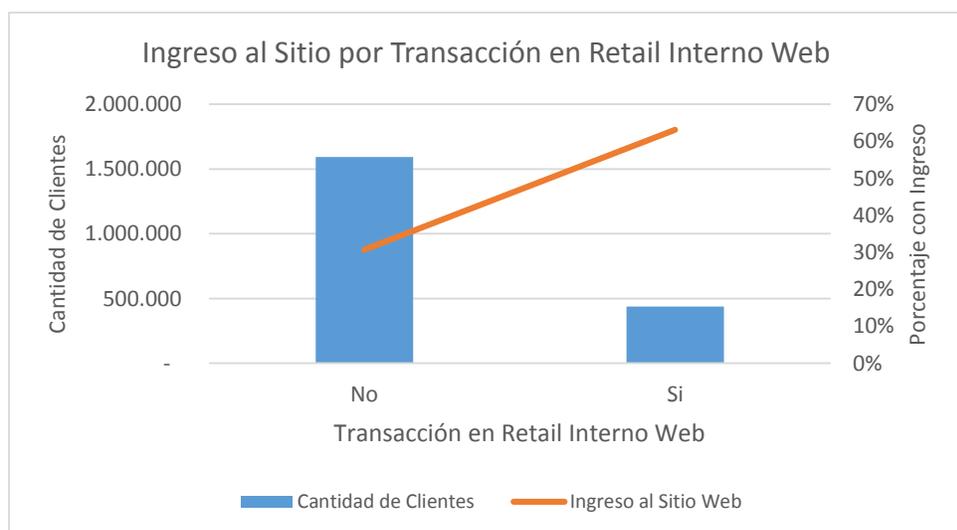
Luego, al comparar el ingreso al sitio web con el resto de las variables, se obtiene que las que más información entregan corresponden al cupo y las transacciones en el sitio web del formato retail perteneciente al holding.

Como se muestra en la **Ilustración 11**, existe una correlación positiva entre mayor cupo e ingresos al sitio, por lo que se espera obtener un efecto de poder económico como factor relevante en el estudio.



*Ilustración 11: Relación entre ingresos al sitio web y el cupo del cliente.  
Fuente: Elaboración propia.*

Finalmente, como se muestra en la **Ilustración 12**, los clientes que han realizado al menos una transacción en el sitio web del retail interno tienen una tasa bastante superior al resto de los clientes, lo que se puede explicar debido a que quienes ya compraron mediante el canal digital se encuentran más familiarizados, por lo que ingresar al sitio web de la empresa se vuelve natural.



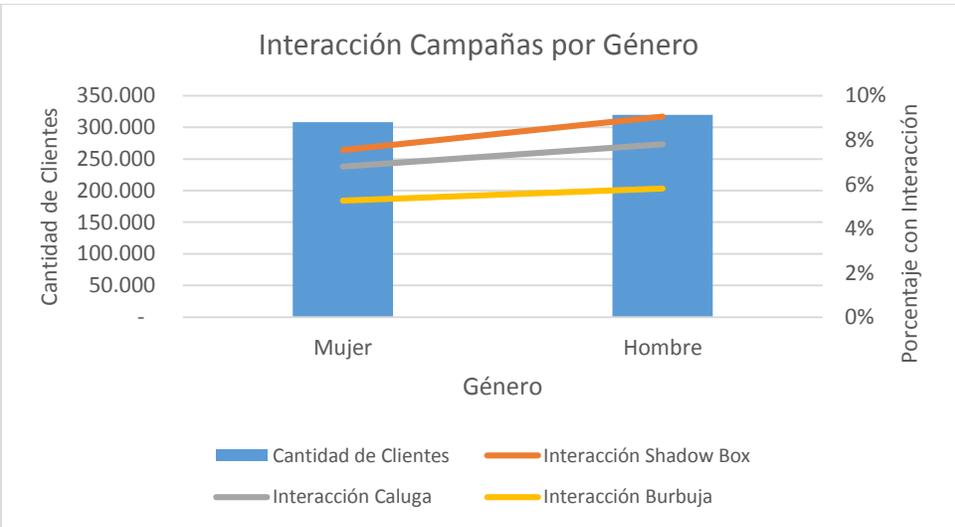
*Ilustración 12: Relación entre ingresos al sitio web y transacciones en la web del retail interno.  
Fuente: Elaboración propia.*

En base a estos resultados, se observó que en general los clientes que ingresan al sitio web presentan características bastante intuitivas, como ser en general adultos jóvenes, con altos cupos y ya familiarizados al canal, sin embargo, contar con un respaldo cuantitativo de esta intuición se considera como información relevante para posteriores tomas de decisiones.

**6.3.2 Interacción con campañas de publicidad<sup>7</sup>**

Para el caso de la interacción con la publicidad web por medio de clics en Shadow Box, Calugas y Burbujas, a priori se estudia cómo se relacionan con las principales características demográficas, tales como género y edad, en particular esta última al tratarse de un fenómeno digital.

En primer lugar, se puede apreciar que en la diferencia entre el género no representa mayores diferencias en términos del porcentaje de clientes que realiza una interacción, como se muestra en la **Ilustración 13**:



*Ilustración 13: Relación entre interacción con campañas y el género del cliente. Fuente: Elaboración propia.*

Luego, en relación con la edad de los clientes, se tiene que no existe una correlación clara, como se muestra en la **Ilustración 14**, a pesar de que a partir de los 50 años en adelante comienza a haber una baja, no se puede observar mayor tendencia.

<sup>7</sup> Anexos VII al IX muestran relación con otras variables relevantes.

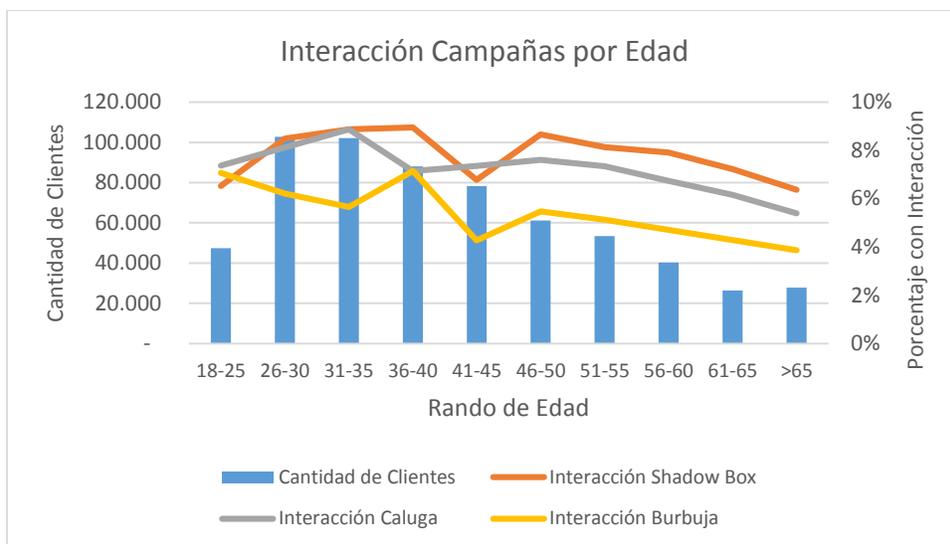


Ilustración 14: Relación entre interacción con campañas y la edad del cliente.  
Fuente: Elaboración propia.

Ahora, con respecto a las variables transaccionales incluidas en el análisis, se muestran las más relevantes.

Se tiene que existe una correlación alta entre haber realizado una transacción en el sitio web del negocio de retail del holding y haber interactuado con Shadow Box y Calugas en particular, como se muestra en la **Ilustración 15**. Para el caso de Burbujas no se ve una relación clara.

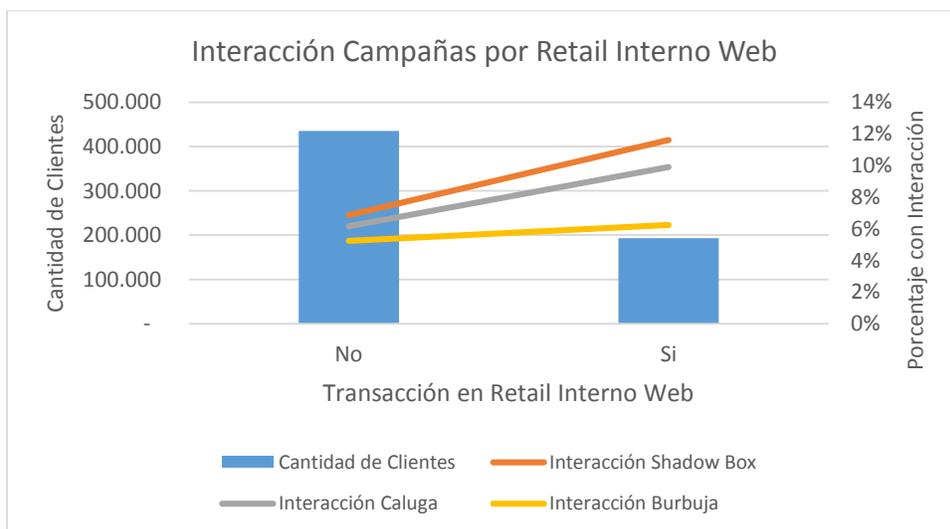


Ilustración 15: Relación entre interacción con campañas y transacciones en la web del retail interno.  
Fuente: Elaboración propia.

Se produce una relación similar entre la interacción con Shadow Box y Calugas y haber realizado una transacción en el rubro de viajes externo (**Ilustración 16**) y el rubro

de restaurants (**Ilustración 17**), es decir, se ve un aumento en el porcentaje de clientes con interacción en los casos que efectivamente realizan transacciones en los rubros mencionados. Para el caso de Burbujas, no se ve una correlación clara.

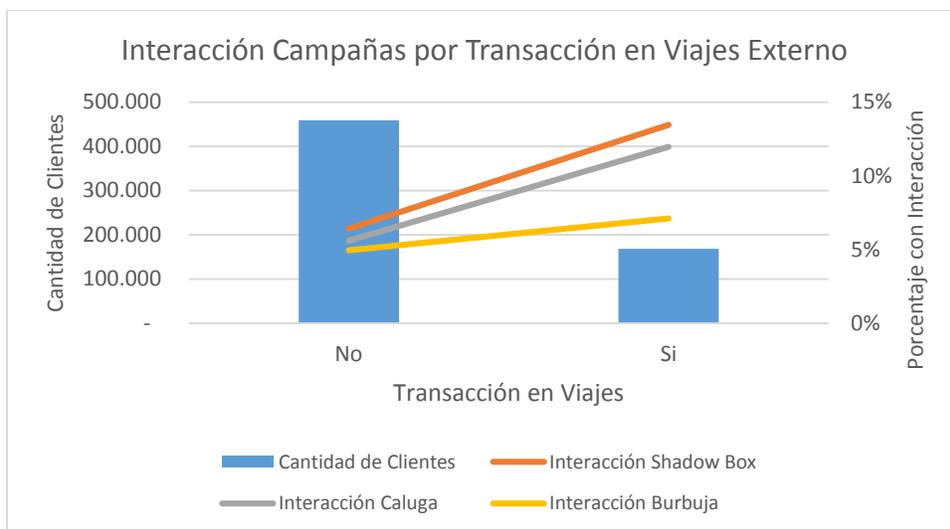


Ilustración 16: Relación entre interacción con campañas y transacciones en viajes externo.  
Fuente: Elaboración propia.

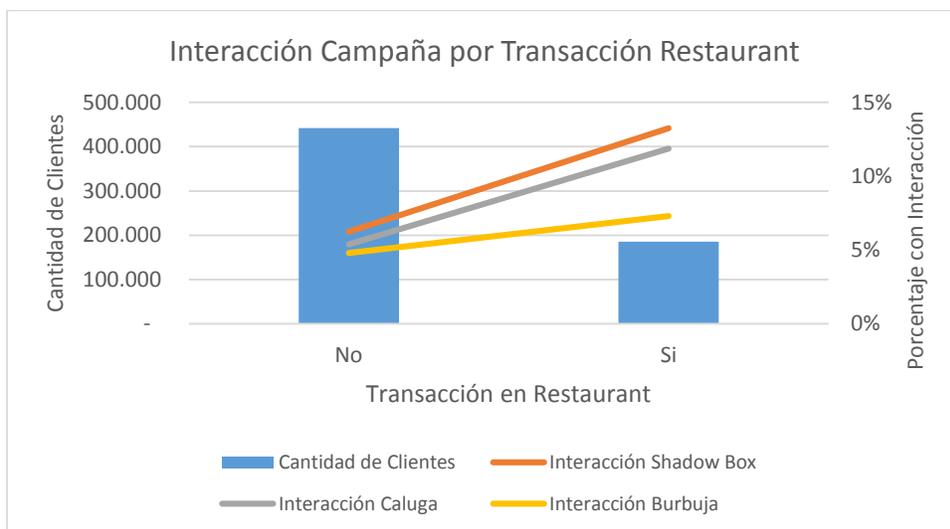


Ilustración 17: Relación entre interacción con campañas y transacciones en restaurant.  
Fuente: Elaboración propia.

En base a este análisis, se puede apreciar a priori, que clientes que clickean Shadow Box y Calugas son bastante similares, o incluso los mismos clientes. Por otro lado, se tiene que el cliente que presenta un comportamiento transaccional diferente.

## 6.4 Modelos de Propensión

En esta etapa de la metodología, se pretende modelar la interacción del cliente con el sitio web de la empresa, ya sea el ingreso al sitio web o el clic en las diferentes campañas de publicidad que se exhiben en él, para lo cual se utiliza la información obtenida en las etapas previas de análisis para determinar qué variables incluir en los posibles modelos.

Los modelos que se utilizarán para describir los diferentes fenómenos en estudio, corresponden a Regresión Logística (Logit), herramienta estadística proveniente de la econometría, y Árboles de Decisión, algoritmo proveniente de la minería de datos. Ambos modelos utilizarán el mismo conjunto de clientes para medir cómo afectan las distintas variables.

Dada la naturaleza de los fenómenos en estudio, la cual corresponde a problemas de clasificación en la que no se tienen igual cantidad de casos en cada clase, es decir, se tiene una base de datos desbalanceada, y con el fin de evitar la paradoja de la precisión (Accuracy Paradox), es decir, que la precisión sea alta por solo reflejar la clase que tiene la mayor distribución, es que se estudian posibles métodos para eliminar este sesgo.

De acuerdo a la literatura, se tienen principalmente dos opciones para balancear una base de datos de ésta índole. Una opción corresponde a realizar un modelo ponderado que resuelva el efecto mencionado anteriormente, mientras que la segunda opción corresponde a realizar un balance de la base de entrenamiento. En este caso, se utiliza el método del balance, el cual utiliza muestreos aleatorios para obtener una proporción cercana al 50/50 entre las diferentes clases que se tienen, que corresponden a casos positivos y negativos para el caso de los modelos de propensión que se muestran a continuación.

Asimismo, otra problemática que tienen los métodos de clasificación corresponde al punto de corte, el cual corresponde al valor umbral que define a partir de qué probabilidad se considera un caso positivo o negativo, el cual por defecto corresponde a 50%, sin embargo, dependiendo del problema puede ser mayor o menor. Para obtener un punto óptimo de corte, se realiza para cada modelo un cruce entre las curvas de especificidad y sensibilidad, siendo la intersección el óptimo. En general, la mayoría de los modelos presenta puntos óptimos de cortes muy cercanos al 50%, pero siempre menores, esto debido al balance que se realiza.

Así, recapitulando, antes de modelar cada fenómeno mediante Regresión Logística o Árbol de Decisión, se separa la base de datos en dos muestras aleatorias, una (75%) para entrenar los modelos, ya que son modelos supervisados, y una segunda

base (25%) para validar los resultados que entrega. Luego, sobre la base de entrenamiento se realiza el balance, obteniéndose una base con una distribución de clases cercana al 50/50 entre casos positivos y negativos, mientras que la base de validación o prueba mantiene su distribución original. Con una base analítica procesada y los puntos óptimos de cortes definidos se procede a modelar los diferentes fenómenos.

## 6.4.1 Ingreso al Sitio Web

Por modelo de ingreso al sitio web se refiere al fenómeno de ingresar a la web de la empresa e iniciar sesión, por lo que la variable dependiente será 1 en el caso de haber iniciado sesión durante el período en estudio, 0 en caso contrario. A continuación, se explica el modelo y sus principales resultados.

### 6.4.1.1 Regresión Logística

Uno de las metodologías propuestas para explicar el fenómeno en estudio corresponde a la regresión logística binaria, la cual se explica en la sección (5), y con el fin de obtener un buen ajuste, se realizaron diferentes configuraciones de las variables.

Los principales resultados del modelo con mayor ajuste y capacidad interpretativa se explican a continuación:

Variable	Coefficiente	Error estándar	Exp(Coef)
Edad	-0,062*	0,000	0,940
Categoría Alta	0,601*	0,021	1,825
Categoría Media	0,269*	0,008	1,309
Logaritmo del Cupo	0,310*	0,002	1,363
Transacción en Combustible	0,302*	0,006	1,352
Transacción en Clínicas	0,138*	0,008	1,148
Transacción en Restaurant	0,314*	0,007	1,369
Transacción en Telefonía	0,529*	0,008	1,697
Transacción en Viajes Externo	0,507*	0,007	1,660
Transacción en Supermercados Externo	0,116*	0,006	1,123
Transacción en Retail Web Interno	0,919*	0,006	2,507
Transacción en Mej. del Hogar Web Interno	0,944*	0,012	2,569
Enganchado Puntos	0,248*	0,006	1,282
Constante	-1,979*	0,032	0,138

Tabla 5: Resultados regresión logística modelo de ingreso al sitio web<sup>8</sup>.  
Fuente: Elaboración propia.

<sup>8</sup> NS corresponde a No Significativo, \* corresponde a un coeficiente significativo a  $P < 0,05$ .

Como se puede ver en la **Tabla 5**, existe un efecto negativo de la edad sobre el ingreso al sitio web, es decir, clientes más jóvenes son más propensos a ingresar, lo que va en línea con la intuición acerca del uso de herramientas digitales y del análisis exploratorio.

Se puede apreciar también un efecto positivo de pertenecer a las categorías alta y media, y también del efecto de estar enganchado con el club de puntos, por lo que los clientes que acumulan puntos con la empresa tienen mayor probabilidad de ingresar al sitio web, lo que puede estar relacionado a conocer el catálogo de productos canjeables o simplemente revisar sus puntos dentro de la página.

Asimismo, existe un efecto positivo de contar con cupos más altos, y también de haber realizado transacciones en los rubros de viajes externo, por lo que corresponde a clientes con poder económico y que utilizan su tarjeta de crédito para financiar rubros de gasto sobre el promedio.

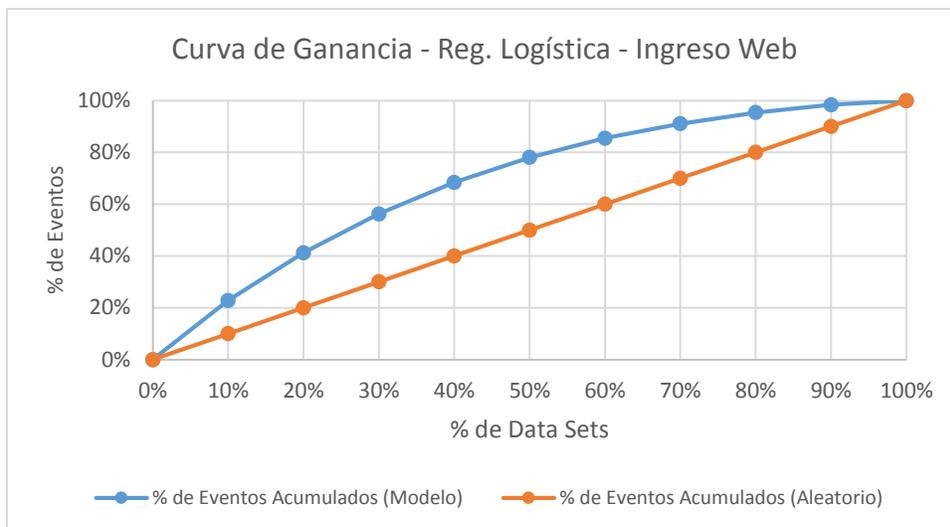
Finalmente, se tiene un efecto positivo importante relacionado a transacciones en sitios web del holding, tales como el sitio web del negocio de retail y de mejoramiento del hogar. Dado esto, se tiene que clientes que son compradores en línea, tienen mayor propensión a también ingresar a la web de la empresa.

