



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

**“DESARROLLO DE UN MODELO DE ATRIBUCIÓN DE CONVERSIONES *CROSS-CHANNEL* PARA UN RETAIL ONLINE”**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

CAMILA PAZ PINEDA PARRA

PROFESOR GUÍA:  
ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN:  
MARCEL GOIC FIGUEROA  
CAROLINA SEGOVIA RIQUELME

SANTIAGO DE CHILE  
2017

## **RESUMEN DE LA MEMORIA PARA OPTAR AL**

**TÍTULO DE:** Ingeniera Civil Industrial

**POR:** Camila Paz Pineda Parra

**FECHA:** 20/11/2017

**PROFESOR GUÍA:** Alejandra Puente Chandía

### **RESUMEN EJECUTIVO**

La industria del *retail* debe adaptarse a los cambios en las necesidades de los clientes y del mercado. En particular, la empresa en la que se desarrolla el estudio se posiciona como la compañía líder en Chile, por lo cual debe saber responder a este dinamismo, para ello cuenta con diversas estrategias entre las cuales destaca una amplia gama de canales para contactar al cliente. Ante esto es necesario conocer cómo se mueve el cliente entre los distintos medios, y así comprender el rol que cumple cada uno y su relevancia en el proceso de compra. Actualmente se atribuye el 100,0% de una conversión al último canal que visita el cliente previo a la compra, lo cual se denomina modelo de atribución *last touch*, este modelo ignora todo el proceso previo, por lo que no es muy adecuado para canales que no participan en estas últimas etapas.

El objetivo principal de la memoria es proponer un modelo de atribución de conversiones *cross channel*, que incorpore información del negocio, para así conocer los aportes de cada canal considerando su participación no solo en la última etapa. Para la realización del modelo se realizó en primera instancia, una exploración de datos, luego se desarrolló el modelo basado en cadenas de Markov de hasta tercer orden y posteriormente se validaron los resultados obtenidos para determinar el modelo que mejor se adecuaba a la situación de la empresa. Se complementaron estos resultados con la realización de un análisis descriptivo para el canal de Facebook, indagando en las interacciones dentro de este medio y las discrepancias existentes.

Por las características de los datos de la empresa, el modelo que mejor describe la atribución de conversiones es un modelo basado en cadenas de Markov de tercer orden, con este modelo fue posible constatar que algunos canales como Directo eran un tanto sobreestimados con la utilización del modelo de atribución *last touch*, mientras que otros resultaban estar subestimados por esta forma de atribuir. Dada la naturaleza de los canales, se recomienda observar la atribución *last touch* y *first touch* como complemento del modelo que se propone, de esta forma es posible obtener una visión mucho más completa de los aportes que realiza cada uno de los medios con los cuales opera la empresa.

A modo de trabajo futuro se propone en primer lugar, realizar el estudio considerando un modelo basado en Markov de cuarto orden, este modelo debiese ser más preciso que el modelo realizado. Además se sugiere calcular el modelo para los otros países en los que opera la empresa y así disponer de modelos construidos de acuerdo a las rutas de los clientes en cada país. Por otro lado si se tuviese información más desagregada a nivel usuario, sería interesante caracterizar a los clientes que realizan ciertas rutas, y así segmentar e implementar acciones que incentiven a esos clientes a completar sus caminos habituales hasta concretar la compra.

## AGRADECIMIENTOS

Me gustaría agradecer en primer lugar a mi familia. A mis padres por enseñarme a luchar por lo que queremos, a continuar a pesar de las adversidades y a aprender de los errores. Les agradezco sobre todo por apoyarme en cada etapa de mi vida, siempre con mucho cariño y preocupación. A mi hermana Fabiola también le quiero agradecer por estar dispuesta siempre a ayudarme, por la empatía, la comprensión, el cariño y el apoyo especialmente en esta última etapa.

También quisiera mencionar a mis amigos de plan común, con los cuales compartí los primeros años de la universidad y quienes me ayudaron a distraerme en esos difíciles semestres. Gracias por los buenos momentos.

A mis compañeros de carrera, especialmente a mis amigos Andrew y Martin. Gracias por el apoyo durante los primeros años de la carrera, sin duda nunca olvidaré las tardes en las que nos dedicábamos a hacer tareas o las mañanas de estudio bajo presión antes de los controles de las 14:00 hrs. Tratando de ser productivos y asociar los contenidos muchas veces terminábamos asociándolo a algo chistoso.

A mis amigos más cercanos de la universidad, Byron, Juan, Guillermo y Javiera. Muchas gracias por compartir tantas anécdotas divertidas que sin duda amenizaron mi paso por la universidad. Les agradezco también por escucharme y apoyarme en los momentos difíciles.

A mi profesora, Alejandra Puente, por guiarme y siempre estar dispuesta a ayudarme de la mejor forma cuando estaba acoirazada. A mi profesor Marcel Goic, por la paciencia y la ayuda en aquellos momentos en los cuales no tenía muy claro cómo seguir avanzando en mi proyecto.

Al equipo de Falabella.com, por siempre ser gentiles y ayudarme en lo que fuese necesario. Me gustaría agradecer especialmente al equipo de SEO SEM. En primer lugar a Edurne Benito, por guiarme en el proceso de la memoria mostrando siempre buena disposición ante mis dudas. A Carolina Cifuentes, por enseñarme todo durante mi práctica, con la amabilidad que la caracteriza; y a Andrés Echavarría, por ayudarme a resolver mis inquietudes durante el proceso de la memoria.

Agradezco además a todas aquellas personas con las que compartí este proceso, las que me acompañaron y que me dieron ánimos en esta última etapa.

# TABLA DE CONTENIDO

1. INTRODUCCIÓN .....	1
1.1 TENDENCIAS DEL MERCADO DEL RETAIL .....	1
1.2 DESCRIPCION DE LA EMPRESA .....	2
2. DESCRIPCIÓN DEL ESTUDIO.....	3
2.1 JUSTIFICACIÓN DE LA OPORTUNIDAD .....	3
2.2 OBJETIVOS.....	4
2.3 ALCANCES .....	5
3. MARCO CONCEPTUAL.....	5
3.1 DEFINICIONES PREVIAS.....	5
3.1.1. MÉTRICAS PROPIAS DEL E-COMMERCE.....	5
3.1.2. MODELOS DE ATRIBUCIÓN.....	8
3.2 MARCO TEÓRICO .....	9
3.2.1. CADENAS DE MARKOV .....	9
3.2.2. EFICIENCIA DE UN MODELO .....	13
4. DESARROLLO METODOLÓGICO .....	15
4.1 DESARROLLO DEL MODELO.....	15
4.1.1. SELECCIÓN DE DATOS.....	15
4.1.2. PREPROCESAMIENTO DE DATOS.....	16
4.1.3. EXPLORACIÓN PRELIMINAR MODELO .....	17
4.1.4. REALIZACIÓN DEL MODELO .....	21
4.2 RESULTADOS .....	22
4.3 VALIDACIÓN DEL MODELO.....	27
4.4 PROPUESTA DEL MODELO DE ATRIBUCIÓN .....	32
4.5. ANÁLISIS COMPLEMENTARIO SOBRE FACEBOOK .....	35
5. RECOMENDACIONES Y CONCLUSIONES .....	44
5.1 RECOMENDACIONES PARA LA EMPRESA .....	44
5.2 PRINCIPALES CONCLUSIONES .....	44
5.3 TRABAJO FUTURO .....	45
6. BIBLIOGRAFÍA .....	47
7. ANEXOS .....	49

# 1. INTRODUCCIÓN

## 1.1 TENDENCIAS DEL MERCADO DEL RETAIL

El mercado del *retail* ha experimentado cambios debido a transformaciones de las necesidades de los clientes y a variaciones del entorno en el que están inmersas estas empresas. La tecnología ha sido la gran protagonista de estos cambios, dándole dinamismo a un sector que se encuentra en una etapa de desaceleración desde el 2014 hasta la fecha (Cámara de Comercio de Santiago, 2016).

Una de las transformaciones más relevantes en el último tiempo ha sido el auge del comercio electrónico. Actualmente, el 70,0% de los usuarios de internet visita sitios de *retail* (Economía y Negocios, 2015) y 4 millones de personas en Chile realiza compras *online* de forma habitual<sup>1</sup>, esto ilustra la importancia de este canal tanto para apoyar decisiones de compra como también para realizarlas.

Hace algunos años las personas comenzaron a utilizar los dispositivos móviles a tal punto que hoy, la penetración de estos aparatos alcanza cifras del 98,0% a nivel mundial (The World Bank, 2016). Debido a esto, parte importante del tráfico online comenzó a migrar a diversos dispositivos, tales como celulares, *tablets* e incluso consolas de juegos, modificando también los procesos de compra de los usuarios hacia una perspectiva *multidevice*<sup>2</sup>.

Hoy, uno de los principales focos de las empresas del rubro no solo es saber responder a los desafíos vinculados a las tecnologías, sino que también competir tanto a nivel nacional como a nivel mundial. Este panorama ha llevado a las compañías a innovar cada vez más en nuevas formas de contactar a los clientes para iniciar un proceso de compra. Actualmente entre los desafíos que enfrenta el mercado se encuentra el correcto manejo de una amplia variedad de dispositivos y canales (multicanalidad) entre los cuales navega el cliente, lo cual implica desde saber combinarlos de forma sinérgica hasta conocer la importancia de cada canal en el proceso de compra. Esto último se denomina problema de atribución de conversiones y consiste en determinar una forma de atribuir a los canales las distintas conversiones que se llevan a cabo en un sitio web, de esta forma es posible ajustar las inversiones en cada medio de acuerdo a la relevancia que tienen en los diferentes procesos de compras que realizan los clientes.

---

<sup>1</sup> Información proporcionada por RRHH de Falabella Retail.

<sup>2</sup> En el presente estudio se considerará *device* como los distintos tipos de dispositivos por los cuales puede ingresar un cliente a un sitio web, por otro lado al mencionar *channel* se referirá a todos los medios por los cuales puede ocurrir una interacción cliente-empresa.

## 1.2 DESCRIPCION DE LA EMPRESA

Falabella es una empresa multinacional chilena perteneciente a la industria del *retail*, actualmente esta compañía se desempeña en varias líneas de negocio, siendo la principal y más característica, la de tiendas por departamentos.

Con 127 años en el rubro, esta empresa se puede considerar como la iniciadora de las tiendas por departamento en Chile. Hoy su presencia abarca seis países de Sudamérica [Anexo 1], alcanzando 476 tiendas en total y cuyos ingresos para el año 2016 superaron los 4.300 millones de dólares solo considerando tiendas por departamento (Memoria Anual Falabella, 2016).

Falabella crea valor a sus clientes al brindar una oferta innovadora de productos en diversas categorías y líneas, ofreciendo soluciones para distintos aspectos de sus vidas. En su rubro de tienda por departamento gran parte de este valor es percibido por medio de las marcas exclusivas que ofrece Falabella, así como también el surtido y variedad en ciertas categorías.

A lo largo de toda su trayectoria Falabella se ha caracterizado por liderar en el sector de las multitiendas. Este hecho se evidencia desde que comenzó a operar siendo la primera sastrería en Chile, luego transformándose en la única tienda por departamentos del país, y posteriormente creando la tarjeta CMR, primer instrumento de crédito para segmentos hasta entonces calificados como riesgosos. Este carácter innovador y la oportuna adaptación a los cambios en las preferencias de los clientes y del entorno, representan su ventaja competitiva, la cual permite a Falabella posicionarse como líder de mercado con un 38,0% de los ingresos de la industria, a lo cual le sigue Cencosud con un 27,0% y Ripley, alcanzando un 21,0% (Economía y Negocios., 2015.). [Anexo 2]

Hoy en día ese carácter innovador se evidencia en el interés que tiene Falabella en el mercado electrónico. Debido al auge que ha tenido este mercado en los últimos años la compañía ha construido un sitio web de gran calidad que recibe más de 400.000 visitas a diario.<sup>3</sup> En dos años el personal dedicado a trabajar en Falabella internet se ha triplicado. Sumado a esto, han implementado herramientas para posicionar el sitio y para mejorar la experiencia al navegar; además la mayoría de sus campañas de marketing se lanzan tanto en canales convencionales como también utilizando una amplia gama medios digitales.

La gestión del correcto funcionamiento del sitio web demanda un equipo capacitado en diversos aspectos. Una de las áreas requeridas es la de SEO SEM, la cual se encarga de dos aspectos importantes del sitio, el SEO o *Search Engine Optimization* y el SEM o *Search Engine Marketing*, ambos tienen como objetivo lograr posicionamiento del sitio web pero se diferencian en la forma de obtención de éste, el primero lo hace de forma orgánica y el segundo se relaciona con publicidad online que se realiza en los buscadores

---

<sup>3</sup> Información entregada por el área de SEO/SEM de Falabella.com

de Google. En la práctica, esta área está encargada además de gestionar todos los medios digitales<sup>4</sup>, implementar medidas para captar más tráfico al sitio por medio de otras fuentes, crear estrategias de optimización para potenciar visitas y ventas; y negociar posicionamiento y publicidad online con proveedores de marcas en el sitio. En resumen, el equipo que conforma esta área debe preocuparse por derivar tráfico al sitio desde distintas fuentes.

El presente trabajo es solicitado por el área de trabajo descrita, con el propósito de abordar el problema de atribución de conversiones que existe hoy en Falabella.com. En la empresa, actualmente cuando se realiza una conversión esta se atribuye a la última interacción antes de la venta sin considerar las visitas previas que pudo haber hecho el cliente usando otros canales, este es el modelo de atribución *last touch* o de último toque, el cual se cree que podría no estar reflejando el real aporte de cada canal a lo largo del proceso de compra de un determinado cliente.

## 2. DESCRIPCIÓN DEL ESTUDIO

### 2.1 JUSTIFICACIÓN DE LA OPORTUNIDAD

Considerando la amplia y constantemente creciente oferta en el comercio online, las firmas han debido implementar diversas formas de captar y cautivar visitantes para conducirlos hacia la conversión. Dado esto, los clientes son expuestos a múltiples estímulos desde distintos dispositivos y medios, mientras navegan por internet. Algunos de estos estímulos generan interacciones, las cuales muchas veces contribuyen a que el cliente convierta a través de un sitio web. Dada la gran cantidad de canales y dispositivos por los cuales se pueden generar estas interacciones, resulta complejo saber realmente cual medio contribuyó a generar una conversión. De esto se trata el problema de atribución, una temática recurrente en los *e-commerce* y que repercute en que como muchos desconocen cuánto aporta cada medio a la venta, tampoco tienen certeza de lo efectiva que es la distribución actual de su inversión en cada medio.

La atribución puede estudiarse a través de canales de marketing (*cross channel*) y también a través de dispositivos (*cross device*). Lo primero permite saber la contribución que realiza cada canal en la ruta de un cliente hasta la conversión y el segundo tipo de estudio entrega información sobre las migraciones entre los diferentes dispositivos y los roles que cumplen a lo largo del proceso de compra.

Muchas empresas utilizan el modelo de atribución *last touch*, sin embargo este modelo no da crédito a las interacciones que se realizan al inicio o en el medio del proceso de compra, y que son igualmente relevantes. Por ejemplo, si una persona ingresa al sitio por medio de un anuncio orgánico en Google, dos horas después ingresa al sitio por medio

---

<sup>4</sup> Medios Digitales incluye Facebook, anuncios en la red *display*, anuncios en gmail, en sitios con alto tráfico y asociados.

de un mail y en la noche convierte por medio de un anuncio en Facebook, toda la venta fue gatillada por este último anuncio bajo el modelo de *last touch*, mientras que los dos primeros toques igualmente participaron en el proceso. De la misma forma este modelo puede considerar dos ventas del mismo producto por un cierto medio como iguales, cuando podrían ser completamente distintas en el proceso de compra involucrado en cada una. [Anexo 3]

Las acciones que está realizando Falabella en temas de atribución de conversiones no son muchas. Actualmente utilizan el modelo de último toque tanto para canales como también para dispositivos, pero declaran que existe un aporte a la conversión distinto al que pueden ver con este modelo, sobre todo en el canal de Facebook y con los dispositivos móviles, que dado que participan en las primeras etapas del proceso de compra (Fuenzalida, 2016) terminan siendo subestimados por esta forma de atribuir.

Además del modelo de atribución mencionado también existen otros modelos que asignan aportes a la conversión considerando otras distribuciones, sin embargo siempre lo más recomendable es utilizar un modelo de atribución basado en los datos de la empresa, pues de esta forma el modelo le asigna los aportes a los canales de acuerdo a la contribución que realizan a lo largo de todo el proceso de compra, teniendo información más precisa para un determinado negocio que puede permitir una mejor distribución de los esfuerzos entre todos los canales.

## **2.2 OBJETIVOS**

### **2.2.1. OBJETIVO GENERAL**

Proponer un modelo de atribución de conversiones considerando los distintos canales de marketing que utiliza un *retail* online.

### **2.2.2. OBJETIVOS ESPECÍFICOS**

- Describir los *customer journey*, identificando las secuencias más frecuentes y caracterizando los canales involucrados.
- Modelar la migración de tráfico entre los distintos canales de marketing.
- Evaluar el modelo propuesto con variaciones del mismo y con los modelos disponibles en Omniture (*first touch* y *last touch*).
- Estudiar la contribución de las interacciones de Facebook a la conversión del mismo canal como también a la que se realiza por medio de otros canales.
- Entregar recomendaciones sobre acciones que se pueden realizar considerando el aporte de los canales de marketing.

## 2.3 ALCANCES

El proyecto que se describe se remitirá sólo a Chile en una primera instancia, considerando información de Google Analytics correspondiente al año 2016 y 2017. Además abordará modelos basados en cadenas de Markov de hasta tercer orden, el modelo de cuarto orden queda propuesto debido al tiempo de procesamiento que requiere, considerando el número de canales con los cuales trabaja la empresa y el número de iteraciones que esto implica.

El modelo que se propone se limita a abordar la atribución de conversiones con información *cross channel*, esto debido a que la disponibilidad de datos en Google Analytics no permite incorporar en el modelo información *cross device*, pues no se puede acceder a los datos desagregados a nivel persona e identificar los dispositivos por los cuales se contacta a un determinado cliente.

Como complemento del modelo de atribución se incluirá un análisis particular del canal de Facebook, se hará especial énfasis en este medio debido al interés de la gerencia en este canal en particular<sup>5</sup>. En este análisis se estudiarán las migraciones entre dispositivos dentro de este canal, pudiendo realizar algunos alcances *cross device* de lo que sucede en Facebook.

## 3. MARCO CONCEPTUAL

### 3.1 DEFINICIONES PREVIAS

#### 3.1.1. MÉTRICAS PROPIAS DEL E-COMMERCE

En el comercio online se utilizan determinadas métricas para poder medir el desempeño de los sitios web y de las campañas que se realizan, a continuación se señalan algunos de esos indicadores:

**Visitas:** Se trata de una sesión en el sitio que dura como máximo 30 minutos, en el caso que un usuario esté conectado por más tiempo se consideran como dos visitas distintas.

**Entradas:** Son todas las veces que el usuario ingresó al sitio en un tiempo determinado, independiente de la duración de su sesión.

**Ventas:** Corresponden a las ventas realizadas por medio de la plataforma web en un tiempo determinado.

---

<sup>5</sup> Se tiene un interés de estudiar este canal ya que existe la sospecha que este medio está siendo subvalorado con el modelo de atribución *last touch*.

**Órdenes:** Son todos los pedidos que se realizan en un tiempo determinado, independiente el número de artículos que ellos contengan.

**Impresiones:** Son las veces que se muestra un determinado anuncio en una plataforma publicitaria.

**Clics:** Se refiere a todas las veces en las que un cliente hace clic en un anuncio.

**Tasa de Conversión:** Es un indicador que mide rendimiento comparando cuantos objetivos se alcanzan con respecto al total de visitas que se registraron. Lo más común en *e-commerce* es que estos objetivos sean órdenes, por lo cual en estos casos la tasa de conversión se calcula de la siguiente forma:

$$TC = \frac{\text{órdenes}}{\text{visitas}}$$

**Ticket promedio:** Es el valor de la venta promedio de un cliente que realiza una orden, se calcula:

$$\overline{TX} = \frac{\text{ventas}}{\text{órdenes}}$$

**CPC o Costo por clic:** Se refiere al costo que tiene para cierta plataforma conseguir un clic en un enlace.

**CPM o Costo por Mil Impresiones:** Es el costo que tiene conseguir mil impresiones en cierta plataforma publicitaria.

**CTR o *Clic Through Rate*:** Hace referencia al nivel de calidad que tiene un anuncio en cierta plataforma, específicamente cuantas impresiones se traducen en clics, se calcula:

$$CTR = \frac{\text{clics}}{\text{impresiones}}$$

**Checkout:** Corresponde al proceso en el cual el cliente se dirige al pago de su orden.

**Page Views:** Es el total de páginas vistas en un determinado tiempo.

**Products Views:** Es el total de páginas correspondientes a fichas de productos que un cliente ve en un determinado tiempo.

**Bounces o Rebotes:** Métrica que indica las veces que un usuario abandona el sitio después de haber visitado el sitio menos de 30 segundos. Es un buen indicador de la calidad de un sitio y lo relevante que es su contenido para los usuarios.

