



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

ANÁLISIS DE LOS FACTORES DETERMINANTES DE LA EFECTIVIDAD DE
PROMOCIONES GEO LOCALIZADAS A NIVEL INDIVIDUAL A TRAVÉS
DE SMARTPHONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

DANIEL EDUARDO OLCAY ESPINOZA

PROFESOR GUÍA
MARCEL GOIC FIGUEROA

MIEMBROS DE LA COMISIÓN
TODD PEZZUTI
JOSÉ GUAJARDO ANDRADES

SANTIAGO DE CHILE
2016

ANÁLISIS DE LOS FACTORES DETERMINANTES DE LA EFECTIVIDAD DE PROMOCIONES GEO LOCALIZADAS A NIVEL INDIVIDUAL A TRAVÉS DE SMARTPHONES

Chile es uno de los países con mayor crecimiento de dispositivos móviles, pasando de ser un 52% en marzo del 2014 a 64% de conexiones a marzo del 2015. Más aun, informes de la subsecretaría de telecomunicaciones indican que cerca del 77% de las conexiones totales a internet del país se realizan efectivamente a través de teléfonos celulares. Lo anterior configura un escenario propicio para que las compañías busquen este canal de comunicación para interactuar de manera más directa con los usuarios y, para de esta forma conseguir un mayor alcance y presencia en el mercado nacional.

Izit es una compañía pionera en el país, cuya competencia central es el envío de promociones exclusivas en el *lugar adecuado y en el momento justo*. Haciendo unos de tecnologías GPS/Bluetooth, la empresa busca disminuir la percepción de invasividad del general de las empresas de avisaje enviando estímulos basados en la localización de los usuarios.

La interacción del usuario con la aplicación pasa por 4 hitos principales: envío, apertura, aceptación y canje de la promoción. Sin embargo, la compañía no posee claridad respecto al comportamiento de los usuarios en las distintas fases de decisión de estos, por lo que el presente estudio tiene como propósito el determinar qué factores moderan la apertura de envíos y en qué medida lo hacen.

Para el desarrollo del proyecto se utilizó un enfoque de modelos de elección discreta, en particular los modelos logit heterogéneos y jerárquicos bayesianos. Dichos modelos consideran, en primera instancia, solo información relacionada a la bandeja de entrada de los usuarios, para luego incluir el efecto de la distancia de estos a los *fences* al recibir una promoción.

Las principales conclusiones del estudio apuntan a que uno de los factores más importantes en cuanto a la apertura de promociones es el historial de envíos y aperturas que el usuario “almacena” a lo largo de su interacción con la aplicación, además de la importancia que tienen los mensajes asociados al club de fidelización de la compañía. El estudio logra identificar dos clases latentes: un segmento leal que prefiere envíos asociados a centros comerciales con poca sensibilidad ante la acumulación de envíos/aperturas previas y otro con mayor sensibilidad ante las métricas anteriores que prefieren recibir estímulos asociados a centros nocturnos y lugares de dispersión.

Finalmente, un estudio aproximado de la distancia entre los usuarios y los *fences* revela que existe heterogeneidad en cuanto a las preferencias de los usuarios por dicha métrica, existiendo usuarios que no son del todo sensible a la geo localización.

Tabla de Contenido

1.	Introducción.....	8
1.1.	Location Based Advertising.....	8
1.2.	Izit.....	9
2.	Descripción del estudio	12
2.1.	Planteamiento del problema y justificación.....	12
2.2.	Revisión bibliográfica.....	13
2.3.	Objetivos.....	14
2.3.1.	Objetivo general	14
2.3.2.	Objetivos específicos	14
2.4.	Alcances.....	14
3.	Procesamiento de datos	16
3.1.	Consolidación de datos	16
3.1.1.	Inbox.....	16
3.1.2.	Users.....	16
3.1.3.	Whispers.....	16
3.1.4.	Categorías	16
3.1.5.	Campaigns	17
3.1.6.	Fences	17
3.1.7.	User Position	17
3.1.8.	Brand	17
3.1.9.	User Whispers Gifts.....	17
3.2.	Pre-procesamiento de datos	18
3.2.1.	Creación de variables.....	19
	Edad.....	19
	Tiempo desde el registro	20
	Canje de <i>bananas</i>	20
	Envíos, apertura y aceptación histórica.....	20
	Envíos y aperturas acumuladas	20
3.3.	Filtro de datos	20
4.	Análisis descriptivo de datos	24
4.1.	Métricas generales	24
4.2.	Usuarios	24
4.2.1.	Género	25
4.2.2.	Edad.....	25
4.2.3.	Contador de bananas.....	27
4.2.4.	Tiempo desde registro	28
4.2.5.	Envíos y aperturas históricas	29
4.3.	Promociones	29
4.4.	Fences	34
4.5.	Acumulación de envíos y aperturas	37

5. Marco Teórico	40
5.1. Modelo logit.....	40
5.2. Modelo logit con heterogeneidad discreta	40
5.3. Modelo mixed logit.....	41
5.4. Modelo mixed logit jerárquico bayesiano.....	41
6. Desarrollo metodológico	43
6.1. Modelo logit simple	44
6.1.1. Modelo logit considerando envíos y aperturas acumuladas	44
6.2. Modelo logit con heterogeneidad discreta	45
6.2.1. Modelo logit con dos clases latentes	45
6.3. Modelo mixed logit jerárquico bayesiano.....	47
7. Desarrollo metodológico – Inclusión de medida de distancia.....	52
7.1. Modelo logit simple	54
7.2. Modelo con heterogeneidad discreta	55
7.3. Modelo mixed logit.....	55
7.4. Modelo mixed logit jerárquico bayesiano.....	56
8. Discusión y recomendaciones	59
8.1. Modelo logit con clases latentes (sin distancia).....	59
8.2. Modelo jerárquico bayesiano.....	62
8.3. Modelo jerárquico bayesiano (incluida variable distancia)	67
9. Conclusiones y trabajos futuros.....	70
9.1. Conclusiones.....	70
9.2. Trabajos futuros	71
10. Bibliografía.....	73
11. Anexos	75

Índice de Tablas

Tabla 1 - Lista de intereses desplegados por la aplicación.....	9
Tabla 2 - Registros considerados en inbox.....	19
Tabla 3 - Tablas mezcladas con inbox	19
Tabla 4- Filtros inbox 1	21
Tabla 5 – Filtros inbox 2	23
Tabla 6 - Radio de fences	35
Tabla 7 - Marcas por fences	37
Tabla 8 - Resultados modelo logit con envios y apertura acumulada	44
Tabla 9 - Resultados modelo 2 clases latentes	46
Tabla 10 - Valores promedio Δ	50
Tabla 11 - Resultados modelo logit con distancia.....	55
Tabla 12 - Valores promedio Δ (distancia)	58
Tabla 13 - Caracterización de clases latentes	59
Tabla 14 - Variables demográficas promedio por segmento.....	60
Tabla 15 - Variables demográficas promedio por segmento (cluster jerárquico)	65
Tabla 16 - Resultados logit considerando indicadores históricos escalados [anexos]	81
Tabla 17 - Resultados modelo logit con indicadores acumulados escalados [anexos]	82
Tabla 18 - Desviación estándar Δ [anexos]	82
Tabla 19 - Promedio preferencias de usuarios (jerárquico) [anexos].....	82
Tabla 20 - Resultados modelo logit 3 clases latentes [anexos]	83
Tabla 21 - Resultados modelo logit 2 clases latentes (distancia) [anexos]	83
Tabla 22 - Resultados modelo mixed logit (distancia) [anexos]	84
Tabla 23 - Promedio preferencias usuarios (jerárquico-distancia) [anexos]	86

Índice de Ecuaciones

Ecuación 1 - Criterio de filtro en tasas de apertura	22
Ecuación 2: Probabilidad de elección logit simple.....	40
Ecuación 3: Probabilidad de elección modelo clases latentes	41
Ecuación 4: Probabilidad de elección - mixed logit	41
Ecuación 9: Probabilidad de elección condicional	41
Ecuación 10: Distribución de coeficientes aleatorios modelo jerárquico.....	42
Ecuación 11: Distribución posterior de hiperparámetros	42
Ecuación 5 - Especificación general utilidad modelo logit	43
Ecuación 6 - Especificación de utilidad modelo logit considerando comportamiento acumulado	44
Ecuación 7 - Probabilidad de apertura modelo logit considerando comportamiento acumulado .	44
Ecuación 8 - Especificación utilidad clases latentes	45
Ecuación 12 - Especificación de utilidad modelo jerárquico	47
Ecuación 13 - Jerarquía en modelo mixed logit	47
Ecuación 14 - Especificación modelo logit (distancia)	54