Para medir el desempeño del modelo en términos cuantitativos, se presentan a continuación la matriz de confusión (**Tabla 6**) y la curva de ganancia de información (**Ilustración 18**). En base a la matriz de confusión se puede apreciar que el modelo clasifica a los clientes de manera correcta a más del 70% de los clientes, lo que se considera un nivel adecuado de sensibilidad y especificidad.

		Pronosticado			
		No Ingresa	Ingresa	Porcentaje correcto	
Observado	Entrenamiento	No Ingresa	427.724	147.740	74,3%
		Ingresa	162.433	411.949	71,7%
		Porcentaje Global			73,0%
	Validación	No Ingresa	235.009	81.751	74,2%
		Ingresa	54.660	137.195	71,5%
		Porcentaje Global			73,2%

Tabla 6: Matriz de confusión de regresión logística de modelo de ingreso al sitio web.  
Fuente: Elaboración propia.

Luego, como se muestra en la curva de ganancia de información en la **Ilustración 18**, tomando los tres primeros deciles con mayor probabilidad entregada por el modelo, se logra identificar poco menos de un 60% de los clientes que realmente ingresaron al sitio web.



*Ilustración 18: Curva de ganancia de información regresión logística de modelo de ingreso al sitio web. Fuente: Elaboración propia.*

Complementariamente, utilizando la curva COR (**Anexo X**) generada por las probabilidades estimadas por el modelo, se tiene un área bajo la curva (AUC) de 80,5% sobre los casos clasificados correctamente.

#### **6.4.1.2 Árbol de Decisión**

El método de árbol de decisión utilizado para modelar el ingreso al sitio web corresponde al CART, lo cual fue determinado principalmente por el desempeño que presentaba.

Como se muestra en la **Ilustración 19**, el árbol de decisión presenta como primera variable de corte a la edad del cliente, separando clientes entre menores y mayores de 44,5 años, obteniéndose una propensión de 57,4% para los clientes más jóvenes. Luego, para el caso de los clientes menores se tiene que la segunda variable discriminante corresponde a la transacción en la web del retail perteneciente al holding, mientras que para los clientes mayores la variable discriminante corresponde al estar enganchado con el club de puntos.

En estas ramas superiores se obtiene un nodo final, correspondiente al de clientes menores a 44,5 años y que han realizado transacciones en la web del retail interno, los

cuales presentan una propensión a ingresar al sitio web de 77,8%, lo que representa un lift de 2,06 para este nodo.

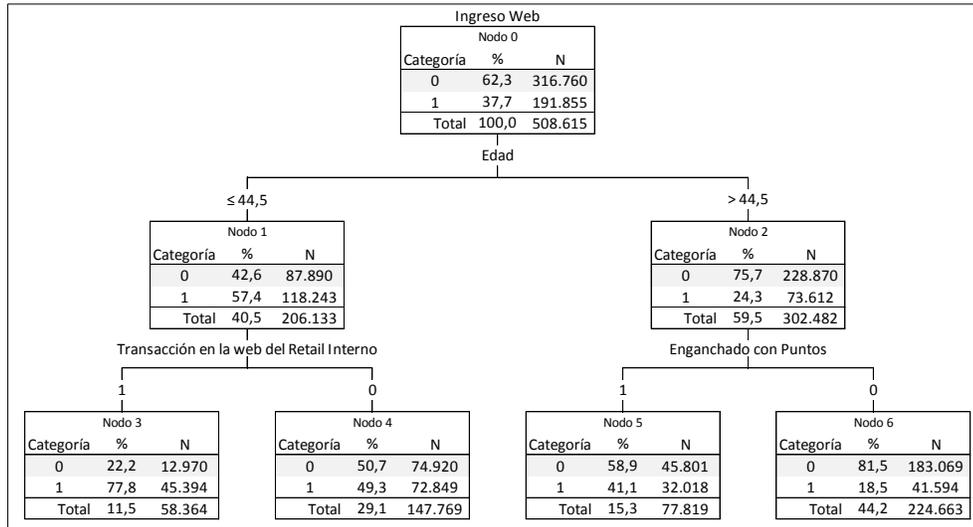


Ilustración 19: Ramas superiores árbol de decisión modelo de ingreso al sitio web.  
Fuente: Elaboración propia.

Luego, como se ve en la **Ilustración 20**, en las ramas inferiores del árbol de decisión, se incorporan como variables de corte las transacciones en el rubro de viajes, la edad y la compra en la web del retail interno nuevamente y el cupo del cliente, obteniendo en el nodo más propenso una probabilidad de 71,5%, lo que representa un lift de 1,90.

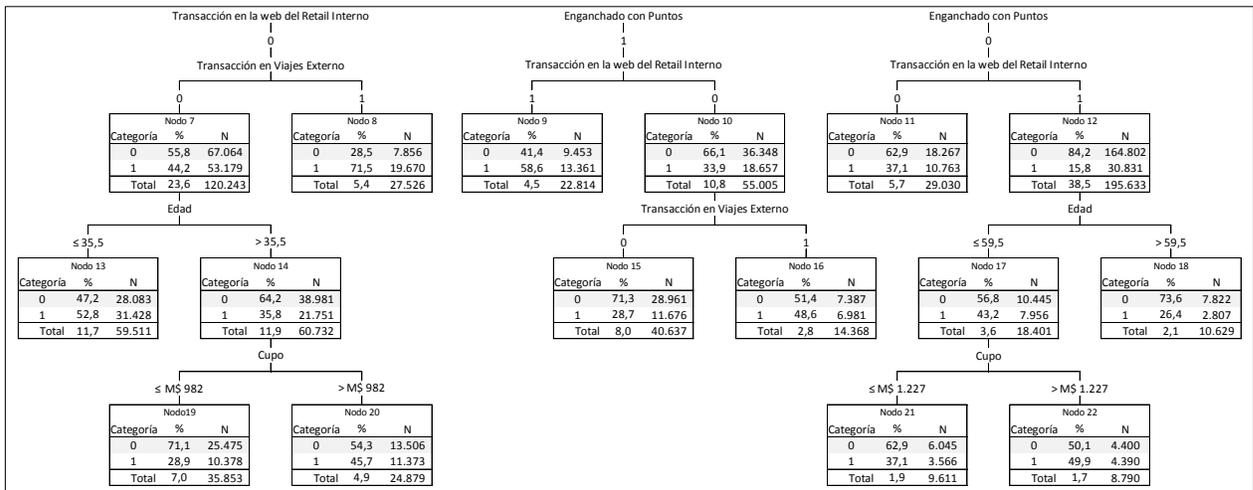


Ilustración 20: Ramas inferiores árbol de decisión modelo de ingreso al sitio web.  
Fuente: Elaboración propia.

Al igual que en el caso de la regresión logística, se utilizan como métricas de desempeño la matriz de confusión (**Tabla 7**) y la curva de ganancia de información (**Ilustración 21**).

En base a la matriz de confusión, se tiene que se está clasificando de manera correcta a cerca de un 70%, bastante similar a lo que entrega la regresión logística.

		Pronosticado			
		No Ingresa	Ingresa	Porcentaje correcto	
Observado	Entrenamiento	No Ingresa	423.649	151.815	73,6%
		Ingresa	176.629	397.753	69,2%
		Porcentaje Global			71,4%
	Validación	No Ingresa	233.105	83.655	73,6%
		Ingresa	59.258	132.597	69,1%
		Porcentaje Global			71,9%

Tabla 7: Matriz de confusión de árbol de decisión de modelo de ingreso al sitio web.  
Fuente: Elaboración propia.

Asimismo, la curva de ganancia de información no presenta mayores diferencias con el modelo de regresión logística, tomando el primer decil con mayor probabilidad se logra identificar más del 20% de los clientes que realmente ingresaron.

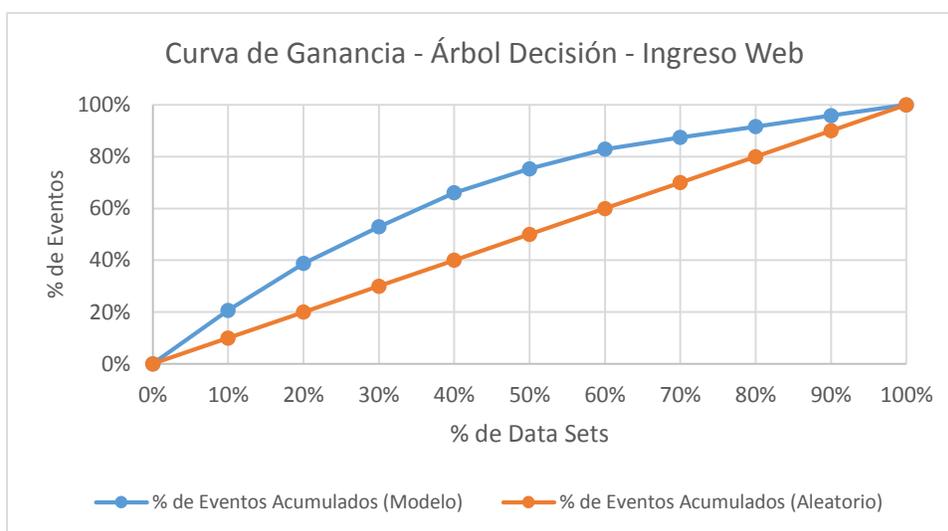


Ilustración 21: Curva de ganancia de información árbol de decisión de modelo de ingreso al sitio web.  
Fuente: Elaboración propia.

Complementariamente, se utilizando la curva COR (**Anexo XI**) generada por las probabilidades estimadas por el modelo, se tiene un área bajo la curva (AUC) de 76,2%, menor a lo que entrega la regresión logística.

## 6.4.2 Publicidad Web

Por modelo de publicidad web se refiere al fenómeno de hacer clic en cualquier tipo de campaña publicitaria que se exhiba en el sitio web, sin importar el formato que sea, por lo tanto, la variable dependiente será 1 en el caso de haber clicado cualquier publicidad web, 0 en caso contrario. A continuación, se explica cómo se realizó este modelo y los principales resultados.

### 6.4.2.1 Regresión Logística

Con el fin de obtener un modelo que entregue un ajuste aceptable, es que se realizaron distintas iteraciones de la regresión logística agregando e incluyendo diferentes variables y transformaciones de variables (por ejemplo, utilizar el logaritmo del cupo). Se seleccionó el modelo que entrega un rendimiento superior a la mayoría y que permita que la interpretación sea intuitiva.

Los principales resultados se explican a continuación:

Variable	Coefficiente	Error estándar	Exp(Coef)
<b>Categoría Alta</b>	1,002*	0,021	2,723
<b>Categoría Media</b>	0,553*	0,021	1,739
<b>Transacción en Farmacias</b>	0,078*	0,011	1,081
<b>Transacción en Restaurant</b>	0,263*	0,011	1,301
<b>Transacción en Retail Externo</b>	0,129*	0,012	1,137
<b>Transacción en Supermercado Externo</b>	0,099*	0,012	1,104
<b>Transacción en Retail Web Interno</b>	0,293*	0,009	1,341
<b>Solicita Avance en Efectivo</b>	0,342*	0,010	1,407
<b>Solicita Crédito de Consumo</b>	0,290*	0,013	1,337
<b>Enganchado Puntos</b>	0,221*	0,010	1,247
<b>Constante</b>	0,244*	0,023	1,276

Tabla 8: Resultados regresión logística modelo de publicidad web<sup>9</sup>.

Fuente: Elaboración propia.

En primer lugar, se tiene que la característica de tener una categoría alta o media tiene un efecto positivo sobre hacer clic en publicidad web, asimismo, dado que todas las variables consideradas corresponden a dicotómicas, tienen una alta magnitud, por lo que el nivel base (ser cliente de categoría baja), disminuye en gran medida la propensión.

<sup>9</sup> NS corresponde a No Significativo, \* corresponde a un coeficiente significativo a  $P < 0,05$ .

Luego, se puede ver que solicitar financiamiento, ya sea por medio de un avance o un crédito de consumo, tiene impacto positivo en la propensión de clickear, al igual que la métrica de enganche con el club de puntos o lealtad, lo que habla de clientes que están de cierta manera fidelizados por la empresa, ya que prefiere solicitar este financiamiento mediante la empresa antes que una entidad bancaria u otra casa comercial que entregue el mismo beneficio y el enganche con el club de lealtad habla de un cliente que acumula puntos en la empresa.

Asimismo, haber realizado transacciones en los rubros de restaurants tiene un impacto positivo en la propensión, al igual que comprar en el sitio web del retail interno, los cuales presentan magnitudes similares.

Finalmente, haber realizado transacciones en retail y supermercados externos también presenta signo positivo, sin embargo, en términos de magnitud, no es tan relevante en comparación con el resto de las variables.

Para medir el desempeño del modelo, se utiliza la matriz de confusión (**Tabla 9**) y la curva de ganancia de información (**Ilustración 22**).

En base a la matriz de confusión, se tiene un balance entre sensibilidad y especificidad, que sobrepasan el 60%, es decir, se clasifica de manera correcta a más del 60% de los clientes.

		Pronosticado			Porcentaje correcto
		No Ingresa	Ingresa		
Observado	Entrenamiento	No Ingresa	71.768	44.897	61,5%
		Ingresa	44.463	69.987	61,2%
		Porcentaje Global			61,3%
	Validación	No Ingresa	80.065	49.711	61,7%
		Ingresa	10.346	16.736	61,8%
		Porcentaje Global			61,7%

Tabla 9: Matriz de confusión de regresión logística de modelo de publicidad web.  
Fuente: Elaboración propia.

Por otro lado, de acuerdo a la curva de ganancia de información, se puede apreciar que tomando el primer decil ordenado según la probabilidad entregada por el modelo, se identifica a más de un 20% de los clientes que efectivamente clickean alguna publicidad.

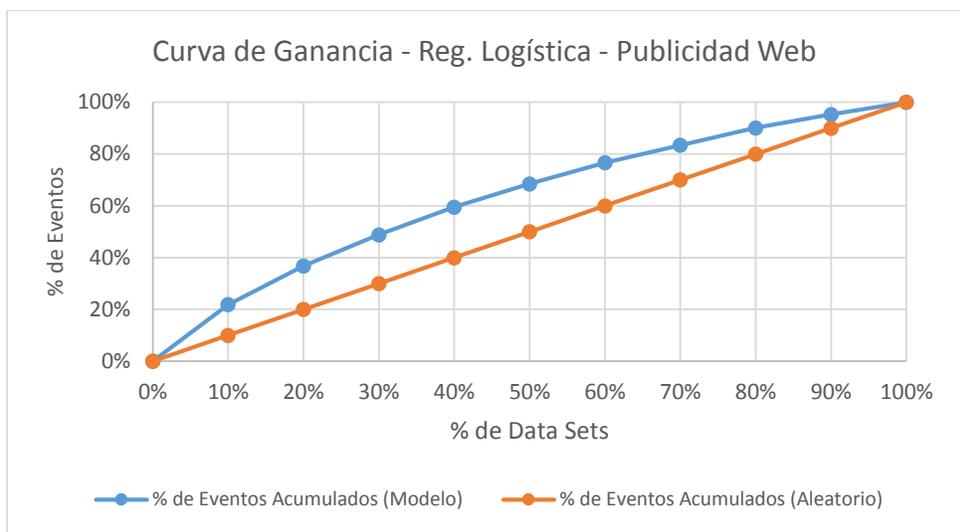


Ilustración 22: Curva de ganancia de información de árbol de decisión modelo de publicidad web.  
Fuente: Elaboración propia.

Complementariamente, en este caso, se tiene un área bajo la curva (AUC) de 65,8% (Anexo XII).

### 6.4.2.2 Árbol de Decisión

En el caso del modelo de publicidad web, el algoritmo CHAID, seleccionado en base al desempeño y la interpretación que se puede hacer del árbol. Se tiene que en las ramas superiores (Ilustración 23), la primera variable de corte corresponde a la categoría, siendo los clientes con alta categoría los más propensos. Luego, aparecen la transacción en restaurant, compra en la web del retail interno y solicitud de avance en efectivo.

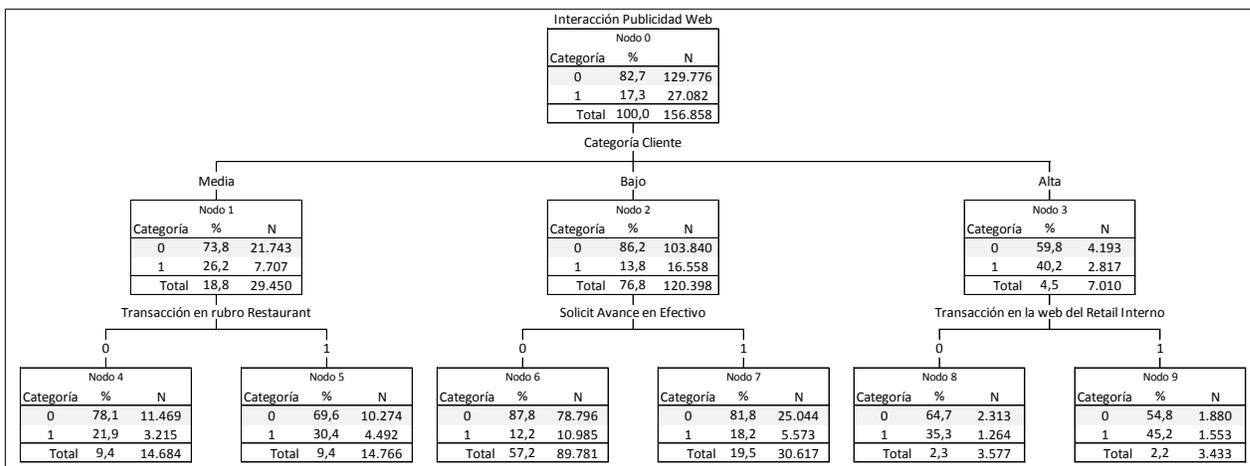


Ilustración 23: Ramas superiores árbol de decisión de modelo de publicidad web.  
Fuente: Elaboración propia.

Luego, como se ve en la **Ilustración 24**, en las ramas inferiores del árbol de decisión, se incorporan las variables de transacción en viajes, en supermercados externo, restaurant y el enganche con el club de puntos, obteniéndose que el mejor nodo tiene una propensión de 45,2%, lo que representa un lift de 2,61.