**Bounce Rate o Tasa de Rebote:** Corresponde a la cantidad de rebotes que suceden para cierto número de visitas, es decir cuántas de esas visitas son efectivamente rebotes.

$$\overline{TX} = \frac{\text{bounces}}{\text{visitas}}$$

**Customer Journey:** Se refiere al conjunto de todas las interacciones cliente marca a las cuales está sometido un usuario a lo largo de todo su proceso de compra hasta llegar a una conversión, también se utiliza este concepto para referirse al proceso de compra del cliente.

**Canales de marketing:** Son todos los medios por los cuales se contacta al cliente o por los cuales llega el usuario a un sitio web. Esta definición es especialmente importante debido a que este será el significado del concepto que se utilizará en el estudio, el cual podría ser distinto al utilizado en otras fuentes.

Los canales utilizados por la empresa son:

- *Email:* Correos electrónicos.
- *RRSS Pago:* Anuncios de pago en Redes Sociales<sup>6</sup>.
- *RRSS Orgánico:* Anuncios en Redes Sociales sin pago asociado (Publicaciones orgánicas de las cuentas de la marca).
- *Display:* Anuncios dispuestos en la red *display* de Google que se traduce en banners en sitios web con alto tráfico.
- *Anuncios SEM:* Anuncios que aparecen sobre los resultados de búsqueda en Google.
- *Resultados SEO:* Resultados orgánicos que aparecen al realizar una búsqueda en Google.
- *Directo:* Escritura de la URL del sitio web en la barra de navegación de un buscador. Normalmente se autocompleta gracias a las *cookies* del navegador.
- *RMKT Dinámico:* Consiste en mostrar anuncios en base a productos que un usuario ya ha visto en un determinado sitio web.

Entre todos los canales se distinguen aquellos que son iniciados por la firma (*Firm Initiated*) y los que son iniciados por el cliente (*Customer Initiated*). En el primer grupo se encuentran Email, RRSS Pago, RRSS Orgánico y RMKT Dinámico, mientras que en la segunda clasificación se encuentra el resto: SEO, SEM y Directo. Esta denominación

---

<sup>6</sup> Específicamente lo conforman los anuncios que se administran desde la plataforma de Facebook pero que pueden ser anunciados por medio de Facebook o Instagram.

será relevante en el rol que cumplen los distintos canales a lo largo del proceso de compra.

**Dispositivo:** El tráfico online puede llegar al sitio desde distintos tipos de artefactos capaces de conectarse a internet, a eso justamente apunta la definición de este concepto que se utilizará en el estudio, los que utiliza la empresa son los siguientes:

- *Other:* En la empresa, se refiere al tráfico que llega desde usuarios que navegan por *desktop*.
- *Tablet:* Corresponden a usuarios que están navegando por *tablets* o *ipad's*.
- *Mobile Phone:* Tráfico que proviene desde celulares.
- *Gaming Console:* Tráfico que viene desde consolas de juego que se conectan a internet.
- *Media Player:* Son los usuarios que visitan la página desde un reproductor de música.
- *Televisión:* Se refiere al tráfico que llega desde televisores (*Smart TV*).
- *E-Reader:* Es el tráfico que proviene desde dispositivos que funcionan como lectores de libros electrónicos.
- *Set top box:* Corresponde al tráfico que llega al sitio desde dispositivos decodificadores de señal de televisión que se pueden conectar a internet.

### 3.1.2. MODELOS DE ATRIBUCIÓN

En el contexto de los *e-commerce*, un modelo de atribución es un conjunto de reglas que determinan la forma de asignar un porcentaje de la conversión a los diferentes medios que están involucrados en el proceso de compra hacia la conversión. Normalmente se estudia la atribución entre distintos canales de marketing (*cross channel*) o considerando los distintos tipos de dispositivos (*cross device*). Los estudios de estos tipos de atribuciones indican como aporta cada canal de marketing y cada tipo de dispositivo en la conversión final, respectivamente.

Para las empresas resulta de bastante ayuda estos estudios, ya que permiten entre otras cosas, visualizar el camino completo que realizan los clientes hasta realizar una conversión, de esta manera es posible detectar cuantas interacciones en promedio son necesarias para realizar una conversión en el sitio. Por otro lado al estudiar la atribución *cross channel* se podría cuantificar el aporte de cada canal y con esto ajustar los presupuestos de inversión en cada medio. Con respecto a la atribución *cross device*, su estudio permite conocer los roles que cumplen los distintos dispositivos y como se complementan a lo largo del proceso de compra.

Según las diferentes distribuciones que se le asignan a los toques con el cliente, es posible distinguir los siguientes modelos [Anexo 4]:

- *Último toque o last touch*: es el modelo que por defecto utiliza Google Analytics y se basa en asignar toda la atribución a la última interacción previa a la etapa de conversión.
- *Primer toque o first touch*: asigna toda la atribución a la primera interacción del proceso de compra que derivó en una conversión.
- *Último toque indirecto*: También atribuye toda la conversión al último toque sin embargo no considera el medio Directo.
- *Distribución Lineal*: Le otorga un porcentaje igual de la conversión a todos los canales con los que interactúa el usuario a lo largo del proceso de compra.
- *Tiempo de decaimiento o Time Decay*: Asigna un porcentaje de la atribución que se incrementa para los medios que contactaron al cliente en etapas más cercanas a la conversión.
- *Distribución por Posición*: Atribuye un mayor porcentaje de la venta al primer y último toque, y el porcentaje restante se distribuye en partes iguales para los medios que interactuaron en etapas intermedias.
- *Basado en un modelo*: Corresponde a una distribución de la atribución que se construye a partir de los datos de la empresa y no responde a una regla predefinida. Este método se basa en un modelo que se acomoda al proceso de compra propio del negocio, se determinan los aportes de cada canal de acuerdo a la contribución que realizan estos y se determina así la atribución a cada medio.

Uno de los modelos utilizados en el estudio de la atribución es el basado en las cadenas de Markov, a continuación se describe este tópico haciendo énfasis en la aplicación de este tema al problema que se está abordando.

## **3.2 MARCO TEÓRICO**

### **3.2.1. CADENAS DE MARKOV**

Una cadena de Markov es un proceso estocástico con la característica de que la probabilidad de que suceda un evento depende únicamente del estado en el que se encuentra el sistema, el cual resume la información de las etapas previas. De esta forma cada una de las etapas de la cadena se relaciona entre sí, a diferencia por ejemplo de eventos independientes como lanzar una moneda al aire  $n$  veces. Las cadenas de Markov están compuestas por algunos elementos principales, los cuales se detallan en la sección de anexos [Anexo 5].

Se utilizan los procesos de Markov para describir las probabilidades de transición entre canales justamente debido a que la dependencia del evento anterior o propiedad Markoviana se adecua bastante bien a lo que sucede con las migraciones de tráfico online, en las cuales la evolución del sistema, es decir el canal por el cual se realizará el contacto con el cliente dependerá de las interacciones previas. En este caso todo el

*customer journey* sería representado por una cadena de Markov, donde los estados que puede tomar el sistema serían los distintos canales entre los cuales se mueven los clientes, y las transiciones serían justamente las migraciones entre un canal y otro.

La Figura 1 ilustra el proceso de migración de clientes entre los distintos canales hasta que finalizan su proceso de compra. Al conjunto de estados determinados por los canales se les ha añadido la etapa de Inicio y Fin del proceso, el cual puede estar determinado por una conversión o bien, simplemente puede terminar con el hecho de que el cliente no compre el producto. En el estudio que se describe se trabajará solo con procesos que finalizan en conversiones pues la fuente de datos que se está utilizando no es capaz de registrar aquellas rutas que no finalizan en ventas en el sitio.

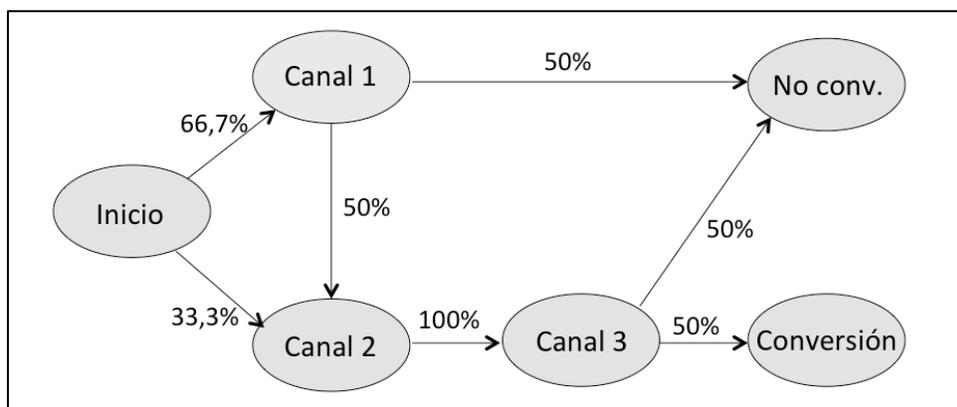


Figura 1: Proceso de atribución de conversiones representado como una cadena de Markov<sup>7</sup>.  
(Anderl, Becker, von Wangenheim, & Schumann, 2016)

Como se anticipó, de acuerdo a las cadenas de Markov, el estado en el cual se encontrará el sistema en un periodo siguiente dependerá de los estados previos. Esta dependencia puede implicar distinto número de estados anteriores, lo cual determina el orden de la cadena de Markov con la cual se estará modelando el fenómeno de estudio. Una cadena de orden uno indica que la probabilidad de que ocurra un evento depende solo del estado actual, pues éste resume la información de los estados previos. Por otro lado, una cadena de orden dos señala que la probabilidad mencionada depende esta vez tanto de la etapa actual como también de la etapa previa.

En la Figura 2 se muestra una representación de la atribución de conversiones como una cadena de Markov de orden uno y dos, específicamente para el estado de “conversión”. En el estudio que se describe se considerará Markov hasta orden cuatro, con el objetivo de poder determinar cuál de ellos describe mejor la situación de atribución de conversiones para Falabella.

---

<sup>7</sup> En la imagen, C1, C2 y C3 son los canales por los cuales pasa el cliente, *start* es donde inicia el proceso, *null* donde termina el proceso sin conversiones y *conv* se refiere a la etapa en la cual el cliente convirtió.

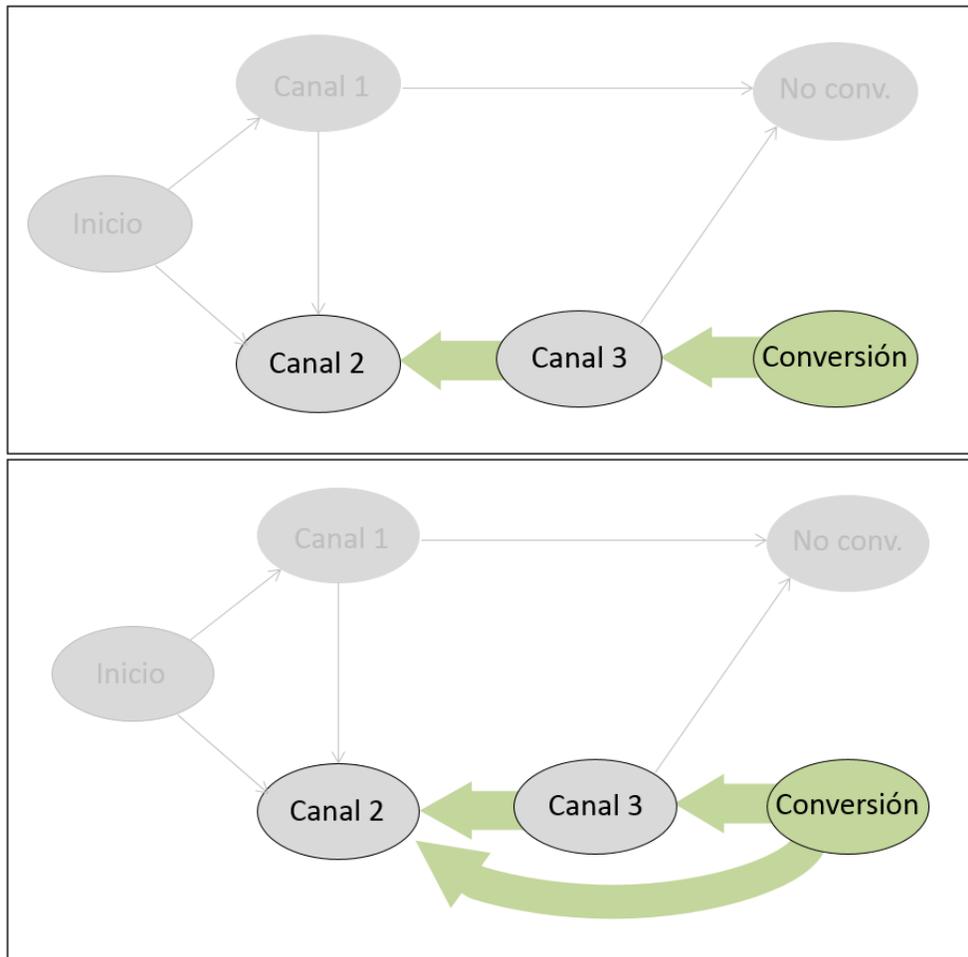


Figura 2: Proceso de atribución de conversiones representado como una cadena de Markov de primer y segundo orden para el estado “Conversión”. (Elaboración propia)

Para cuantificar la importancia que tiene un canal específico en el proceso de compra se utilizará el *Removal Effect* “definido como el cambio en la probabilidad de lograr la conversión cuando se remueve un estado (canal) del grafo” (Anderl, Becker, von Wangenheim, & Schumann, 2016). El *Removal Effect* se calcula extrayendo el canal de interés para así identificar la cantidad de conversiones que se realizan en este nuevo escenario. En la Figura 3 se ilustra este proceso para el ejemplo anteriormente mencionado (ver Figura 1).

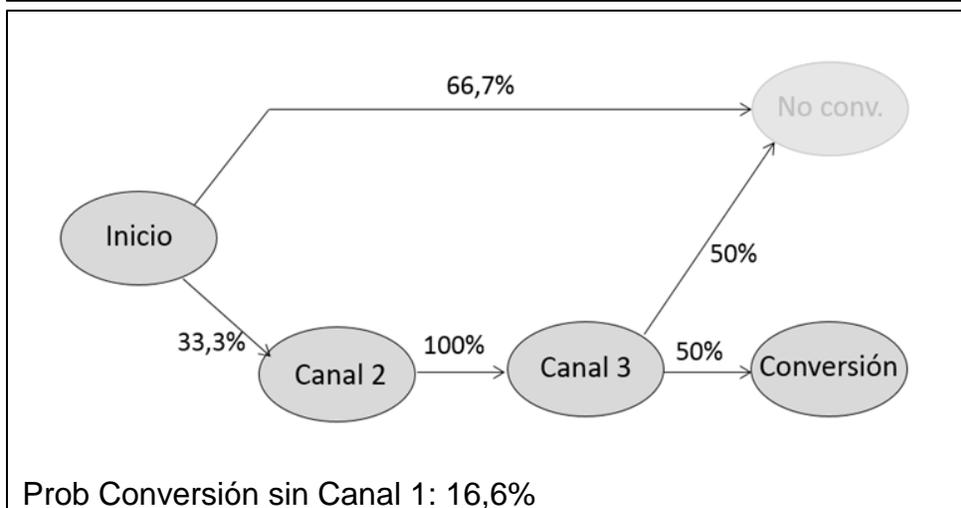
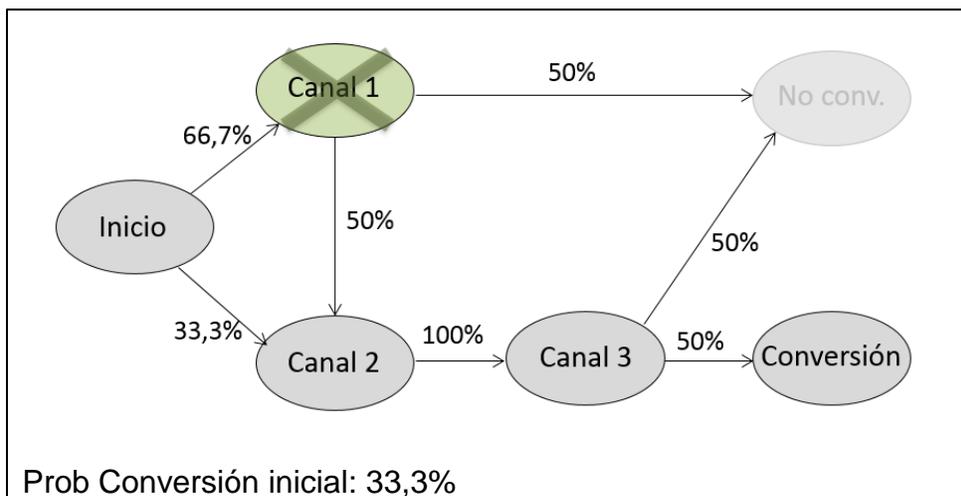


Figura 3: Representación gráfica del *Removal Effect*. (Elaboración Propia)

A partir del cálculo del *Removal Effect* se obtiene un valor que puede estar entre 0 y la probabilidad de conversión inicial (sin modificar el set de canales). Para conocer la atribución de conversiones correspondiente a cada canal se debe multiplicar este valor por el total de conversiones. A continuación se señala el procedimiento empleado.

Sean,

$P_T$ : Probabilidad de conversión Total

$P_c$ : Probabilidad de conversión habiendo extraído el canal  $c$ .

$C$ : Conjunto de canales con los que se está trabajando.

En primer lugar se calcula  $P_T$ , correspondiente a la probabilidad de convertir del escenario que se está abordando sin extraer ningún canal. Esto se cuantifica identificando todos los posibles caminos hacia la conversión y calculando la probabilidad de ocurrencia de cada uno de estos, una vez obtenida esta probabilidad, se deben agregar todos los valores correspondientes a todas las rutas posibles.

En el ejemplo que se presentó en la Figura 1 la probabilidad de convertir es de 33,3%.

$P_c$  se calcula de una forma análoga pero considerando la situación en la cual se ha extraído un canal del escenario inicial, esta modificación cambia las rutas hacia la conversión, por lo cual es necesario recalculer las probabilidades de ocurrencia de cada uno de los caminos por los cuales el cliente convierte. Para el ejemplo presentado en la Figura 3 el valor de  $P_c$  es de 16,7%.

Posteriormente se define  $v_c = \frac{P_T - P_c}{P_T}$  como un valor auxiliar para cada canal y luego de calcular los  $v_c$  para los demás medios, se estima el *Removal Effect* para cada canal  $c$  de la siguiente forma:

$$RE_c = \frac{v_c}{\sum_1^c v_c}$$

En el ejemplo que se mencionó anteriormente los valores de  $v_c$  y  $RE_c$  serían los siguientes:

$$\begin{array}{ll} v_{c1} = 0,5 & RE_{c1} = 0,2 \\ v_{c2} = 1 & RE_{c2} = 0,4 \\ v_{c3} = 1 & RE_{c3} = 0,4 \end{array}$$

Por último se procede a calcular la atribución para el canal  $c$ :  $Atrib_c = RE_c * Conv$ . Si para el ejemplo asumimos que existe 1 conversión, la atribución toma los mismos valores de los  $RE_c$  que ya han sido presentados.

### 3.2.2. EFICIENCIA DE UN MODELO

La eficiencia de un modelo se puede cuantificar utilizando una serie de métricas que se desprenden a partir de una tabla denominada matriz de confusión, la cual permite comparar los resultados que un modelo entrega (resultados predichos) con la realidad de la situación que se está representando por medio del modelo.

	Soleado	Nublado
Soleado según modelo	13 (T.P.)	1 (F.P.)
Nublado según modelo	1 (F.N.)	15 (T.N.)

Tabla 1: Ejemplo de matriz de confusión. (Elaboración propia)

En la tabla que se adjunta se muestra la matriz de confusión para el caso de un modelo que predice las condiciones meteorológicas de un día en la ciudad de Santiago. El modelo

resulta ser bastante acertado pues falla solo en un par de predicciones. En la tabla además se incluyen las iniciales de las calificaciones que se suelen utilizar para los distintos cuadrantes de la matriz de confusión, a continuación se detalla cada una de ellas:

T.P. (*True Positive*): Se refiere a todos aquellos casos en los cuales el modelo predice un resultado positivo y en la realidad también lo es.

F.N. (*False Negative*): Son todos los casos en los que el modelo indica un resultado negativo pero en la realidad ese caso es positivo.

F.P. (*False Positive*): Este cuadrante representa los casos negativos en los cuales el modelo predice erradamente resultados positivos.

T.N. (*True Negative*): Corresponden a los casos que son negativos y el modelo efectivamente los predice como negativos.

A partir de estas denominaciones para cada una de las situaciones es posible calcular algunas métricas que indican el desempeño de un modelo. El *accuracy* es una de ellas y es justamente el indicador que se utilizará para evaluar los modelos obtenidos.