Índice de Ilustraciones

Ilustración 1 - Bandeja de entrada aplicación.....	10
Ilustración 2 - Mapa de fences en Santiago	11
Ilustración 3 - Diagrama tablas IZIT	18
Ilustración 4 - Densidad tasa de apertura por usuarios	22
Ilustración 5 - Envíos por usuarios	23
Ilustración 6 - Métricas agregadas.....	24
Ilustración 7 – Apertura y envíos por género	25
Ilustración 8 - Histograma de edades usuarios	26
Ilustración 9 - Apertura promedio por edad	27
Ilustración 10 - Apertura promedio por contador de bananas	28
Ilustración 11 - Apertura promedio por tiempo desde registro	29
Ilustración 12 - Apertura y envíos por marca.....	31
Ilustración 13 - Apertura promedio por marca	32
Ilustración 14 - Apertura y envíos por categoría	32
Ilustración 15 - Apertura promedio por categoría	33
Ilustración 16 - Apertura promedio por categoría banana.....	34
Ilustración 17 - Muestra de fences considerados.....	35
Ilustración 18 - Apertura promedio por fence	36
Ilustración 19 - Apertura promedio por stock de envíos	38
Ilustración 20 – Apertura promedio por stock de aperturas	39
Ilustración 22 - Promedio de posterior de coeficientes	49
Ilustración 23 - Ejemplo de aproximación de distancia	52
Ilustración 24 - Distribución de distancias por fence	53
Ilustración 25 - Promedio de aperturas por distancia	54
Ilustración 26 - Promedio de posterior de coeficientes (distancia)	57
Ilustración 27 - Comparación envíos/aperturas históricas por segmentos	61
Ilustración 28 - Comparación envíos/aperturas acumuladas por segmento (clases latentes).....	61
Ilustración 29 - Tasas de apertura por categoría y segmentos.....	62
Ilustración 30 - Distribución de envíos acumulados (muestra clientes).....	63
Ilustración 31 - Distribución acumulación de envío y aperturas (muestra usuarios)	64
Ilustración 33 - Comparación envíos/aperturas históricas por segmentos (cluster jerárquico).....	65
Ilustración 34 - Comparación envíos/aperturas acumuladas por segmentos (cluster jerárquico) .	66
Ilustración 35 - Comparación envíos/aperturas acumuladas por segmentos - 2 clases (cluster jerárquico).....	67
Ilustración 36 - Distribución de distancia (jerárquico) (muestra clientes)	68
Ilustración 37 - Distribución de variables modelo con distancia (jerárquico) (muestra usuarios)	68
Ilustración 38 - Test medias indicador de apertura y género [anexos].....	77
Ilustración 39 - Test medias indicador de apertura y fences [anexos].....	77
Ilustración 40 - Box plot envíos acumulados vs contador de bananas [anexos]	77
Ilustración 41 - Box plot aperturas acumulados vs contador de bananas [anexos]	78
Ilustración 42 - Box plot envíos acumulados vs tiempo desde registro [anexos].....	78
Ilustración 43 - Box plot aperturas acumuladas vs tiempo desde registro [anexos].....	79
Ilustración 44 - Apertura promedio vs envíos históricos [anexos].....	80

Ilustración 45 - Test medias indicador de apertura y bananas [anexos].....	80
Ilustración 46 - Apertura promedio vs apertura histórica.....	81
Ilustración 47 - Evolución log verosimilitud modelo jerárquico (sin distancia) [anexos]	84
Ilustración 48 - Comparación tiempo desde registro entre segmentos[anexos]	85
Ilustración 49 - Comparación contador de bananas entre segmentos[anexos].....	85
Ilustración 50 - Distribución de apertura acumulada (muestra clientes) [anexos]	86
Ilustración 51 - Distribución fence Metro Baquedano (muestra clientes) [anexos]	87
Ilustración 52 - Distribución fence Metro Tobalaba (muestra clientes) [anexos]	87
Ilustración 53 - Distribución fence Plaza Vespucio (muestra clientes) [anexos]	88
Ilustración 54 - Distribución fence Patio Bellavista (muestra clientes) [anexos].....	88
Ilustración 55 - Representación gráfica de clusters (jerárquico) [anexos]	89
Ilustración 56 - Comparación tiempo desde registro entre segmentos (jerárquico)[anexos]	89
Ilustración 57 - Comparación contador de bananas entre segmentos (jerárquico) [anexos].....	90
Ilustración 58 - Representación gráfica de clusters 2 clases (jerárquico) [anexos].....	90

1. Introducción

1.1. Location Based Advertising

El uso de smartphones se ha masificado a nivel mundial y Chile no es la excepción, siendo uno de los países con mayor penetración de dispositivos celulares e internet móvil en la región. Según el estudio realizado por Pew Research Center en el año 2014 [1], Chile se encuentra entre las naciones emergentes con mayor uso de internet y teléfonos móviles, alcanzando un 91% de penetración en el país.

Lo anterior abre valiosas oportunidades en cuanto al uso de información generada por dispositivos móviles, entre las cuales se encuentra el uso de Location Based Services (servicios basados en localización). En primera instancia, un LBS (por sus siglas en inglés) corresponden a servicios que usan información respecto a la posición de los usuarios para entregar una oferta de valor determinada. Así, servicios tales como Waze, Foursquare y Google Maps corresponden a servicios basados en localización.

En específico, y a partir de los LBS, nace el concepto de Location Based Advertising (LBA), cuya propuesta de valor de valor consiste en entregar servicios publicitarios/comerciales a usuarios, basándose en la posición de éstos, por medio de dispositivos móviles

LBA es una nueva herramienta para las empresas, la cual se configura como una manera innovadora de interactuar de manera individual con los clientes, y por ende, es imperante el comprender cómo los usuarios se comportan ante los estímulos enviados. Por un lado, los usuarios podrían percibir grandes beneficios en el envío de información comercial en sus dispositivos móviles, pero también dichos mensajes pueden ser considerados intrusivos y molestos.

La gran diferencia que se vislumbra entre estos servicios y herramientas tradicionales de comunicación electrónica es que estos últimos tienen un mayor riesgo de generar reacciones negativas por parte de los usuarios debido a la ausencia de oportunidad y contexto en el mensaje. Así, el núcleo de los LBA es enviar el anuncio indicado, en el momento y lugar indicado, pues de esta forma la publicidad enviada se alinea dinámicamente con los intereses del usuario. A modo de ejemplo, el enviar una publicidad de zapatos a un potencial cliente que se encuentra en un centro comercial será mucho más efectivo en comparación al envío del mismo anuncio mientras se encuentra en su casa en familia. Lo anterior refleja la relevancia de la modalidad LBA en comparación con los canales usuales de marketing.

Finalmente, las proyecciones económicas para el avisaje mediante teléfonos móviles son bastante auspiciosas, vaticinando gastos del orden de 100 billones de dólares para el año 2016, abarcando el 51% del mercado digital y siendo Estados Unidos y China los principales países en cuanto a su participación en la escena móvil [2]. Esto último implica que esta nueva plataforma no solo abre un mundo de posibilidades para las firmas en relación a la forma de realizar publicidad,

sino que también puede significar un negocio altamente lucrativo debido a su rápido crecimiento y alto valor.

1.2. Izit

El proyecto se enmarca en la empresa Izit, desarrolladora de una aplicación para teléfonos inteligentes que envía promociones dependiendo de las preferencias declaradas por el usuario al descargar la aplicación. Actualmente, cuenta con alrededor de 270 mil usuarios registrados y más de 2000 campañas realizadas para cerca de 300.000 clientes. Izit se define como un canal promocional cuyas actividades claves son: micro-segmentación, generación de tráfico, presencia de marca, fidelización, experiencia de usuario y aumento de probabilidad de conversión [3]. Su propuesta de valor se basa en 4 pilares fundamentales que apuntan a la oferta de beneficios exclusivos y personalizados, envío de promociones sólo cuando se esté cerca de una tienda, funcionamiento completamente autónomo que requiere sólo del registro del usuario, mantener activado el GPS del teléfono móvil y seleccionar las categorías preferidas al descargar el programa [4].

Al descargar la aplicación se solicita al usuario seleccionar sus intereses favoritos de un total de 34 posibles (es posible también seleccionar todos los intereses de manera automática), lo cual tiene como objetivo perseguir el primer foco de la propuesta de valor de la compañía, esto es, el envío de promociones personalizadas. Más aún, lo anterior promueve el aumento del valor percibido por los usuarios frente a las promociones enviadas, lo cual apunta a disminuir la posibilidad de percepción de *spam* por parte de los usuarios de la aplicación.