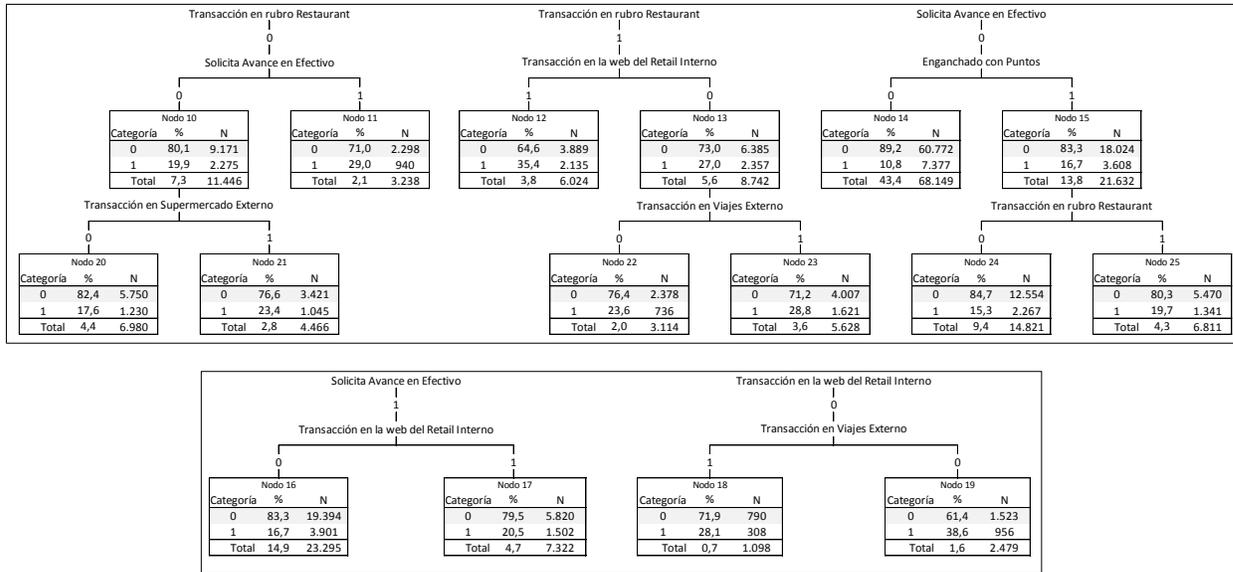


Ilustración 24: Ramas inferiores árbol de decisión de modelo de publicidad web.  
Fuente: Elaboración propia.

Al igual que en los casos anteriores, se tiene la matriz de confusión (**Tabla 10**) y la curva de ganancia de información (**Ilustración 25**).

En base a la matriz de confusión, se tiene un desempeño bastante similar al de la regresión logística, pero con menor sensibilidad y mayor especificidad.

		Pronosticado			Porcentaje correcto
		No Interactúa	Interactúa		
Observado	Entrenamiento	No Interactúa	74.553	42.112	63,9%
		Interactúa	47.913	66.537	58,1%
		Porcentaje Global			61,0%
	Validación	No Interactúa	82.757	47.019	63,8%
		Interactúa	11.328	15.754	58,2%
		Porcentaje Global			62,8%

Tabla 10: Matriz de confusión árbol de decisión de modelo de publicidad web.  
Fuente: Elaboración propia.

La curva de ganancia de información entrega el mismo desempeño en comparación a la regresión logística en términos prácticos.

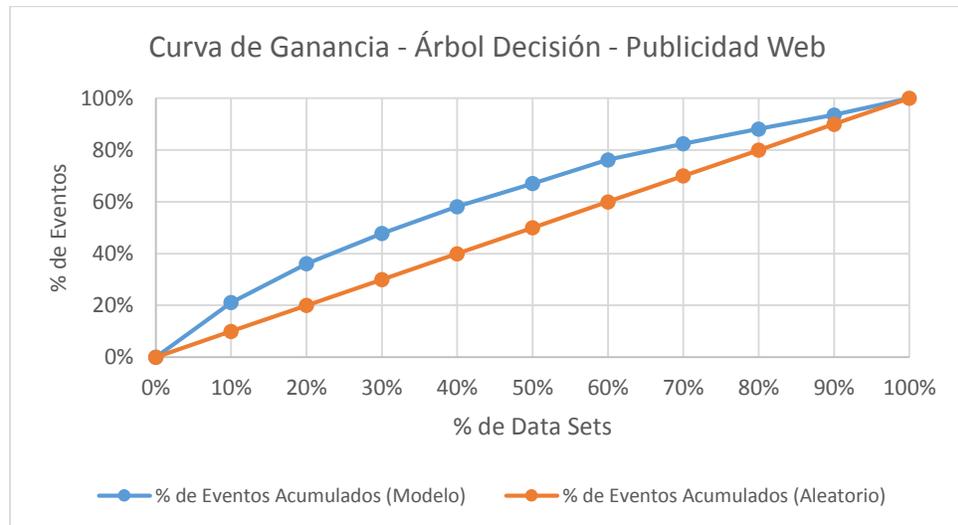


Ilustración 25: Curva de ganancia de información de árbol de decisión de modelo de publicidad web.  
Fuente: Elaboración propia.

Complementariamente, se tiene que la curva COR (**Anexo XIII**) posee un área bajo la curva (AUC) de 64,9% de casos correctamente especificados.

### 6.4.3 Calugas

Por modelo de Calugas se refiere al fenómeno de hacer clic en la publicidad que se exhibe solamente en formato Caluga, por lo tanto, la variable dependiente será 1 en el caso de haber clickeado una Caluga, 0 en caso contrario. A continuación, se explica cómo se realizó este modelo y los principales resultados.

#### 6.4.3.1 Regresión Logística

Al igual que en los modelos anteriores, se utilizaron diferentes combinaciones de variables y sus respectivas transformaciones con el fin de obtener el mejor modelo que se pueda con la información que se cuenta.

Los principales resultados se presentan a continuación:

Variable	Coficiente	Error estándar	Exp(Coef)
<b>Categoría Alta</b>	0,938*	0,023	2,555
<b>Categoría Media</b>	0,485*	0,023	1,625
<b>Transacción en Farmacias</b>	0,126*	0,012	1,135
<b>Transacción en Retail Externo</b>	0,120*	0,013	1,127
<b>Transacción en Educación</b>	0,148*	0,012	1,159
<b>Transacción en Restaurant</b>	0,252*	0,013	1,287
<b>Transacción en Viajes Externo</b>	0,247*	0,013	1,280
<b>Transacción en Retail Web Interno</b>	0,292*	0,011	1,339
<b>Transacción en Mej. Hogar Web Interno</b>	0,207*	0,019	1,230
<b>Solicita Avance en Efectivo</b>	0,221*	0,012	1,247
<b>Solicita Crédito de Consumo</b>	0,179*	0,015	1,196
<b>Enganchado Puntos</b>	0,253*	0,012	1,288
<b>Constante</b>	0,043 <sup>NS</sup>	0,025	1,044

*Tabla 11: Resultados regresión logística modelo de Calugas<sup>10</sup>.  
Fuente: Elaboración propia.*

En primer lugar, al igual que en modelos anteriores, se puede apreciar un efecto positivo y de gran magnitud en el hecho de pertenecer a categorías alta y media, en comparación al nivel base (categoría baja).

Luego, en el caso de Calugas se ve que haber realizado transacciones en los sitios web de los formatos de retail y mejoramiento del hogar pertenecientes al holding entrega una mayor propensión, por lo que los clientes que interactúan con este formato ya se encuentran familiarizados con el canal digital.

Por otro lado, también el haber realizado transacciones en los rubros de viajes y restaurant presentan un efecto positivo y una magnitud importante, por lo que los clientes que interactúan con Calugas corresponden a clientes de experiencias.

Finalmente, se tiene que el estar enganchado con el club de puntos presenta un efecto positivo importante también, lo que también se relaciona con que son clientes cercanos a la empresa.

En base a la matriz de confusión (**Tabla 12**), este modelo clasifica de manera correcta a más de un 60% de los clientes.

<sup>10</sup> NS corresponde a No Significativo, \* corresponde a un coeficiente significativo a  $P < 0,05$ .

		Pronosticado			
		No Interactúa	Interactúa	Porcentaje correcto	
Observado	Entrenamiento	No Interactúa	55.784	31.285	64,1%
		Interactúa	32.368	53.664	62,4%
		Porcentaje Global			63,2%
	Validación	No Interactúa	93.032	52.367	64,0%
		Interactúa	4.212	7.247	63,2%
		Porcentaje Global			63,9%

Tabla 12: Matriz de confusión de regresión logística de modelo de Calugas.  
Fuente: Elaboración propia.

Luego, en base a la curva de ganancia de información (**Ilustración 26**), se tiene que el primer decil de clientes ordenados por la probabilidad entregada por el modelo identifica a un 27% de los clientes que efectivamente han clicado este formato.

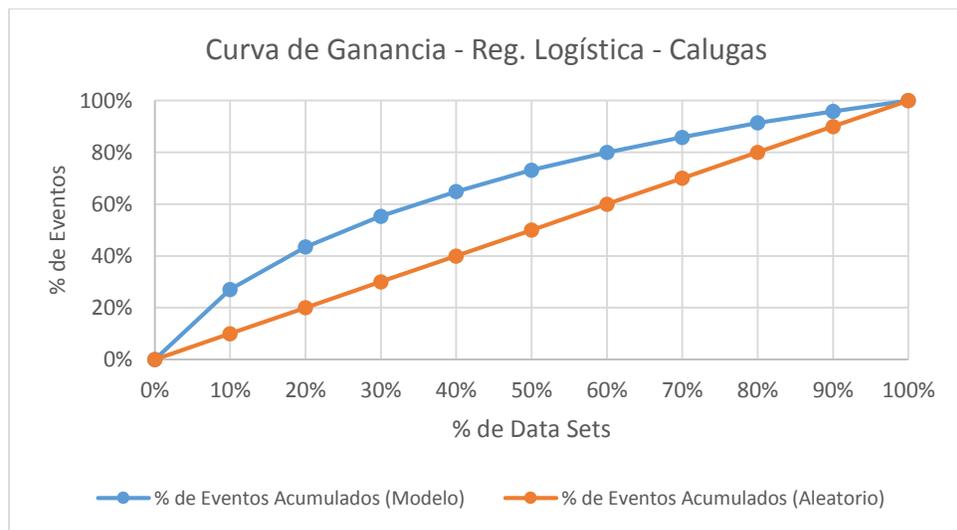


Ilustración 26: Curva de ganancia de información de regresión logística modelo de Calugas.  
Fuente: Elaboración propia.

Complementariamente, la curva ROC (**Anexo XIV**) posee un área bajo la curva (AUC) de 68% sobre los casos clasificados correctamente.

### 6.4.3.2 Árbol de Decisión

En el caso del modelo de Calugas, el algoritmo utilizado corresponde al tipo CART, utilizando los mismos criterios que para casos anteriores.

Se tiene que, para las ramas superiores del árbol, como se muestra en la **Ilustración 27**, la primera variable discriminante corresponde a la categoría a la que pertenece el cliente, siendo las categorías más altas las que presentan mayor aumento de la propensión. Luego, aparece como variables de corte las transacciones en restaurant para las categorías medias y altas y el estar enganchado en puntos para el caso de la categoría baja.

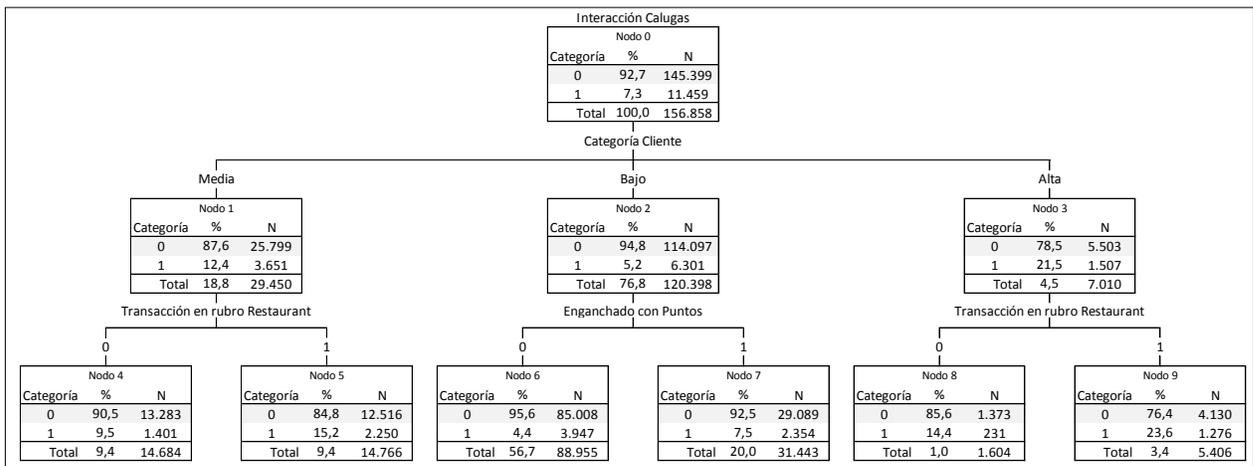


Ilustración 27: Ramas superiores árbol de decisión modelo de Calugas.  
Fuente: Elaboración propia.

Luego, en las ramas inferiores (**Ilustración 28**), se puede ver que aparecen las variables de corte de transacción en la web del retail interno, transacción en viajes, transacción en farmacias, transacción en educación, solicita avance en efectivo y se repiten las variables de enganche con puntos y de transacción en restaurant, obteniéndose que el mejor nodo tiene una probabilidad de 26,9%, lo que representa un lift de 3,68.

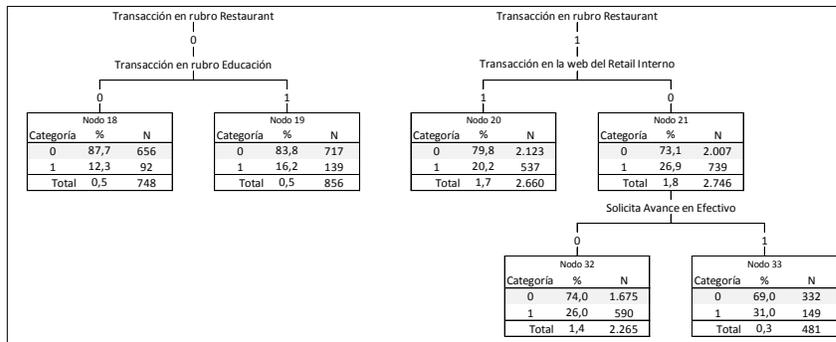
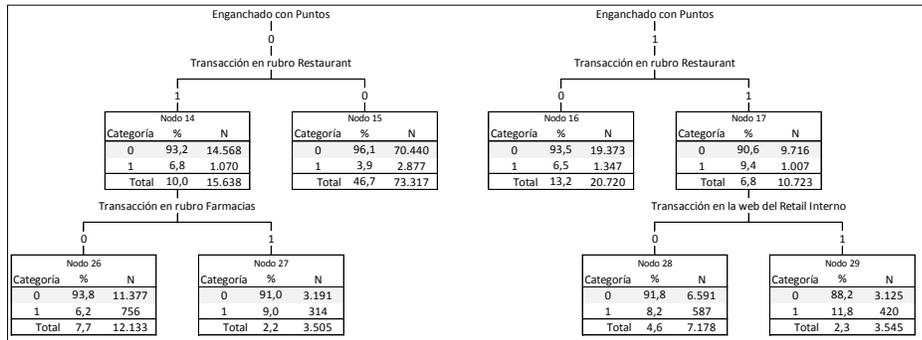
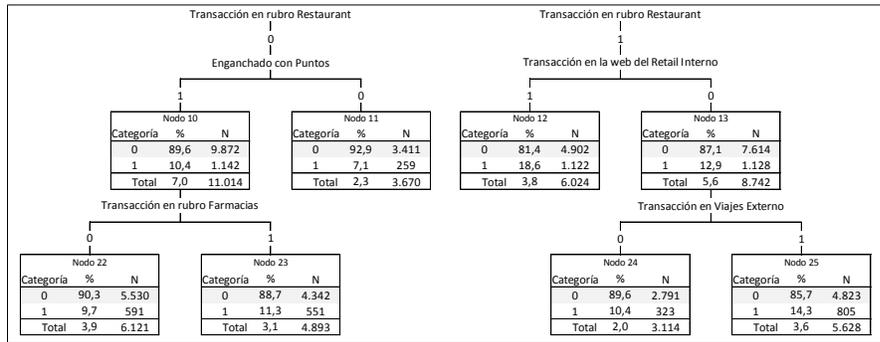


Ilustración 28: Ramas inferiores árbol de decisión modelo de Calugas.  
Fuente: Elaboración propia.

Al revisar la matriz de confusión (**Tabla 13**), se tiene un desempeño menor al de la regresión logística, en términos de la sensibilidad, sin embargo, en términos de precisión mejora, lo que se produce debido al desbalanceo de la base de validación.

		Pronosticado			
		No Interactúa	Interactúa	Porcentaje correcto	
Observado	Entrenamiento	No Interactúa	59.934	27.135	68,8%
		Interactúa	36.918	49.114	57,1%
		Porcentaje Global			63,0%
	Validación	No Interactúa	99.769	45.630	68,6%
		Interactúa	4.820	6.639	57,9%
		Porcentaje Global			67,8%

Tabla 13: Matriz de confusión de árbol de decisión de modelo de Calugas.  
Fuente: Elaboración propia.

Por otro lado, al ver la curva de ganancia de información (**Ilustración 29**), se tiene que es muy similar al caso de la regresión logística, siendo el primer decil de clientes ordenados en base a la probabilidad del modelo un 26% de quienes efectivamente han clicado este formato, por lo que en términos del aporte en información no hay un modelo que supere al otro en la práctica.

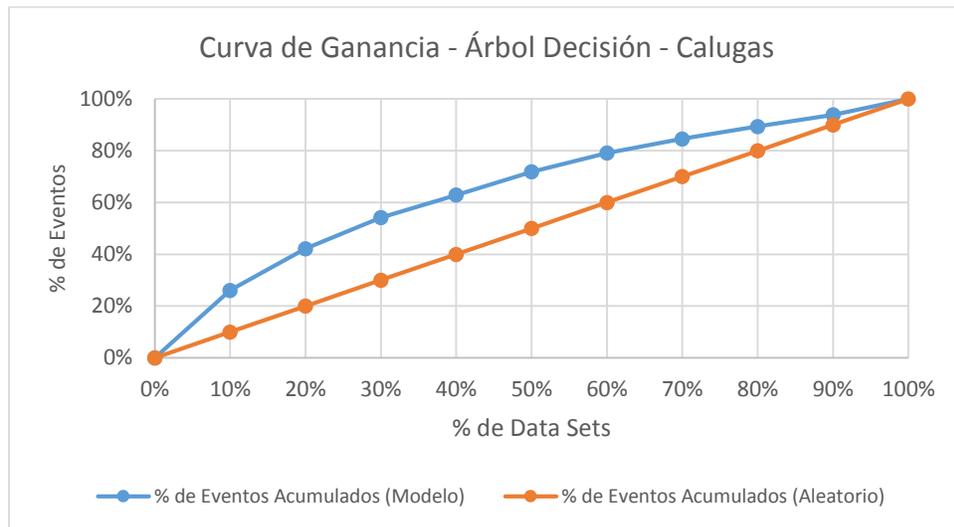


Ilustración 29: Curva de ganancia de información de árbol de decisión modelo de Calugas.  
Fuente: Elaboración propia.

Complementariamente, se tiene que la curva COR (**Anexo XV**) tiene un área bajo la curva (AUC) de 66,9% de los casos clasificados correctamente.

#### 6.4.4 Shadow Box

Por modelo de Shadow Box se refiere al fenómeno de hacer clic en la publicidad que se exhibe solamente en formato Shadow Box, por lo tanto, la variable dependiente será 1 en el caso de haber clickeado un Shadow Box, 0 en caso contrario. A continuación, se explica cómo se realizó este modelo y los principales resultados.

##### 6.4.4.1 Regresión Logística

Nuevamente, se utilizaron diferentes combinaciones de variables y sus respectivas transformaciones.