*Accuracy*: Es la proporción del total de predicciones realizadas correctamente con respecto al total de casos predichos. Indica la exactitud de un modelo.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

Para el ejemplo que se adjunta el *accuracy* se calcula de la siguiente forma:

$$Accuracy = \frac{15 + 13}{15 + 1 + 1 + 13} = 0,933 \Rightarrow 93,333\%$$

En este caso el modelo tiene un 93,3% de exactitud lo cual resulta ser bastante bueno.

## 4. DESARROLLO METODOLÓGICO

El objetivo principal del proyecto que se describe es desarrollar un modelo de atribución de conversiones basado en cadenas de Markov, este modelo se confeccionará utilizando el camino hacia la conversión que realizan los clientes cuando adquieren productos de forma online.

De forma complementaria a los resultados del modelo de atribución se han planteado las hipótesis que se señalan a continuación, estas se relacionan tanto con la exploración previa a la construcción del modelo como también con un análisis adicional de Facebook que se incluye como resultado complementario al trabajo realizado.

H1: En el contexto de la empresa, los canales del tipo *firm initiated* participan más en las primeras etapas del proceso de compra mientras que los *customer initiated*, lo hacen más en las etapas finales.

H2: En particular el canal de Facebook contribuye a la conversión de algunos canales *firm initiated*, sobretodo de SEO y directo. Además el aporte de este canal no está siendo evidenciado por el modelo de atribución *last touch*.

La metodología que se utiliza para abordar la temática señalada es KDD. A continuación se describen las diferentes etapas del estudio.

### 4.1 DESARROLLO DEL MODELO

#### 4.1.1. SELECCIÓN DE DATOS

En la empresa existen dos fuentes de datos diferentes desde las cuales es posible extraer la información necesaria, se trata de Adobe Analytics y Google Analytics. La primera corresponde a la solución que entrega Adobe Marketing Cloud para reunir, organizar y analizar toda la actividad del cliente a través de un sitio web. En Falabella utilizan a diario esta herramienta, de la cual obtienen información que les permite crear reportes y consultar resultados de diferentes estrategias, esta plataforma les permite además acceder a la información desagregada de las visitas y compras. Con respecto a los modelos de atribución, esta herramienta únicamente muestra las opciones de *first* y *last touch*, pudiendo acceder al *customer journey* completo pero sólo contratando una versión Premium de Analytics (con un pago adicional), lo cual no se encuentra en las prioridades actuales de la empresa.

Por otro lado, Google Analytics es la herramienta gratuita de Google que ofrece información sobre el tráfico online que llega a un sitio web. Esta plataforma permite tener acceso a los procesos de compra de forma agregada, señalando el número de órdenes que se llevaron a cabo por medio de una ruta específica [Anexo 6]. En la empresa no utilizan esta herramienta a diario, dado que tienen contratado el plan básico de la suite

de Adobe que les permite acceder a información más desagregada que la que se muestra en la propuesta de Google<sup>8</sup>.

Cada una de las fuentes de datos señaladas utilizan diferentes formas de medición, por lo cual los resultados que muestran ambas plataformas no son comparables entre sí, por ejemplo si se compara la distribución de las conversiones bajo el modelo de atribución *last touch* se tienen porcentajes distintos para la mayoría de los canales [Anexo 7].

Ante esta discrepancia se hace necesario escoger la fuente de datos más adecuada para el estudio. Lo ideal en este caso sería utilizar Adobe Analytics debido a que esa es la fuente que utiliza la empresa para realizar todo tipo de reportes, sin embargo la información requerida no está disponible en la plataforma. Por esta razón se debió trabajar con la información de Google Analytics.

En esta etapa además se debe definir el periodo de datos a utilizar, para lo cual se debe considerar que la información del sitio comenzó a registrarse en Analytics el día 12 de junio de 2016. Esta herramienta utiliza una ventana de tiempo de 30 días previos a la conversión. De esta manera el 15 de Julio de 2016 es una fecha adecuada para considerar como inicio de la extracción de datos, dado que se están incluyendo las interacciones ocurridas desde que se comenzó a registrar la información.



Figura 4: Explicación ventana al pasado. (Elaboración Propia)

El periodo completo que se estudiará será desde el 15 de Julio del 2016 hasta el 31 de marzo del 2017. Esta información se consolidó en una base de datos de 5.844 rutas distintas con un total 1.454.398 conversiones<sup>9</sup>.

#### 4.1.2. PREPROCESAMIENTO DE DATOS

El proceso de preparación de los datos se enfocó en la limpieza de estos. Al total de los datos se le quitaron las conversiones realizadas los días correspondientes a CyberDay de Noviembre, pues durante estos días se registran cifras de ventas y visitas mayores a las habituales<sup>10</sup> [Anexo 8], por lo cual podrían sesgar las métricas que se calculen. La base de datos que se obtuvo con esta operación es de 1.316.180 conversiones y 5.664 rutas distintas.

---

<sup>8</sup> Considerando la información que utilizan a diario, no la que se requiere para el presente estudio.

<sup>9</sup> Una ruta específica puede utilizarse para generar varias conversiones (compras).

<sup>10</sup> De hecho se registran conversiones diarias que representan más de 20 veces un día normal.

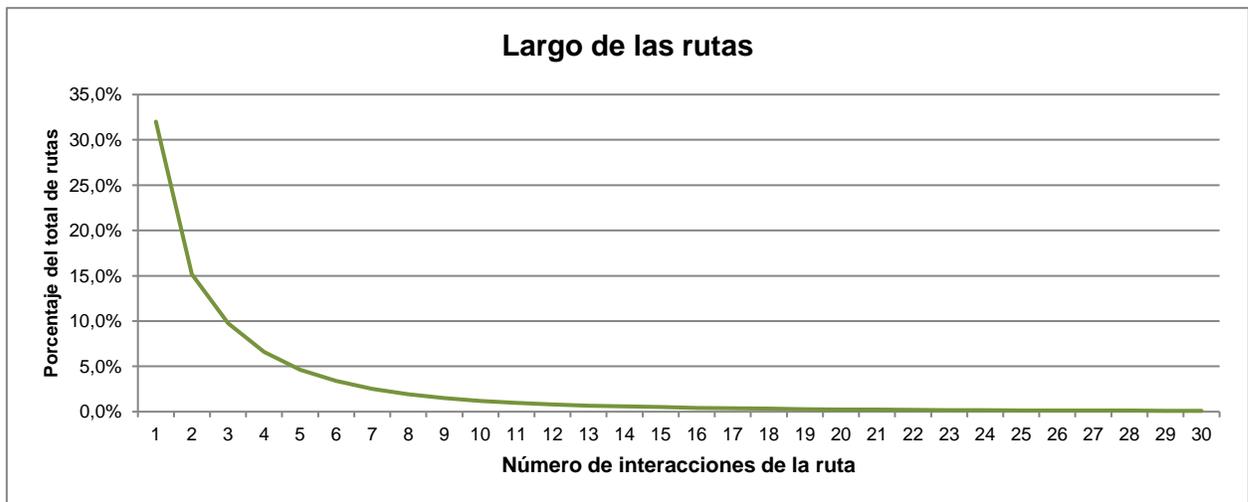
Acerca de la integración y transformación de los datos, no se realizaron acciones relacionadas con estos aspectos más que la consolidación de las dos bases de datos provenientes de las dos extracciones de datos realizadas desde Google Analytics.

### 4.1.3. EXPLORACIÓN PRELIMINAR MODELO

En esta etapa tiene lugar la construcción del modelo de atribución, para eso se han definido 3 etapas que permiten cumplir dicho objetivo. En primer lugar es necesario realizar una exploración de datos que permita calcular métricas generales, e incluir alcances descriptivos sobre la información que se tiene. En una segunda instancia tiene lugar la realización del modelo de atribución en el software R, y en una tercera etapa se realiza la validación que asegura la fiabilidad de los resultados.

La exploración de datos permite conocer la información que se tiene y además poder identificar algunos *outliers* que no se incluirán en el modelo. Para comenzar, se estudió el largo de las rutas y la frecuencia de éstas (número de interacciones que la componen). Posterior a eso se caracterizaron las rutas, indicando los canales que las componen, cómo son las más frecuentes y la distribución de canales bajo el modelo *last touch* y *first touch*, considerando la muestra con la que se trabajará.

Con respecto al largo de la ruta, es decir la cantidad de interacciones que componen un *customer journey*, en promedio el largo era de 6 interacciones, sin embargo existían rutas de hasta 222 toques con muy baja frecuencia. Por esta razón se procedió a utilizar el *elbow test* con el objetivo de determinar hasta qué ruta es importante incluir en el análisis. Como puede verse en la figura 5, a partir de las rutas con 30 interacciones el aporte marginal es mínimo en comparación con los aportes de las rutas previas, por lo cual se decidió considerar solo las rutas con estas características abarcando un 97,0% de la muestra inicial ya que los datos que se están excluyendo son secuencias que presentan



muy baja frecuencia y además bajas ventas.

Figura 5: Distribución del largo de rutas de Google Analytics. (Elaboración Propia)

Relacionado con lo anterior, también se estudió la frecuencia de las rutas, ya que ocurre que hay ciertas rutas muy frecuentes que abarcan gran porcentaje del total de las órdenes, mientras que otras rutas son más bien casos particulares. Por ejemplo, existen 5 rutas con una frecuencia tal que alcanzan el 58,0% del total de casos. En el gráfico de la Figura 6 se presenta el porcentaje de la muestra total que alcanza cada una de las primeras 30 rutas más frecuentes.

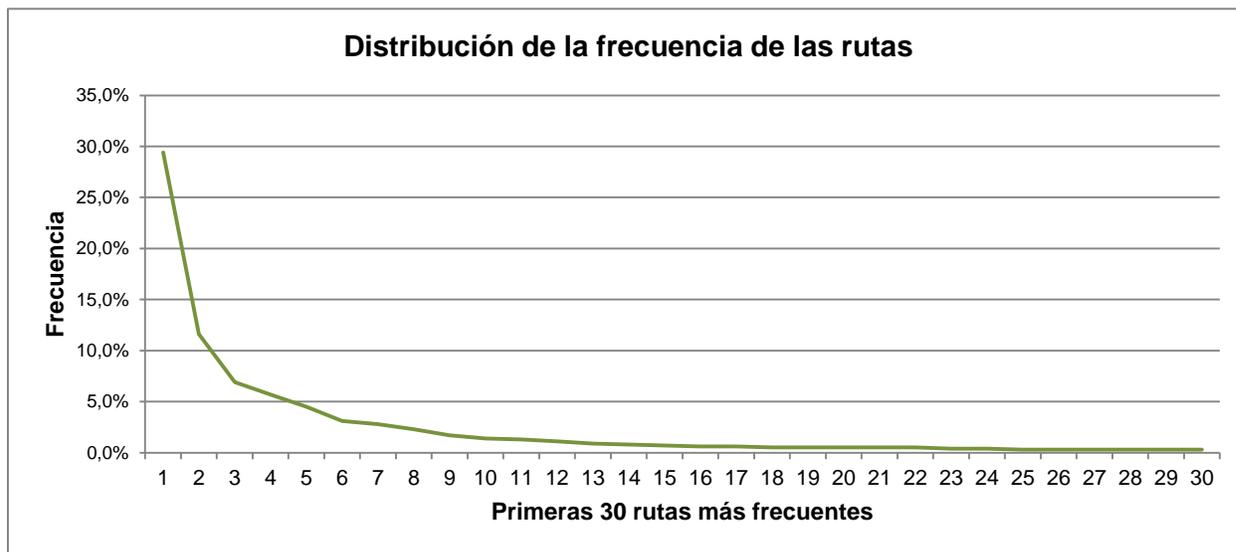


Figura 6: Distribución de la frecuencia de las rutas de Google Analytics. (Elaboración Propia)

Para enfocarse en las secuencias de mayor frecuencia en una primera instancia se consideró utilizar el *elbow test* también. Sin embargo debido a que estas rutas presentan muy poca variabilidad en cuanto a los canales presentes en ellas, es conveniente realizar un análisis de sensibilidad y ver como varía la presencia de los distintos canales cuando se consideran rutas adicionales. Una muestra de las primeras 200 rutas distintas aborda un 90,0% de la muestra total, incluyendo participación de todos los canales, con menos rutas se deja de incluir la totalidad de los medios, y con más rutas no aumenta mayormente la participación de estos [Anexo 9].

Ahora bien, pasando a la caracterización de las rutas, ya se mencionó que existen algunos caminos que concentran gran parte de las conversiones totales de la base de datos. En la tabla 2 se puede ver la forma de las primeras 10 rutas más frecuentes, las que representan un 69,4% del total [Anexo 10].

En las rutas que se adjuntan se puede ver gran presencia del canal Directo que es justamente el canal al que Google Analytics asigna mayor participación en el modelo de atribución *last touch*. Esto es bastante lógico debido a la naturaleza de este canal, el cual por ser del tipo *customer initiated* participa cerrando las ventas cuando el cliente ya está más avanzado en el proceso de compra.

Marketing Channels	Conversiones	Porcentaje
Inicio > Directo > Fin	387.326	29,4%
Inicio > Directo > Directo > Fin	152.367	11,6%
Inicio > Directo > Directo > Directo > Fin	90.848	6,9%
Inicio > SEO > Fin	74.695	5,7%
Inicio > Directo > Directo > Directo > Directo > Fin	58.658	4,5%
Inicio > Directo > Directo > Directo > Directo > Directo > Fin	40.453	3,1%
Inicio > SEO > Directo > Fin	36.757	2,8%
Inicio > Directo > Directo > Directo > Directo > Directo > Directo > Fin	30.466	2,3%
Inicio > Directo > Fin	22.796	1,7%
Inicio > SEO > Directo > Directo > Fin	18.398	1,4%

Tabla 2: Primeras 10 rutas más frecuentes de Google Analytics. (Elaboración Propia)

Como ya se anticipó, los canales que participan en las rutas se pueden clasificar en dos grupos, *Customer Initiated* (C.I.) y *Firm Initiated* (F.I.). Los primeros son aquellos en los cuales la comunicación se inicia por parte del cliente y los canales pertenecientes al segundo grupo son aquellos en los que la interacción es iniciada por la empresa. Esta distinción es relevante ya que habla del interés de cada una de las partes involucradas en el proceso de compra. En el estudio que se está desarrollando, estos son conceptos importantes para concluir acerca de los procesos de compra de los usuarios y las migraciones de tráfico entre canales.

Los medios que se utilizarán en el modelo se pueden clasificar bajo este criterio de la siguiente forma:

Customer Initiated (C.I.)	{	Directo SEO SEM
Firm Initiated (F.I.)	{	Redes Sociales Orgánicas Redes Sociales de Pago Email Remarketing Tiendas Display Afiliados

Con respecto a los canales que más aparecen en las rutas (independiente su posición en la ruta) lidera Directo con un 93,0% de presencia en las interacciones, a lo cual le sigue SEO con un 4,8%, SEM con un 1,1% y luego Email con un 0,7%. A pesar de la gran participación de Directo en las interacciones, este canal se dejará a un margen del estudio en las recomendaciones, es decir, se considerará para el cálculo del modelo pero no se

incluira mayormente en los análisis y en las conclusiones ya que no se puede realizar ninguna acción sobre este canal que tenga repercusión en el corto plazo.

La posición en la ruta para la interacción con cierto canal, también es un aspecto relevante en el estudio de la atribución, esto debido a que hay algunos canales que cumplen la función de captar a los clientes e iniciar los procesos de compra, mientras que otros funcionan mejor “cerrando las ventas” (Goic, Jerath, & Kalyanam, 2014). Por esta razón se estudiaron las contribuciones de los canales bajo el modelo de atribución *first touch* y *last touch* considerando la muestra que se está trabajando.

Para el modelo de atribución *first touch*, como se puede ver en la Tabla 3, se le atribuye un 81,1% al canal Directo, a lo cual le sigue SEO con un 14,0%, SEM con un 2,3% e Email con un aporte del 1,6% [Anexo 11]. Bajo el modelo *last touch*, Google Analytics, le entrega un 89,7% de las conversiones al canal Directo, a lo cual le sigue SEO con un 7,5% y SEM, con un 1,0% [Anexo 12].

Canales	<i>First Touch</i>		<i>Last Touch</i>	
	Conversiones	Porcentaje	Conversiones	Porcentaje
Directo	957.215	81,1%	1.059.136	89,7%
SEO	165.145	14,0%	88.347	7,5%
SEM	27.127	2,3%	11.958	1,0%
Email	18.392	1,6%	8.951	0,8%
Otros	5.265	0,5%	7.075	0,6%
RMKT	2.430	0,2%	2.586	0,2%

Tabla 3: Contribución a la conversión de los principales canales bajo el modelo de atribución *first touch* y *last touch*. (Elaboración Propia)

#### 4.1.4. REALIZACIÓN DEL MODELO

Como se mencionó anteriormente, el modelo de atribución de conversiones que se propone en este estudio está basado en cadenas de Markov. En este caso los procesos de compra pueden ser representados por un grafo donde los estados serían los canales y las transiciones corresponderían a las migraciones entre los distintos medios.

Para calcular el modelo de atribución basado en la información de la empresa se utilizó el software R trabajando con una base de datos compuesta por las primeras 200 rutas más frecuentes de todo el periodo seleccionado<sup>11</sup>, abarcando el 90,0% de la muestra inicial<sup>12</sup>. Los caminos hacia la conversión podían estar compuestos por diferentes combinaciones de los 11 canales con los cuales opera la empresa y además estaban conformados por un número variable de interacciones del cliente con los distintos canales.

A partir de las interacciones mencionadas fue posible calcular la matriz de transición entre los medios involucrados. Esta matriz señala las probabilidades de migración entre todas las combinaciones de canales posibles, de esta forma una vez que el cliente tiene contacto en un canal, es posible determinar qué tan probable es que en el siguiente estado, el cliente se encuentre en un canal específico. En la Tabla 4 se muestra un ejemplo de una matriz de transición cuando se tienen 4 canales con los cuales puede interactuar el cliente.

	Canal 1	Canal 2	Canal 3	Canal 4
Canal 1	0,58	0,21	0,05	0,16
Canal 2	0,13	0,48	0,27	0,12
Canal 3	0,22	0,31	0,19	0,28
Canal 4	0,51	0,34	0,07	0,08

Tabla 4: Ejemplo de matriz de transición. (Elaboración propia)

Para los datos de la empresa, las matrices de transición estaban conformadas por 13 filas de probabilidades, debido a los 11 canales con los que trabaja la compañía, más los estados de Inicio y Fin (Compra) que se debieron añadir para luego calcular el modelo de atribución con el método de la estimación de *removal effect*.

Una vez obtenidas las matrices de transición se procedió a cuantificar la importancia de cada canal, para esto se estimaron las conversiones que se llevaban a cabo en distintos escenarios determinados por la extracción de cada uno de los canales, esto es justamente lo que se conoce como *removal effect*. A partir de este cálculo fue posible

---

<sup>11</sup> Julio 2016 a Marzo 2017.

<sup>12</sup> Se decidió trabajar con este subconjunto de la muestra para facilitar el cálculo computacional, además con este número de rutas se abarca un 90,0% del total de las conversiones y al agregar más rutas distintas (que tienen frecuencia marginal) se conservan los porcentajes de la participación de cada canal.

estimar el aporte de cada canal en la conversión y obtener el modelo de atribución basado en la información de la empresa.

El desarrollo se realizó en primer lugar considerando modelos basados en cadenas de Markov de hasta cuarto orden, sin embargo este último modelo tardaba demasiado en compilar por lo cual se acotó el estudio solo a modelos de orden 1, 2 y 3, quedando el modelo de Markov de orden 4 en la propuesta de trabajo futuro. Las matrices de transición de cada orden se calcularon de forma independiente y análoga al cálculo del modelo de primer orden. La comparación de estos modelos se realizó considerando el *accuracy* de cada uno y se describe en la sección de validación del modelo.

## 4.2 RESULTADOS

Como se mencionó en la sección anterior, se incluyó en el estudio el cálculo de los modelos de Markov orden 1, 2 y 3, para los cuales se calculó la matriz de probabilidades de transición considerando el total de la muestra. A continuación se adjuntan los resultados obtenidos para cada uno de los modelos mencionados.

### Modelo basado en Cadenas de Markov de Orden 1.

A partir de la construcción del modelo de atribución basado en cadenas de Markov de primer orden se obtuvo una matriz de probabilidades de 13x13. Esta matriz contiene las probabilidades de transición entre todas las combinaciones de canales posibles. Es importante mencionar que algunas duplas eran completamente improbables<sup>13</sup> por lo cual no se incluyen en la Tabla 5 que contiene dichos resultados.