Intereses			
Alimentos	Casinos y Juegos de Azar	Infantil	Restaurantes
Aseo y Limpieza	Concursos	Libros, artes y joyas	Salud y Cuidado Personal
Audio	Decohogar	Maletería	Tecnología
Automotriz	Deportes y Outdoor	Mascotas	Telefonía
Bancos	Electrohogar	Moda Hombre	TV y Video
Bebé	Entretenimiento	Moda Mujer	Viajes
Bebestibles	Fotografía	Música	VideoJuegos
Belleza	Hogar	Regalos	Vinos y Licores

Tabla 1 - Lista de intereses desplegados por la aplicación

El funcionamiento de la aplicación resulta intuitivo debido a la simpleza de interfaces que despliega. Una vez descargada, la aplicación despliega 4 interfaces principales: promociones aceptadas, promociones enviadas y que aún no se aceptan, lista de amigos en Facebook que poseen la aplicación y una interfaz de ajustes (la cual permite entre otras funciones, editar las preferencias mencionadas anteriormente en cualquier momento).

En la Ilustración 1 se presenta la bandeja de entrada de la aplicación, en donde se puede apreciar la interfaz que muestra los mensajes recibidos y aún no aceptados.



Ilustración 1 - Bandeja de entrada aplicación

Una vez instalada y encendida la aplicación, el usuario está habilitado para comenzar a recibir promociones o *whispers*. Un *whisper* se define como un mensaje promocional enviado por la aplicación y a medida que se van enviando a la bandeja de entrada del usuario, se apilan desde el más actual al más antiguo y desaparecen gradualmente en caso de expirar.

La interacción entre el usuario y cada promoción enviada por la compañía pasa por cuatro hitos, de los cuales tres implican una decisión directa por parte del usuario que recibe el mensaje:

- **Envío de promoción:** Esta decisión es tomada por parte de la aplicación y depende de las preferencias declaradas por el usuario y su cercanía a la tienda que la ofrece y las condiciones indicadas por el avisador como el segmento etario al que desea apuntar o el sexo target. Un envío se puede realizar de cuatro formas:
 - **Envío geo localizado:** El envío del mensaje se realiza en función de la ubicación del cliente y para ello se define un área geográfica llamada *fence* tal que, cuando la aplicación identifica que un cliente ingresa a dicha área se gatilla un mensaje. La identificación de clientes se realiza tanto por tecnología GPS como triangulación de antenas WIFI, permitiendo la identificación de la localización de clientes con una precisión de 200 o 300 m. Adicionalmente, existen *superfences* que a través de señales de sonido o *bluetooth* permite identificar la ubicación de clientes con una mayor precisión que típicamente bordea los treinta metros. Estos se utilizan para iluminar sectores más reducidos, como pasillos de malls, entradas de tiendas, etc. La Ilustración 2 muestra la distribución de *fences* de la compañía a lo largo de Santiago, mostrando que la gran mayoría se encuentran en el sector centro/oriente de la capital.

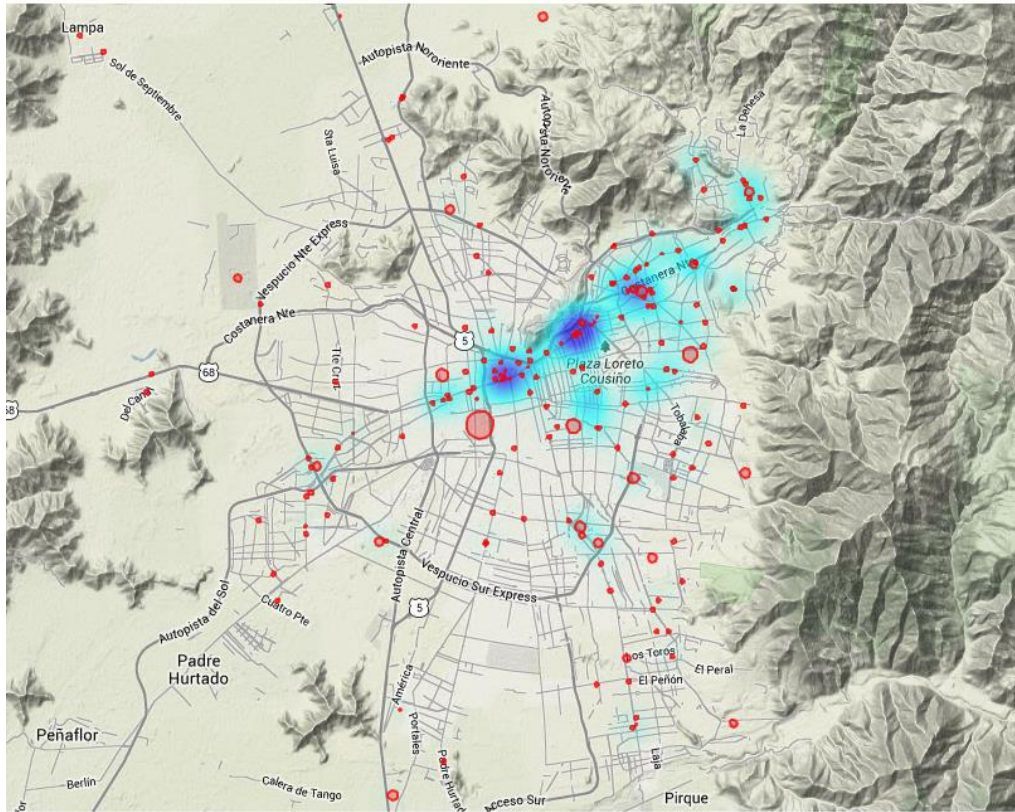


Ilustración 2 - Mapa de fences en Santiago

- **Envío no geo localizado:** En este caso, el envío no considera la ubicación del usuario y solo toma en cuenta la segmentación del cliente y las preferencias declaradas por el usuario. Si bien la propuesta de valor de la compañía es la relevancia de los estímulos promocionales, el envío no geo localizado surge como una alternativa para aumentar el alcance de las promociones, de modo de llegar a un público más masivo.
- **Mensajes de Facebook:** Algunas promociones están habilitadas para realizar envíos solo a los usuarios que hayan puesto *me gusta* en el mensaje alusivo a la promoción en el muro de Izit. De igual manera, la compañía habilita con cierta regularidad promociones disponibles en el *fanpage* de la aplicación en dicha red social. Este tipo de envío no toma en cuenta la ubicación del usuario ni la segmentación o preferencias indicadas al descargar la aplicación.
- **Apertura de la promoción:** Una vez que el usuario recibe la promoción, aparecerá una notificación en su teléfono con el fin de dar aviso y entregar una descripción de la promoción y las condiciones de canje. En este punto, el usuario debe decidir si entra a la aplicación y abre el mensaje para saber más detalles sobre la promoción o si lo deja pasar.
- **Aceptación de la promoción:** Una vez abierto el mensaje, el usuario se enfrentará una pantalla con detalles de la promoción, un botón para aceptarla y el tiempo que resta para que la

promoción expire. Si una promoción expira, puede ser revisada si es que aún no se ha abierto, pero no puede ser aceptada.

- **Uso de la promoción:** Finalmente, luego de aceptar la promoción, se puede hacer uso de esta en la tienda asociada. La aplicación entregará un código que se deberá entregar al dependiente de la tienda al momento de hacer uso de la promoción.
- **Reenvío:** Una cierta promoción recibida puede ser reenviada a la lista de amigos de un usuario en su red social. En este caso, la ubicación del usuario no es un requisito para poder reenviar una promoción determinada.

La interacción de los usuarios con los reenvíos se encuentra moderada en cierta parte por el programa de fidelización de Izit asociado a *bananas*, las cuales al acumularse pueden canjearse por distintos beneficios para los usuarios.

Cabe destacar que, por diseño de la base transaccional, todo envío debe estar asociado a un *fence*, y por tanto las promociones que no son geo localizadas, son asignadas a un *fence* ficticio llamado *Izitland*.

2. Descripción del estudio

2.1. Planteamiento del problema y justificación

Actualmente, la tasa de apertura y aceptación¹ agregadas de las promociones enviadas por Izit bordean el 28% y el 8% respectivamente, por lo que un entendimiento del comportamiento de los clientes mediante herramientas econométricas y analíticas puede mejorar de manera considerable estas métricas. Así, el objetivo de la investigación propuesta es evaluar cuales son los factores que moderan en mayor medida la decisión de los usuarios de responder positivamente a las promociones enviadas por la compañía.

En un trabajo anterior se analizó la efectividad de las promociones nivel agregado, concluyendo que a nivel de envíos la geo localización tiene un rol clave, determinando el alcance de la promoción [5]. Sin embargo, el estudio concluye que la apertura de éstas no se encuentra moderada en mayor medida por la geo localización y, más aún, en la fase de aceptación la información de localización no posee un efecto significativo en las tasas agregadas.