Los principales resultados se presentan a continuación:

Variable	Coefficiente	Error Estándar	Exp(Coef)
<b>Categoría Alta</b>	0,897*	0,024	2,452
<b>Categoría Media</b>	0,452*	0,024	1,571
<b>Logaritmo del cupo</b>	0,182*	0,007	1,199
<b>Tiene Visa</b>	0,330*	0,017	1,390
<b>Transacción en Supermercado Interno</b>	0,183*	0,011	1,201
<b>Transacción en Educación</b>	0,137*	0,012	1,147
<b>Transacción en Restaurant</b>	0,213*	0,013	1,238
<b>Transacción en Telefonía</b>	0,179*	0,015	1,196
<b>Transacción en Viajes</b>	0,201*	0,013	1,223
<b>Transacción en Retail Web Interno</b>	0,317*	0,011	1,373
<b>Solicita Crédito de Consumo</b>	0,205*	0,015	1,228
<b>Enganchado Puntos</b>	0,261*	0,013	1,298
<b>Constante</b>	-2,682*	0,101	0,068

Tabla 14: Resultados regresión logística modelo de Shadow Box<sup>11</sup>.  
Fuente: Elaboración propia.

En primer lugar, nuevamente se tiene que pertenecer a las categorías alta y media presentan un efecto positivo y de mayor magnitud con respecto al resto de las variables dicotómicas incluidas.

Luego, al igual que en el modelo de Calugas, se tiene que el haber realizado transacciones en la web del retail interno aumenta la propensión, al igual que las transacciones en los rubros de viajes y restaurant.

<sup>11</sup> NS corresponde a No Significativo, \* corresponde a un coeficiente significativo a  $P < 0,05$ .

A diferencia del modelo de Calugas, el modelo de Shadow Box presenta un efecto económico que aumenta la probabilidad de interactuar con este formato, dado por la tenencia de Visa y del cupo, este último siendo la única variable de naturaleza continua, por lo que, pese a no ser una magnitud de las mayores, en conjunto con el logaritmo del cupo puede aportar en mayor medida a la propensión. Asimismo, en el caso de Shadow Box, se tiene que la solicitud de crédito de consumo tiene un efecto positivo en la probabilidad de clicar este formato, por lo que son clientes que tienen acceso a mayores montos de financiamiento en comparación a los clientes de Calugas, quienes solicitan avances en efectivo.

Finalmente, el estar enganchado con el club de puntos nuevamente cumple un rol de aportar a la propensión con una de las magnitudes más altas.

Al revisar la matriz de confusión (**Tabla 15**), se tiene una que el modelo clasifica de manera correcta entre un 60% a un 70% de los clientes.

		Pronosticado			
		No Interactúa	Interactúa	Porcentaje correcto	
Observado	Entrenamiento	No Interactúa	59.673	26.250	69,4%
		Interactúa	34.674	51.674	59,8%
		Porcentaje Global			64,6%
	Validación	No Interactúa	99.439	44.434	69,1%
		Interactúa	5.125	7.860	60,5%
		Porcentaje Global			68,4%

Tabla 15: Matriz de confusión de regresión logística de modelo Shadow Box.  
Fuente: Elaboración propia.

En el caso de la curva de ganancia de información (**Ilustración 30**), se tiene que el primer decil de los clientes ordenados de acuerdo a la probabilidad entregada por el modelo identifica a un 28% de quienes efectivamente han clicado este formato.

Complementariamente, la curva COR (**Anexo XVI**) presenta un área bajo la curva (AUC) de 69,7% de los casos correctamente clasificados.

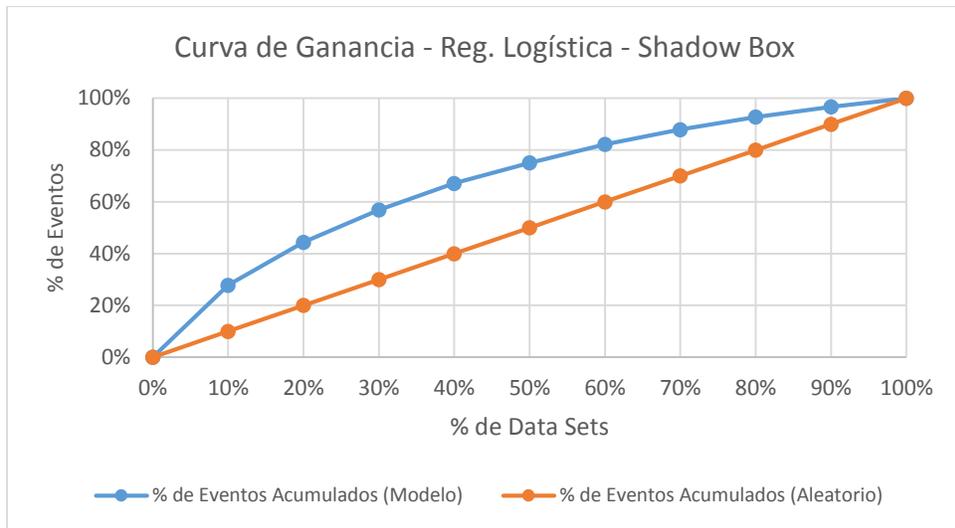


Ilustración 30: Curva de ganancia de información de regresión logística modelo de Shadow Box.  
Fuente: Elaboración propia.

#### 6.4.4.2 Árbol de Decisión

Utilizando un árbol de decisión CART para modelar el fenómeno en estudio, se puede apreciar que la primera variable que divide a los clientes corresponde a la categoría de éstos, como se muestra en la **Ilustración 31**, y luego, en el segundo nivel de corte, se utiliza la variable de estar enganchado con el club de puntos y nuevamente la categoría del cliente. Se tiene que los clientes de categorías más altas presentan mayor propensión de interactuar con este formato.

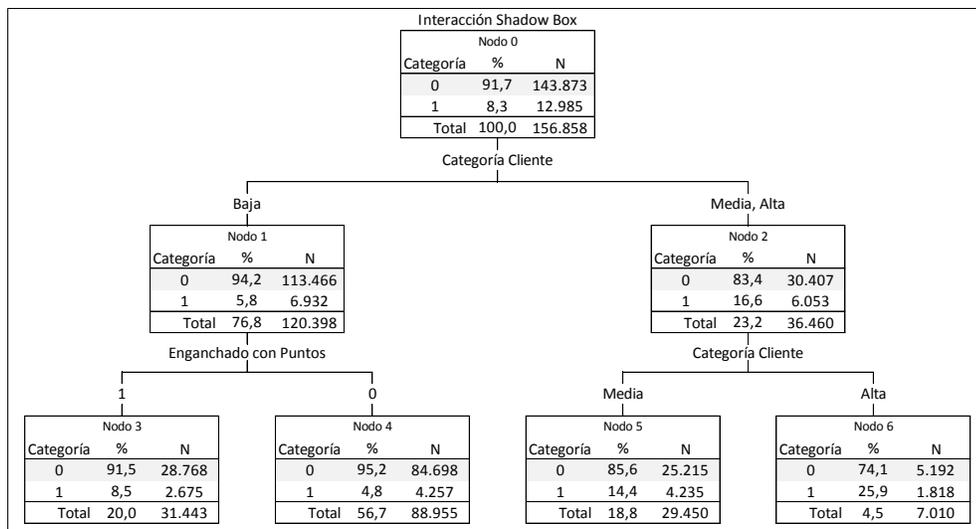


Ilustración 31: Ramas superiores de árbol de decisión de modelo de Shadow Box.  
Fuente: Elaboración propia.

Luego, como se muestra en la **Ilustración 32**, en las ramas inferiores del árbol de decisión, aparecen como variables de corte el cupo del cliente, transacciones en restaurant, transacciones en el sitio web del retail interno, transacciones en viaje y la tenencia de Visa, obteniéndose que el mejor nodo tiene una propensión de 29,9%, lo que representa un lift de 3,60.

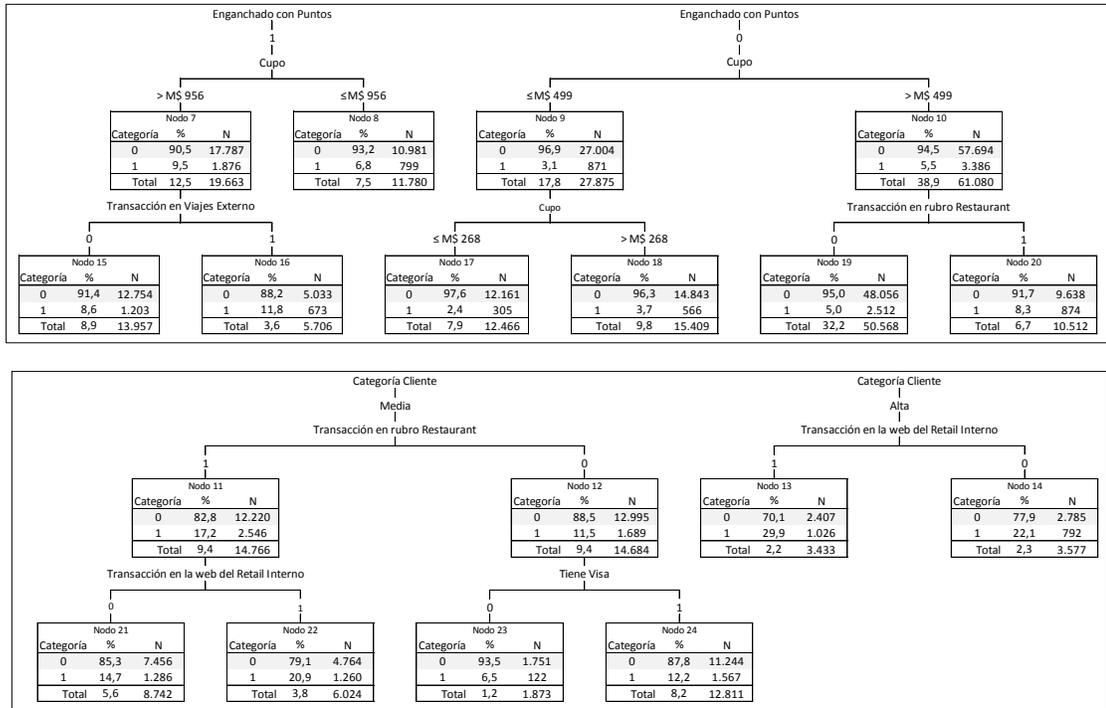


Ilustración 32: Ramas inferiores de árbol de decisión de modelo de Shadow Box.  
Fuente: Elaboración propia.

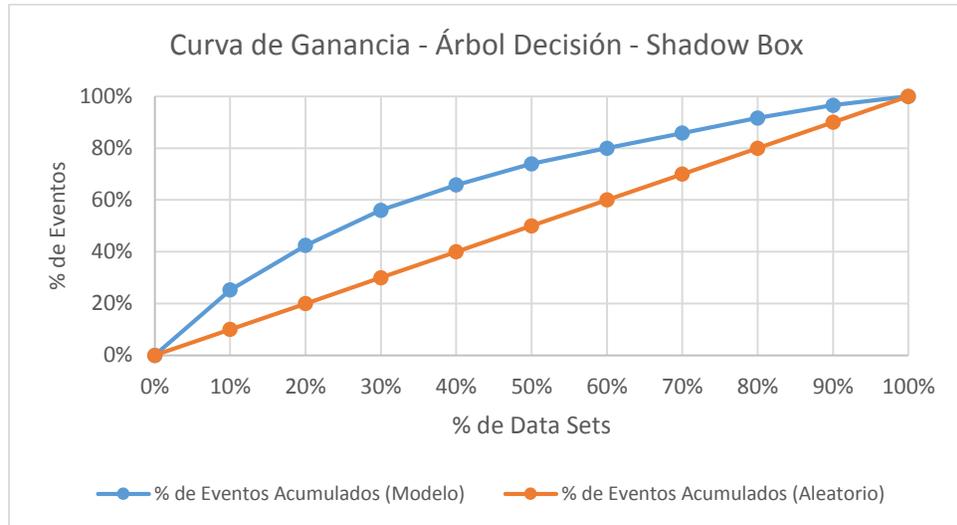
Nuevamente al revisar la matriz de confusión (**Tabla 16**), se tiene un desempeño relativamente similar al caso de la regresión logística, teniendo una mejor clasificación de los clientes que efectivamente interactúan.

		Pronosticado			Porcentaje correcto
		No Interactúa	Interactúa		
Observado	Entrenamiento	No Interactúa	56.062	29.861	65,2%
		Interactúa	31.158	55.190	63,9%
		Porcentaje Global			64,6%
	Validación	No Interactúa	93.510	50.363	65,0%
		Interactúa	4.711	8.274	63,7%
		Porcentaje Global			64,9%

Tabla 16: Matriz de confusión de árbol de decisión de modelo de Shadow Box.  
Fuente: Elaboración propia.

Ahora, revisando la curva de ganancia de información (**Ilustración 33**), se tiene que el primer decil de clientes ordenados por la probabilidad entregada por el modelo permite identificar el 25% de quienes efectivamente han clicado este formato, lo que pese a ser menor a la regresión logística, no genera una diferencia relevante.

Complementariamente, la curva COR (**Anexo XVII**) tiene un área bajo la curva (AUC) de 69,2% de los casos correctamente clasificados, bastante similar al caso de regresión logística.



*Ilustración 33: Curva de ganancia de información de árbol de decisión modelo de Shadow Box.  
Fuente: Elaboración propia.*

## 6.4.5 Burbuja

Por modelo de Burbuja se refiere al fenómeno de hacer clic en la publicidad que se exhibe solamente en formato Burbuja, por lo tanto, la variable dependiente será 1 en el caso de haber clicado una Burbuja, 0 en caso contrario. A continuación, se explica cómo se realizó este modelo y los principales resultados.

### 6.4.5.1 Regresión Logística

Nuevamente, se utilizaron diferentes combinaciones de variables y sus respectivas transformaciones con el fin de obtener el mejor modelo que se pueda con las variables disponibles.

Los principales resultados se presentan a continuación:

Variable	Coefficiente	Error estándar	Exp(Coef)
<b>Categoría Alta</b>	0,560*	0,024	1,751
<b>Categoría Media</b>	0,320*	0,024	1,378
<b>Antigüedad</b>	-0,028*	0,001	0,973
<b>Transacción en Combustible</b>	0,114*	0,012	1,120
<b>Transacción en Farmacias</b>	0,054*	0,012	1,056
<b>Transacción en Retail Externo</b>	0,076*	0,013	1,079
<b>Transacción en Supermercado Externo</b>	0,115*	0,014	1,122
<b>Transacción en Restaurant</b>	0,099*	0,013	1,104
<b>Transacción en Telefonía</b>	0,246*	0,014	1,278
<b>Transacción en Retail Web Interno</b>	0,155*	0,011	1,168
<b>Solicita Avance en Efectivo</b>	0,661*	0,011	1,936
<b>Solicita Crédito de Consumo</b>	0,429*	0,014	1,535
<b>Enganchado Financiamiento</b>	0,411*	0,013	1,508
<b>Constante</b>	0,068*	0,027	1,070

*Tabla 17: Resultados regresión logística de modelo de Burbujas<sup>12</sup>.  
Fuente: Elaboración propia.*

En primer lugar, como se ha visto en los modelos anteriores, se tienen que el hecho de pertenecer a las categorías alta y media tiene un efecto positivo, sin embargo, no presentan magnitudes tan altas como en casos previos.

Luego, se tiene que la antigüedad tiene un efecto negativo, por lo que clientes nuevos son más propensos a clicar este formato.

Por otro lado, se tiene que solicitar avances en efectivo y créditos de consumo tienen efecto positivo en la propensión, al igual que el estar enganchado con financiamiento, por lo que los clientes que cliclean Burbujas gustan o necesitan endeudarse.

Finalmente, los rubros que presentan mayor propensión corresponden a telefonía, variable que no ha sido relevante en modelos anteriores, y también la transacción en la web del retail interno, lo que es transversal a todos los modelos.

Al revisar la matriz de confusión (**Tabla 18**), el modelo de Burbujas clasifica de manera correcta un 59% de los que interactúan y 67% de los que no.

<sup>12</sup> NS corresponde a No Significativo, \* corresponde a un coeficiente significativo a  $P < 0,05$ .

		Pronosticado			
		No Interactúan	Interactúan	Porcentaje correcto	
Observado	Entrenamiento	No Interactúan	58.831	29.765	66,4%
		Interactúan	36.602	52.187	58,8%
		Porcentaje Global			62,6%
	Validación	No Interactúan	98.527	49.635	66,5%
		Interactúan	3.582	5.114	58,8%
		Porcentaje Global			66,1%

Tabla 18: Matriz de confusión de regresión logística de modelo de Burbujas. Fuente: Elaboración propia.

Luego, de acuerdo a la curva de ganancia de información (**Ilustración 34**), se tiene que el primer decil ordenado de acuerdo a la probabilidad entregada por el modelo logra identificar a un 23% de los clientes que efectivamente han clicado este formato.

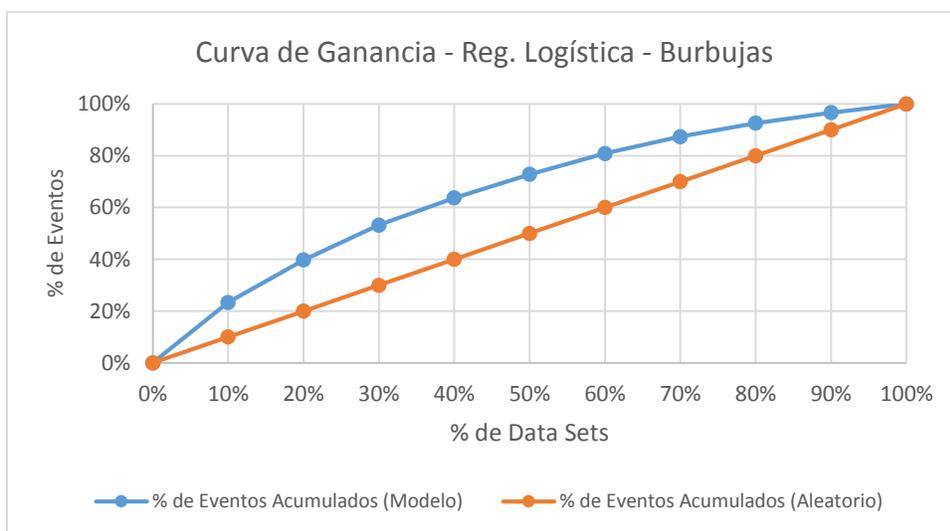


Ilustración 34: Curva de ganancia de información de regresión logística modelo de Burbujas. Fuente: Elaboración propia.

Complementariamente, la curva ROC (**Anexo XVIII**) presenta un área bajo la curva (AUC) de 67,1% de casos correctamente clasificados.

#### 6.4.5.2 **Árbol de Decisión**

Utilizando un árbol de decisión tipo CHAID para modelar este formato en particular, se tiene que la primera variable discriminante corresponde a haber solicitado un avance en efectivo, asignando mayor propensión a quienes sí lo solicitan. Luego, se incorporan

como variables de corte el estar enganchado con financiamiento y transacciones en supermercados.

Luego, como se muestra en la **Ilustración 35**, las variables que también dividen clientes en este fenómeno, corresponden a la categoría del cliente y haber solicitado un crédito de consumo, obteniéndose que el mejor nodo una probabilidad de 13,1%, lo que representa un lift de 2,38.