	Directo	SEO	SEM	EmailMkt	RRSSOrganico	Remarketing	Display	Afiliados	Otros	Fin
Inicio	81,07%	13,99%	2,30%	1,56%	0,12%	0,21%	0,16%	0,05%	0,45%	0,00%
Directo	70,47%	0,79%	0,38%	0,18%	0,01%	0,14%	0,00%	0,00%	0,14%	27,89%
SEO	52,79%	0,23%	1,21%	0,38%	0,12%	0,39%	0,00%	0,00%	0,41%	44,48%
SEM	68,61%	0,61%	4,84%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	25,94%
EmailMkt	51,28%	6,23%	0,00%	12,47%	0,00%	0,00%	0,00%	0,00%	0,00%	30,02%
RRSSOrganico	17,85%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	82,15%
Remarketing	66,59%	2,60%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	30,82%
Display	71,52%	10,65%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	17,83%
Tiendas	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
RRSSPago	67,36%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	32,64%
Afiliados	38,64%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	30,68%	0,00%	30,68%
Otros	31,79%	5,44%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	1,87%	60,90%

Tabla 5: Matriz de probabilidades de transición para Markov orden 1. (Elaboración propia)

Tal como se puede observar en la matriz que se adjunta, se obtuvieron varias celdas en las cuales la probabilidad de transición resultante fue nula o muy baja, para explicar esto es necesario recordar las características de los datos que se tienen. En la base de datos

<sup>13</sup> La columna que representaba las transiciones desde un canal hasta el estado de inicio, y la fila de las probabilidades entre el estado de Fin y un canal específico, no se incluyen en la tabla.

existen algunos canales que poseen muy pocas apariciones en las rutas, mientras que otros participan en la mayoría de los caminos hacia la conversión (por ejemplo Directo), es por esto que dado que estos últimos canales son más frecuentes en los procesos de compra, las probabilidades de transición desde y hacia ellos serán mayores que las de los otros medios mencionados. El hecho de no tener tanta información para ciertos canales de alguna forma sesga los resultados pues es complejo predecir con exactitud cuando se tiene escasa información<sup>14</sup>. Este es un aspecto importante a considerar en la validación del modelo propuesto.

A pesar del sesgo mencionado es posible señalar las siguientes particularidades de los resultados. La mayoría de las rutas inician con el canal directo (81,1%)<sup>15</sup>, esto debido a la gran participación de este canal en toda etapa de las rutas registradas.

Del conjunto de los canales del tipo *customer initiated* (Directo<sup>16</sup>, SEO y SEM), SEO efectivamente es más probable que interactúe con el cliente en etapa final (44,5%) que en etapas iniciales del proceso de compra (13,9%).

Con respecto a los canales del tipo *firm initiated*, estos presentan baja frecuencia<sup>17</sup> en las rutas, en comparación con los canales *customer initiated* mencionados; sin embargo el canal Email es un buen medio<sup>18</sup> para la observación de resultados. En la tabla 5 se puede ver que para este canal existe mayor probabilidad de participar en la ruta previo al canal Directo (51,3%), que finalizando el proceso de compra<sup>19</sup> (30,0%), lo cual apoya la teoría de que este tipo de canales no son muy utilizados para convertir.

## **Modelo basado en Cadenas de Markov de Orden 2.**

La matriz correspondiente al modelo basado en cadenas de Markov de segundo orden se construyó de manera análoga a la de primer orden. De igual forma en la tabla que se adjunta se omiten las celdas que por lógica deben tener probabilidad nula. Los valores

---

<sup>14</sup> Al incluir más rutas en el cálculo, no aumentaba la participación de estos canales con respecto al total. El sesgo viene de una falta de información de ciertos canales, independiente el número de rutas que se considere.

<sup>15</sup> No se puede afirmar algo análogo para el final de la ruta debido a que como ya se mencionó, la poca información que se tiene para ciertos canales sesga los resultados.

<sup>16</sup> No es posible afirmar algo similar del canal Directo debido a su amplia participación en la mayoría de las rutas (92,6%). Además como se mencionó anteriormente este no es un canal que se abordará con detalle en el estudio pues por la naturaleza orgánica del canal no permite implementar acciones con repercusión en el corto plazo.

<sup>17</sup> La participación promedio de este tipo de canales es de 0,3% debido principalmente a la gran presencia de Directo que se lleva un 92,6% de participación en los procesos de compra.

<sup>18</sup> Email es un canal que presenta una de las mayores participaciones (0,7%) en las rutas dentro del conjunto de canales del tipo *firm initiated*.

<sup>19</sup> Es decir previo al estado "Fin".

nulos que se obtuvieron corresponden a combinaciones de canales<sup>20</sup> que no ocurren en la base de datos con la que se está trabajando, debido a la poca información que existe para algunos medios<sup>21</sup>.

	Directo	SEO	SEM	EmailMkt	RRSSOrganico	Remarketing	Afiliados	Otros	Fin
InicioDirecto	55,89%	1,85%	0,75%	0,37%	0,05%	0,27%	0,00%	0,32%	40,46%
InicioSEO	51,48%	0,28%	1,45%	0,45%	0,14%	0,46%	0,00%	0,50%	45,23%
InicioSEM	65,32%	1,03%	8,23%	0,00%	0,00%	0,00%	0,00%	0,00%	25,42%
InicioEmailMkt	46,59%	10,11%	0,00%	16,18%	0,00%	0,00%	0,00%	0,00%	27,12%
InicioRRSSOrganico	26,62%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	73,38%
InicioRemarketing	60,99%	8,97%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	30,04%
InicioDisplay	67,12%	12,30%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	20,59%
InicioRRSSPago	67,36%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	32,64%
InicioAfiliados	55,75%	0,00%	0,00%	0,00%	0,00%	0,00%	44,25%	0,00%	0,00%
InicioOtros	37,69%	11,99%	0,00%	0,00%	0,00%	0,00%	0,00%	4,12%	46,19%
DirectoDirecto	76,70%	0,45%	0,23%	0,13%	0,00%	0,10%	0,00%	0,07%	22,32%
DirectoSEO	62,89%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	37,11%
DirectoSEM	76,13%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	23,87%
DirectoEmailMkt	70,85%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	29,15%
DirectoRemarketing	72,69%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	27,31%
DirectoOtros	32,14%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	67,86%
SEODirecto	55,99%	0,00%	0,63%	0,00%	0,00%	0,00%	0,00%	0,34%	43,04%
SEOSEM	66,06%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	33,94%
SEOEmailMkt	46,74%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	53,26%
SEORemarketing	43,02%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	56,98%
SEMDirecto	55,65%	0,00%	0,75%	0,00%	0,00%	0,00%	0,00%	0,00%	43,60%
SEMSEM	62,99%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	37,01%
EmailMktDirecto	48,85%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	51,15%
EmailMktSEO	46,58%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	53,42%
EmailMktEmailMkt	38,78%	0,00%	0,00%	19,96%	0,00%	0,00%	0,00%	0,00%	41,26%
RemarketingDirecto	46,08%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	53,92%
DisplayDirecto	39,50%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	60,50%
RRSSPagoDirecto	34,56%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	65,44%
OtrosDirecto	39,99%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	60,01%
OtrosSEO	35,60%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	64,40%

Tabla 6: Matriz de probabilidades de transición para Markov orden 2. (Elaboración propia)

Las filas que registran valores muy extremos en las probabilidades de transición hacia otros canales (por ejemplo filas que tienen una celda con una probabilidad del 100,0% y el resto 0,0% de ocurrencia) posiblemente corresponden a combinaciones para las cuales se tienen muy pocos datos, por lo cual las probabilidades se calculan usando solo esta información y se obtienen valores con estas características. Estos casos han sido

<sup>20</sup> Recordar que dado que se trata de una matriz de transición de orden 3 cada combinación indica una tupla de tres canales.

<sup>21</sup> Las columnas de Tiendas y RRSSPago han sido omitidas.

omitidos en la Tabla nº 6 pero se incluyen en la matriz que se adjunta en anexos [Anexo 13].

En la tabla, los mayores porcentajes se encuentra en las columnas de Directo y Fin. Debido a las características de la base de datos con la que se está trabajando, la probabilidad de que un cliente entre al sitio por medio de Directo dado que ya ha visitado una determinada dupla de canales será alta la mayoría de las veces. Por otro lado, con respecto a las transiciones hacia el estado Fin, es posible notar que aquellas celdas en las cuales se registran porcentajes importantes corresponden en su mayoría, a canales del tipo *customer initiated* (indicadas en la Tabla nº6 en naranja).

### **Modelo Basado en Cadenas de Markov de Orden 3.**

La matriz de orden 3 también se construyó de forma análoga a la matriz de Markov para primer orden. A continuación se señalan los resultados obtenidos donde igualmente se omiten las celdas ya mencionadas en las matrices previas, además de aquellas combinaciones que no ocurren en la base de datos y que por tanto tienen probabilidad nula<sup>22</sup>.

---

<sup>22</sup> Las columnas de Tiendas y RRSSPago han sido omitidas.

	Directo	SEO	SEM	EmailMkt	Remarketing	Otros	Fin
InicioDirectoDirecto	68,80%	1,24%	0,63%	0,36%	0,28%	0,21%	28,48%
InicioDirectoSEO	61,06%	0,00%	0,00%	0,00%	0,00%	0,00%	38,94%
InicioDirectoSEM	76,61%	0,00%	0,00%	0,00%	0,00%	0,00%	23,39%
InicioDirectoEmailMkt	72,32%	0,00%	0,00%	0,00%	0,00%	0,00%	27,68%
InicioDirectoRemarketing	72,81%	0,00%	0,00%	0,00%	0,00%	0,00%	27,19%
InicioDirectoOtros	43,85%	0,00%	0,00%	0,00%	0,00%	0,00%	56,15%
InicioSEODirecto	55,57%	0,00%	0,78%	0,00%	0,00%	0,42%	43,23%
InicioSEOSEM	66,06%	0,00%	0,00%	0,00%	0,00%	0,00%	33,94%
InicioSEOEmailMkt	46,74%	0,00%	0,00%	0,00%	0,00%	0,00%	53,26%
InicioSEORemarketing	43,02%	0,00%	0,00%	0,00%	0,00%	0,00%	56,98%
InicioSEMDirecto	56,70%	0,00%	1,34%	0,00%	0,00%	0,00%	41,96%
InicioSEMSEM	62,99%	0,00%	0,00%	0,00%	0,00%	0,00%	37,01%
InicioEmailMktDirecto	53,93%	0,00%	0,00%	0,00%	0,00%	0,00%	46,07%
InicioEmailMktSEO	46,58%	0,00%	0,00%	0,00%	0,00%	0,00%	53,42%
InicioEmailMktEmailMkt	38,94%	0,00%	0,00%	24,93%	0,00%	0,00%	36,12%
InicioRemarketingDirecto	50,07%	0,00%	0,00%	0,00%	0,00%	0,00%	49,93%
InicioDisplayDirecto	48,59%	0,00%	0,00%	0,00%	0,00%	0,00%	51,41%
InicioRRSSPagoDirecto	34,56%	0,00%	0,00%	0,00%	0,00%	0,00%	65,44%
InicioOtrosDirecto	41,49%	0,00%	0,00%	0,00%	0,00%	0,00%	58,51%
InicioOtrosSEO	35,60%	0,00%	0,00%	0,00%	0,00%	0,00%	64,40%
DirectoDirectoDirecto	79,59%	0,26%	0,14%	0,07%	0,06%	0,04%	19,84%
DirectoDirectoSEO	65,56%	0,00%	0,00%	0,00%	0,00%	0,00%	34,44%
DirectoDirectoSEM	77,37%	0,00%	0,00%	0,00%	0,00%	0,00%	22,63%
DirectoDirectoEmailMkt	69,32%	0,00%	0,00%	0,00%	0,00%	0,00%	30,68%
DirectoDirectoRemarketing	72,57%	0,00%	0,00%	0,00%	0,00%	0,00%	27,43%

DirectoDirectoOtros	18,93%	0,00%	0,00%	0,00%	0,00%	0,00%	81,07%
DirectoSEODirecto	59,52%	0,00%	0,00%	0,00%	0,00%	0,00%	40,48%
DirectoSEMDirecto	55,98%	0,00%	0,00%	0,00%	0,00%	0,00%	44,02%
DirectoEmailMktDirecto	50,29%	0,00%	0,00%	0,00%	0,00%	0,00%	49,71%
DirectoRemarketingDirecto	48,54%	0,00%	0,00%	0,00%	0,00%	0,00%	51,46%
DirectoOtrosDirecto	38,25%	0,00%	0,00%	0,00%	0,00%	0,00%	61,75%
SEODirectoDirecto	60,62%	0,00%	0,00%	0,00%	0,00%	0,00%	39,38%
SEODirectoSEM	50,53%	0,00%	0,00%	0,00%	0,00%	0,00%	49,47%
SEOSEMDirecto	46,91%	0,00%	0,00%	0,00%	0,00%	0,00%	53,09%
SEMDirectoDirecto	55,68%	0,00%	0,00%	0,00%	0,00%	0,00%	44,32%
SEMSEMDirecto	49,72%	0,00%	0,00%	0,00%	0,00%	0,00%	50,28%
EmailMktDirectoDirecto	48,90%	0,00%	0,00%	0,00%	0,00%	0,00%	51,10%
EmailMktSEODirecto	35,57%	0,00%	0,00%	0,00%	0,00%	0,00%	64,43%
EmailMktEmailMktDirecto	25,66%	0,00%	0,00%	0,00%	0,00%	0,00%	74,34%
EmailMktEmailMktEmailMk	38,14%	0,00%	0,00%	0,00%	0,00%	0,00%	61,86%
RemarketingDirectoDirecto	41,36%	0,00%	0,00%	0,00%	0,00%	0,00%	58,64%
DisplayDirectoDirecto	39,20%	0,00%	0,00%	0,00%	0,00%	0,00%	60,80%
OtrosDirectoDirecto	32,43%	0,00%	0,00%	0,00%	0,00%	0,00%	67,57%

Tabla 7: Matriz de probabilidades de transición para Markov orden 3. (Elaboración propia)

En este caso igualmente se obtuvieron bastantes filas con valores de probabilidades 0,0% y 100,0%, los cuales se cree que corresponden a tuplas con muy poca ocurrencia en la base de datos [Anexo 14].

En esta matriz también se obtienen los mayores valores de probabilidades para las duplas que se dirigen al canal Directo y al estado Fin. Lo primero como ya se mencionó ocurre por la alta presencia de Directo en las rutas, de hecho se registran porcentajes bastante altos en las transiciones que van de Directo a Directo (se indican en la Tabla n°7 con color azul).

Por otro lado, de los porcentajes más altos que se observan en la última etapa de los procesos de compra, la mayoría corresponde a rutas que finalizan utilizando un canal del tipo *customer initiated*, lo cual se relaciona con los objetivos secundarios del estudio (se indican en la Tabla n°7 con naranja).

### 4.3 VALIDACIÓN DEL MODELO

Con el objetivo de garantizar la fiabilidad del modelo basado en cadenas de Markov y por tanto de sus resultados, se debe primero evaluar la propuesta; para esto se compararon modelos de Markov de distinto orden con diferentes conjuntos de datos, y así poder determinar el orden que mejor describía la situación, independiente de la información que se ingresaba al modelo sugerido para la empresa.

La validación cruzada fue la herramienta que se utilizó para realizar estas comparaciones con distintos sets de datos. Este método permite realizar distintas combinaciones de la

información que se tiene y así generar diferentes combinaciones de conjuntos para testear y entrenar el modelo. Tal como se muestra en la Figura 7, se empleó una validación cruzada de 4 iteraciones ya que el total de datos se dividió en 4 subconjuntos (SC), los cuales se agruparon en información para entrenamiento y testeo, de acuerdo al orden que en la imagen se señala.

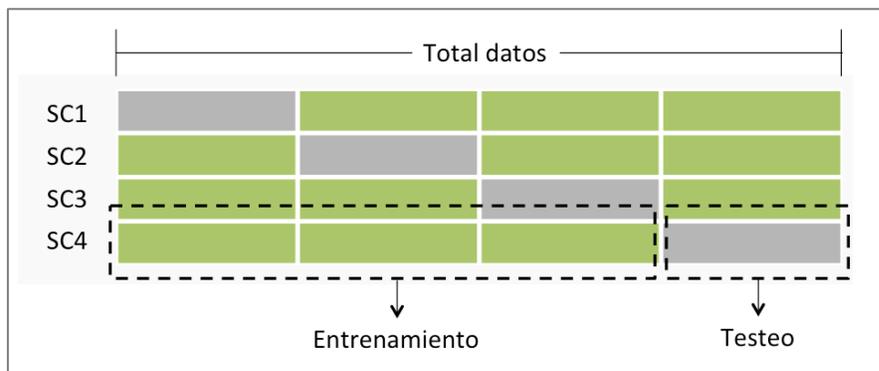


Figura 7: Representación de la validación cruzada de 4 iteraciones. (Elaboración Propia)

Con cada uno de los conjuntos (SC) señalados se calculó el modelo basado en cadenas de Markov para orden 1, 2 y 3.

Se determinó posteriormente un valor denominado probabilidad límite, por medio del cual fue posible discriminar si el resultado del modelo se inclinaba más hacia una conversión realizada o hacia una conversión no realizada. En instancias iniciales se consideró la misma probabilidad límite para todos los órdenes de los modelos, sin embargo debido a los resultados obtenidos se realizó una prueba considerando valores distintos de probabilidad límite para cada uno de los modelos de distinto orden. De esta forma se calculó el número de conversiones que el modelo estimaba tanto para el mismo conjunto de entrenamiento como también para el conjunto de datos de testeo<sup>23</sup>. Con este valor fue posible calcular la métrica del *accuracy* (contrastando conversiones predichas con conversiones reales) para cada uno de los modelos y así compararlos cuantitativamente.

Como ya se anticipó, se realizaron diferentes pruebas modificando las probabilidades límites utilizadas y también el conjunto de canales con los cuales se trabajaba. A continuación se detalla cada una de las pruebas realizadas especificando las medidas de exactitud obtenidas tanto para la validación dentro como fuera de la muestra de entrenamiento.

## Prueba 1

---

<sup>23</sup> Se incluyó validación dentro y fuera de la muestra.

Se calculó el modelo basado en cadenas de Markov para orden 1,2 y 3 considerando una probabilidad límite de 0,5 para cada uno de los tres modelos. La base de datos con la cual se trabajó estaba compuesta por todos los canales (11 canales más los estados Inicio y Fin).

Los resultados obtenidos no fueron muy satisfactorios pues se obtuvieron valores de *accuracy* bastante bajos sobre todo para el modelo de Markov de orden 1, a medida que aumentaba el orden del modelo, la precisión de éste aumentaba, sin embargo se decidió realizar pruebas adicionales para evaluar como variaban estos indicadores cuando se consideraban otras probabilidades límites.

## Prueba 2

Dada la situación ocurrida en la primera prueba se decidió considerar una probabilidad límite de 0,3 para todos los modelos. Esta vez se obtuvieron mayores valores para la precisión de los modelos, tanto para validación dentro (Acc In) como fuera de la muestra de entrenamiento (Acc Out). Para el modelo de orden 3 se obtuvo el mejor valor, registrando una precisión del 52,1%, independiente la muestra, lo cual podría indicar que este es el modelo que mejor describe el problema que se está abordando, de todas formas se realizaron otras pruebas que se muestran en lo que sigue, para confirmar lo expuesto. En esta instancia también se trabajó con la base de datos completa, es decir incluyendo los 13 canales.

Subconjunto	Orden del Modelo	Acc In	Acc Out
1	Markov 1	9,3%	9,3%
1	Markov 2	47,8%	47,8%
1	Markov 3	52,1%	52,1%
2	Markov 1	9,3%	9,3%
2	Markov 2	47,8%	47,7%
2	Markov 3	52,1%	52,2%
3	Markov 1	9,3%	9,2%
3	Markov 2	47,8%	47,7%
3	Markov 3	52,1%	52,1%
4	Markov 1	9,3%	9,3%
4	Markov 2	47,8%	47,8%
4	Markov 3	52,1%	52,2%

Tabla 8: Valores para el indicador de *accuracy* obtenido para cada uno de los Modelos de Markov en todas las muestras de la validación cruzada, para la Prueba 2. (Elaboración Propia).

## Prueba 3

En la tercera prueba que se realizó se utilizaron distintas probabilidades límite, esto motivado por el hecho que las probabilidades de convertir de cada modelo son diferentes en promedio, por lo cual considerar un mismo valor para todos los órdenes podría no ser la mejor opción.

Para realizar esto se observaron los valores de las probabilidades de conversión para cada uno de los modelos, y de acuerdo a esto se fijaron los siguientes valores de probabilidad límite:

PL1: 0,3

PL2: 0,5

PL3: 0,6

Subconjunto	Orden del Modelo	Acc In	Acc Out
1	Markov 1	9,3%	9,3%
1	Markov 2	2,2%	2,2%
1	Markov 3	45,6%	45,7%
2	Markov 1	9,3%	9,3%
2	Markov 2	2,2%	2,2%
2	Markov 3	45,6%	45,7%
3	Markov 1	9,3%	9,2%
3	Markov 2	2,2%	2,2%
3	Markov 3	45,6%	45,6%
4	Markov 1	9,3%	9,3%
4	Markov 2	2,2%	2,2%
4	Markov 3	45,6%	45,7%

Tabla 9: Valores para el indicador de *accuracy* obtenido para cada uno de los Modelos de Markov en todas las muestras de la validación cruzada, para la Prueba 3. (Elaboración Propia).

Los resultados obtenidos en esta prueba resultaron ser de menor precisión que los obtenidos en la prueba anterior. En este caso igualmente se obtienen mejores valores para el modelo de orden 3, en todos los subconjuntos, sin embargo dados estos resultados se recomienda utilizar una probabilidad límite de 0,3 para todos los modelos.