Si bien el trabajo descrito anteriormente permite explicar el rendimiento de las promociones a nivel agregado, no permite a la compañía el tomar decisiones a nivel de usuarios, obviando una serie de factores que determinan el rendimiento de una promoción como lo son: variables demográficas, preferencias individuales de cada usuario con una cierta categoría de promoción,

¹ Condicional a la apertura

lealtad del usuario frente a ciertas categorías, aprendizaje o stock de mensajes recibidos/abiertos/aceptados por categoría, entre otros.

La información recopilada por la aplicación permite realizar un estudio individual respecto al comportamiento de los usuarios en las distintas fases de interacción con la compañía, de manera de comprender el actuar de éstos y las motivaciones que tienen para responder positivamente a los estímulos enviados por Izit.

Así, el realizar un análisis a nivel individual permite comprender e identificar las variables que promueven y moderan el uso de la aplicación en las distintas fases de decisión del cliente, por lo que da la oportunidad a Izit de tomar decisiones de manera más informada. De esta forma, la compañía podría enviar promociones personalizadas a cada cliente (o segmento de clientes), considerando no solo los intereses declarados por éste, sino también considerando qué tan propenso es a responder positivamente ante el estímulo enviado, por lo que la investigación propuesta posee potencial valor en la ayuda en la toma de decisiones de la compañía.

Con los datos disponibles se estimará la probabilidad individual que un cliente abra una promoción determinada, controlando por distintos factores, entre ellos demográficos como el sexo, edad y nivel de actividad histórica con la aplicación, como también factores dinámicos como lo son el stock de envíos y aperturas acumulados a través del tiempo. Lo anterior busca comprender cuáles son las determinantes de la dinámica de apertura de los usuarios, de modo de posteriormente tomar decisiones en base al conocimiento obtenido.

Una de las propuestas de valor de la compañía en lo que al avisaje respecta es el uso de tecnología GPS/Bluetooth para localizar la posición del usuario y así enviar estímulos pertinentes. En ese sentido, una pregunta de investigación válida es evaluar en qué medida dicha característica modera las tasas de apertura de los clientes de Izit.

Así, en fases posteriores de la investigación, se incluirá la distancia de los usuarios al *fence* que despliega la promoción para determinar la efectividad de la distancia en cuanto a tasas de apertura respecta.

2.2. Revisión bibliográfica

Respecto a la literatura relacionada, el impacto de los servicios basados en la localización de los individuos se ha estudiado bajo distintas aristas. Así, por ejemplo, Banerjee y Dholakia [6] muestran que los consumidores desarrollan mayor intención de compra al interactuar con campañas publicitarias cercanas (en comparación con las más lejanas). Spiekermann et al. [7], en el contexto del canje de cupones promocionados de manera directa, concluye que la probabilidad de canje aumenta cuando el usuario recibe el cupón cerca de la tienda física y en la misma línea, Ghose et al. [8] determina que la distancia entre un usuario y la tienda física aumenta la cantidad de clicks en el canal web. Finalmente, y en un escenario muy distinto, Andrews et al. [9] concluye que, los individuos son más propensos a responder de manera positiva una campaña publicitaria móvil mientras mayor es el hacinamiento en espacios reducidos.

Sin embargo, todos los estudios anteriormente mencionados responden a la modalidad de envíos mediante SMS (Short Message Service), por lo que no es posible analizar una serie de

conductas que en el contexto de IZIT si es posible, como lo son el despliegue de información progresivo y la posibilidad de múltiples contactos del cliente con la aplicación.

La principal diferencia del estudio propuesto respecto a la literatura es que, los datos provenientes de la compañía responden a una aplicación real con clientes recibiendo múltiples estímulos que permiten analizar efectos dinámicos, mientras que los estudios anteriormente mencionados corresponden a estudios experimentales de campo en donde se simulan situaciones hipotéticas para intentar extraer conocimiento respecto al comportamiento de los usuarios.

2.3. Objetivos

2.3.1. Objetivo general

Estimar el comportamiento de clientes frente a promociones usando información a nivel individual.

2.3.2. Objetivos específicos

- Determinar la probabilidad que, un cliente abra una promoción.
- Identificar los factores con mayor influencia en la apertura de promociones a nivel individual.
- Describir la heterogeneidad de los clientes en su interacción con la aplicación.
- Generar propuestas de mejora a la compañía en base a la situación modelada.

2.4. Alcances

Respecto a los alcances del proyecto, solo se considerará el análisis de tasas de apertura de promociones enviadas por la compañía. Esto responde principalmente a dos razones:

- En primera instancia, los datos disponibles respecto a aceptación de promociones y reenvío de estas están poco poblados, lo cual dificulta la estimación econométrica y no permite identificar los efectos de interés.
- Adicional a lo anterior y dado que se desea medir el efecto de la distancia de los usuarios al *fence* como moderadora de la tasa de apertura de estos, el estudio realizado por Toledo (2015) evidencia que la geo localización no posee efecto significativo en fases posteriores a la apertura de promociones.

Sobre los datos disponibles, para la estimación de los modelos que se propondrán solo se considerarán los 5 *fences* que posean mayor cantidad de envíos (excluyendo *Izitland*). Esto debido a que existe una gran cantidad de *fences* con poca actividad que dificultan la estimación. Dichos *fences* son Costanera Center, Plaza Vespucio, Metro Baquedano y Patio Bellavista.

Finalmente, no se considerará la fase de canje de promociones debido a que la información relativa es deficiente y poco fidedigna. Además, dado que IZIT centra su interés en las etapas de

apertura y aceptación, el estudio responde a la necesidad imperante de la compañía por entender los factores que moderan las fases tempranas de la interacción de los usuarios con la aplicación.

3. Procesamiento de datos

3.1. Consolidación de datos

Los datos usados para el desarrollo del estudio provienen de información relativa a la bandeja de entrada de 173.712 usuarios de la aplicación a lo largo de 3.400.000 observaciones entre las fechas 01/01/2015 y 20/06/2015 (previo a ningún filtro).

La información entregada por la compañía se despliega en 9 tablas las cuales se describen a continuación.

3.1.1. Inbox

Es la tabla más importante que maneja la compañía, puesto que contiene la información de la bandeja de entrada de todos los usuarios que posee la aplicación. Es en dicha tabla donde se almacena la actividad detallada de cada usuario, indicando qué mensaje recibió, la fecha en la que lo recibió e indicadores binarios de acuso de apertura, aceptación, reenvío y recibimiento mediante reenvío.

3.1.2. Users

En esta tabla se almacena información relativa al usuario y su actividad reciente con la compañía (al momento de realizar la consulta). Contiene el género del usuario, tipo de dispositivo, contador de *bananas* a la fecha, primera fecha de inicio de sesión, entre otras. Las *bananas* corresponden a un club de fidelización de la compañía, el cual premia con puntos (*bananas*) a los usuarios que más interacción tienen con la aplicación.

3.1.3. Whispers

Tabla que contiene información respecto a los *whispers* o envíos que realiza la compañía. En esta tabla se despliega la información del *whisper* que se visualiza en la bandeja de entrada del usuario al momento de recibir un envío, como por ejemplo el encabezado y el contenido o mensajes específico que se muestra al usuario tanto al momento de recibir como abrir la promoción subyacente.

3.1.4. Categorías

A partir de la tabla de *whispers* se generó la tabla que a cada envío le asocia una categoría predefinida. Las categorías consideradas son las siguientes:

- **dxu**: Envíos asociados a promociones 2x1, esto es, por la compra del producto asociado a la promoción, el usuario se lleva 2 ejemplares de este.
- **reb**: Envíos asociados a promociones que poseen rebajas en monto monetario (en pesos chilenos).
- **ptge**: Envíos asociados a promociones que poseen descuentos porcentuales.
- **conc**: Promociones asociadas a concursos.
- **reg**: Envíos asociados a promociones que regalan algún producto/servicio por el solo hecho de aceptar el *whisper*.

- **aviso:** Mensajes de aviso enviados por la compañía a los usuarios.
- **aviso_ban:** Avisos relacionados al club de fidelización de la compañía.
- **canje_ban:** Envíos que promueven el canje de *bananas* mediante la aceptación del *whisper*.

Notar que se realiza una distinción entre los envíos asociados a rebajas en monto y a rebajas porcentuales, pues la literatura sugiere que los usuarios perciben de manera distinta estos formatos de promoción [10]. Además, se cree que la inclusión del programa de fidelización de la compañía tiene un efecto distintivo en cuanto al proceso de apertura de los usuarios, pues éstos podrían seguir motivaciones distintas al momento de interactuar con estos envíos en comparación con una promoción usual.