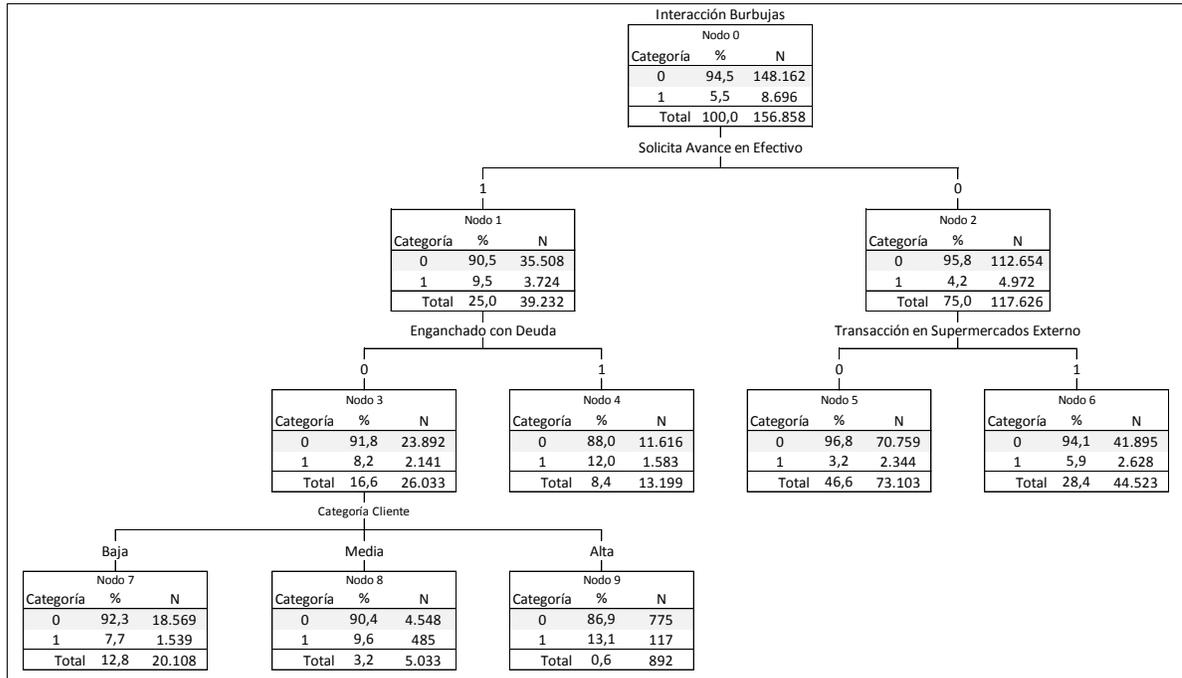


Ilustración 35: Árbol de decisión modelo de Burbujas.  
Fuente: Elaboración propia.

Revisando la matriz de confusión (**Tabla 19**), utilizando árbol de decisión se tiene un desempeño levemente menor al que se obtiene al utilizar una regresión logística.

		Pronosticado			
		No Ingresa	Ingresa	Porcentaje correcto	
Observado	Entrenamiento	No Ingresa	61.797	26.799	69,8%
		Ingresa	41.693	47.096	53,0%
		Porcentaje Global			61,4%
	Validación	No Ingresa	103.350	44.812	69,8%
		Ingresa	4.132	4.564	52,5%
		Porcentaje Global			68,8%

Tabla 19: Matriz de confusión de árbol de decisión de modelo de Burbujas.  
Fuente: Elaboración propia.

Luego, observando la curva de ganancia de información (**Ilustración 36**), se puede ver que tomando el primer decil de clientes ordenados según la probabilidad entregada por el modelo, se logra identificar a un 21% de quienes efectivamente cliclean este formato, lo que es bastante similar a lo que se gana con la regresión logística.

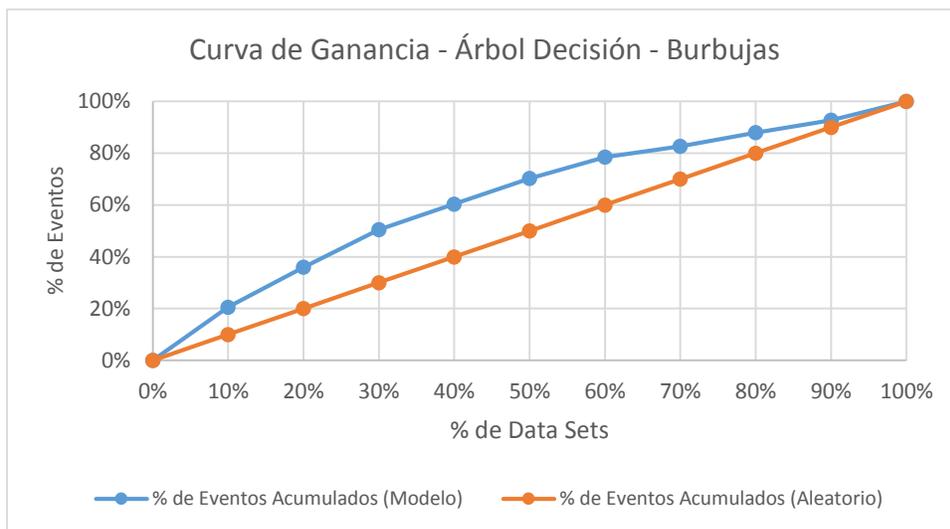


Ilustración 36: Curva de ganancia de información de árbol de decisión modelo de Burbujas.  
Fuente: Elaboración propia.

Complementariamente, la curva ROC (**Anexo XIX**) presenta un área bajo la curva de 64,5% de casos correctamente clasificados.

#### 6.4.6 Campaña a clientes que han acumulado niveles de canje (Puntos)

Por modelo de Puntos se refiere al fenómeno de hacer clic en la publicidad que se exhibe solamente en formato Caluga y presenta contenido que se exhibe solo a clientes que han alcanzado un nivel de puntos suficientes para poder realizar canjes por productos y/o viajes, por lo tanto, la variable dependiente será 1 en el caso de haber clickeado una de estas campañas, 0 en caso contrario. A continuación, se explica cómo se realizó este modelo y los principales resultados.

En este caso en particular, se trabaja con una base de clientes la cual es un subconjunto de las anteriores, incorporando solamente a quienes se le haya mostrado las campañas mencionadas. Al igual que en los casos anteriores, se procedió a balancear la base y generar un grupo de entrenamiento y otro de validación.

### 6.4.6.1 Regresión Logística

Utilizando la misma lógica que para los modelos anteriores, se presentan a continuación los principales resultados obtenidos:

Variable	Coefficiente	Error estándar	Exp(Coef)
Logaritmo del Cupo	-0,176*	0,023	0,838
Transacción en Autos/Accesorios	0,175*	0,034	1,192
Transacción en Combustible	0,214*	0,034	1,239
Transacción en Educación	0,148*	0,036	1,159
Solicita Avance en Efectivo	0,210*	0,035	1,233
Solicita Crédito de Consumo	0,204*	0,039	1,226
Enganchado con Puntos	0,350*	0,041	1,419
Constante	1,802*	0,329	6,063

Tabla 20: Resultados regresión logística modelo campaña de puntos<sup>13</sup>.

Fuente: Elaboración propia.

En primer lugar, se puede ver que el cupo tiene un impacto negativo en estos clientes, por lo que los clientes con mayor poder económico están menos interesados en los canjes que se promocionan.

Como se esperaba, el enganche con el club de puntos genera mayor propensión con las campañas de puntos, lo que tiene sentido, ya que esta métrica de la empresa busca medir la intención de un cliente al utilizar su tarjeta de crédito, es decir, alguien interesado en acumular puntos también está interesado en las campañas relacionadas al canje de éstos.

Existe un efecto positivo de haber realizado transacciones en combustible, autos y accesorios y educación, rubros que en general no se encuentran disponibles como canjes en el catálogo de puntos, lo que genera la posibilidad de atacar a estos clientes de una forma diferente, innovando el tipo de canjes que se realizan por el canal digital, por ejemplo.

Finalmente, se tiene que solicitar un avance en efectivo y solicitar un crédito de consumo tienen un efecto positivo en la propensión, por lo que también se abre la oportunidad de relacionar estas campañas a productos financieros.

<sup>13</sup> NS corresponde a No Significativo, \* corresponde a un coeficiente significativo a  $P < 0,05$ .

Al observar la matriz de confusión (**Tabla 21**), el modelo alcanza a clasificar de manera correcta entre 58% a 69% de los clientes que clickean este tipo de campañas.

		Pronosticado			
		No Interactúa	Interactúa	Porcentaje correcto	
Observado	Entrenamiento	No Interactúa	59.934	27.135	68,8%
		Interactúa	36.918	49.114	57,1%
		Porcentaje Global			63,0%
	Validación	No Interactúa	99.769	45.630	68,6%
		Interactúa	4.820	6.639	57,9%
		Porcentaje Global			67,8%

Tabla 21: Matriz de confusión regresión logística modelo de Puntos.  
Fuente: Elaboración propia.

Luego, al revisar la curva de ganancia de información (**Ilustración 37**), se puede ver que el primer decil de los clientes ordenados según las probabilidades entregadas por el modelo identifican solamente a un 14% de quienes han clickeado estas campañas.

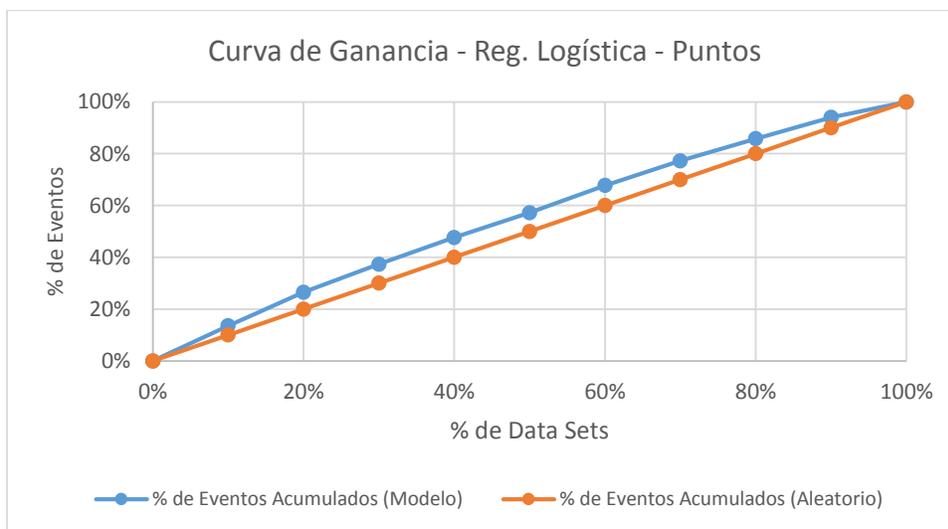


Ilustración 37: Curva de ganancia de información de regresión logística modelo de Puntos.  
Fuente: Elaboración propia.

Complementariamente, la curva ROC (**Anexo XX**) presenta un área bajo la curva (AUC) de 57,7%, lo que es un porcentaje bajo con respecto a modelos anteriores.

### 6.4.6.2 Árbol de decisión

El algoritmo utilizado para este caso corresponde a un árbol tipo CART, el cual presenta como primera variable de corte la compra en combustible, siendo más propensos quienes efectivamente compran. Luego aparecen la solicitud de avance en efectivo y el enganche con el club de puntos, como se ve en la **Ilustración 38**.

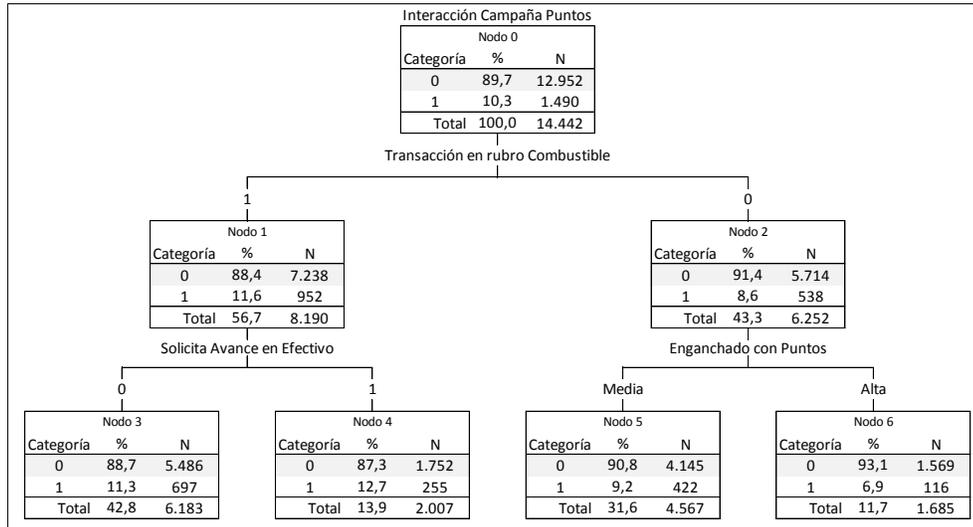


Ilustración 38: Ramas superiores árbol de decisión modelo de Puntos.  
Fuente: Elaboración propia.

Luego, como se muestra en la **Ilustración 39**, en las ramas inferiores del árbol se incluyen las variables de cupo, solicita crédito de consumo, ticket promedio y compra en el rubro de autos, y también se repite la variable de solicitud de avance en efectivo. Se obtiene que el mejor nodo tiene una probabilidad de 12,7%, lo que representa un lift de 1,23.

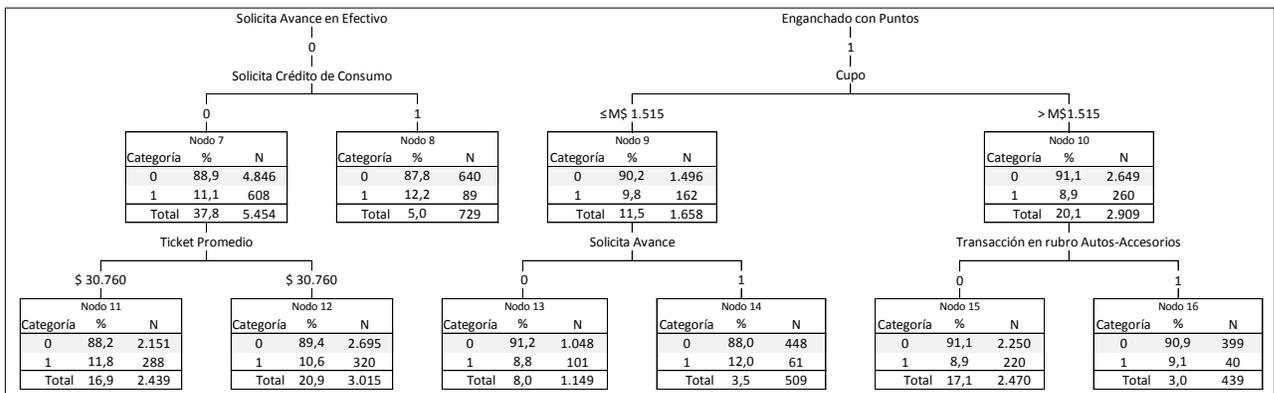


Ilustración 39: Ramas inferiores árbol de decisión modelo de Puntos.  
Fuente: Elaboración propia.

Ahora, observando la matriz de confusión (**Tabla 22**), se tiene que el nivel de clientes que clasifica correctamente es bastante bajo, no alcanzando el 60%, y en comparación al modelo de regresión logística tiene un desempeño menor.

		Pronosticado			
			No Interactúa	Interactúa	Porcentaje correcto
Observado	Entrenamiento	No Interactúa	5.583	4.215	57,0%
		Interactúa	4.118	5.032	55,0%
		Porcentaje Global			56,0%
	Validación	No Interactúa	7.167	5.785	55,3%
		Interactúa	719	771	51,7%
		Porcentaje Global			55,0%

Tabla 22: Matriz de confusión de árbol de decisión de modelo de Puntos.  
Fuente: Elaboración propia.

En términos de la ganancia de información que representa este modelo, como se ve en la **Ilustración 40**, ésta es casi nula, teniendo que el primer decil de clientes ordenados en base a la probabilidad que entrega el modelo identifica solamente a un 12% de quienes efectivamente hicieron clic en este tipo de campañas publicitarias, incluso menor que el caso de la regresión logística.

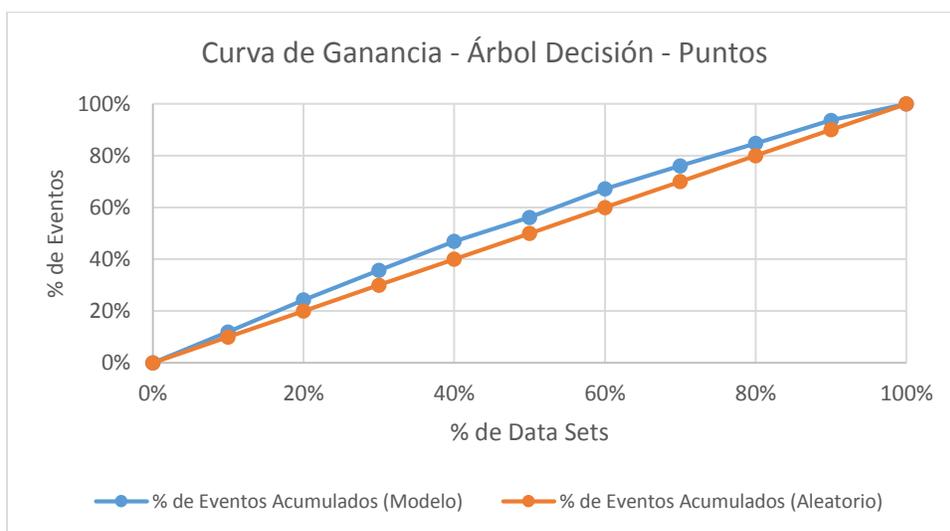


Ilustración 40: Curva de ganancia de información de árbol de decisión modelo de Puntos.  
Fuente: Elaboración propia.

Complementariamente, la curva ROC (**Anexo XXI**) presenta un área bajo la curva (AUC) de 56,9%, bastante similar a la regresión logística, por lo que ninguno de los modelos cumple con tener un desempeño que permita tomar acciones para gestionar a estos clientes.

## 6.5 Conclusión Modelos de Propensión

Tras haber realizado los modelos de regresión logística y árbol de decisión para cada tipo de interacción con el sitio web, la primera comparación tiene que ver con los desempeños de ambos mecanismos. De acuerdo a esto, se puede apreciar que no existen mejoras claras, de acuerdo a las curvas de ganancia de información y la curva ROC, entre uno y otro enfoque.

Asimismo, entre los diferentes fenómenos en estudio que se modelan, se tiene que el mejor desempeño lo tiene el Ingreso Web, seguido del modelo de Shadow Box, luego por el modelo de Calugas, Burbujas, Publicidad Web y Puntos, en ese orden.

Dado que no se tienen mayores beneficios por seleccionar un tipo de modelo (regresión logística o árbol de decisión) por sobre otro, es que se opta por elegir el árbol de decisión para el diseño de medidas de acción. Esto permite a la empresa tener mayor comprensión de los perfiles de mayor propensión para cada tipo de publicidad.

A continuación, se detallan los perfiles de clientes obtenidos por los modelos, en el caso de los modelos con mejor desempeño y que permitan accionar sobre éstos, correspondientes a Ingreso Web, Shadow Box, Calugas y Burbujas:

### 6.5.1 Perfil de Ingreso Web

Para el caso del ingreso al sitio web, se tiene que las características que definen a los perfiles más propensos corresponden a las siguientes:

- ✓ Clientes con edad menor o igual a 45 años
- ✓ Compra en la web del retail interno
- ✓ Compra en el rubro de viajes
- ✓ Enganchado con el club de puntos

Caracterizando a este perfil de clientes, se obtiene que poseen un cupo promedio cercano al MM\$ 1,4. Además, tienen una edad promedio de 34 años, un 52% son hombres, solo un 6% son ABC1, un 22% C2 y 27% C3.