#### Prueba 4

Como ya se mencionó en secciones anteriores, dentro de los datos que se tienen existen algunos canales que poseen muy baja participación en los procesos de compra, por lo cual se cree que para estos canales el modelo podría predecir de forma no muy acertada debido a la escasa información. Es por esta razón que en la cuarta prueba del modelo que se realizó, se trabajó con un subconjunto del total de canales, específicamente con los 6 canales que tenían mayor presencia en las rutas<sup>24</sup>.

Con respecto a las probabilidades límites que se usaron en esta prueba, se consideraron igual que en la prueba 2 (0,3), debido a que con este valor se obtuvieron los mejores resultados.

---

<sup>24</sup> Estos canales son Directo, SEO, SEM, Email, Remarketing y Otros.

A continuación se muestran los valores de las medidas de *accuracy* que se obtuvieron en esta prueba.

Subconjunto	Orden del Modelo	<i>Acc In</i>	<i>Acc Out</i>
1	Markov 1	8,3%	9,1%
1	Markov 2	47,6%	47,6%
1	Markov 3	51,9%	51,9%
2	Markov 1	9,1%	9,1%
2	Markov 2	47,6%	47,6%
2	Markov 3	51,9%	51,9%
3	Markov 1	8,3%	9,1%
3	Markov 2	47,5%	47,6%
3	Markov 3	51,9%	51,9%
4	Markov 1	9,1%	9,0%
4	Markov 2	47,6%	47,7%
4	Markov 3	51,9%	51,9%

Tabla 10: Valores para el indicador de *accuracy* obtenido para cada uno de los Modelos de Markov en todas las muestras de la validación cruzada, para la Prueba 4. (Elaboración Propia).

En este caso los valores de precisión que se obtuvieron no resultaron ser mejores que los obtenidos en pruebas anteriores, esto puede deberse a que, a pesar de que se está dejando de lado los canales para los cuales se tiene información acotada, igualmente la gran participación de Directo en la mayoría de las rutas sesga bastante el modelo y no permitiría estimar la atribución con una precisión mayor al 52,1%, la cual se obtiene con Markov de orden 3.

## Prueba 5

Para la cuarta prueba que se realizó se utilizó una base de datos que contenía un menor número de canales, seleccionando esta vez sólo los 4 canales más frecuentes (Directo, SEO, SEM, Email). Las probabilidades límites que se emplearon también fueron de 0,3 para todos los modelos. Los resultados obtenidos se muestran en la Tabla 11, en la cual se puede ver que los mejores resultados igualmente se consiguieron al utilizar un modelo basado en Markov de tercer orden.

Subconjunto	Orden del Modelo	Acc In	Acc Out
1	Markov 1	7,6%	8,4%
1	Markov 2	47,2%	47,1%
1	Markov 3	51,4%	51,3%
2	Markov 1	8,4%	8,3%
2	Markov 2	47,2%	47,2%
2	Markov 3	51,4%	51,3%
3	Markov 1	7,6%	8,4%
3	Markov 2	47,2%	47,2%
3	Markov 3	51,4%	51,3%
4	Markov 1	8,4%	8,4%
4	Markov 2	47,1%	47,3%
4	Markov 3	51,3%	51,5%

Tabla 11: Valores para el indicador de *accuracy* obtenido para cada uno de los Modelos de Markov en todas las muestras de la validación cruzada, para la Prueba 5. (Elaboración Propia).

De forma análoga a la Prueba 4, en la Tabla 11 se obtienen resultados que tampoco superan en precisión a los de la Prueba 2. A pesar de restringir la base de datos a un subconjunto de canales para los cuales se tiene mucha más información que para otros medios, no es posible predecir con una mayor precisión, esto se debe principalmente a las características de la base de datos con la cual se está trabajando. De todas formas no es recomendable restringir tanto el conjunto de datos con el cual se está trabajando pues se estaría modificando el problema que se quiere resolver.

#### 4.4 PROPUESTA DEL MODELO DE ATRIBUCIÓN

Luego de la validación del modelo calculado, es posible proponer un modelo de atribución de conversiones para los canales con los cuales opera la empresa. Para esto se ha seleccionado el modelo basado en Markov de tercer orden, dado que con esta distribución se obtenían los mejores resultados de exactitud, independiente de los subconjuntos de la validación cruzada que se estuviesen usando, o el número de canales que se considerara.

Por medio de la utilización del concepto de *Removal Effect* fue posible cuantificar la importancia de cada canal, determinando el número de conversiones que se dejaban de concretar al eliminar un canal del conjunto de medios con el cual se estaba trabajando. A continuación se presenta la atribución de conversiones que se propone para el total de canales con los cuales opera la empresa.

Canales	Conversiones	Porcentaje
Directo	330.982	28,0%
SEO	161.211	13,7%
SEM	116.454	9,9%
EmailMkt	100.321	8,5%
RRSS Orgánico	58.530	5,0%
Remarketing	104.442	8,8%
Display	55.022	4,7%
Tiendas	46.236	3,9%
RRSS Pago	49.952	4,2%
Afiliados	48.094	4,1%
Otros	109.535	9,3%

Tabla 12: Modelo de atribución de conversiones basado en cadenas de Markov de tercer orden. (Elaboración propia).

Para conocer que tan distinto es este modelo de la atribución que usualmente utiliza la compañía, se adjunta una comparación del modelo propuesto con la atribución de conversiones *first touch* y *last touch*.

Canales	Modelo basado en Markov	Modelo First Touch	Diferencia del Modelo basado en Markov con respecto al First Touch	Modelo Last Touch	Diferencia del Modelo basado en Markov con respecto al Last Touch
Directo	28,03%	81,07%	-53,04%	89,70%	-61,67%
SEO	13,65%	13,99%	-0,34%	7,48%	+6,17%
SEM	9,86%	2,30%	+7,56%	1,01%	+8,85%
EmailMkt	8,50%	1,56%	+6,94%	0,76%	+7,74%
RRSSOrganico	4,96%	0,12%	+4,84%	0,15%	+4,81%
Remarketing	8,85%	0,21%	+8,64%	0,22%	+8,63%
Display	4,66%	0,16%	+4,50%	0,03%	+4,63%
Tiendas	3,92%	0,12%	+3,80%	0,00%	+3,92%
RRSSPago	4,23%	0,09%	+4,14%	0,03%	+4,20%
Afiliados	4,07%	0,05%	+4,02%	0,02%	+4,05%
Otros	9,28%	0,45%	+8,83%	0,60%	+8,68%

Tabla 13: Comparación del modelo propuesto con los modelos de atribución *First Touch* y *Last Touch*. (Elaboración propia).

En la Tabla 13 es posible visualizar que el modelo propuesto es bastante distinto a las distribuciones que plantean los modelos tradicionales. Esto se explica por la forma en la que se calculó el modelo (utilizando *Removal Effect*), este procedimiento empleado permite identificar que tan relevante es cada canal en la realización de las ventas, independiente si participa en el inicio o en el final del proceso de compra.

En general, los porcentajes de ciertos canales que son menos frecuentes<sup>25</sup> aumentan bastante su aporte a la conversión en el modelo que se postula, llevándose porcentaje de la participación de Directo. Es importante recordar que en el modelo basado en Markov se está considerando información de la ruta completa hacia la conversión, lo cual los modelos *first touch* y *last touch* ignoran. Por esta razón, el modelo basado en cadenas de Markov podría resultar ser una visión mucho más completa de lo que sucede antes de la compra. Sin embargo también se debe considerar que dadas las características de los datos que se tienen, la atribución de estos canales ha sido calculada con información acotada y que por tanto se alcanzó una precisión del modelo de un 52,1% como máximo.

A continuación se describe lo que podría estar sucediendo con la atribución de conversiones de cada uno de los canales utilizados por la empresa.

Las mayores diferencias se registran para el canal Directo, tanto comparando con *first touch* como con *last touch*. Para este canal se propone una atribución muy por debajo de las cifras que utiliza la empresa a diario, esto se explica tanto por las características de la construcción del modelo como también por la naturaleza *customer initiated* de este canal. Como ya se ha mencionado estos tipos de medios participan más concretando las conversiones, y por lo tanto si se observa sólo la última interacción se le atribuirán más conversiones que si se considera todo el proceso de compra, tal como lo hace el modelo de atribución basado en Markov.

Con respecto a SEO, el modelo que se propone asigna un porcentaje similar al determinado por el modelo *first touch* pero muy por sobre la atribución asignada bajo *last touch*, esto podría indicar que este medio se encuentra subestimado en este último modelo, debido a que más que participar en las últimas etapas del proceso de compra, lo hace en otros momentos de las rutas, por ejemplo al inicio o en instancias intermedias, lo cual resulta ser un tanto inesperado considerando la naturaleza del canal.

SEM también muestra una diferencia importante entre el porcentaje de conversiones propuestas y el que utiliza la empresa. Al parecer, este canal también estaría siendo subvalorado por el modelo de atribución tanto *first touch* como *last touch* lo que podría indicar que el aporte de dicho medio tiene lugar en las interacciones intermedias del proceso de compra. Sucede algo similar con el canal compuesto por correos electrónicos, para el cual el modelo propuesto asigna mucha más conversión que los modelos tradicionales. Dada la naturaleza de este último canal (*firm initiated*), su aporte no se ve reflejado en las últimas interacciones, de hecho al ver el porcentaje asignado a este canal en cada modelo, es posible afirmar que este medio participa más que nada en las etapas intermedias de las rutas.

Los canales Remarketing y Otros presentan importantes diferencias entre el modelo propuesto y los que utiliza la empresa, lo cual indica que su participación en los procesos de compra se ve un tanto sesgada si se considera solo modelos del tipo *first touch* o *last*

---

<sup>25</sup> Canales como RRSSOrgánico, Remarketing, Display, Tiendas, RRSSPago, Afiliados y Otros.

*touch*. Display, Redes Sociales Orgánicas y Afiliados igualmente resultan subvalorados bajo la visión de los modelos tradicionales de atribución.

Además de los medios mencionados es importante destacar lo que sucede con el canal RRSS Pago, para este canal el modelo de atribución *last touch* asigna un 0,03% de las conversiones, mientras que el modelo propuesto le atribuye un 4,23%. La diferencia de un 4,20% a simple vista es bastante alta pero a diferencia de otros medios que también presentan discrepancias de magnitud semejante, en este canal sucede algo muy particular. La plataforma por la cual se administra toda la inversión en RRSS Pago es Facebook, de la cual se han sabido casos en los cuales existen incongruencias entre la información reportada por esta plataforma y las herramientas de análisis de métricas online como lo son Google Analytics u Omniture. Es por esta razón que el área de SEO SEM de Falabella.com tiene especial interés en averiguar si efectivamente está sucediendo esto, debido principalmente a que las inversiones en este medio no son marginales y al parecer esto no se está reflejando en los resultados *last touch*. En la siguiente sección se adjunta un análisis complementario que se realizó sobre este canal.

#### **4.5. ANÁLISIS COMPLEMENTARIO SOBRE FACEBOOK**

La necesidad de indagar en este medio en particular surge por una inquietud del área de SEO SEM de Falabella.com. El equipo está interesado en conocer si la plataforma que utilizan normalmente (Omniture<sup>26</sup>) está reportando el real aporte del canal de Facebook, en los valores obtenidos usando el modelo de atribución *last touch*. Esto motivado por el contraste entre los grandes esfuerzos tanto en inversión como en dedicación que se dirigen al medio, y los bajos niveles de ventas que al parecer genera este canal según esta fuente [Anexo 15]. Mediante este análisis además se pretende abordar el efecto *cross device* que se cree que se da en este canal, este aspecto resulta interesante pues más del 90,0% del tráfico en el canal de Facebook proviene de dispositivos móviles lo cual contrasta con los bajos niveles de conversión *mobile* que se registran en el sitio [Anexo 16].

Gracias a la incorporación de un *tag* que fue instalado este año,<sup>27</sup> se tienen las herramientas para acceder a información más detallada de todas las interacciones que tiene un cliente con los anuncios del canal de Facebook [Anexo 17], de esta forma se cuenta con el escenario propicio para desarrollar este análisis y complementarlo con el modelo de atribución que se está realizando. Las variables más relevantes que entrega este reporte denominado Order Level Report, son:

- ID que identifica la orden a la cual corresponde una interacción.

---

<sup>26</sup> En este caso si se utiliza Omniture ya que se dispone de acceso a la información que se requiere en esta plataforma.

<sup>27</sup> La información se comenzó a registrar el 28 de Mayo del 2017

- Dispositivo por el cual se realiza la interacción, se distingue entre *Mobile* (iOS, Android) y *Desktop*.
- Tipo de interacción, es decir, indica si la interacción es un clic o una impresión.
- Tipo de *placement* del anuncio en Facebook, esto es el lugar en el cual aparece el aviso, puede ser en el *feed* de Facebook o en el costado, por ejemplo.

Para llevar a cabo el análisis mencionado se seleccionó información del mes de Abril debido a que es un mes que no presenta estacionalidades como si puede presentar el mes de mayo o junio, por ejemplo<sup>28</sup>. La base de datos creada con información desde el 1 al 30 de Abril está constituida por 1.577.533 filas, cada una de ellas representa distintas interacciones con el canal de Facebook, las cuales corresponden a 54.856 órdenes distintas que no necesariamente cerraron su venta utilizando dicho canal.

A partir de la base de datos mencionada se eliminaron aquellas órdenes que dentro de su ruta contenían algún ID's con error de medición o con algún formato inusual, además se extrajeron todas las rutas que tenían al menos una interacción durante los días 10,11, 12 y 19 de Abril<sup>29</sup>, esto debido a que durante estos días se realizaron eventos de precio en el sitio, los cuales involucraron considerables inversiones en Facebook elevando las ventas y visitas por sobre un día normal. Con estas modificaciones los datos se reducen a 264.735 interacciones distintas que corresponden a 28.921 órdenes únicas [Anexo 18].

Considerando los largos de las rutas contenidas en la base de datos fue posible identificar aquellas rutas más frecuentes y excluir todas aquellas rutas cuya secuencia presentaba muy poca frecuencia en la base, a estos últimos se les consideró como *outliers* de la base de datos y fue posible determinarlos por medio del *elbow test* [Anexo 19]. Todas aquellas secuencias con un largo mayor a 40 interacciones se clasificaron como *outliers*, con esto se alcanzó el 96,0% del total de las órdenes, obteniendo una base de datos de 135.095 interacciones y 27.795 órdenes distintas. A partir de esta información se extrajo una muestra aleatoria de 400 órdenes<sup>30</sup>, necesaria para los análisis más detallados sobre las características de los *customer journey* en el canal de Facebook.

Sumado a los datos de Facebook también se incluyó en el análisis la información disponible en Omniture sobre estas órdenes en particular, de esa forma fue posible acceder a información adicional como complemento de las interacciones visibles en Facebook. Variables como el ticket promedio, el canal por el cual convertían y la hora en la cual lo hacían, fueron parte de esta información extra con la que se trabajó.

---

<sup>28</sup> El mes de mayo posee estacionalidad ya que durante la primera quincena los *retailers* se enfocan en vender regalos para el día de la madre, por otro lado a fin de mes tiene lugar CyberDay. El mes de Junio normalmente comienzan las rebajas de la temporada de invierno, por lo que aumentan las ventas en comparación a un mes que no sucede esto.

<sup>29</sup> Un total de 21.957 ID's tienen en su *customer journey* alguna interacción dentro de los días especificados.

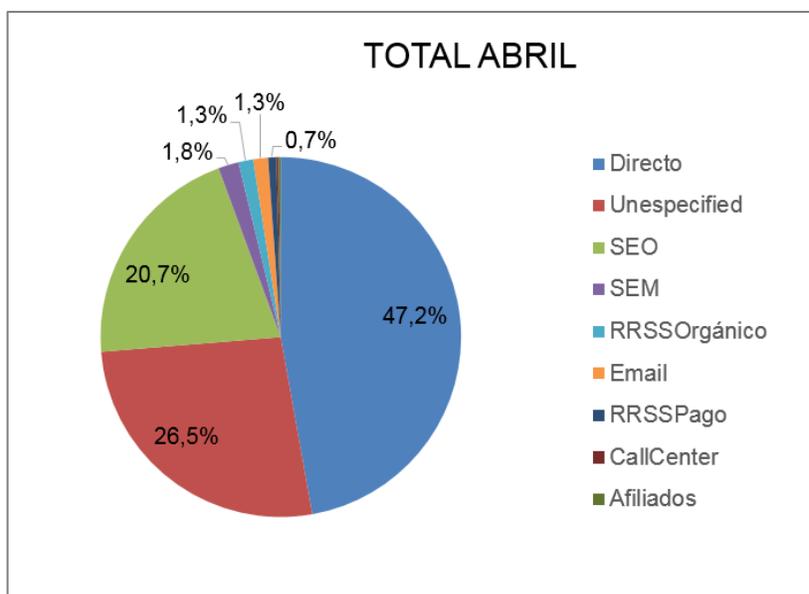
<sup>30</sup> Se calculó el mínimo muestral de órdenes para un N de 27.795, lo cual entregó un n = 379, se utilizó SPSS Statistics para extraer la muestra.

A continuación se incluyen los *insights* más relevantes sobre el análisis realizado y la justificación que llevó al estudiante a esas conclusiones.

I1: *Facebook funciona más como un canal que asiste las conversiones que como un canal para convertir. Al ser un canal del tipo firm initiated, participa más en las primeras e intermedias etapas, que en las últimas etapas del proceso de compra (cuando el cliente concreta la compra).*

Como ya se mencionó anteriormente, la contribución a la conversión del canal RRSS de Pago bajo el modelo de atribución *last touch* es bastante bajo (0,1%), esto podría estar indicando que esta forma de reportar no es la más adecuada para este canal, o que efectivamente este es el aporte que realiza el medio. Para averiguar qué es lo que está sucediendo se utilizó la información disponible en Omniture en conjunto con las órdenes que se extrajeron del Order Level Report. En primer lugar se calculó que del total de órdenes realizadas en el mes de estudio, un 26,0% posee al menos una interacción en Facebook dentro de su *customer journey*. En otras palabras, 1 de cada 4 conversiones tiene al menos un toque en este canal. Esto claramente contrasta con la cifra del aporte *last touch* de solo un 0,1%.

Para seguir indagando en este tópico se identificaron los canales de conversión de los ID's que registraban interacciones en Facebook, para identificar así si este canal estaba ayudando a la conversión de otros medios. Efectivamente Facebook asiste conversiones que terminan realizándose en su mayoría en otros canales (Directo y SEO<sup>31</sup>).



<sup>31</sup> No se considerará el canal "unespecified" ya que este corresponde a todas aquellas conversiones que no están clasificadas en un canal determinado, si no que el sistema de Omniture es incapaz de asignarlas a un canal de marketing.

Figura 8: Representación gráfica de la asistencia de conversiones desde Facebook a otros canales. (Elaboración Propia)

Dado que Facebook no asiste grandes conversiones a si mismo (0,7%) y tampoco registra un porcentaje importante de visitas (3,4%) y ventas (0,1%) *last touch*<sup>32</sup>, es posible afirmar que si bien este es un canal que participa en un cuarto del total de conversiones, no lo hace en las últimas etapas del proceso de compra.

*I2: Omniture efectivamente no está reportando la real participación de Facebook. Dada la naturaleza del modelo last touch, éste no refleja el aporte de un canal como Facebook (firm initiated) que no contribuye mayormente en las últimas etapas del proceso de compra (sólo un 3,4% en visitas last touch).*

La justificación de este *insight* está muy relacionado con lo que se incluyó en el apartado anterior. El hecho que Facebook no aporte bajo el modelo de atribución *last touch* no quiere decir que no aporte en lo absoluto, si no que su contribución no se refleja en este modelo. Considerando esta situación es necesario buscar otras maneras de cuantificar el aporte de estos canales (*firm initiated*), en especial de Facebook, de tal forma que sea posible visualizar la contribución que estos canales están dando a la conversión. Una herramienta útil para esto último es el modelo de atribución basado en cadenas de Markov que se propone en la siguiente sección, así como también algunas métricas que pueden servir para tener una mayor certeza del aporte de Facebook, por ejemplo el porcentaje de rutas en las cuales este canal registra interacciones con el cliente en comparación con el total de conversiones, entre otros.

*I3: Existe un gap entre lo reportado por Omniture y la información que entrega Facebook. Más específicamente, existen algunas conversiones que se realizan por RRSSPago, pero que en Omniture no están siendo atribuidas a ese canal. Es importante conocer este gap para tenerlo en cuenta al momento de cuantificar el aporte de Facebook.*

Con la información extraída de la API de Facebook es posible detectar situaciones en las cuales la conversión se le atribuyó a un canal determinado (C2 en la Figura 9) cuando en realidad el último toque previo a la conversión fue con el canal de Facebook; estos son algunos casos que Omniture es incapaz de visualizar. Estas situaciones muestran como cambia la atribución a Facebook, cuando se incorpora información del Order Level Report, a la cual no se puede acceder al considerar únicamente la fuente Omniture.

---

<sup>32</sup> Considerando información extraída desde Omniture para todo 2017.