3.1.5. Campaigns

Cada *whisper* está asociado a una campaña promocional, las cuales pueden contener varias promociones distintas o algunas promociones que se envían más de una vez. Así, esta tabla contiene información sobre el nombre de la campaña y el cliente asociado a ésta. Por ejemplo, campaña 2x1 Mc Colosso asociada al cliente Mc'Donalds.

3.1.6. Fences

Información relacionada con los *fences* (lugar geográfico con un radio definido que hace factible el envío de promociones cuando un usuario toma contacto con él). Los registros de la tabla contienen información sobre la ubicación (latitud y longitud) de cada *fence*, como también el radio de alcance y el nombre de cada uno. Dicha tabla es de importancia debido a que se desea modelar las tasas de apertura con un subconjunto de *fences*, por lo que será necesario tomarla en cuenta al momento de proceder con la discriminación del set de estimación.

3.1.7. User Position

Tabla que indica información respecto a la posición instantánea de los usuarios entre los meses de Febrero y Marzo de 2015. Entre los registros más importantes de esta tabla se encuentra la distancia del usuario al *fence* asociado al envío que recibió. Cabe destacar que no existe una forma directa de realizar la unión de la tabla *inbox* con esta tabla, por lo que más adelante al considerar dicha métrica como una variable relevante, se imputarán gran cantidad de datos con miras a aproximar la distancia de un usuario al *fence* respectivo.

3.1.8. Brand

Contiene los nombres de las marcas asociadas a los envíos del horizonte temporal considerado.

3.1.9. User Whispers Gifts

Esta tabla contiene información respecto al reenvío de promociones por parte de los usuarios. En ella se registra todos los usuarios que reenviaron y que recibieron promociones mediante esta modalidad de envío.

A continuación, en la Ilustración 3, se presenta un diagrama que resume gráficamente la información descrita. Cabe destacar que, como primer procesamiento de las tablas, se realizó el cruce entre *Whispers* y *Brand* mediante la llave *Whisper Campaign*.

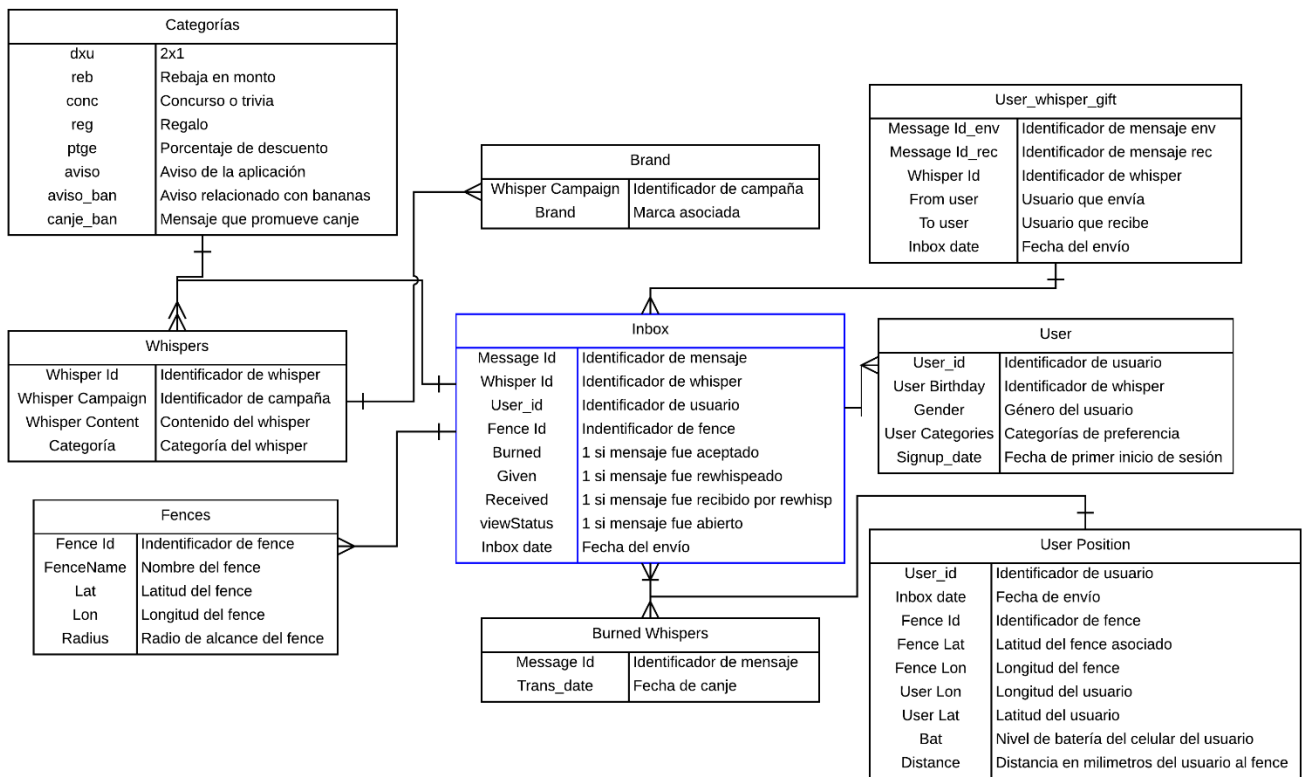


Ilustración 3 - Diagrama tablas Izit

Para obtener una definición más detallada del significado de cada registros de las tablas anteriormente mencionadas, ver sección de anexos.

3.2. Pre-procesamiento de datos

En esta sección se describen los principales filtros realizados al set de datos consolidado. Para consolidar dicho set de datos se realizaron una serie de procedimientos en el software R [11] que permitieron cruzar las distintas tablas con *inbox* según sus respectivas llaves primarias.

Las columnas consideradas en la base de datos *inbox* se muestran en la Tabla 3, donde es posible apreciar que el primer registro corresponde a la llave primaria de dicha base y los tres registros siguientes son indicadores foráneos que servirán para realizar el cruce con los datos de las bases restantes.

Registros	Breve Descripción
Message Id	Identificador de registro
Whisper Id	Identificador de envío
Fence Id	Identificador de fence
User Id	Identificador de usuario
Inbox Date	Fecha del envío
View Status	1 si abre, 0 si no
Burned	1 si acepta, 0 si no
Given	1 si reenvía, 0 si no
Received	1 si recibe por reenvío, 0 si no

Tabla 2 - Registros considerados en inbox

Así, se procedió a realizar el cruce de *inbox* con las bases de datos que se consideraron de interés. Más específicamente, la Tabla 3 muestra los registros efectivamente mezclados.

Tabla	Registros
Fence	Fence id
	Fence name
Whispers	Whisper id
	Whisper campaign
	Campaign name
	Category
	Brand
User	User id
	User birthday
	Gender
	Signup date
	Bananas count

Tabla 3 - Tablas mezcladas con inbox

3.2.1. Creación de variables

Antes de comenzar con los procesos de filtro y análisis descriptivo de los datos, se decidió crear algunas variables de interés. A continuación se describe exhaustivamente el proceder para su creación y la metodología usada para cada una de ellas.

Edad

Como es posible observar en la Tabla 3, la tabla *Users* posee entre sus registros la fecha de nacimiento de cada usuario. En base a esta columna se calculó la edad en años fijando como fecha de referencia el día 27/10/2015².

² Fecha de la última actualización de la función que calcula la edad

Tiempo desde el registro

La base de datos *Users* posee entre tus registros la fecha de registro de cada usuario de la aplicación. Así, se calculó el tiempo en semanas entre la fecha del primer envío recibido en la base de datos *inbox* y la fecha de registro del usuario.

Canje de *bananas*

Se sospecha que ante el envío de una promoción que motive el acumular *bananas* tiene un efecto diferenciador en la decisión de abrir una promoción por parte de los usuarios. Es por esto que, a partir de la categorización realizada en 3.1.4, se creó una variable binaria que indica 1 si el envío asociado es de la categoría **canje_ban**. Esto es, no se consideraron los envíos con avisos sobre el programa de fidelización sino solo las promociones que gatillan la aceptación del *whisper* por parte los usuarios.

Envíos, apertura y aceptación histórica

Con la base de datos *inbox* completa, se consolidó la cantidad de envíos recibidos, la cantidad de envíos abiertos y la cantidad de envíos aceptados para cada usuario.

Envíos y aperturas acumuladas

Como hipótesis de investigación se sospecha que los individuos a lo largo de su interacción con la aplicación “aprenden” y se familiarizan con ella. Es por esto que se crearon variables dinámicas de acumulación de envíos y aperturas las cuales se van actualizando a medida que el usuario recibe y/o abre promociones tanto en los *fences* considerados para la estimación como en los demás *fences*.