### 6.5.2 Perfil de Shadow Box

Para el caso de Shadow Box, se tiene que las características que definen a los perfiles más propensos corresponden a las siguientes:

- ✓ Categoría del cliente alta o media
- ✓ Tarjeta de crédito tipo Visa
- ✓ Compra en la web del retail interno
- ✓ Compra en el rubro de restaurant
- ✓ Compra en el rubro de viajes

Caracterizando a este perfil de clientes, se obtiene que poseen un cupo promedio de más de MM\$ 2,7, lo cual es bastante alto, tomando en cuenta que más de un 50% de los clientes presentan cupo de más de MM\$ 2. Además, tienen una edad promedio de 42 años, un 57% son mujeres, un 22% de ellos son ABC1, un 35% son C2 y un 23% son C3.

### **6.5.3 Perfil de Calugas**

Para el caso de Calugas, se tiene que las características que definen a los perfiles más propensos corresponden a las siguientes:

- ✓ Categoría del cliente alta o media
- ✓ Compra en la web del retail interno
- ✓ Compra en el rubro restaurant
- ✓ Compra en el rubro de educación y servicios básicos
- ✓ Solicita avance en efectivo

Caracterizando a este perfil de clientes, se obtiene que poseen un cupo promedio superior a MM\$ 2,7, lo cual es bastante alto. Además, presentan una edad promedio de 42 años, un 57% son mujeres, un 21% son ABC1, un 35% son C2 y un 23% son C3.

### **6.5.4 Perfil de Burbujas**

Para el caso de Burbujas, se tiene que las características que definen a los perfiles más propensos corresponden a las siguientes:

- ✓ Solicita avance en efectivo
- ✓ Solicita crédito de consumo
- ✓ Compra en el rubro de supermercados externo
- ✓ Enganchado con financiamiento

Caracterizando a este perfil de clientes, se obtiene que poseen un cupo promedio cercano a MM\$ 1,3, lo que es más bajo que en el resto de formatos. Además, tienen una edad promedio de 40 años, un 51% son hombres, solo un 6% son ABC1, un 20% son C2, un 26% son C3 y un 25% son D.

### **6.5.5 Comparación Perfiles de Shadow Box, Calugas y Burbujas**

Tras haber obtenido los perfiles de los tres formatos en que se muestra la publicidad web en el sitio de la empresa, se tiene que el perfil de Shadow Box y el de Calugas es bastante similar, en términos de las características que definen a un cliente propenso como en términos de su cupo, proporción de género, edad y grupo socioeconómico.

Por otro lado, Shadow Box y Calugas se diferencian de Burbujas por las características que definen a un cliente propenso, siendo el perfil de Burbujas un cliente con menor cupo, y pertenecientes en su mayoría a grupos socioeconómicos de clases media baja e incluso vulnerable, mientras que para el perfil de Shadow Box y Calugas corresponden a clases alta y media.

## **6.6 Líneas de Acción**

Con los perfiles de los clientes con mayor propensión a interactuar con el sitio web definidos, se desarrolla una propuesta para que la empresa implemente estos resultados en su gestión de clientes.

Para esto, en conjunto con la empresa se generó una estrategia para, en primer lugar, migrar clientes que no ingresan al sitio web hacia éste, y luego, entregarles contenido de acuerdo a las características que los hacen más propensos, lo que permita que el sitio web tenga un mayor flujo de clientes, y estos clientes se interesen más por las campañas que se realizan, forjando una mejor relación entre el cliente y la empresa y aumentando las métricas de desempeño utilizadas por la empresa.

### **6.6.1 Aplicación Comercial**

La estrategia a desarrollar debe cumplir con ser lo suficiente clara y fácil de comprender y aplicar para que sea ejecutable por parte de la empresa, además se debe tener en consideración que los resultados obtenidos permiten entregar recomendaciones asociadas a características demográficas o rubros de comercios que tengan mayor relación con que un cliente interactúe, por lo que en términos de la forma del mensaje que se le entregue al cliente se mantendrá a decisión del equipo de diseño de la empresa.

Esta estrategia se diseñará en dos etapas: una etapa de migración de clientes hacia el canal digital, la cual utilizará la información entregada por el modelo de Ingreso Web, y una segunda etapa de personalización de las campañas web, dentro de la cual se utilizarán los formatos de Caluga y Shadow Box, dado que permiten entregar un mensaje que contenga información acerca de productos relacionados a rubros específicos.

El caso de Burbujas, dado el formato en que entrega información, no se utilizará para una estrategia. En el caso del modelo de Puntos, al no entregar un desempeño suficiente para generar estrategias, se opta por enfocar la estrategia a nivel de formato.

### 6.6.1.1 Migración a Canal Digital

Para generar que los clientes que no utilizan la web de la empresa migren hacia el canal digital, se propone el desarrollo del siguiente piloto, el cual consiste en el envío de un correo electrónico invitando a los clientes a utilizar el sitio web:

- ✓ Conversión a mejorar

La variable dependiente de este experimento corresponde al ingreso al sitio web y su posterior inicio de sesión.

- ✓ Hipótesis a testear

*“Gestionar a los clientes más propensos a ingresar al sitio web a través de un correo electrónico, aumentará el número de clientes activos en la web”.*

- ✓ Variables experimentales

Las variables independientes y sus niveles para este experimento se muestran en la **Tabla 23**.

Variable	Niveles
<b>Contenido</b>	1. Incluye promoción relacionada al perfil propenso <sup>14</sup> 2. Solo informativo
<b>Asunto</b>	1. Incluye nombre del cliente 2. No incluye nombre del cliente

*Tabla 23: Variables independientes y sus niveles para el experimento de migración a canal digital.  
Fuente: Elaboración propia.*

<sup>14</sup> La promoción en particular se debe negociar con el área de marketing, entre las que definen al perfil propenso.

Utilizando la primera variable independiente de contenido, permite determinar si efectivamente las características que identifican a un cliente propenso de acuerdo a los modelos de propensión realizados generan un mayor impacto en la métrica que se desea incrementar.

Por otro lado, la variable independiente de asunto, se incorpora, en primer lugar, porque mantiene la simpleza del experimento y, por otro lado, ayuda a comprender si más personalización aumenta la conversión deseada.

✓ Target de clientes

El target de clientes corresponde a clientes que cumplan las características que identifican al perfil propenso del modelo de Ingreso Web y que además no hayan ingresado al sitio web. Estas características corresponden a clientes menores a 45 años, que hayan comprado en la web del retail interno y en el rubro de viajes y se consideren enganchados al club de puntos.

Para obtener este target se utilizó un criterio acordado en conjunto con la empresa, el cual corresponde al siguiente:

- Clientes con mayor probabilidad de respuesta:  $Lift \geq 1,25$

Utilizando estos filtros para seleccionar, se obtiene que el grupo target cuenta con 280.069 clientes.

✓ Grupo de control

Con el fin de poder medir la acción de marketing realizada, se seleccionará un 20% de clientes pertenecientes al grupo target utilizando un Muestreo Aleatorio Simple. Así, se tiene que el grupo de control cuenta con 56.014 clientes.

✓ Medición resultados

Para medir el impacto de este piloto, se utilizará el ingreso al sitio web y su posterior inicio de sesión como métrica, y el plazo que se utilizará para evaluar estos ingresos será el correspondiente a la política de toques que tiene el holding, es decir, a 3 días desde que se envía el correo [8].

### 6.6.1.2 Personalización de Campañas Web

Para generar que los clientes que ya ingresan al sitio aumenten su interacción con las campañas de publicidad que se les muestran, se propone realizar un experimento estilo A/B test, en el cual se muestra a un grupo de tratamiento una publicidad relacionada al perfil más propenso de cada formato, es decir, de Shadow Box y de Calugas, y simultáneamente se le muestra al grupo de control una publicidad de un rubro asociado a una baja propensión.

✓ Conversión a mejorar

La variable dependiente de este experimento corresponde al clickeo en la respectiva campaña.

✓ Hipótesis a testear

*“Gestionar a los clientes más propensos a clickear determinado formato de publicidad web a través de contenidos enfocados en sus transacciones anteriores, aumentará el clickeo en las campañas”.*

✓ Variables experimentales

Las variables independientes y sus niveles para este experimento se muestran en la **Tabla 24**.

Variable	Niveles
Contenido	1. Promoción relacionada al perfil propenso <sup>15</sup> 2. Promoción relacionada a clientes no propensos.

*Tabla 24: Variables independientes y sus niveles para experimento de personalización de campañas.  
Fuente: Elaboración propia.*

✓ Target de clientes

El target de clientes corresponde a clientes que cumplan las características que identifican al perfil propenso del modelo de Shadow Box y Calugas, respectivamente.

Al igual que en el experimento anterior, se utilizó un criterio acordado en conjunto con la empresa para obtener este target:

- Clientes con mayor probabilidad de respuesta: Lift  $\geq 1,25$

<sup>15</sup> La promoción en particular se debe negociar con el área de marketing, entre las que definen al perfil propenso.

En el caso particular de este experimento, se tiene que los perfiles de clientes propensos a Shadow Box y Calugas resultan ser bastante similares, tal como se explicó anteriormente, por lo que con fines de evitar que clientes sean gestionados por ambos formatos, a los clientes que cumplen con el criterio expuesto para ambos, se le asigna el formato en el que tenga mayor lift.

En base a lo anterior, se tiene que el grupo target para el formato Shadow Box cuenta con 80.499 clientes y para el formato Caluga con 109.247 clientes.

✓ Grupo de control

Nuevamente, con el fin de medir la acción de marketing realizada, se seleccionará el 20% de clientes pertenecientes a cada grupo target utilizando un Muestreo Aleatorio Simple. Así, se tiene que el grupo de control para Shadow Box cuenta con 16.100 clientes y para Calugas cuenta con 21.849 clientes.

✓ Medición de resultados

Se utilizará el clic en las distintas publicidades para medir las diferencias entre grupos. Esto se hará en un plazo de un mes, ya que las campañas que se exhiben en el sitio web se definen mensualmente y, debido a la naturaleza del sitio web, no es un sitio web en el que los clientes ingresen muchas veces al mes.

## 6.6.2 Evaluación Económica

En base a la estrategia propuesta, es necesario medir en términos cuantitativos los beneficios económicos que implica la personalización de los contenidos del sitio web. Para esto, el análisis de estos beneficios se realizará en función del incremento en gasto de un cliente si migra a la web (caso modelo de ingreso al sitio web) o comienza a interactuar con campañas de publicidad web (caso modelo de Calugas, Shadow Box y Burbujas). Teniendo el incremento en gasto individual por cliente, se puede obtener el beneficio económico total que genera cada acción de marketing que se decida mediante la siguiente ecuación:

$$BE = \sum_{i \in \{Target\}} (IG_i \cdot Prob(Interactuar) - C_i)$$

En donde **BE** corresponde al beneficio económico, **IG<sub>i</sub>** el incremento en gasto del cliente **i**, **Prob(Interactuar)** la probabilidad de que un cliente interactúe con el sitio web, ya sea ingresando al sitio web o clicar algún tipo de campaña de publicidad y **C<sub>i</sub>** el costo de gestionar al cliente **i**. Este beneficio económico, al estar ponderado por la

probabilidad de interacción, es un valor esperado, sin embargo, este caso considera condiciones ideales, es decir, que el cliente efectivamente sea expuesto a la gestión por parte de la empresa, por lo que es un caso ideal.

Luego, el incremento en gasto de un cliente se define como el mínimo entre el gasto potencial del cliente ( $GP_i$ ), el cual se calcula internamente en la empresa utilizando características propias del individuo, y el gasto del perfil target ( $GPT$ ) el cual se calcula como el promedio de gasto que tienen los clientes que actualmente ingresan al sitio web o bien clickean las campañas, menos el gasto actual del cliente ( $GA_i$ ), como se muestra en la siguiente ecuación:

$$IG_i = \min\{GP_i, GPT\} - GA_i$$

Es decir, si se tiene que un cliente al gestionarlo cambia su comportamiento e ingresa al sitio o comienza a clickear campañas, su aumento en gasto debería llegar a niveles del grupo de clientes al que migró, sin embargo, esto está limitado por su potencial de gasto. Luego, a este gasto se le debe restar lo que ya gasta, obteniéndose el gasto incremental.

Con el fin de explicitar este método, se explica un ejemplo:

*“Cliente A tiene un potencial de gasto de \$2.000.000 anuales, el grupo al que se quiere migrar tiene un gasto promedio de \$1.500.000 anuales y gastó en el período \$500.000.*

*Luego, se tiene que su potencial es mayor al gasto del perfil target, por lo que el incremento en gasto corresponde a \$1.500.000 - \$500.000 = \$1.000.000.*

*Si su probabilidad de interactuar es 10%, el cliente genera un incremental esperado de \$100.00.”*

Como alcance, este procedimiento solo es aplicable a clientes que cambien de comportamiento, es decir, que pasen de ser clientes que no web a clientes web o que pasen de ser clientes que no clickean a clickear las campañas, lo que en el caso del Ingreso Web incluye a todo el target, pero en el caso de Shadow Box y Calugas se tiene una cantidad menor.

Otra consideración a tomar en cuenta, es que las acciones de marketing propuestas no implican un costo incremental, ya que, en el caso del envío de correo electrónico, éste sería un caso de los correos que actualmente la empresa envía a sus clientes, mientras que en el caso de Shadow Box y Calugas, éstos formatos se exhiben siempre en el sitio web, por lo que no hay un costo extra por esta propuesta en particular.

Luego, utilizando este método para la estrategia propuesta, se tiene que la ganancia promedio es mayor en el caso de los clientes de Shadow Box, seguido por Calugas y luego el Ingreso Web, sin embargo, al ser este último el piloto que gestiona a una mayor cantidad de clientes, es el que presenta la mayor ganancia, como se muestra en la **Tabla 25**.

Estrategia	Clientes Target	Promedio Gasto	Mínimo Gasto	Máximo Gasto	Gasto Grupo MM\$	% del Gasto Anual
<b>Ingreso Web</b>	280.069	\$ 937.830	\$ 0	\$ 2.058.286	\$ 156.164	5,2%
<b>Calugas</b>	94.268	\$ 1.127.545	\$ 0	\$ 3.946.212	\$ 12.740	0,4%
<b>Shadow Box</b>	68.289	\$ 1.302.566	\$ 0	\$ 3.921.526	\$ 11.697	0,4%
<b>Total</b>					<b>\$ 180.600</b>	<b>6,1%</b>

*Tabla 25: Beneficios económicos por la gestión de clientes propuesta.  
Fuente: Elaboración propia.*

Por lo tanto, la evaluación económica del proyecto realizado genera una ganancia en el mejor escenario de \$ 180 mil millones anuales, equivalente al 6,1% del gasto anual del universo de clientes estudiados.

Dado que es la primera prueba de este tipo que se realiza en la compañía y considerando la complejidad de que todos los clientes del target, en particular en el caso de clientes que no utilizan la web de la empresa, se considera un escenario de efectividad del 5% de lo esperado, obteniéndose beneficios por **\$ 9 mil millones anuales**, equivalentes al 0,3% del gasto anual.

## **7 RESULTADOS COMPLEMENTARIOS**

Dentro de lo realizado en la empresa durante el trabajo de título, se generaron indicadores para gestionar y monitorear el comportamiento de clientes dentro del sitio web, el primero relacionado al comportamiento dentro del sitio web (Clasificación por Comportamiento de Navegación), un segundo relacionado al ingreso (Intensidad Web) y otro relacionado al nivel de enganche o lealtad que presenta con el sitio web (Score Engagement Web). Ambos indicadores presentan criterios impuestos por expertos de la empresa en base a análisis descriptivos previos.

### **7.1 Clasificación por Comportamiento de Navegación**

Como menciona la literatura, se tiene que el comportamiento de navegación puede ocurrir independientemente de la conversión en el sitio web [7], que se puede interpretar como la interacción con campañas de publicidad web, por lo que conocer qué realiza un cliente dentro del sitio web genera interés, tanto en términos de conocimiento como en términos de cómo gestionar a los clientes.

En base a esto, se utiliza la información disponible por la empresa, la cual corresponde a qué secciones del sitio web un cliente cliquea, por lo que se realiza un análisis descriptivo con que, en conjunto con expertos de la empresa, se define que lo más relevante para el sitio web corresponde al pago de la tarjeta de crédito y la interacción con los productos financieros, tales como avances en efectivo y crédito de consumo.

Luego, teniendo en cuenta lo anterior, se definen criterios en base a cortes determinísticos, los cuales se explican en la **Tabla 26**, para los cuales se crea la variable Tasa Pago, para medir si el cliente ingresa efectivamente al sitio para pagar su tarjeta de crédito:

$$Tasa\ Pago = \frac{Número\ de\ Meses\ con\ Pago}{Número\ de\ Meses\ que\ Ingresa}$$

Asimismo, para el caso de los productos financieros, se crean dos variables asociadas a este uso del sitio web, uno que mide la intención de obtener un avance en efectivo o crédito de consumo y otra variable que mide si un cliente ya ha solicitado un avance en efectivo o un crédito de consumo:

$$Intención\ PF = Número\ de\ Ingresos\ a\ Sección\ de\ Productos\ Financieros$$

$$Conversión\ PF = Número\ de\ Productos\ Financieros\ Solicitados$$

En base a estas variables, se tienen los siguientes criterios:

Dimensión	Criterio
Pago	$Tasa\ Pago \geq 50\%$
Productos Financieros <sup>16</sup>	$Intención\ PF \geq 2 \vee Conversión\ PF \geq 1$
Otros	$Tasa\ Pago < 50\% \wedge Intención\ PF < 2 \wedge Conversión\ PF < 1$

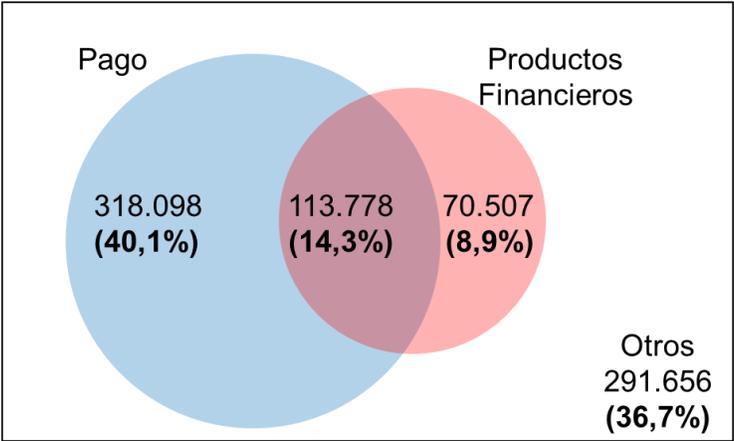
Tabla 26: Criterios clasificación de clientes web.

Fuente: Elaboración propia.

Estos criterios se desarrollaron en conjunto con expertos de la empresa, y tienen sentido en términos de que logran incluir a clientes que solamente ingresan pagar, pero no utilizan su tarjeta de crédito todos los meses en el caso del pago, y también se hace cargo de la frecuencia con que se solicitan los productos financieros que la empresa ofrece, dándole mayor relevancia al hecho de haber solicitado ya el producto en comparación a la intención simplemente.

<sup>16</sup> Productos Financieros corresponden a Avance en Efectivo y Crédito de Consumo.

En particular, se tiene que entre los clientes que califican para pago también puede haber clientes que califican para productos financieros, por lo que esta intersección se puede mostrar mediante un diagrama de Venn (**Ilustración 41**), quedando en evidencia que el grupo de clientes que califican por ambos criterios es mayor que los clientes que solamente se interesan por avances en efectivo y crédito de consumo dentro del sitio web.