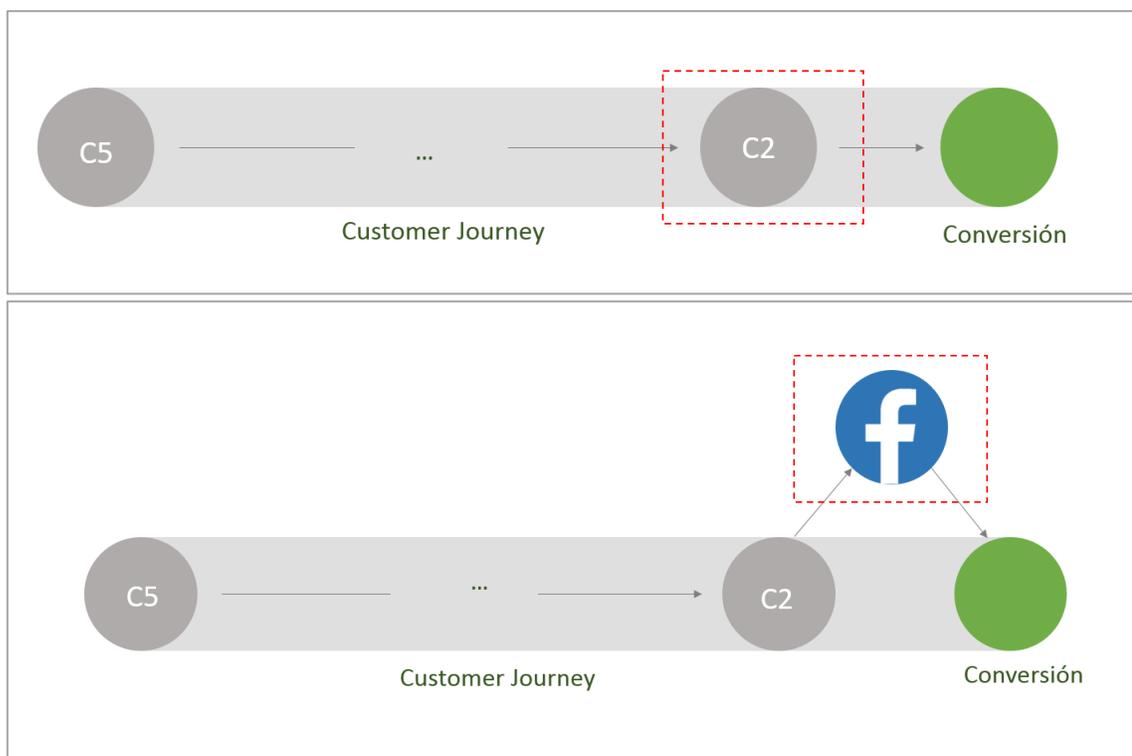


Figura 9: Representación del cambio de la atribución a Facebook cuando se incorpora información del Order Level Report. (Elaboración Propia)

Este fenómeno ocurre por discordancias entre los sistemas, las cuales se explican por situaciones en las cuales no es posible reportar correctamente la última interacción de la conversión<sup>33</sup>, esto sucede cuando por ejemplo el clic en un anuncio en Facebook ocurre en un dispositivo o en un navegador diferente al previo a la conversión (*last touch*).

Al calcular cuánto cambia la atribución a Facebook considerando esta información, es posible determinar un indicador denominado multiplicador, que señala por cuanto es necesario ponderar las atribuciones a Facebook según Omniture, para así tener una visión más completa de esta cifra. Para realizar este cálculo se debía trabajar con los *timestamp* de las interacciones que se registraban en Facebook y en Omniture, valores que resultaban muy engorrosos de extraer tanto desde Falabella como también por parte de Facebook. Por esta razón se debió trabajar con una muestra reducida de datos correspondientes al mes de interés, a partir de estos datos se comparó el número de conversiones atribuidas según Omniture a Facebook, con el número que se obtenía incorporando información del Order Level Report; a partir de esta comparación se pudo obtener una primera aproximación de lo que pudiese ser el multiplicador.

Considerando los datos de Omniture para el mes de Abril, sólo dos conversiones son atribuidas a Facebook. Sin embargo cuando se incorpora la información del Order Level Report esta cifra sube a 4 conversiones adicionales por medio de este canal, las cuales

<sup>33</sup> Información proporcionada por ejecutivos de Facebook

Omniture no asignó correctamente. Con estas cifras es posible calcular el multiplicador de la siguiente forma:

M: Multiplicador

C<sub>OM</sub>: Conversiones atribuidas a Facebook en Omniture

C<sub>OLR</sub>: Conversiones atribuidas a Facebook con información del Order Level Report

$$M = \frac{C_{OM} + C_{OLR}}{C_{OM}} \Rightarrow \frac{2 + 4}{2} = 3$$

Con los datos obtenidos del mes de Abril, el multiplicador es 3, lo cual está dentro de los rangos esperables según información de los encargados de Facebook, sin embargo igualmente se recomienda calcular los multiplicadores de otros meses y con mayor información cuando sea posible extraerla de una manera más fácil, para poder corroborar esta cifra a modo de trabajo futuro.

*14: En Facebook predominan las interacciones mobile (en su mayoría impresiones), sin embargo cuando los clientes pasan al last touch (etapa previa a la conversión), migran a otros canales y a desktop.*

A partir de la información del Order Level Report, es posible identificar las rutas *cross device* más frecuentes dentro del canal de Facebook. Si se observan estas rutas de inmediato llama la atención que más del 50,0% más frecuente está compuesto completamente por interacciones en *mobile* (Tabla 14).

Rutas más frecuentes	Frecuencia	Porcentaje
Mobile	134	33,5%
Mobile, Mobile	81	20,3%
Desktop	27	6,8%
Mobile, Mobile, Mobile	21	5,3%
Desktop, Desktop	9	2,3%

Tabla 14: Rutas más frecuentes indicando dispositivos empleados en ellas. (Elaboración Propia) [Anexo 20]

Dado que los dispositivos móviles participan más en las primeras etapas del proceso de compra (Fuenzalida, 2016) era esperable que en el canal de Facebook también sucediera esto, es decir, que en las primeras etapas de la ruta existiera más participación *mobile* que en las últimas, mientras que de forma análoga se esperaba que *desktop* tuviese más presencia hacia el final de la ruta. Motivados por esta teoría se estudió la distribución de los distintos dispositivos, especialmente de *mobile* y *desktop* a lo largo de la ruta realizada por los clientes dentro del canal de Facebook. En primer lugar se evaluó si entre la primera y la última interacción existía gran diferencia en los dispositivos utilizados, el resultado obtenido fue contra intuitivo pues no se encontraron diferencias lo suficientemente importantes como se hubiese esperado (Figura 10).

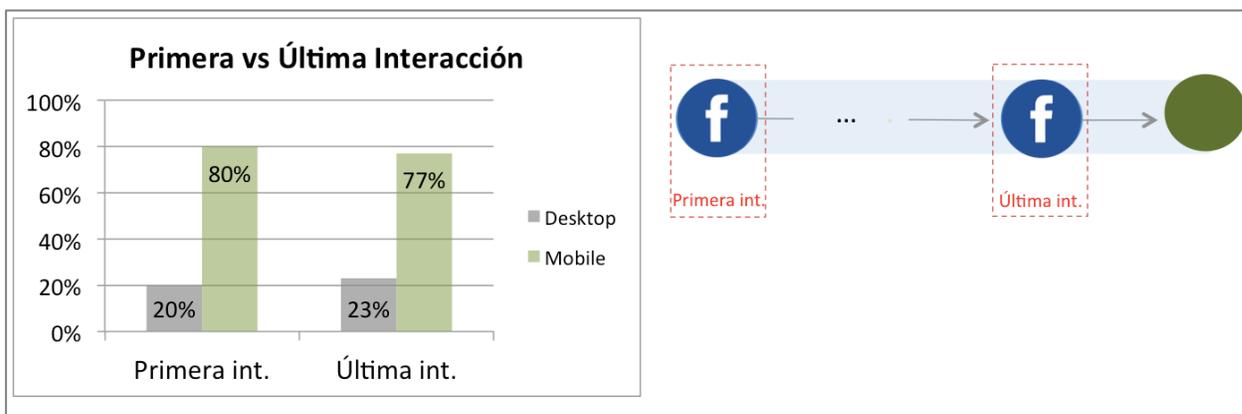


Figura 10: Representación de las diferencias de participación *desktop – mobile* entre la primera y última interacción dentro de la ruta en Facebook. (Elaboración Propia)

De forma análoga y para continuar con el estudio de las rutas, se realizó el mismo ejercicio pero esta vez se consideró no solo la primera y última interacción, si no que se comparó el inicio con el final de la ruta, definiendo inicio como el primer tercio del proceso de compra y final como el último tercio<sup>34</sup>. Los resultados tampoco mostraron grandes diferencias (sólo de un 6,0%) (Figura 11).

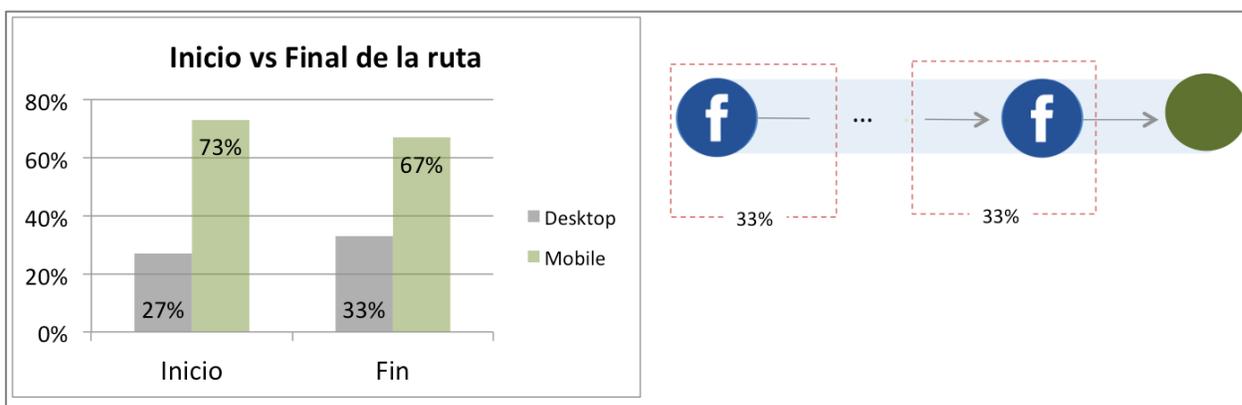


Figura 11: Representación de las diferencias de participación *desktop – mobile* entre el inicio y el final de la ruta en Facebook. (Elaboración Propia)

Esto indica que a lo largo de la ruta que realizan los clientes no hay mayores migraciones desde los dispositivos *mobile* a *desktop*. De hecho es posible calcular la probabilidad de las transiciones entre estos dos medios y efectivamente es sólo de un 5,3%. [Anexo 21]. Ante estos resultados se indagó en lo que pasaba en el último toque (inmediatamente previo a la conversión) y se comprobó que esta es la instancia en la cual los clientes se cambian de dispositivo. Efectivamente, de la muestra que se utilizó para caracterizar las rutas solo un 26,0% convertía por *mobile*, mientras que un 67,0% lo hacía utilizando dispositivos *desktop*.

<sup>34</sup> Esto se definió utilizando el criterio experto por medio de una consulta al equipo de Falabella.com

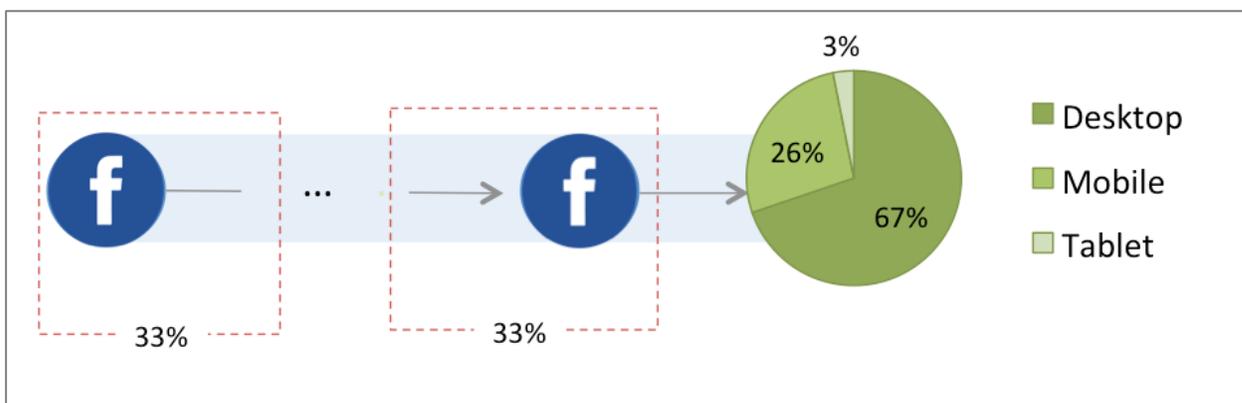


Figura 12: Dispositivos en los cuales convierten los ID's de la muestra en Facebook.  
(Elaboración Propia)

Con respecto al tipo de interacción, es interesante notar la gran diferencia que existe entre los porcentajes de clics e impresiones. Tanto para la base de datos sin *outliers* como también para la muestra, el porcentaje de clics no supera el 20,0% del total de interacciones (Tabla 15), mientras que el resto lo conforman impresiones de los anuncios en Facebook.

Tipo de Interacción	Base sin Outliers	Muestra
Clics	4,7%	18,0%
Impresiones	95,3%	81,9%

Tabla 15: Comparación de clics e impresiones para distintas selecciones de la bases de datos.  
(Elaboración propia)

*16: A medida que aumenta el ticket de una orden, no varían las rutas más frecuentes escogidas por los clientes dentro del canal de Facebook (en términos de dispositivos). Sin embargo si varía el dispositivo que escogen para convertir.*

Para concluir esto se realizó un análisis en el cual se identificaron las rutas más frecuentes para distintos rangos de ticket. Se clasificaron las órdenes de la muestra dentro de cuatro rangos de ticket distintos<sup>35</sup>, menores a \$50.000, entre \$50.000 y \$100.000, entre \$100.000 y \$300.000, y mayores de \$300.000. Este último rango se dejó fuera del estudio dado que se contaba con muy pocas observaciones<sup>36</sup>, con lo cual es probable que no se obtuviesen conclusiones válidas. El primero de los niveles era el más frecuente, siguiéndole a continuación el tercero mencionado, de \$100.000 a \$300.000.

<sup>35</sup> Esto fue decidido en conjunto con el equipo de SEO-SEM, considerando los rangos de ticket que se utilizan en los reportes de Falabella.com

<sup>36</sup> Dentro de este rango estaba incluido solo un 3,0% de la muestra.

Rangos	Porcentajes de la muestra
0-\$50.000	52,0%
\$50.000-\$100.000	17,7%
\$100.000-\$300.000	26,9%
> \$300.000	3,4%

Tabla 16: Rangos de tickets determinados para evaluar eventuales diferencias en la ruta en Facebook. (Elaboración Propia)

Para cada uno de los tres rangos se identificó, además de la ruta que seguían los clientes, el canal y el dispositivo por el cual convertían, para así poder identificar diferencias entre cada uno de los niveles de tickets.

Los resultados obtenidos indican que las rutas con mayor frecuencia son iguales para los tres rangos de órdenes considerados. Estas rutas siempre son:

- *Mobile*
- *Mobile -> Mobile*
- *Desktop*

Sin embargo a medida que crece el ticket aumenta la conversión por medio de *desktop*, esto es esperable ya que las personas son reacias a comprar productos de precios muy elevados a través de una interfaz más pequeña y muchas veces no tan amigable con el usuario. Por otro lado Directo es el medio más utilizado para convertir, independiente de cuánto dinero se esté gastando, sin embargo este canal predomina mayormente usando los dispositivos *desktop* en los tickets más altos.

## 5. RECOMENDACIONES Y CONCLUSIONES

### 5.1 RECOMENDACIONES PARA LA EMPRESA

Habiendo desarrollado un modelo de atribución basado en los datos de la empresa, se recomienda al área de SEO SEM, implementar el hábito de visualizar las métricas de dicho modelo en complemento con las de los modelos *first touch* y *last touch*, al momento de tomar decisiones sobre inversión en determinados canales. Esto debido a que como se abordó en el estudio, no todos los medios que usa Falabella van a tener participación en la última etapa previa a la conversión, ni tampoco en la primera. De hecho existen varios canales que están subestimados bajo esa forma de atribuir conversiones. Algunos canales participan en las primeras etapas, otros en momentos intermedios del proceso de compra y al ver solo un modelo de atribución no se está obteniendo una visión completa del aporte de cada canal.

Con respecto al análisis de Facebook es posible mencionar las recomendaciones que a continuación se incluyen.

Es necesario para el área de SEO SEM, tener en consideración el sesgo que existe entre Omniture y lo que reporta el Order Level Report para las conversiones atribuidas a Facebook. Para esto se requiere tener en cuenta el multiplicador calculado al momento de determinar el número de compras realizadas por el canal de RRSS de Pago.

Para cuantificar el aporte de Facebook es fundamental ver más allá de las métricas *last touch*. Al ser un canal que asiste conversiones no reportará “buenos resultados” viendo solo ventas *last touch*, si no que será necesario indagar en otros KPI's que permitan visualizar la contribución de este canal.

Además, siendo un poco más específico, se recomienda realizar campañas de Facebook enfocadas en los dispositivos móviles y sobretodo centrados en líneas blandas, específicamente de tickets bajo \$100.000. Además se propone incluir inversión en este medio con énfasis en las campañas de tipo *branding* que se enfocan más en las impresiones que en los clics.

### 5.2 PRINCIPALES CONCLUSIONES

La distribución de conversiones basada en un modelo de Markov de tercer orden resultó ser la que mejor se adecuaba a los datos de la empresa. Sin embargo dada las características de los datos de la empresa, sólo se pudo predecir con un 52,1% de precisión. A pesar de esto, por medio del modelo propuesto fue posible identificar que al observar únicamente el modelo de atribución *last touch* existen varios medios cuyo aporte a la conversión resultó estar subvalorado. Por medio del modelo de atribución que se sugiere es posible tener una visión más completa del aporte de los canales a lo largo de todo el *customer journey* que realizan los clientes hasta la compra.

La importancia de tener distintas visiones de la atribución radica en no sesgarse únicamente a un modelo tan estricto como el *last touch*. Si se conoce más sobre el aporte de los distintos canales a lo largo del proceso de compra, es posible contar con más antecedentes valiosos a la hora de distribuir un presupuesto de una campaña digital en los distintos medios, dependiendo del objetivo de ésta. Por ejemplo, para una inversión en pauta enfocada en dar a conocer una nueva marca o colección se podría recomendar dar énfasis al canal SEO por ejemplo, el cual participa bastante en las primeras e intermedias etapas del proceso de compra. Por otro lado, si se trata de una campaña enfocada en generar conversiones, se recomendaría enfocar los esfuerzos en un canal como Directo que los clientes emplean más para concretar sus compras.

Con respecto a la participación de los distintos tipos de canales a lo largo de las rutas, es posible mencionar que efectivamente de los canales del tipo *customer initiated*, Directo es el que más participa en las últimas etapas del proceso de compra, justo en el momento cuando el cliente ya tiene más desarrollada la decisión y solo se dirige al sitio para adquirir su producto. Este canal por lo tanto reporta los mejores resultados bajo el modelo de atribución *last touch*<sup>37</sup>, sin embargo si se mide su desempeño considerando todo el *customer journey* (es decir con el modelo propuesto), no reporta resultados tan importantes como los antes mencionados.

Relacionado con los canales del tipo *firm initiated*, no se obtuvieron resultados tan contrastados como los señalados en el párrafo anterior, sin embargo es posible señalar que la mayoría de estos medios estaban subestimados bajo el modelo de atribución *last touch*, pues se cree que su participación en la ruta de los clientes ocurría principalmente en instancias intermedias de éstas. Es importante recordar que para estos canales se tenía información mucho más limitada que para los canales *customer initiated*, por lo cual predecir para estos medios es mucho más complejo.

A partir del análisis complementario del canal Facebook se comprobó que este canal actúa principalmente asistiendo conversiones que luego se concretan en otros medios, en su mayoría del tipo *customer initiated* como Directo, por lo cual al ver los resultados bajo el modelo *last touch* este medio no reporta buenos desempeños. Con los resultados del modelo propuesto y el análisis que se realizó se pudo constatar que existen discrepancias entre los sistemas de Facebook y Omniture para lo cual es necesario tomar consideraciones extras como el cálculo del multiplicador, y de esta forma tener mayor certeza del aporte de este canal.

### 5.3 TRABAJO FUTURO

Como trabajo futuro se propone en primer lugar realizar el modelo basado en cadenas de Markov de orden 4. Se espera que estos resultados sean mucho más precisos que

---

<sup>37</sup> 89,7% para Directo bajo el modelo *last touch*, en comparación con un 28,0% que reportó este medio bajo el modelo de atribución propuesto.

los obtenidos con el modelo de orden 3, pero implicarían mucho más tiempo de procesamiento que los modelos realizados en el presente estudio. Por otra parte se plantea igualmente calcular el modelo para los otros países en los cuales opera Falabella.com. Resulta interesante para la empresa saber cuánto varían los resultados dependiendo del desempeño online de cada país.