El algoritmo que codifica dichas variables se esquematiza de la siguiente forma:

- Ordenar la tabla *inbox* en base a usuarios y fecha en orden creciente. Realizar el mismo procedimiento con la base de datos remanente (es decir, la que no contiene información respecto los *fences* considerados en la estimación).
- Recorrer la tabla *inbox* secuencialmente y la tabla remanente en bloques de submatrices de clientes.
- Si el usuario de *inbox* es igual al usuario de la tabla remanente y la fecha de envío de *inbox* es menor, entonces acumular un envío en stock.
- Además acumular la apertura y aceptación en caso de existir.
- Recorrer todas las observaciones de la submatriz para dicho cliente y sumar envíos, aperturas y aceptación total.

Repetir con todas las observaciones de *inbox*.

3.3. Filtro de datos

En primera instancia, se inició con un set de datos compuesto por 3.400.000 registros usuario-envío, de los cuales un porcentaje despreciable posee registros erróneos en cuanto a la fecha del envío (fecha “00-00-00 00:00:00”), por lo que se decidió eliminarlos de la base de datos al considerarlos como problemas de almacenamiento por parte de la compañía.

Por otro lado, y dada la intención de incluir variables demográficas en la modelación, se decidió consultar con la contraparte de la empresa respecto a un límite razonable de edad de los usuarios de la aplicación, debido a que, dado que los usuarios deben ingresar su fecha de nacimiento, existen múltiples registros con fechas incoherentes y fuera de rango. Luego de deliberar con personal de Izit, se decide que un umbral de edad adecuado corresponde a los 50 años. Así, la Tabla 4 muestra el porcentaje de reducción de la base de datos total respecto a los filtros aplicados.

Luego de realizar el filtro anterior, se procedió a separar el data set de estimación del set de datos remanente. Los datos para realizar la estimación provienen de los *fences*: Costanera Center, Mall Plaza Vespucio, Metro Tobalaba, Metro Baquedano y Patio Bellavista. La elección de éstos responde a que, sin considerar el *fence Izitland*, son los que poseen mayor densidad de envíos y equivalen al 22% de los envíos totales de la base de datos.

Filtro	Nro. Registros	Porcentaje
Inicial	3400000	100%
Filtro fechas	3355915	99%
Filtro edad	3281929	97%

Tabla 4- Filtros inbox 1

Una vez separado el conjunto de datos para la estimación, se procedió a graficar las tasas de apertura por usuarios para verificar la existencia de algún comportamiento anómalo por parte de ciertos clientes en el data set usado para la estimación. La Ilustración 4 muestra el gráfico de lo anteriormente mencionado.

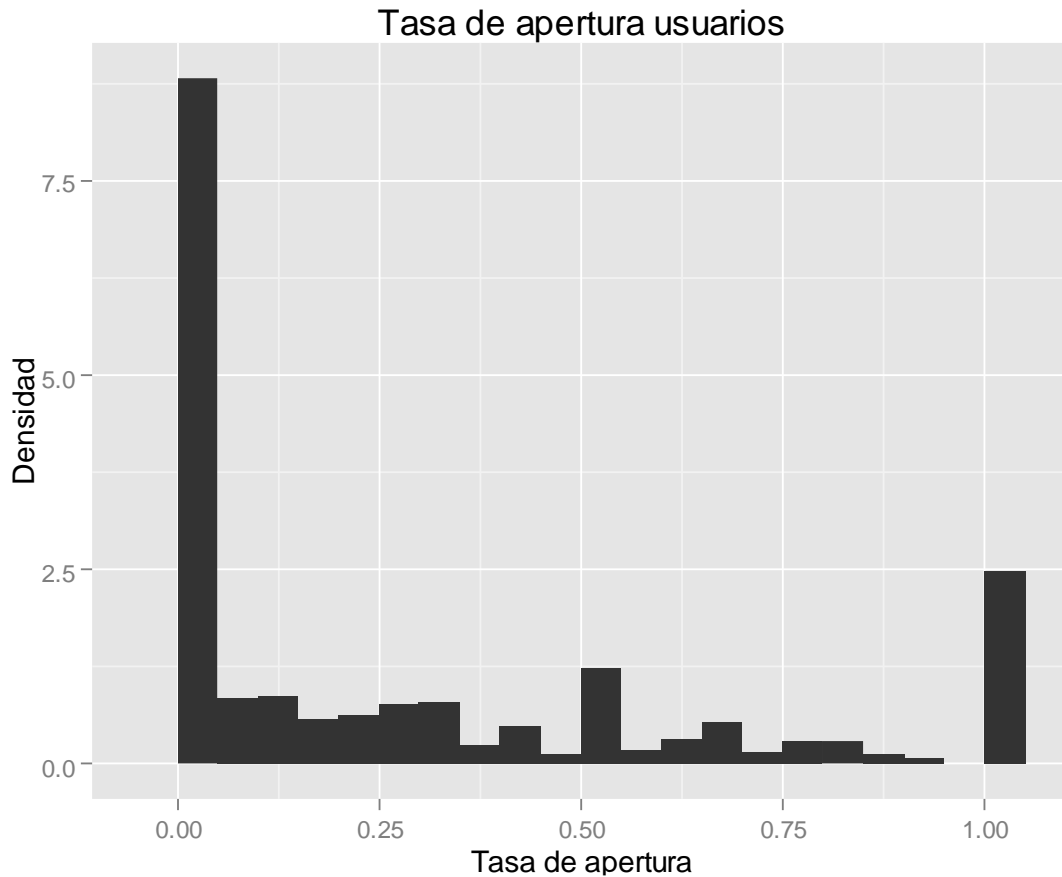


Ilustración 4 - Densidad tasa de apertura por usuarios

Es posible observar que existe una gran cantidad de usuarios que no posee ninguna apertura a lo largo del horizonte temporal considerado, como también existe una cantidad no menor de usuarios que abren todas las promociones a las cuales se enfrentan.

Si bien es posible que la acumulación de respuestas negativas hacia la apertura se deba al horizonte temporal reducido de set de datos considerado para la estimación, y en consecuencia sea esperable una gran cantidad de tasas nulas, se decidió eliminar dichos registros debido a que dificultan los procedimientos econométricos y sesgan los resultados hacia la no apertura de promociones. Por otro lado, si bien los usuarios que abren todos los envíos que reciben son menores en cantidad comparados con los que no abren, de igual forma se decidió excluirlos del set de datos de estimación debido a que existe la probabilidad que dichas personas respondan a motivaciones distintas para con la compañía (usuarios relacionados con la empresa, por ejemplo).

Finalmente, el criterio de exclusión de usuarios considerando tasas de apertura fue:

$$2,5\% \leq \text{aper}_{rate} \leq 97,5\%$$

Ecuación 1 - Criterio de filtro en tasas de apertura

Al realizar el filtro considerado en la Ecuación 1, la muestra de clientes disminuye en un 55% en comparación con el tamaño original sin filtros.

Respecto a los envíos por usuarios, en la Ilustración 5 se logra apreciar que existe una gran cantidad de personas con un solo envío. Específicamente, estos envíos corresponde al 19.4% del total enviado en el horizonte temporal considerado.