*Ilustración 41: Distribución de clientes de acuerdo a clasificación por navegación.  
Fuente: Elaboración propia.*

Como se muestra, en Otros se agrupa una gran proporción de clientes, por lo cual se tiene que 100.715 (12,7%) al menos realizan alguna acción en el sitio web, ya sea que no califiquen bajo los criterios expuestos, o bien, algún otro uso del sitio web, y 190.941 (24,0%) de los clientes que o no hace nada el sitio web, simplemente ingresa, o lo que hace no está siendo tomado en cuenta por los eventos que la empresa está recolectando.

Teniendo esta marca por cliente sobre cómo interactúa con el sitio web, permite a la empresa gestionar las campañas de publicidad web de manera más enfocada, entregando la oportunidad también de realizar el cruce entre los grupos que entrega esta clasificación y los perfiles de clientes más propensos a interactuar en base a las probabilidades entregadas por los modelos de la sección anterior.

**7.2 Intensidad Web**

Este indicador tiene el objetivo de diferenciar a los clientes en base a cuántas veces ingresan al sitio web de la empresa durante un mes. Los principales usos tienen que ver con monitorear a los clientes y gestionarlos, por ejemplo, acciones de marketing que requieran que ingrese más de una vez al sitio web.

Cada mes se tienen la cantidad de ingresos al sitio web por cada cliente, en base a eso se le asigna a cada cliente una etiqueta de Nulo (no ingresa), Bajo (un ingreso), Medio (dos o tres ingresos) y Alto (4 o más ingresos), luego, por cada trimestre móvil se calcula la moda, con lo que se obtiene una etiqueta final.

Como resultado de este indicador se tiene que entre los clientes que efectivamente ingresan al sitio web, se tiene que entre 51%-53% son de Intensidad Baja, entre 30%-31% son de Intensidad Media y entre 17%-19% son de Intensidad Alta.

### **7.3 Score Engagement Web (SEW)**

Este indicador tiene el objetivo de determinar si un cliente se encuentra enganchado o es leal al sitio web. Los principales usos tienen que ver con monitorear a los clientes y gestionarlos, por ejemplo, a través de acciones de marketing para quienes están más interiorizados en lo que al sitio web se refiere, o bien, medir si el número de clientes enganchados disminuye en un período dado por problemas en el sitio web.

Metodológicamente, este indicador utiliza una serie de acciones realizables por un cliente dentro y fuera del sitio web, asignándole un puntaje a cada uno y luego sumándolos para obtener su Score final. Las acciones mencionadas hacen referencia a intensidad web, pago en el sitio web, solicitud de avances en efectivo, solicitud de créditos de consumo, canje de puntos y clickeos varios dentro del sitio web, mientras que los puntajes incorporan la exclusividad del canal, es decir, un cliente que realiza todas las acciones nombradas vía web tiene mayor Score que un cliente que realiza algunas de estas acciones presencialmente en tienda.

En base a este indicador se tiene que del total de clientes que ha ingresado al sitio web al menos una vez durante el último año, un 21,7% se encuentra enganchado y un 78,3% no lo está.

## 8 CONCLUSIONES

### 8.1 Conclusiones del Proyecto

El principal objetivo de la presente memoria de título consiste en el diseño de acciones de apoyo a la personalización y gestión de campañas de publicidad web en un retail financiero. La importancia de esto radica en que la fidelidad de un cliente hacia una empresa, y su valor como cliente, está directamente relacionado con generar una experiencia de uso personalizada. Lo anterior se encuentra alineado con la estrategia de omnicanalidad que siguen las empresas del rubro.

Con el fin de resolver el objetivo planteado, se utilizó como principal herramienta el desarrollo de modelos econométricos y minería de datos, los cuales permiten modelar la interacción de los clientes y luego clasificarlos de acuerdo a la probabilidad a que realicen alguna de las interacciones mencionadas.

De acuerdo a esto, los perfiles de los clientes se determinan por:

- ✓ Ingreso Web: Clientes principalmente compuestos por **adultos jóvenes**, relacionados al canal digital al comprar en la **web del retail interno**, además, son activos en compras hedonistas como **viajes** y utilizan la tarjeta principalmente por el **club de puntos**.
- ✓ Shadow Box: Clientes que tienen **categoría media** o **alta**, también familiarizados al canal digital al comprar en la **web del retail interno**, su tarjeta es tipo **Visa** y compran en rubros relacionados con experiencias, como **restaurant** y **viajes**.
- ✓ Calugas: Clientes que tienen **categoría media** o **alta**, nuevamente familiarizados con el canal digital al comprar en la **web del retail interno** y que han tenido transacciones en **servicios básicos**, **educación** y **restaurant**. Asimismo, **solicitan avances en efectivo**.
- ✓ Burbujas: Clientes que utilizan la tarjeta principalmente para buscar financiamiento, ya que solicitan **avances en efectivos** y **créditos de consumo**, y además **compra en cuotas**. Asimismo, ocupan la tarjeta en **supermercados externos**, los cuales son competencia directa de uno de los principales negocios del holding.

A partir de estos perfiles, se tiene que los formatos de publicidad web apuntan a clientes diferentes, ya que por un lado Burbujas se enfocan a clientes que buscan financiamiento, mientras que Shadow Box apunta a clientes que utilizan su tarjeta en rubros relacionados a experiencias, sin embargo, la diferenciación entre Shadow Box y

Calugas no es tan marcada, lo que se puede deber a que ambos formatos presentan campañas de contenidos similares.

Asimismo, de acuerdo a estos perfiles obtenidos se propone desarrollar una estrategia, a mediano plazo, de gestionar a tres grupos de clientes: el primero correspondiente a clientes que no ingresan al sitio web de la empresa y tienen mayor probabilidad de ingreso, el segundo a clientes propensos a interactuar con Shadow Box y el último grupo corresponde a clientes propensos a interactuar con Calugas.

La estrategia apunta a migrar clientes al canal digital en primera instancia, y luego, a los activos en el canal digital a interactuar aún más con los toques que genera la empresa, apuntando a que los clientes mejoren su relación con la empresa, logrando así una experiencia más integrada.

Complementariamente, con el fin de que se mejore la gestión de las campañas de marketing en el sitio web, se tienen las diferentes etiquetas que se le asignan a los clientes de acuerdo al comportamiento dentro del sitio web, cuánto ingresan y qué tan fidelizados se encuentran con el sitio web. Estos indicadores permiten a la empresa tener campañas más focalizadas a cómo el cliente utiliza el sitio web.

Finalmente, para el largo plazo, la empresa debería gestionar periódicamente a los clientes que se encuentran activos en la web de los negocios del holding, lo que se puede lograr mediante alianzas con éstos. Asimismo, de acuerdo a los resultados obtenidos, la empresa debe contactar a clientes jóvenes para que utilicen el sitio web apenas realicen apertura de cuenta. Por otro lado, para las campañas web, la empresa debería utilizar el formato adecuado de acuerdo al contenido de la campaña en función de los perfiles identificados.

En definitiva, este trabajo de título muestra que es posible generar una personalización del contenido web y gestionar las campañas de marketing de forma que la empresa pueda implementar acciones de marketing, permitiendo mejorar la experiencia de uso y la relación con la empresa.

## **8.2 Limitaciones y Trabajos Futuros**

La principal limitación de este trabajo tiene relación con que el estudio de las campañas de publicidad se realiza a nivel de formato y no a nivel de una campaña en particular, o bien de campañas similares. Por lo tanto, se agrupan diferentes formas de comunicar y diferentes objetivos de las campañas, perdiendo el efecto de cada una por separado. En particular, no se incluye el efecto que puede tener un porcentaje de

descuento o si la gráfica utiliza algún sesgo cognitivo<sup>17</sup> sobre el comportamiento de un cliente, lo que ayudaría a validar estas técnicas y estandarizar el uso de ellas dentro de la empresa.

Asimismo, en el estudio no se incluyó variables de interacción con otros canales de ventas, como el presencial o la aplicación de Smartphone, lo que podría ayudar a conocer si existe una correlación entre el canal y cómo interactúa con el sitio web. En esta misma línea, solamente se incluye la interacción con el resto de los sitios web del holding en términos de transacciones, no en términos de interacción con éstos, lo que podría aportar a generar alianzas entre las campañas de marketing de todo el holding.

Una última limitante se relaciona con la información utilizada, en particular, todos los análisis se realizaron a nivel del titular de la cuenta, perdiendo la oportunidad de generar acciones sobre los clientes adicionales, quienes pueden tener necesidades diferentes de los titulares. Esto tiene que ver con que el acceso al sitio web es solamente para titulares, por lo que una recomendación a la empresa es evaluar la posibilidad de entregar acceso a clientes adicionales.

Como trabajo futuro en base a este proyecto y con el fin de comprender y mejorar el sitio web de la empresa, se proponen los siguientes:

- ✓ Experimento Complementario: En las líneas de acción se propone un diseño experimental que utiliza como base los resultados de los modelos de propensión, un trabajo futuro complementario a esto corresponde a la realización de experimentos sin previa segmentación y que utilicen sesgos cognitivos para determinar (1) características de los clientes y (2) acción de marketing que entrega mayor mejora de conversión.
- ✓ Navegación web: En el presente proyecto se utilizaron clics en enlaces del sitio web para definir comportamientos de clientes, este trabajo futuro pretende incorporar nuevas métricas, tales como el tiempo que permanece un cliente en cada sección de la página y el movimiento del puntero en el sitio para generar alertas o triggers que entreguen contenido en base a esto.
- ✓ Eyetracking: Complementando con lo anterior, se puede analizar la efectividad de ciertas campañas e incluso del diseño del sitio web utilizando experimentos con Eyetracking, lo que entrega mapas de calor en donde los clientes ponen mayor atención.

---

<sup>17</sup> Tales como Anchoring Effect, Loss Aversion, Framing Effect, entre otros.

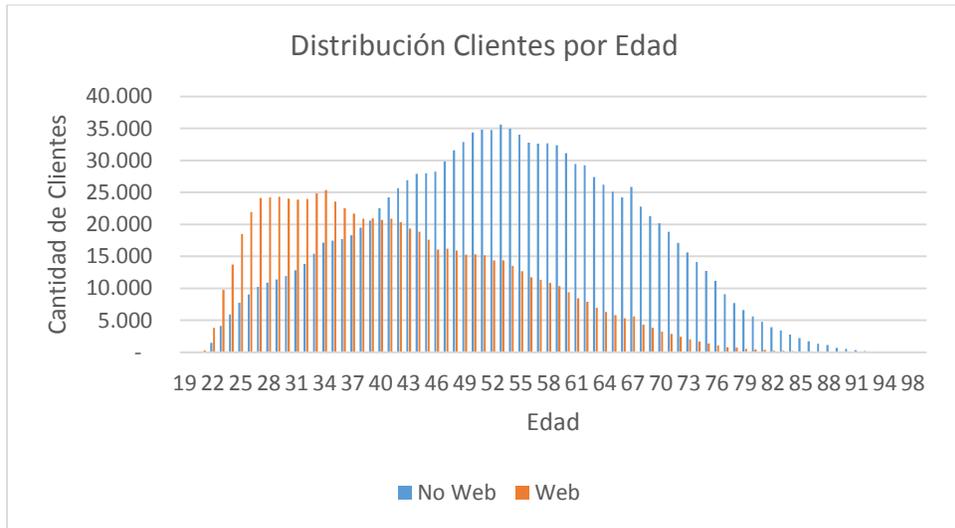
## 9 BIBLIOGRAFÍA

- [1] Instituto Nacional de Estadísticas, “Boletín Sectores Económicos”, septiembre 2015, Santiago, Chile.
- [2] A.T. Kearney, 2015. “Global Retail Development Index 2015”.
- [3] Banco Central de Chile, “Series actividad económica mensual sectorial”, Santiago, 2015.
- [4] WOOLDRIGE, J. 2009. Introductory Econometrics: A Modern Approach, Chapter 17, Limited Dependent Variable Models and Sample Selection Corrections.
- [5] SONG, Y. & LU, Y. 2015. Decision Tree Methods: Applications for Classification and Prediction. Shanghai Archives of Psychiatry, 27(2), 130-135.
- [6] GALLARDO M., C. 2016. Identificación de clientes con patrones de alta interacción con los drivers de una tarjeta de crédito. Memoria para optar al título de Ingeniero Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
- [7] BLOCH, P.H., & RICHINS, M.L. 1983. Shopping without purchase: An investigation of consumer browsing behavior. NA-Advances in Consumer Research, 10.
- [8] ROJAS G., A. 2014. Estudio experimental de automatización de email marketing en un retail online. Memoria para optar al título de Ingeniera Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
- [9] ROJAS S., D. 2013. Configuración de la oferta de productos hogar para clientes residenciales de una empresa de telecomunicaciones. Memoria para optar al título de Ingeniero Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
- [10] FUENZALIDA C., B. 2012. Pronóstico de efectividad de promociones sobre clientes, dadas sus características y respuestas pasadas. Memoria para optar al título de Ingeniera Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

# 10 ANEXOS

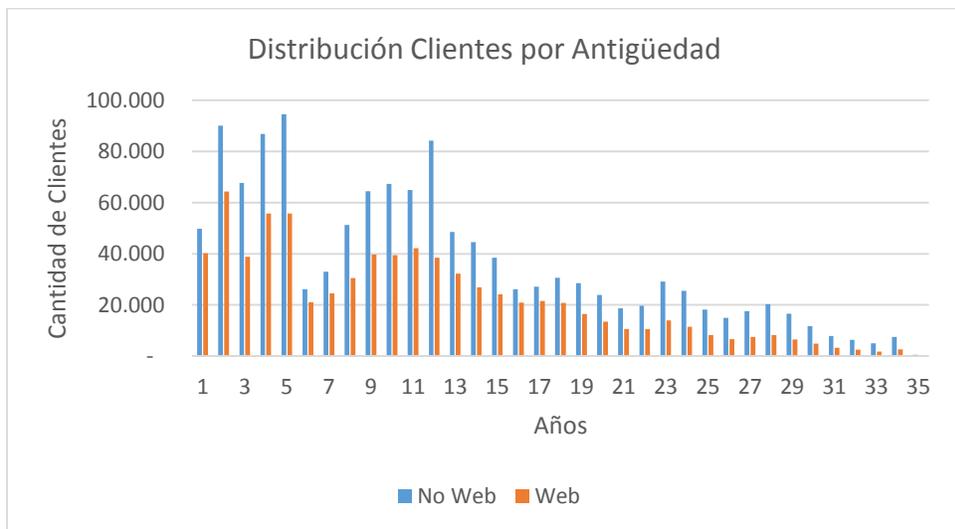
## 10.1 Anexos de Descripción de Variables

### 10.1.1 Anexo I: Distribución de Clientes por Edad



Anexo I: Distribución de clientes según su edad.  
Fuente: Elaboración propia.

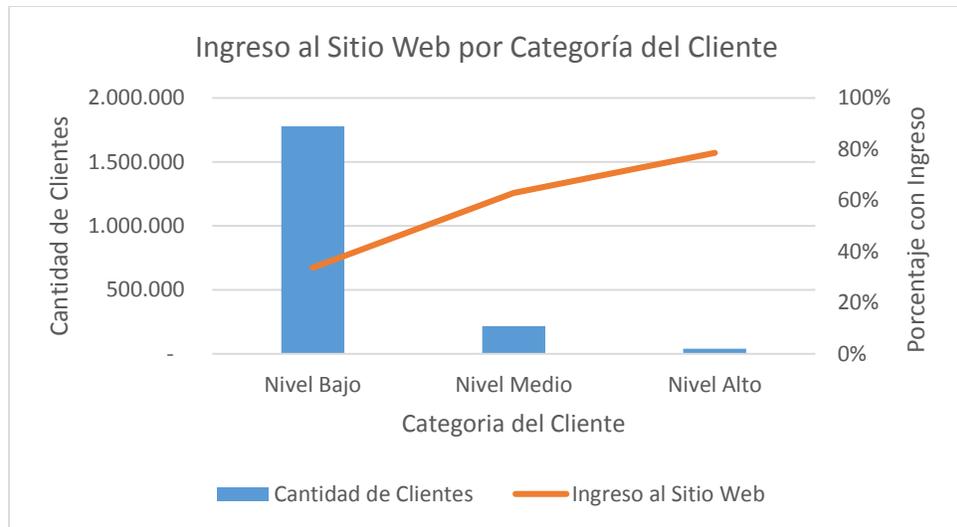
### 10.1.2 Anexo II: Distribución de Clientes por Antigüedad



Anexo II: Distribución de clientes según la antigüedad con la tarjeta de crédito.  
Fuente: Elaboración propia.

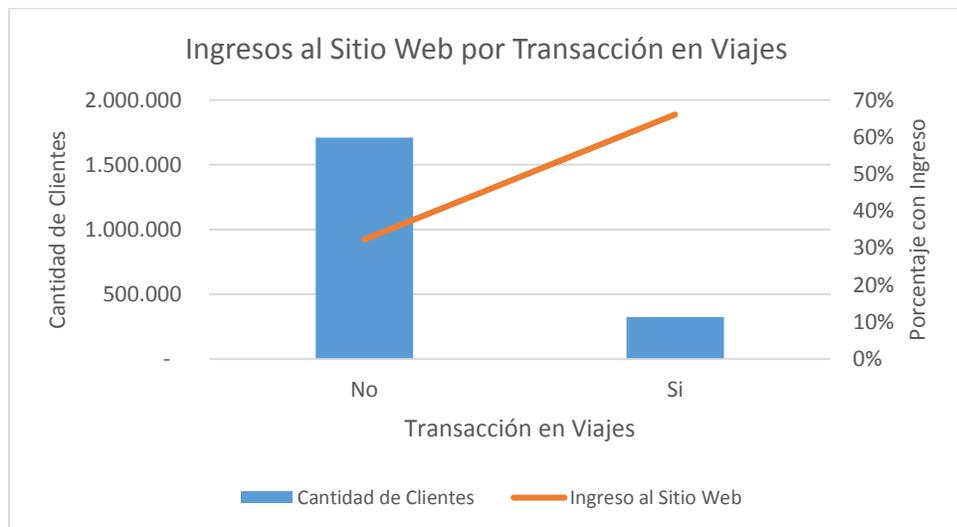
## 10.2 Anexos de Análisis Exploratorio

### 10.2.1 Anexo III: Relación Ingresos al Sitio Web y Categoría del Cliente



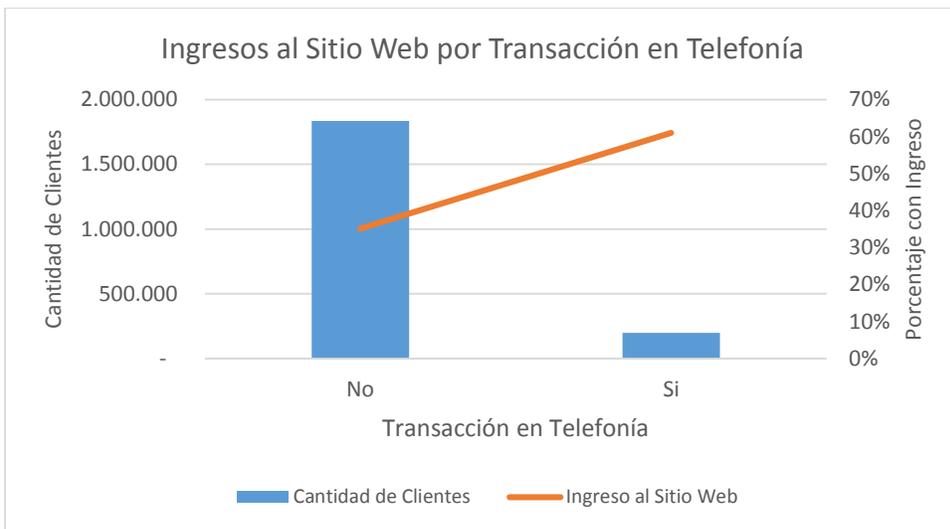
Anexo III: Relación entre ingresos al sitio web y la categoría del cliente.  
Fuente: Elaboración propia.