Además de lo planteado, se propone caracterizar a los clientes que realizan cada una de las rutas que se identificaron. Si se conoce su perfil sería posible identificar distintos segmentos con información y rutas similares para así implementar acciones que conduzcan a determinados clientes por estas rutas habituales, propias de su segmento hasta llevarlos a una conversión.

Si se dispone de la información de las rutas para aquellas compras que no se llevan a cabo también podría incorporarse esa información para conocer donde los clientes abandonan el proceso de compra e implementar medidas en esos instantes, con el objetivo de evitar que los clientes desistan de su compra.

Relacionado con el análisis específico de Facebook se plantea calcular los multiplicadores para otros meses y ver que tanto varía este indicador. Así es posible tener mayor certeza del comportamiento de la cifra al momento de querer usarla en el cálculo de las conversiones atribuidas a Facebook.

Por último, también se propone realizar un análisis descriptivo de otros medios relevantes, como complemento del análisis de Facebook. Considerar por ejemplo las interacciones con el canal de SEO y SEM podría ser de gran utilidad para evaluar la inversión de esfuerzo y presupuesto en estos medios.

## 6. BIBLIOGRAFÍA

Anderl, E., Becker, I., von Wangenheim, F., & Schumann, J. H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing*, 33(3), 457–474.

Blogthinkbig. (n.d.). Tendencias de marketing de atribución digital: plataformas y modelos de atribución. [en línea] <http://blogthinkbig.com/tendencias-de-marketing-de-atribucion-digital-plataformas-y-modelos-de-atribucion/>; [consulta: 17 Marzo 2017]

Bluecaribu. (n.d.). Modelos de atribución en Google Adwords y Analytics. [en línea] <http://www.bluecaribu.com/modelos-atribucion/>; [consulta: 06 Marzo 2017]

Cámara de Comercio de Santiago. (2016). TENDENCIAS DEL RETAIL EN CHILE. [en línea] [https://www.ccs.cl/prensa/2016/01/tendencias\\_retail\\_2016\\_WEB.PDF](https://www.ccs.cl/prensa/2016/01/tendencias_retail_2016_WEB.PDF); [consulta: 05 Marzo 2017]

Desmark. (n.d.). Modelos de atribución, la guía definitiva | desmarkt &gt; Marketing Online para Directivos. [en línea] <http://www.desmarkt.com/modelos-atribucion-guia-definitiva.html>; [consulta : 06 Marzo 2017]

Economía y Negocios, M. P. I. (2015). CCS: Ventas online crecen 15% en 2015. [en línea] <http://www.economiaynegocios.cl/noticias/noticias.asp?id=211769>; [consulta : 15 Noviembre 2016]

Economía y Negocios, & E., J. M. (n.d.). Falabella, Cencosud y Ripley alcanzan el 86% del negocio de las multitiendas del país. [en línea] <http://www.economiaynegocios.cl/noticias/noticias.asp?id=210759>; [consulta : 15 Octubre 2016]

Falabella. (2016). Memoria Anual 2015. [en línea] <http://www.falabella.com/static/staticContent/content/minisitios/Inversionistas/memoriaWeb/2015/MemoriaFalabellaSVS.pdf>; [consulta : 15 Agosto 2017]

Fuenzalida, R. (2016). *Estudio exploratorio del comportamiento de clientes móviles en un sitio web de retail*. Universidad de Chile.

Goic, M., Jerath, K., & Kalyanam, K. (2017). Multichannel Path-to-Purchase: Channels as "Engagers" and "Closers", 10027 (11121434), 1–43.

Li, H. (Alice), & Kannan, P. K. (2014). Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment. *Journal of Marketing Research*, 51, 40–56.

Marketing 4 Commerce. (2015). Qué son los modelos de atribución: por qué el last clic no es suficiente. [en línea] <https://marketing4ecommerce.net/que-son-los-modelos-de-atribucion-y-por-que-el-last-clic-no-es-suficiente/>; [consulta : 16 Marzo 2017]

Sergey Bryl. Attribution model with R (part 1: Markov chains concept) | R-bloggers. [en línea] <https://www.r-bloggers.com/attribution-model-with-r-part-1-markov-chains-concept/>; [consulta : 15 Mayo 2017]

Occam's Razor. (n.d.). Multi-Channel Attribution Modeling: The Good, Bad and Ugly Models - Occam's Razor by Avinash Kaushik. [en línea] <https://www.kaushik.net/avinash/multi-channel-attribution-modeling-good-bad-ugly-models/>; [consulta : 24 Marzo 2017]

The World Bank. (2016). Mobile cellular subscriptions (per 100 people) | Data. [en línea] [http://data.worldbank.org/indicador/IT.CEL.SETS.P2](http://data.worldbank.org/indicator/IT.CEL.SETS.P2); [consulta : 15 Abril 2017]

## 7. ANEXOS

### Anexo 1



Figura 13: Presencia Internacional por línea de negocio Falabella. (Memoria Anual Falabella, 2015)

### Anexo 2

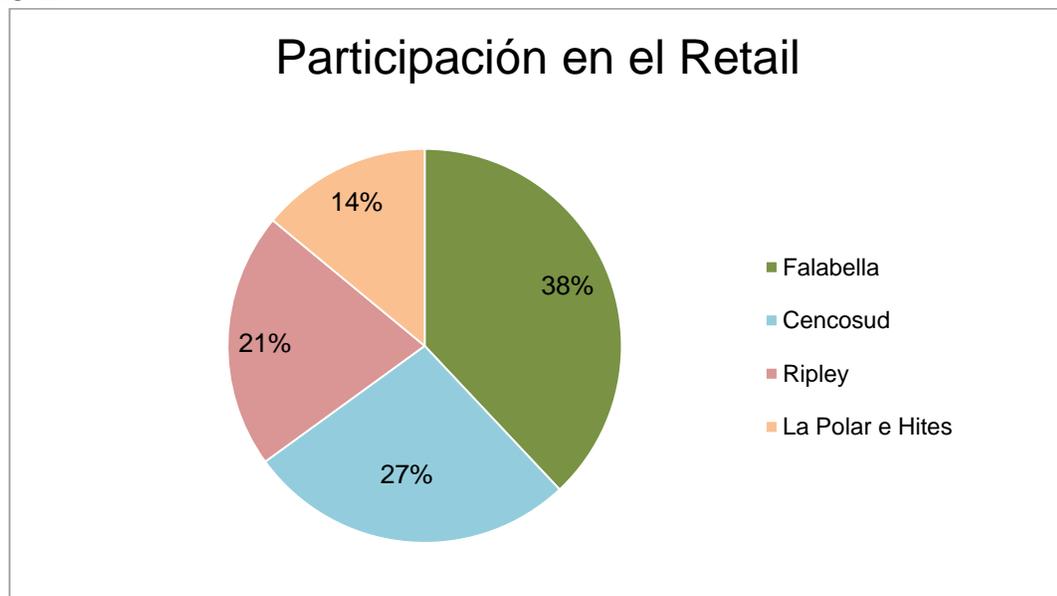


Figura 14: Participación de Falabella en los ingresos del *Retail*. (Economía y Negocios & E, 2015)

### Anexo 3

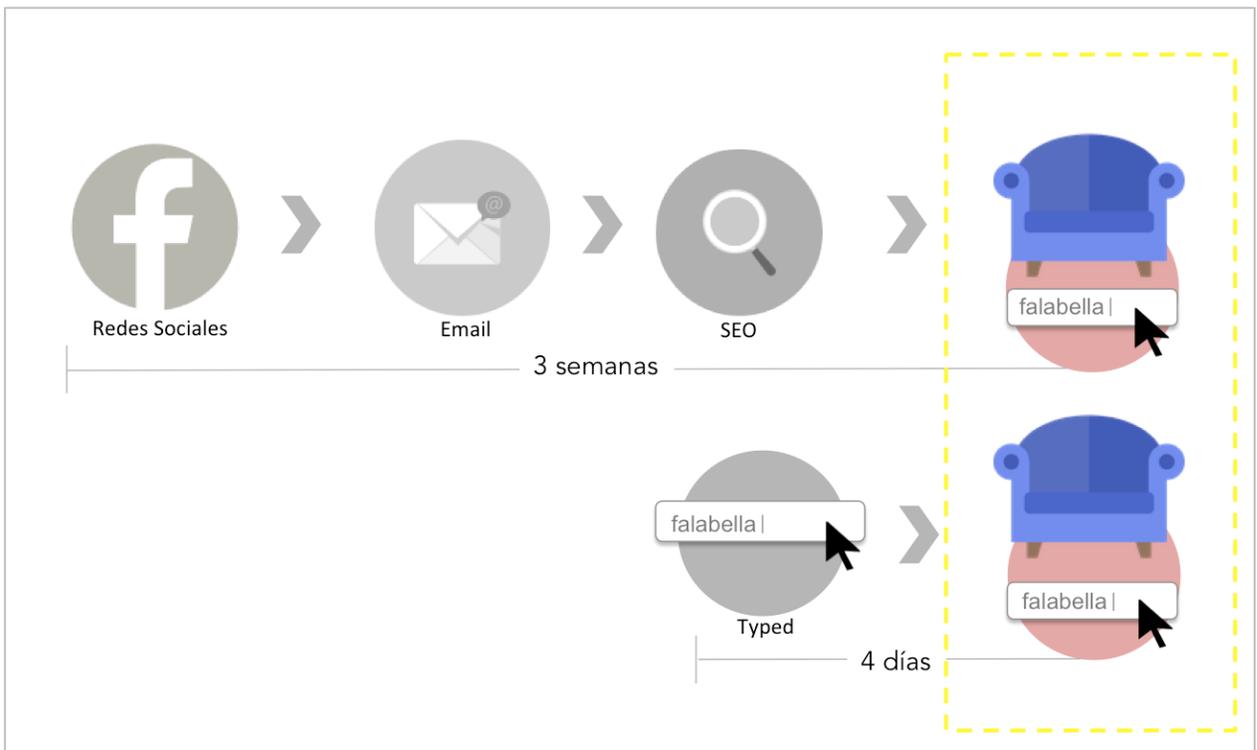


Figura 15: Ejemplo de compras iguales bajo el modelo de atribución *last touch*, pero que presentan distintos procesos de compra. (Elaboración Propia)

### Anexo 4

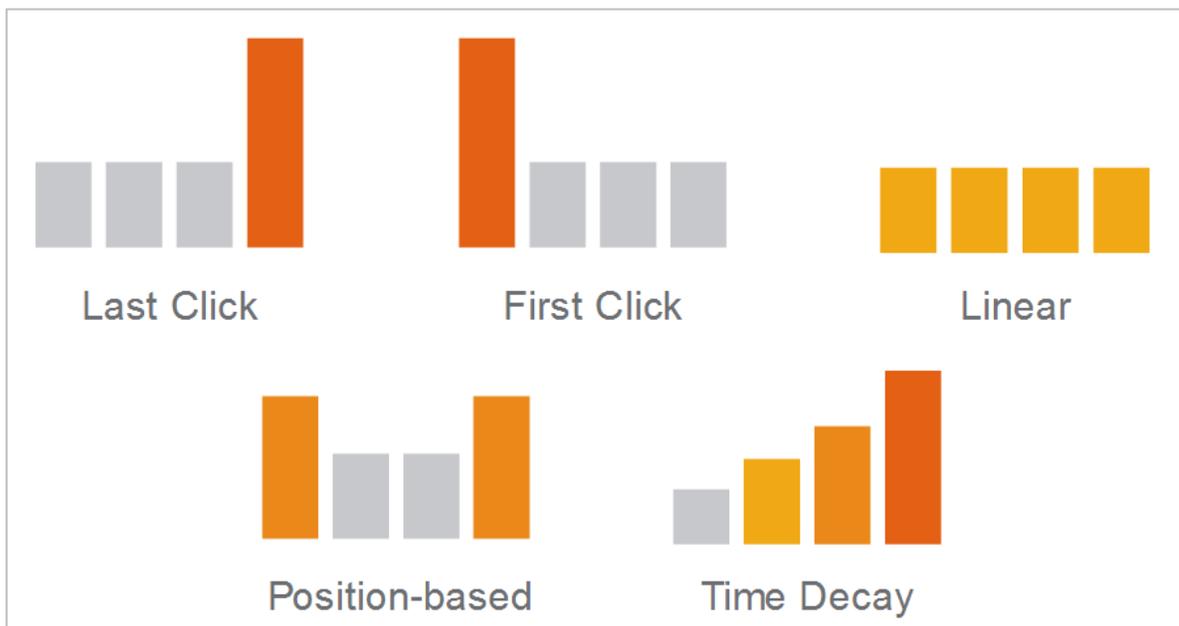


Figura 16: Representación gráfica de los modelos de atribución de conversiones. (Marketing 4 Commerce, 2015)

## Anexo 5

### Elementos Principales de las Cadenas de Markov.

Las cadenas de Markov están compuestas por los elementos que se describen a continuación:

- Un conjunto finito de estados, excluyentes entre sí. El estado de un sistema ( $X_n$ ) en el instante  $n$  es una variable aleatoria que puede tomar distintos valores pertenecientes al conjunto de estados  $E$ . De esta forma es posible describir la situación actual del sistema, si por ejemplo,  $x_n = i$ , entonces el proceso se encuentra en el estado  $i$  para el instante  $n$ .
- Probabilidades de transición entre estados. Estas probabilidades ilustran la posibilidad de que dado que un sistema está en un estado determinado, pase a otro estado que se aloja dentro de los posibles valores que puede tomar el sistema. Específicamente si el sistema se encuentra en un estado  $i$ , la probabilidad de que pase a un estado  $j$  está dada por:

$$P\{x_{n+1} = j \mid x_n = i, x_{n-1} = i_{n-1}, \dots, x_1 = i_1, x_0 = i_0\} = P_{ij}$$

Para todos los estados  $i_0, i_1, i_2 \dots i_{n-1}, i, j$ ;  $n \geq 0$ .

Es importante mencionar también que las probabilidades de transición siempre son no negativas y menores a 1, además satisfacen lo siguiente:

$$P_{ij} \geq 0, \quad i, j \geq 0; \quad \sum_{j=0}^{\infty} P_{ij} = 1, \quad i = 0, 1, \dots$$

Estos valores pueden representarse utilizando una matriz de transición  $P$ , la cual siempre toma la forma de una matriz cuadrada.

$$P = \begin{bmatrix} P_{00} & P_{01} & \dots \\ P_{10} & P_{11} & \dots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

- Distribución actual  $P_0$ . Corresponde a la distribución de las probabilidades de los estados para el periodo actual o el que se considera como inicial, esta matriz permite conocer las probabilidades de los estados posteriores.
- Paso del ciclo de Markov. Es el periodo de tiempo que se utilizará como intervalo para evaluar los estados del sistema.

Además, toda cadena de Markov debe cumplir con la propiedad markoviana, que se describe de la siguiente forma:

Siendo E los estados del sistema,

$$P[x_{n+1} = E_i | X_0 = E_{i_0}, X_1 = E_{i_1}, \dots, X_n = E_{i_n}] = P[X_{n+1} = E_i | X_n = E_{i_n}]$$

La ecuación apunta a que si se conoce la evolución del sistema, el estado presente resume la información relevante para conocer como se comportará dicho sistema en el siguiente estado.

## Anexo 6

Marketing Channels ?	Conversiones ?	↓	Valor de las conversiones ?
1. Directo	190.236 (32,74 %)		20.099.235.863,00 \$ (32,37 %)
2. Directo x 2	66.766 (11,49 %)		7.353.642.385,00 \$ (11,84 %)
3. Directo x 3	37.090 (6,38 %)		4.142.522.636,00 \$ (6,67 %)
4. Directo x 4	22.818 (3,93 %)		2.573.859.000,00 \$ (4,15 %)
5. Directo x 5	15.094 (2,60 %)		1.679.682.858,00 \$ (2,71 %)
6. SEO	12.664 (2,18 %)		1.278.251.794,00 \$ (2,06 %)
7. Directo x 6	11.071 (1,91 %)		1.210.536.162,00 \$ (1,95 %)
8. Directo x 7	8.046 (1,38 %)		894.889.206,00 \$ (1,44 %)
9. SEO Directo	7.897 (1,36 %)		876.094.458,00 \$ (1,41 %)
10. Directo x 8	5.943 (1,02 %)		642.862.663,00 \$ (1,04 %)
11. Directo x 9	4.648 (0,80 %)		515.152.571,00 \$ (0,83 %)
12. SEO Directo x 2	4.077 (0,70 %)		471.673.244,00 \$ (0,76 %)
13. Directo x 10	3.701 (0,64 %)		406.993.545,00 \$ (0,66 %)
14. Directo x 11	2.959 (0,51 %)		327.642.762,00 \$ (0,53 %)
15. Directo x 12	2.436 (0,42 %)		258.593.320,00 \$ (0,42 %)
16. SEM	2.309 (0,40 %)		204.554.829,00 \$ (0,33 %)

Figura 17: Ejemplo de las rutas disponibles en Google Analytics. (Google Analytics).

## Anexo 7

### Información de Google Analytics

Marketing Channels		Última Interacción	
	Conversiones		Valor de las conversiones
1. Directo	890.862,00 (89,89 %)		92.977.280.721,00 \$ (91,01 %)
2. SEO	45.381,00 (4,58 %)		4.528.345.491,00 \$ (4,43 %)
3. SEM	17.587,00 (1,77 %)		1.596.311.809,00 \$ (1,56 %)
4. Email Mkt	12.772,00 (1,29 %)		991.819.375,00 \$ (0,97 %)
5. (Otros)	9.175,00 (0,93 %)		953.497.462,00 \$ (0,93 %)
6. Remarketing	9.142,00 (0,92 %)		566.355.914,00 \$ (0,55 %)
7. RRSS Orgánico	2.609,00 (0,26 %)		287.407.654,00 \$ (0,28 %)
8. RRSS Pago	1.429,00 (0,14 %)		93.570.962,00 \$ (0,09 %)
9. Display	1.103,00 (0,11 %)		62.631.645,00 \$ (0,06 %)
10. Tiendas	652,00 (0,07 %)		67.888.760,00 \$ (0,07 %)

### Información de Adobe Analytics

Last Touch Channel	Ordenes	Ventas
1. None	350.782 34,9%	35.842.636.836 CLP 33,1%
2. Call Center	216.321 21,5%	29.090.332.409 CLP 26,8%
3. Typed	281.173 27,9%	28.270.944.269 CLP 26,1%
4. SEO	123.864 12,3%	12.465.929.889 CLP 11,5%
5. SEM	12.387 1,2%	1.106.921.464 CLP 1,0%
6. Email	12.601 1,3%	866.850.874 CLP 0,8%
7. RRSS Orgánico	4.487 0,4%	390.031.197 CLP 0,4%
8. RRSS Pago	2.288 0,2%	166.109.237 CLP 0,2%
9. RMKT Dinámico	1.379 0,1%	80.325.269 CLP 0,1%
10. Display	596 0,1%	48.992.432 CLP 0,0%
11. Tiendas	298 0,0%	14.061.927 CLP 0,0%
12. Afiliados	67 0,0%	5.050.385 CLP 0,0%

Figura 18: Comparación de los datos obtenidos desde Google Analytics (izquierda) y Adobe Analytics (derecha).