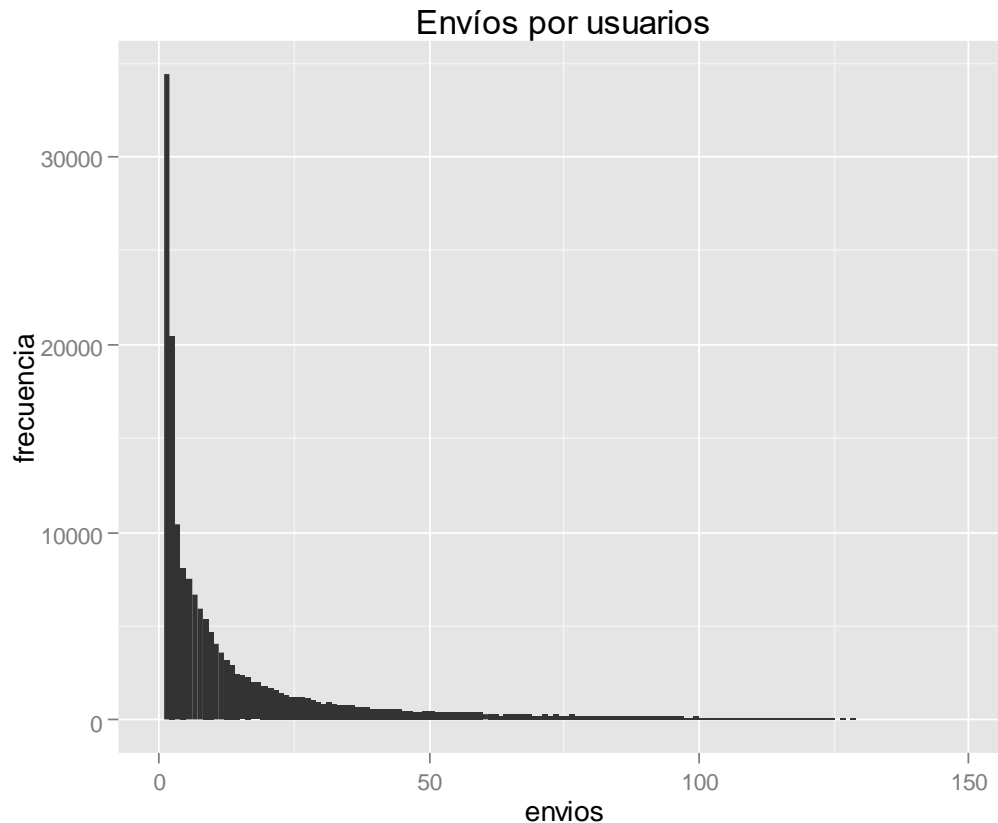


Ilustración 5 - Envíos por usuarios

Así, dado que se busca estudiar el comportamiento dinámico de las variables en relación a las tasas de apertura individuales, se decidió eliminar del data set de estimación a los usuarios con un solo envío.

Finalmente, dado que una de las variables a estudiar es la acumulación de envíos por parte de los usuarios, se decidió imputar *outliers* según el criterio de rango intercuartílico otorgado por el gráfico de cajas de la variable en cuestión.

La Tabla 5 muestra la serie de filtros usados en el data set de estimación y el porcentaje de datos remanentes respecto al inicial.

Filtro	Nro. Envíos	Porcentaje	Usuarios
Inicial	405248	100,0%	40230
Tasas de apertura y envíos³	279658	69,0%	18200
Outliers stock envíos	271548	67,0%	18038

Tabla 5 – Filtros inbox 2

³ Incluye filtro mostrado en Ecuación 1 e impone que la cantidad de envíos sea mayor o igual a 1

4. Análisis descriptivo de datos

A continuación se ilustra el análisis descriptivo de los datos con respecto a la variable de interés, esto es, todo lo realizado se centrará en analizar el impacto de los distintos factores disponibles en función de la tasa de apertura de los usuarios. La sección de análisis descriptivo comenzará entregando métricas generales respecto al estudio, para luego analizar las tasas de apertura promedio en función de usuarios, promociones, *fences* y stock de envíos/apertura.

4.1. Métricas generales

La Ilustración 6 muestra de manera gráfica las métricas agregadas de la compañía en el set de datos considerado. El porcentaje de apertura se calculó considerando el total de envíos en la muestra, mientras que los porcentajes de aceptación y reenvío son condicionales a la apertura.

Como es posible observar, la tasa de aceptación y reenvío es bastante baja, incluso condicionando a la apertura, por lo que si bien en una etapa inicial de la investigación se propuso estudiar estas fases del proceso, debido a lo expuesto en la sección de alcances, el análisis solo se centrará en la apertura.



Ilustración 6 - Métricas agregadas

Según datos otorgados por *Smart Insights* [12], en el 2015 las tasas de apertura asociadas a correos enviados a usuarios en conceptos de *e-commerce* bordean el 17%, por lo que las tasas de apertura de Izit y su propuesta móvil geo localizada están muy por sobre el promedio mundial. Sin embargo, aún se considera que existe espacio de mejora en cuanto al entendimiento de los factores que determinan que un usuario abra o no abra una promoción.

4.2. Usuarios

El set de datos para realizar la estimación contiene información sobre 18.058 usuarios de la aplicación. A continuación se realizará un análisis descriptivo de estos en relación a sus tasas de apertura bajo diferentes aristas.

4.2.1. Género

El comportamiento en cuanto a apertura por género se muestra en la Ilustración 7, en donde se puede apreciar que los hombres reciben una cantidad levemente mayor que las mujeres. Sin embargo, a pesar que el género masculino recibe más promociones respecto a las mujeres, estas últimas poseen tasas de apertura del orden de 32%, que es significativamente menor que la de los hombres que bordea el 27% (p-valor < 0.001).

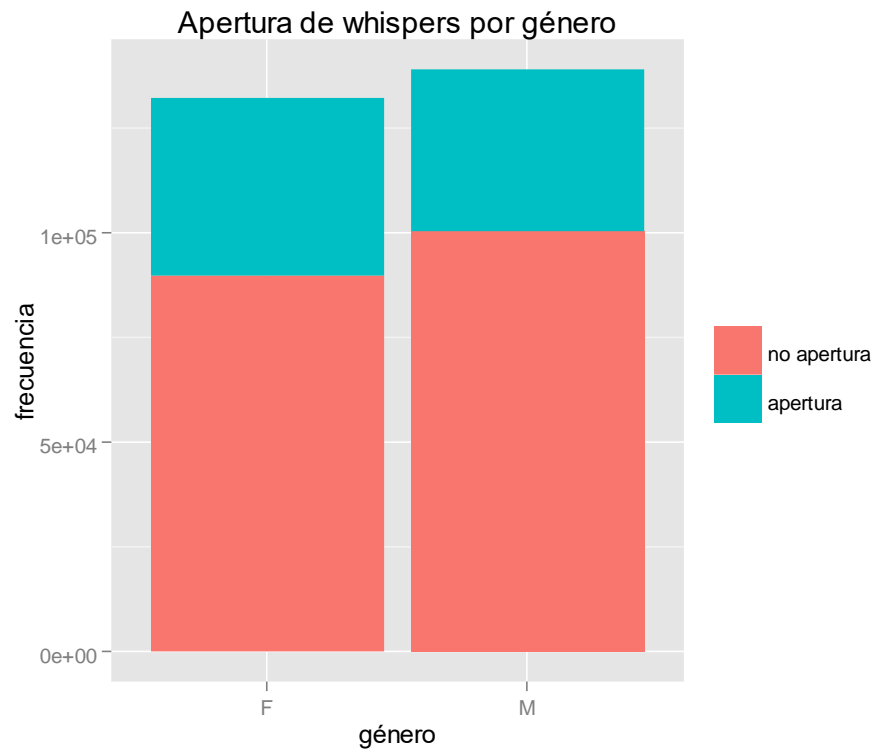


Ilustración 7 – Apertura y envíos por género

Además, el género es una de las variables de segmentación para el envío de promociones por parte de los clientes, por lo que se considera pertinente entender como dicha variable modera las tasas de apertura en los usuarios de la aplicación.

4.2.2. Edad

Además del género, la edad de igual forma es una variable que usan los clientes de Izit para segmentar los usuarios que reciben sus promociones, por lo que es de importancia saber cómo se comporta respecto a la tasa de apertura.

La Ilustración 8 muestra el histograma de las edades de los usuarios, evidenciando que la mayor cantidad de envíos se realizan a personas entre los 20 años y 30 años.

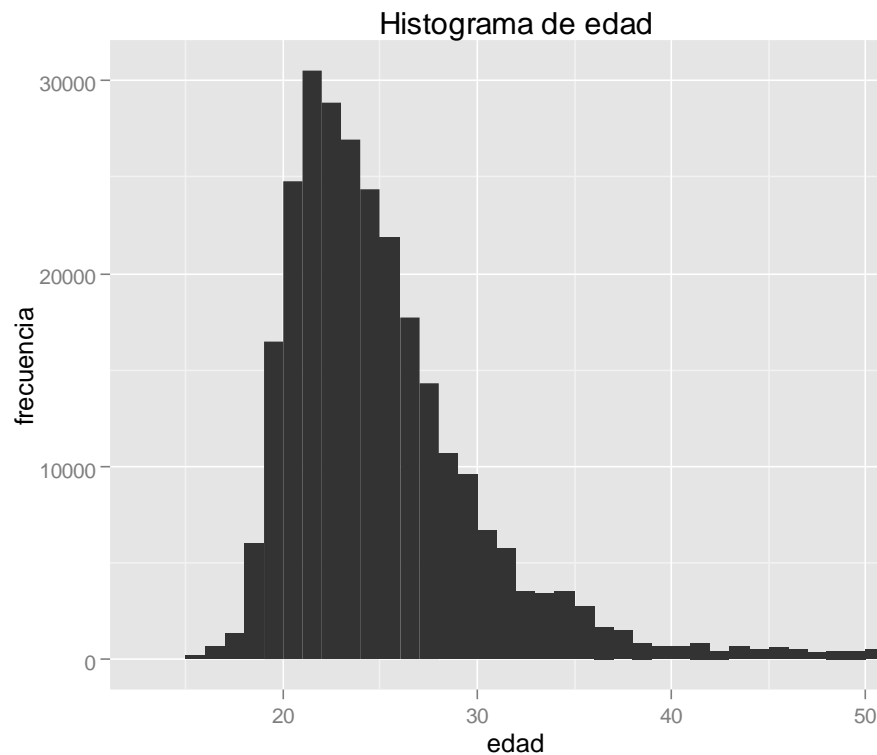


Ilustración 8 - Histograma de edades usuarios

Además, luego de los 40 años se observa una baja sistemática en cuanto a los envíos recibidos por los usuarios. Esto puede deberse a que, para comenzar a recibir promociones el usuario debe activar la aplicación en el teléfono, y dado que la población que se familiariza en mayor medida con la tecnología suelen ser individuos más jóvenes, es esperable que personas de mayor edad reciban menos.