### 10.2.2 Anexo IV: Relación Ingresos al Sitio Web y Transacción en Viajes



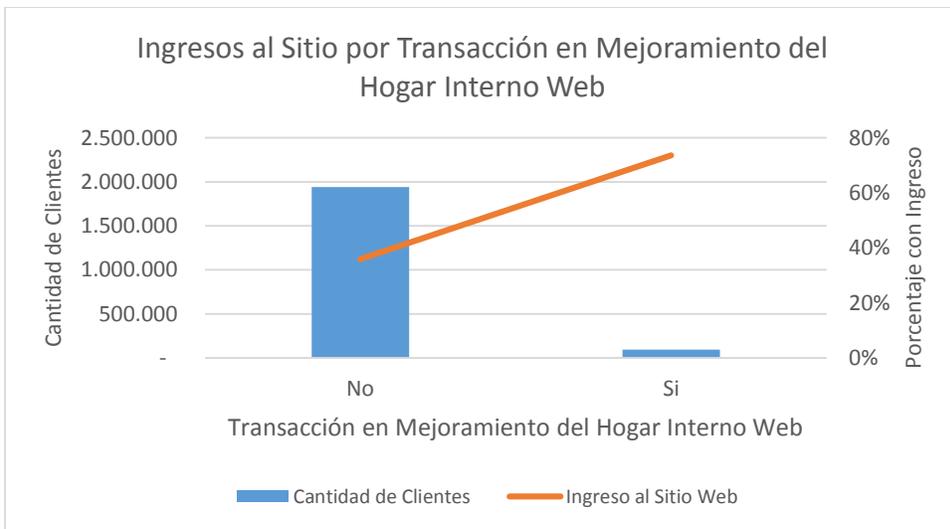
Anexo IV: Relación entre ingresos al sitio web y haber realizado transacciones en el rubro externo de viajes.  
Fuente: Elaboración propia.

### 10.2.3 Anexo V: Relación Ingresos al Sitio Web y Transacción en Telefonía



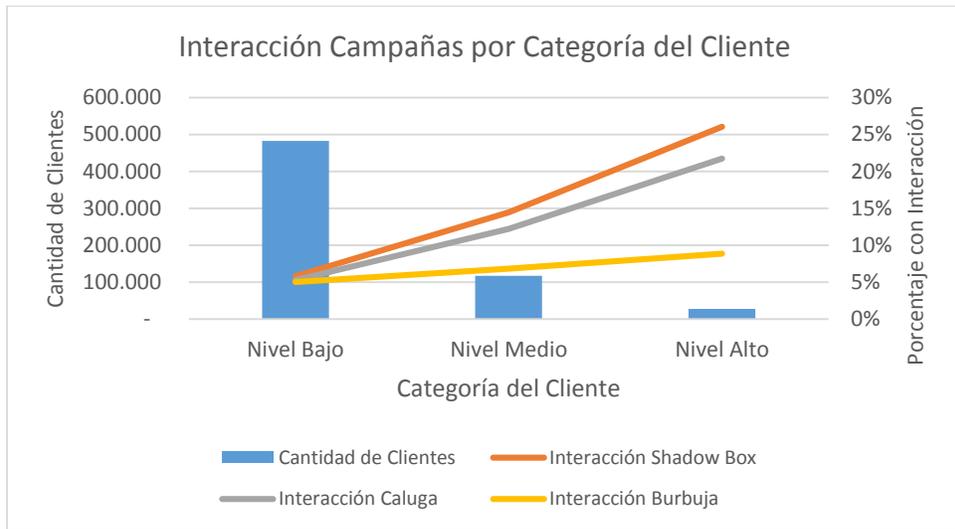
Anexo V: Relación entre ingresos al sitio web y haber realizado transacciones en el rubro de telefonía. Fuente: Elaboración propia.

### 10.2.4 Anexo VI: Relación Ingresos al Sitio Web y Transacción en la Web del Mejoramiento del Hogar Interno



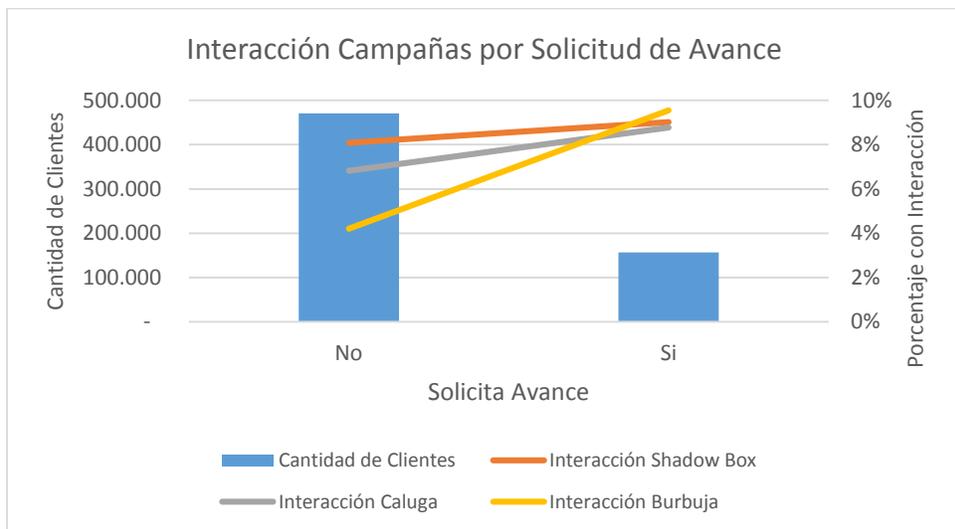
Anexo VI: Relación entre ingresos al sitio web y haber realizado transacciones en la web del mejoramiento del hogar interno. Fuente: Elaboración propia.

### 10.2.5 Anexo VII: Relación Interacción Campañas y Categoría del Cliente



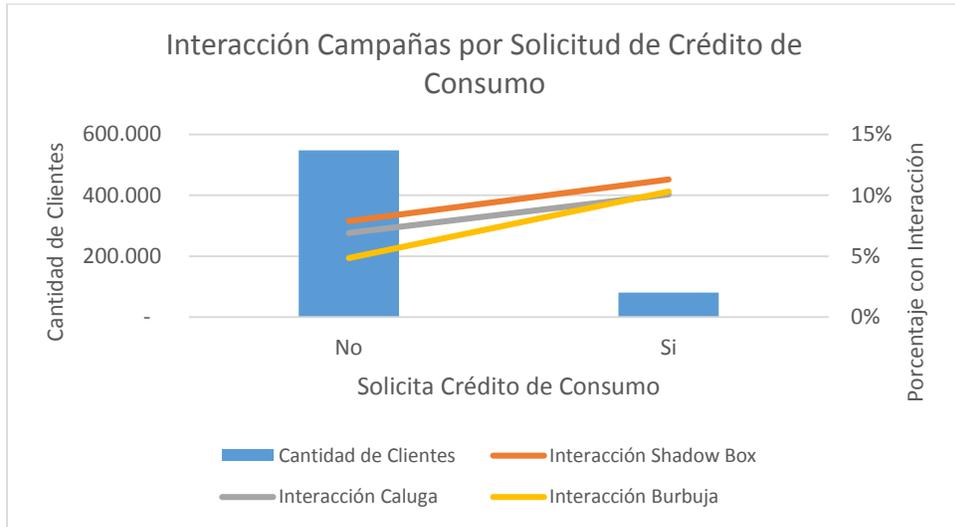
Anexo VII: Relación entre interacción con campañas y la categoría del cliente.  
Fuente: Elaboración propia.

### 10.2.6 Anexo VIII: Relación Interacción Campañas y Solicitud de Avance en Efectivo



Anexo VIII: Relación entre interacción con campañas y haber solicitado un avance en efectivo.  
Fuente: Elaboración propia.

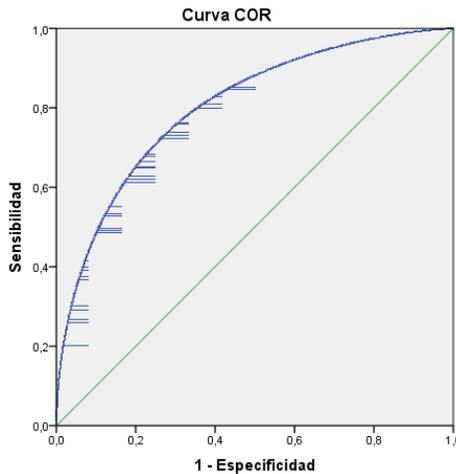
## 10.2.7 Anexo IX: Relación Interacción Campañas y Solicitud Crédito de Consumo



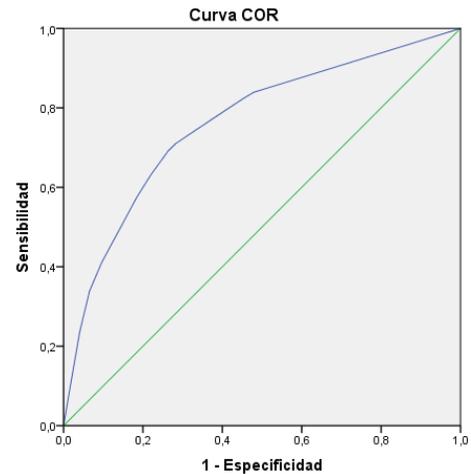
Anexo IX: Relación entre interacción con campañas y haber solicitado un crédito de consumo.  
Fuente: Elaboración propia.

## 10.3 Anexos de Modelos de Propensión

### 10.3.1 Anexo X y XI: Curva COR Modelo Ingreso Web



Los segmentos de diagonal se generan mediante empates.

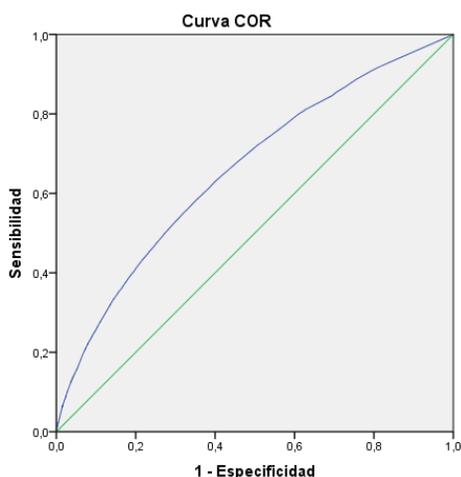


Los segmentos de diagonal se generan mediante empates.

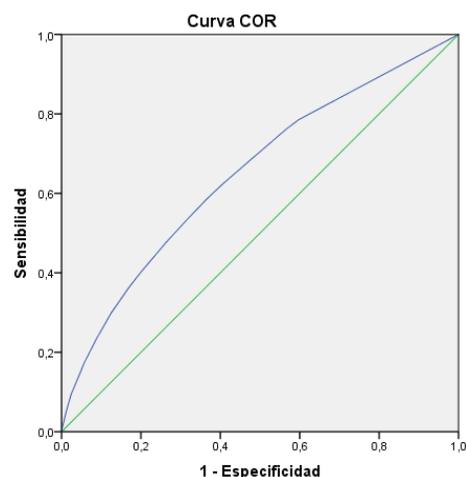
Anexo X: Figura izquierda corresponde a curva COR Regresión Logística Modelo Ingreso Web.  
Fuente: Elaboración propia.

Anexo XI: Figura derecha corresponde a curva COR Árbol Decisión Modelo Ingreso Web.  
Fuente: Elaboración propia.

### 10.3.2 Anexo XII y XIII: Curva COR Modelo Publicidad Web



Los segmentos de diagonal se generan mediante empates.

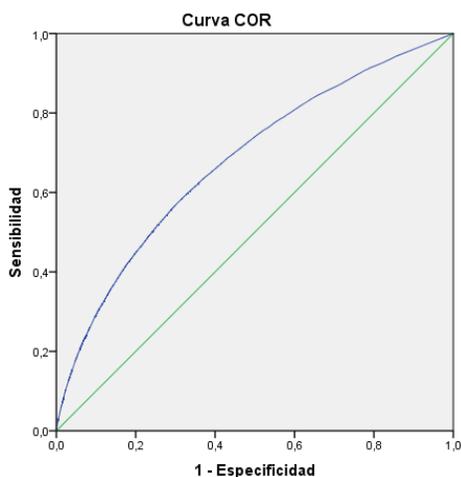


Los segmentos de diagonal se generan mediante empates.

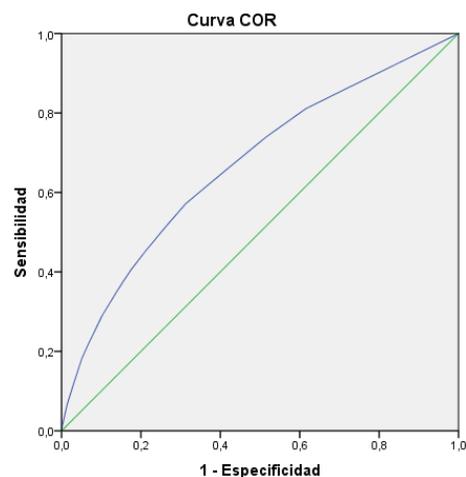
Anexo XII: Figura izquierda corresponde a curva COR Regresión Logística Modelo Publicidad Web.  
Fuente: Elaboración propia.

Anexo XIII: Figura derecha corresponde a curva COR Árbol Decisión Modelo Publicidad Web.  
Fuente: Elaboración propia.

### 10.3.3 Anexo XIV y XV: Curva COR Modelo Calugas



Los segmentos de diagonal se generan mediante empates.

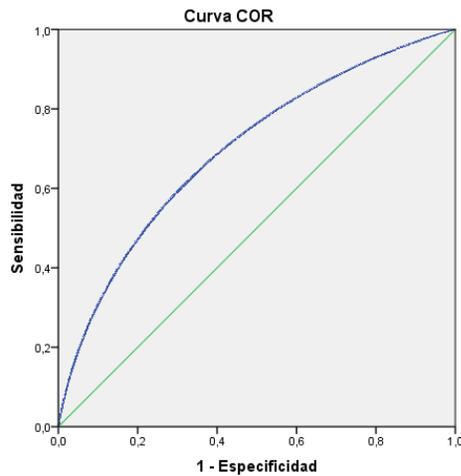


Los segmentos de diagonal se generan mediante empates.

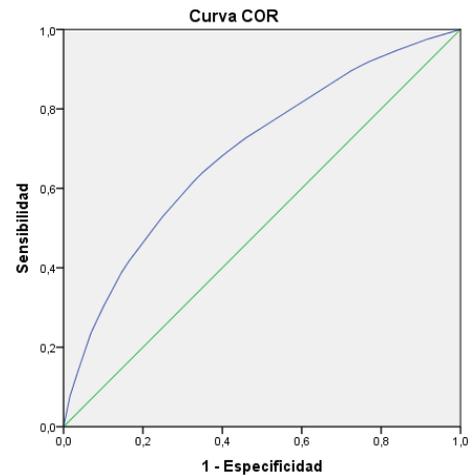
Anexo XIV: Figura izquierda corresponde a curva COR Regresión Logística Modelo Calugas.  
Fuente: Elaboración propia.

Anexo XV: Figura derecha corresponde a curva COR Árbol Decisión Modelo Calugas.  
Fuente: Elaboración propia.

### 10.3.4 Anexo XVI y XVII: Curva COR Modelo Shadow Box



Los segmentos de diagonal se generan mediante empates.

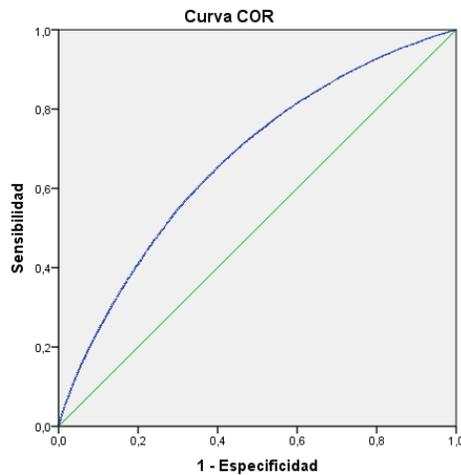


Los segmentos de diagonal se generan mediante empates.

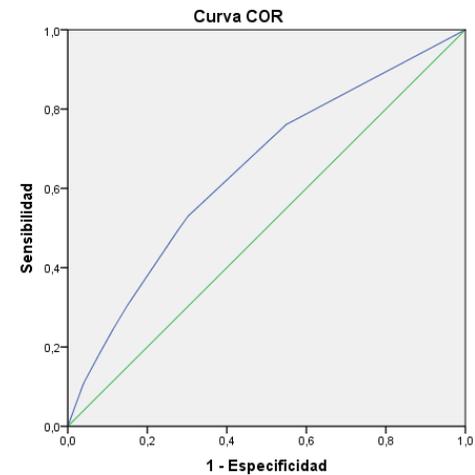
Anexo XVI: Figura izquierda corresponde a curva COR Regresión Logística Modelo Shadow Box.  
Fuente: Elaboración propia.

Anexo XVII: Figura derecha corresponde a curva COR Árbol Decisión Modelo Shadow Box.  
Fuente: Elaboración propia.

### 10.3.5 Anexo XVIII y XIX: Curva COR Modelo Burbujas



Los segmentos de diagonal se generan mediante empates.

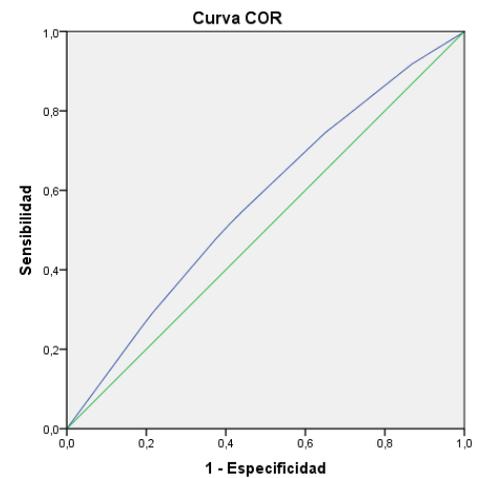
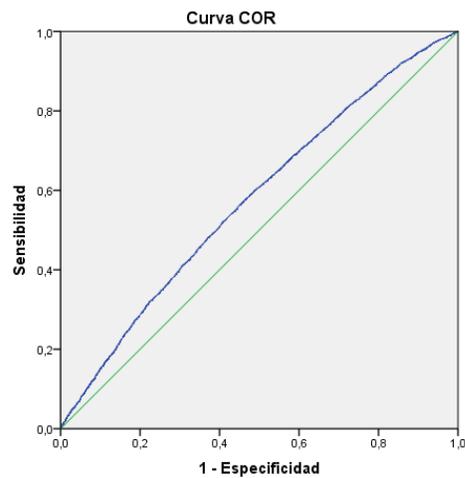


Los segmentos de diagonal se generan mediante empates.

Anexo XVIII: Figura izquierda corresponde a curva COR Regresión Logística Modelo Burbujas.  
Fuente: Elaboración propia.

Anexo XIX: Figura derecha corresponde a curva COR Árbol Decisión Modelo Burbujas.  
Fuente: Elaboración propia.

### 10.3.6 Anexo XX y XXI: Curva COR Modelo Puntos



Anexo XX: Figura izquierda corresponde a curva COR Regresión Logística Modelo Puntos.  
Fuente: Elaboración propia.

Anexo XXI: Figura derecha corresponde a curva COR Árbol Decisión Modelo Puntos.  
Fuente: Elaboración propia.