## Anexo 8

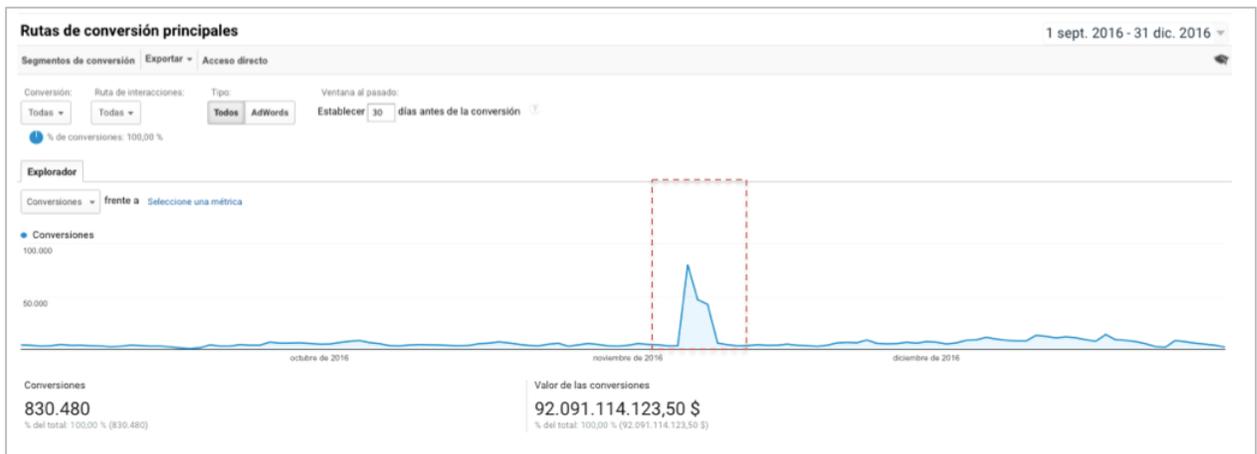


Figura 19: Registro del peak en conversiones durante CyberDay de Noviembre. (Google Analytics)

## Anexo 9

Número de rutas consideradas	30	40	60	100	150	200	1000	4600 (total)
Porcentaje de conversiones alcanzado c/r al total inicial *	82,9%	84,9%	87,3%	90%	91%	92,4%	97,2%	100%
Presencia de cada canal								
Directo	93,1%	94,1%	94,7%	89,2%	85,3%	83,4%	81,3%	81,7%
Afiliados	0,0%	0,0%	0,0%	0,0%	0,0%	0,2%	0,4%	0,7%
SEO	4,3%	3,6%	2,6%	4,0%	5,1%	5,3%	4,9%	4,5%
SEM	1,6%	1,3%	1,4%	3,0%	3,4%	3,8%	3,6%	3,0%
Email	1,1%	0,7%	0,7%	2,1%	2,8%	3,1%	3,9%	3,6%
Display	0,0%	0,0%	0,0%	0,1%	0,3%	0,5%	0,6%	0,7%
Otros	0,0%	0,3%	0,5%	1,0%	1,2%	1,4%	2,1%	3,0%
RRSS Org	0,0%	0,0%	0,0%	0,1%	0,3%	0,3%	0,3%	0,4%
RRSS Pago	0,0%	0,0%	0,0%	0,0%	0,2%	0,2%	0,4%	0,4%
Remarketing	0,0%	0,0%	0,0%	0,5%	1,2%	1,6%	2,1%	1,9%
Tiendas	0,0%	0,0%	0,0%	0,0%	0,1%	0,1%	0,2%	0,1%

\*Total inicial se considera a las 1.316.143 conversiones resultantes de la base de datos sin CyberDay

Tabla 17: Análisis de sensibilidad para la totalidad de rutas a considerar en el modelo.  
(Elaboración Propia)

## Anexo 10

Marketing Channels	Conversiones	Porcentaje
Directo	387.326	29,4%
Directo > Directo	152.367	11,6%
Directo > Directo > Directo	90.848	6,9%
SEO	74.695	5,7%
Directo > Directo > Directo > Directo	58.658	4,5%
Directo > Directo > Directo > Directo > Directo	40.453	3,1%
SEO > Directo	36.757	2,8%
Directo > Directo > Directo > Directo > Directo > Directo	30.466	2,3%
Directo > Directo > Directo > Directo > Directo > Directo > Directo	22.796	1,7%
SEO > Directo > Directo	18.398	1,4%

Directo > Directo	17.741	1,3%
Directo > Directo	14.163	1,1%
Directo > Directo	11.490	0,9%
SEO > Directo > Directo > Directo	10.232	0,8%
Directo > Directo	9.595	0,7%
Directo > Directo	8.018	0,6%
SEM > Directo	7.435	0,6%
Directo > SEO	6.903	0,5%
SEM	6.896	0,5%
Directo > Directo	6.627	0,5%
SEO > Directo > Directo > Directo > Directo	6.023	0,5%
Directo > Directo	5.933	0,5%
Directo > Directo	5.238	0,4%
Email Mkt	4.988	0,4%
Directo > SEO > Directo	4.326	0,3%
Directo > Directo	4.202	0,3%
SEM > Directo > Directo	4.194	0,3%
Directo > Directo	4.028	0,3%
Email Mkt > Directo	3.948	0,3%
SEO > Directo > Directo > Directo > Directo > Directo	3.810	0,3%
Directo > Directo	3.590	0,3%
Directo > Directo	3.034	0,2%





## Anexo 11

Canales	Conversiones	Porcentaje
Directo	957.215	81,1%
SEO	165.145	14,0%
SEM	27.127	2,3%
Email	18.392	1,6%
Otros	5.265	0,5%
RMKT	2.430	0,2%
Display	1.846	0,2%
RRSS Orgánico	1.435	0,1%
RRSS Pago	1.009	0,1%
Afiliados	583	0,1%
Tiendas	330	0,0%

Tabla 19: Contribución a la conversión de los distintos canales bajo el modelo de atribución *first touch*. (Elaboración Propia)

## Anexo 12

Canales	Conversiones	Porcentaje
Directo	1.059.136	89,7%
SEO	88.347	7,5%
SEM	11.958	1,0%
Email	8.951	0,8%
Otros	7.075	0,6%
RMKT	2.586	0,2%
RRSS Orgánico	1.758	0,2%
Display	380	0,0%
RRSS Pago	328	0,0%
Afiliados	258	0,0%
Tiendas	-	0,0%

Tabla 20: Contribución a la conversión de los distintos canales bajo el modelo de atribución *touch*. (Elaboración Propia)

## Anexo 13

	Directo	SEO	SEM	EmailMkt	RRSSOrganico	Remarketing	Afiliados	Otros	Fin
InicioDirecto	55,89%	1,85%	0,75%	0,37%	0,05%	0,27%	0,00%	0,32%	40,46%
InicioSEO	51,48%	0,28%	1,45%	0,45%	0,14%	0,46%	0,00%	0,50%	45,23%
InicioSEM	65,32%	1,03%	8,23%	0,00%	0,00%	0,00%	0,00%	0,00%	25,42%
InicioEmailMkt	46,59%	10,11%	0,00%	16,18%	0,00%	0,00%	0,00%	0,00%	27,12%
InicioRRSSOrganico	26,62%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	73,38%
InicioRemarketing	60,99%	8,97%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	30,04%
InicioDisplay	67,12%	12,30%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	20,59%
InicioTiendas	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
InicioRRSSPago	67,36%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	32,64%
InicioAfiliados	55,75%	0,00%	0,00%	0,00%	0,00%	0,00%	44,25%	0,00%	0,00%
InicioOtros	37,69%	11,99%	0,00%	0,00%	0,00%	0,00%	0,00%	4,12%	46,19%
DirectoDirecto	76,70%	0,45%	0,23%	0,13%	0,00%	0,10%	0,00%	0,07%	22,32%
DirectoSEO	62,89%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	37,11%
DirectoSEM	76,13%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	23,87%
DirectoEmailMkt	70,85%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	29,15%
DirectoRRSSOrganico	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
DirectoRemarketing	72,69%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	27,31%
DirectoDisplay	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
DirectoOtros	32,14%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	67,86%
SEODirecto	55,99%	0,00%	0,63%	0,00%	0,00%	0,00%	0,00%	0,34%	43,04%
SEOSEO	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
SEOSEM	66,06%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	33,94%
SEOEmailMkt	46,74%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	53,26%
SEORRSSOrganico	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
SEORemarketing	43,02%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	56,98%
SEOOtros	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
SEMDirecto	55,65%	0,00%	0,75%	0,00%	0,00%	0,00%	0,00%	0,00%	43,60%
SEMSEO	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
SEMSEM	62,99%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	37,01%
EmailMktDirecto	48,85%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	51,15%
EmailMktSEO	46,58%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	53,42%
EmailMktEmailMkt	38,78%	0,00%	0,00%	19,96%	0,00%	0,00%	0,00%	0,00%	41,26%
RRSSOrganicoDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
RemarketingDirecto	46,08%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	53,92%
RemarketingSEO	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
DisplayDirecto	39,50%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	60,50%
DisplaySEO	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
TiendasDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
RRSSPagoDirecto	34,56%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	65,44%
AfiliadosDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
AfiliadosAfiliados	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
OtrosDirecto	39,99%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	60,01%
OtrosSEO	35,60%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	64,40%
OtrosOtros	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%

Tabla 21: Matriz de probabilidades de transición completa para Markov orden 2. (Elaboración propia)

## Anexo 14

	Directo	SEO	SEM	EmailMkt	Remarketing	Otros	Fin
InicioDirectoDirecto	68,80%	1,24%	0,63%	0,36%	0,28%	0,21%	28,48%
InicioDirectoSEO	61,06%	0,00%	0,00%	0,00%	0,00%	0,00%	38,94%
InicioDirectoSEM	76,61%	0,00%	0,00%	0,00%	0,00%	0,00%	23,39%
InicioDirectoEmailMkt	72,32%	0,00%	0,00%	0,00%	0,00%	0,00%	27,68%
InicioDirectoRRSSOrganico	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioDirectoRemarketing	72,81%	0,00%	0,00%	0,00%	0,00%	0,00%	27,19%
InicioDirectoDisplay	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
InicioDirectoOtros	43,85%	0,00%	0,00%	0,00%	0,00%	0,00%	56,15%
InicioSEODirecto	55,57%	0,00%	0,78%	0,00%	0,00%	0,42%	43,23%
InicioSEOSEO	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioSEOSEM	66,06%	0,00%	0,00%	0,00%	0,00%	0,00%	33,94%
InicioSEOEmailMkt	46,74%	0,00%	0,00%	0,00%	0,00%	0,00%	53,26%
InicioSEORRSSOrganico	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioSEORemarketing	43,02%	0,00%	0,00%	0,00%	0,00%	0,00%	56,98%
InicioSEOOtros	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioSEMDirecto	56,70%	0,00%	1,34%	0,00%	0,00%	0,00%	41,96%
InicioSEMSEO	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioSEMSEM	62,99%	0,00%	0,00%	0,00%	0,00%	0,00%	37,01%
InicioEmailMktDirecto	53,93%	0,00%	0,00%	0,00%	0,00%	0,00%	46,07%
InicioEmailMktSEO	46,58%	0,00%	0,00%	0,00%	0,00%	0,00%	53,42%
InicioEmailMktEmailMkt	38,94%	0,00%	0,00%	24,93%	0,00%	0,00%	36,12%
InicioRRSSOrganicoDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioRemarketingDirecto	50,07%	0,00%	0,00%	0,00%	0,00%	0,00%	49,93%
InicioRemarketingSEO	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioDisplayDirecto	48,59%	0,00%	0,00%	0,00%	0,00%	0,00%	51,41%
InicioDisplaySEO	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioTiendasDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioRRSSPagoDirecto	34,56%	0,00%	0,00%	0,00%	0,00%	0,00%	65,44%
InicioAfiliadosDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioAfiliadosAfiliados	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
InicioOtrosDirecto	41,49%	0,00%	0,00%	0,00%	0,00%	0,00%	58,51%
InicioOtrosSEO	35,60%	0,00%	0,00%	0,00%	0,00%	0,00%	64,40%
InicioOtrosOtros	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
DirectoDirectoDirecto	79,59%	0,26%	0,14%	0,07%	0,06%	0,04%	19,84%
DirectoDirectoSEO	65,56%	0,00%	0,00%	0,00%	0,00%	0,00%	34,44%
DirectoDirectoSEM	77,37%	0,00%	0,00%	0,00%	0,00%	0,00%	22,63%
DirectoDirectoEmailMkt	69,32%	0,00%	0,00%	0,00%	0,00%	0,00%	30,68%
DirectoDirectoRemarketing	72,57%	0,00%	0,00%	0,00%	0,00%	0,00%	27,43%

DirectoDirectoOtros	18,93%	0,00%	0,00%	0,00%	0,00%	0,00%	81,07%
DirectoSEODirecto	59,52%	0,00%	0,00%	0,00%	0,00%	0,00%	40,48%
DirectoSEMDirecto	55,98%	0,00%	0,00%	0,00%	0,00%	0,00%	44,02%
DirectoEmailMktDirecto	50,29%	0,00%	0,00%	0,00%	0,00%	0,00%	49,71%
DirectoRemarketingDirecto	48,54%	0,00%	0,00%	0,00%	0,00%	0,00%	51,46%
DirectoDisplayDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
DirectoOtrosDirecto	38,25%	0,00%	0,00%	0,00%	0,00%	0,00%	61,75%
SEODirectoDirecto	60,62%	0,00%	0,00%	0,00%	0,00%	0,00%	39,38%
SEODirectoSEM	50,53%	0,00%	0,00%	0,00%	0,00%	0,00%	49,47%
SEODirectoOtros	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
SEOSEMDirecto	46,91%	0,00%	0,00%	0,00%	0,00%	0,00%	53,09%
SEOEmailMktDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
SEORemarketingDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
SEMDirectoDirecto	55,68%	0,00%	0,00%	0,00%	0,00%	0,00%	44,32%
SEMDirectoSEM	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
SEMSEMDirecto	49,72%	0,00%	0,00%	0,00%	0,00%	0,00%	50,28%
EmailMktDirectoDirecto	48,90%	0,00%	0,00%	0,00%	0,00%	0,00%	51,10%
EmailMktSEODirecto	35,57%	0,00%	0,00%	0,00%	0,00%	0,00%	64,43%
EmailMktEmailMktDirecto	25,66%	0,00%	0,00%	0,00%	0,00%	0,00%	74,34%
EmailMktEmailMktEmailMk	38,14%	0,00%	0,00%	0,00%	0,00%	0,00%	61,86%
RemarketingDirectoDirecto	41,36%	0,00%	0,00%	0,00%	0,00%	0,00%	58,64%
DisplayDirectoDirecto	39,20%	0,00%	0,00%	0,00%	0,00%	0,00%	60,80%
RRSSPagoDirectoDirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
OtrosDirectoDirecto	32,43%	0,00%	0,00%	0,00%	0,00%	0,00%	67,57%
OtrosSEODirecto	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%

Tabla 22: Matriz de probabilidades de transición completa para Markov orden 3. (Elaboración propia)

## Anexo 15

	Visitas	Ventas
Directo	32,4%	30,4%
SEO	27,1%	14,3%
None	20,3%	27,6%
SEM	9,1%	1,3%
Email	3,8%	0,7%
RRSS Pago	3,4%	0,1%
Tiendas	3,3%	0,0%
RRSS Orgánico	3,2%	0,4%
Display	1,3%	0,0%
Call Center	0,5%	25,0%
RMKT Dinámico	0,3%	0,1%
Afiliados	0,1%	0,0%

Tabla 23: Distribución de las visitas y ventas bajo el modelo de atribución *Last Touch* según Omniture para el periodo de Enero a Julio 2017. (Elaboración Propia)

## Anexo 16

Dispositivo	Tasa de Conversión
Mobile Phone	0,6%
Desktop	2,7%
Tablet	1,7%
Media Player	0,3%
Gaming Console	0,4%
Televisión	0,4%
Set-top Box	0,2%
E Reader	0,0%

Tabla 24: Tasa de conversión por tipo de dispositivo bajo el modelo de atribución *Last Touch* según Omniture para el periodo de Enero a Julio 2017. (Elaboración Propia)

## Anexo 17

A	B	C	D	E	F	G	H	I	J	K	L
Order Timestam	Pixel ID	App ID	Ad Account I	Campaign ID	Adset ID	Ad ID	Action Type	Impression C	Click Timesta	Placement	Device
2017-03-31T00:	5045318219	5,44E+14	view	1,02E+16	2,38E+16	2,38E+16	2,38E+16	Impression	2017-03-30T	MobileFeed	iOS
2017-03-31T00:	5045318219	5,44E+14	view	1,02E+16	2,38E+16	2,38E+16	2,38E+16	Impression	2017-03-30T	MobileFeed	iOS
2017-03-31T00:	5045318332	5,44E+14	view	1,02E+16	2,38E+16	2,38E+16	2,38E+16	Impression	2017-03-30T	Others	Android
2017-03-31T00:	5045318257	5,44E+14	view	1,02E+16	2,38E+16	2,38E+16	2,38E+16	Impression	2017-03-30T	Others	Android
2017-03-31T00:	5045318257	5,44E+14	view	1,38E+15	6,06E+12	6,06E+12	6,06E+12	Impression	2017-03-30T	MobileFeed	Android

Figura 20: Información que entrega el Order ID de Facebook

## Anexo 18

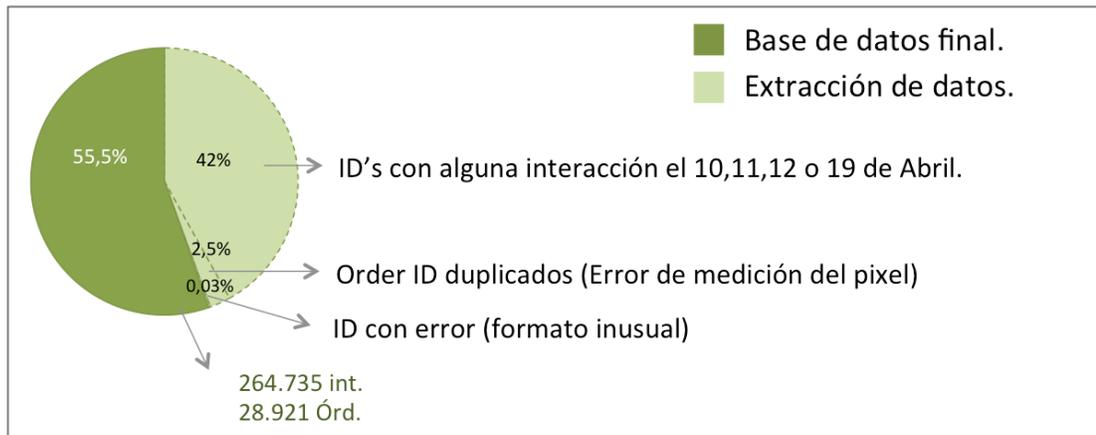


Figura 21: Pre-procesamiento de datos para Análisis de Facebook. (Elaboración Propia)

## Anexo 19

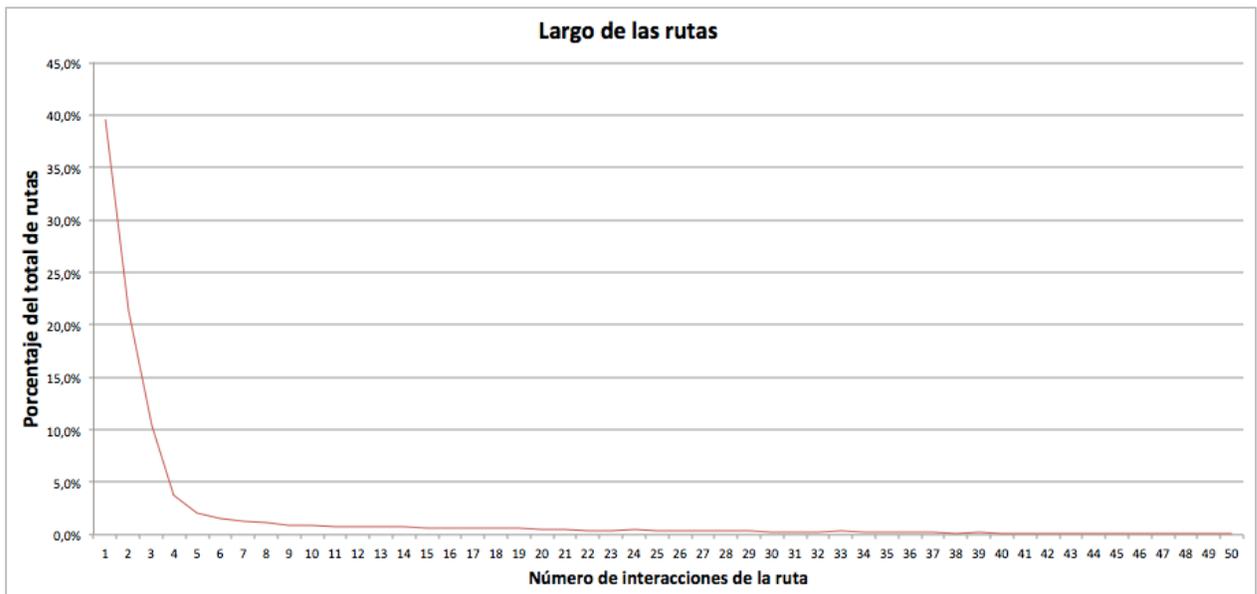


Figura 22: Distribución de los largos de las rutas en la base de datos de Facebook. (Elaboración Propia)

## Anexo 20

Ruta	Frecuencia	Porcentaje
Mobile	134	33,5%
Mobile; Mobile	81	20,3%
Desktop	27	6,8%
Mobile; Mobile; Mobile	21	5,3%
Desktop; Desktop	9	2,3%
Desktop; Desktop; Desktop	7	1,8%
Mobile; Mobile; Mobile; Mobile	7	1,8%
Desktop; Desktop; Desktop; Desktop; Desktop	5	1,3%
Desktop; Mobile	6	1,5%
Mobile; Mobile; Desktop	5	1,3%
Mobile; Mobile	3	0,8%
Desktop; Desktop; Desktop; Desktop; Desktop; Desktop	3	0,8%
Mobile; Mobile	3	0,8%
Mobile; Desktop; Desktop; Desktop; Desktop	2	0,5%
Desktop; Desktop; Desktop; Desktop; Desktop; Desktop; Desktop; Desktop	2	0,5%
Mobile; Mobile	2	0,5%
Mobile; Desktop; Desktop; Desktop	2	0,5%
Mobile; Mobile; Mobile; Mobile; Mobile; Mobile; Mobile; Mobile; Mobile	2	0,5%
Mobile; Mobile	2	0,5%
Mobile; Desktop	3	0,8%

Tabla 25: Rutas indicando dispositivos empleados en ellas. (Elaboración Propia)

## Anexo 21

	Mobile	Desktop
Mobile	64,6%	5,3%
Desktop	3,9%	26,2%

Tabla 26: Probabilidades de transición entre *mobile* y *desktop* para las rutas de la muestra de Facebook. (Elaboración Propia)