Respecto a las tasas de apertura en función de la edad, la Ilustración 9 muestra que a medida que esta aumenta, la tasa de apertura tiende a decrecer, mostrando casos aislados de aumento en la cola derecha debido a la poca cantidad de usuarios recibiendo en ese rango con alta variabilidad. En primera instancia, se esperaba que la tasa de apertura se comportara de manera cuadrática respecto a la edad, esto es, menores tasas a medida que la edad es pequeña y un peak de apertura en edades bordeando los 25 años, para luego descender. Sin embargo, lo anterior no se sustenta mediante el análisis descriptivo.

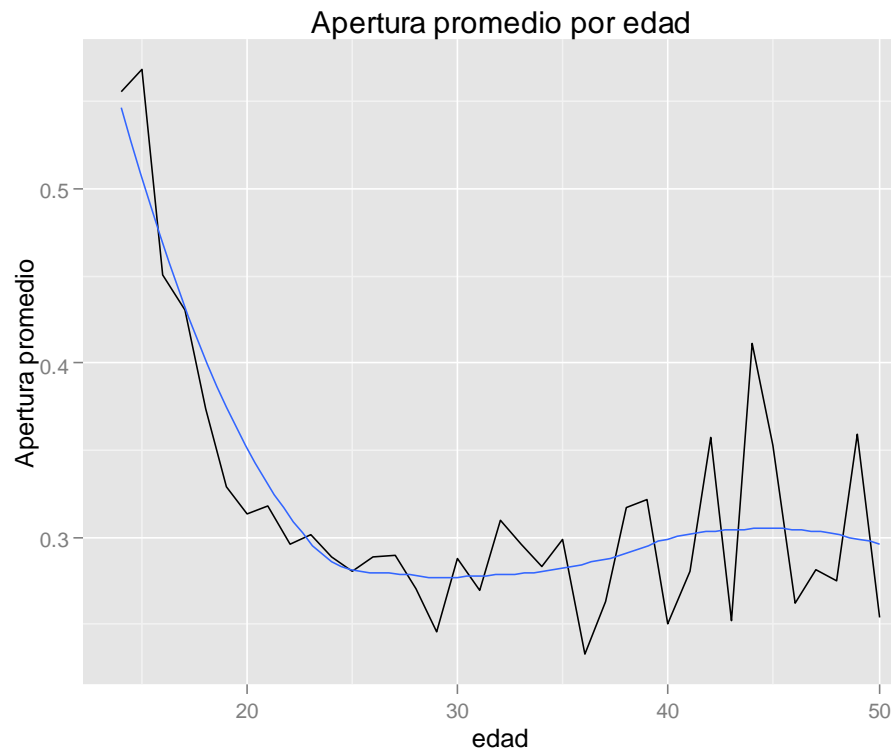


Ilustración 9 - Apertura promedio por edad

4.2.3. Contador de bananas

Como se describió en la sección 1.2, la compañía posee un programa de fidelización que otorga puntos (*bananas*) en caso de realizar ciertas acciones con la aplicación. Algunas de estas acciones son: reenviar una promoción al círculo de redes sociales cercano o aceptar una promoción asociada a una campaña de incentivo del uso de la aplicación mediante este sistema de puntos.

De esta forma, se espera que los usuarios que posean una cantidad importante de *bananas* acumuladas históricamente tengan un grado de interacción distinto con la compañía. Puesto que el programa de lealtad tiene como objetivo el premiar a los usuarios más fieles a la aplicación, se espera que dichos clientes tengan mayores tasas de apertura que la población que es inactivo respecto al programa de fidelización.

Lo anterior se ve reflejado en la Ilustración 10, en donde la cantidad de *bananas* históricas de cada usuario se categorizó en intervalos inducidos por los percentiles de dicha variable. Se observa que a medida que la cantidad de puntos del programa de fidelización de Izit aumenta, en promedio la tasa de apertura de dichos los usuarios es mayor.

Un comportamiento similar se aprecia al graficar los intervalos crecientes de contador de bananas respecto al stock de envíos y, en menor medida, respecto al stock de aperturas (ver anexos), por lo que de manera gráfica se aprecia que la cantidad de *bananas* históricas representa un elemento positivo en cuanto a la actividad de los clientes con la aplicación. Cabe destacar que lo anterior controla solo por correlación, esto es, la direccionalidad podría ser tanto directa (si usuario

abre más, entonces acumula más *bananas*) o inversa (usuario acumula más *bananas* y es por eso que abre más).

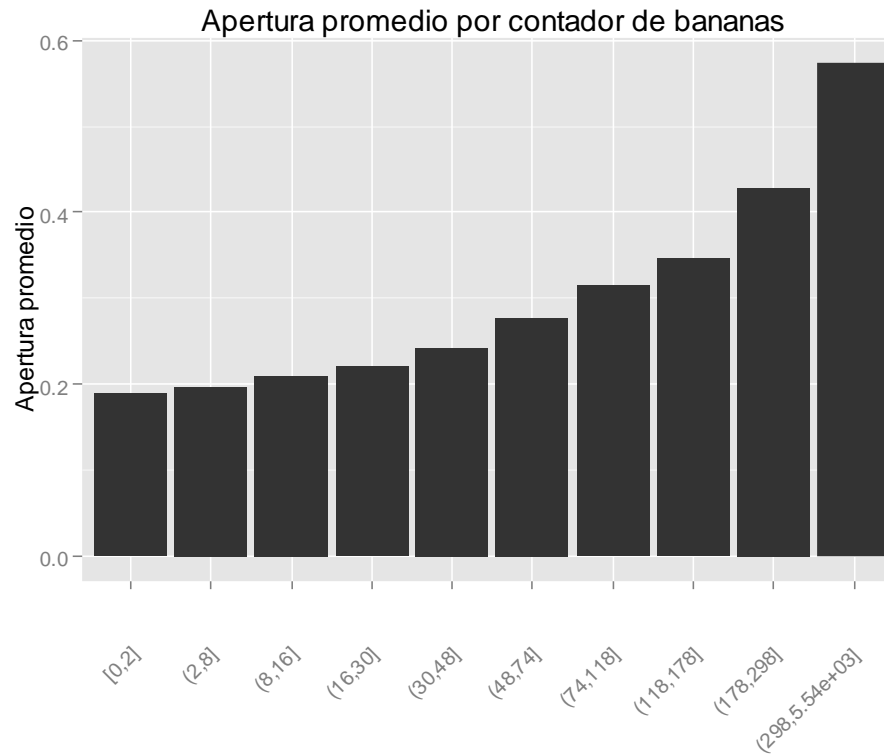


Ilustración 10 - Apertura promedio por contador de bananas

4.2.4. Tiempo desde registro

Respecto al tiempo transcurrido desde el registro de los usuarios hasta el primer envío que poseen en la base de datos, no se considera del todo intuitivo el comportamiento de estos respecto a sus tasas de apertura en función de dicha variable. En primera instancia se podría pensar, que usuarios con menor tiempo con la aplicación instalada resultan más entusiastas respecto a su uso, por lo que sus tasas de aperturas deberían tender a ser mayores; mientras que de igual forma se podría pensar que los usuarios que llevan un tiempo menor con la aplicación instalada poseen un grado de aversión a la participación e interacción con esta, debido a la poca familiarización con la plataforma.

La Ilustración 11 grafica el promedio de apertura para usuarios con distintos niveles de tiempo desde el registro hasta la primera interacción con la aplicación (categorización realizada en base a los percentiles de la variable en cuestión). Se logra apreciar que no existe una tendencia clara del tiempo desde el registro sobre la tasa de apertura de los usuarios, hecho que se refuerza al observar, por ejemplo, cómo influye esta variable tanto en los stock de envíos y apertura de los usuarios considerados en la muestra (Ilustración 40 e Ilustración 41 de anexos).

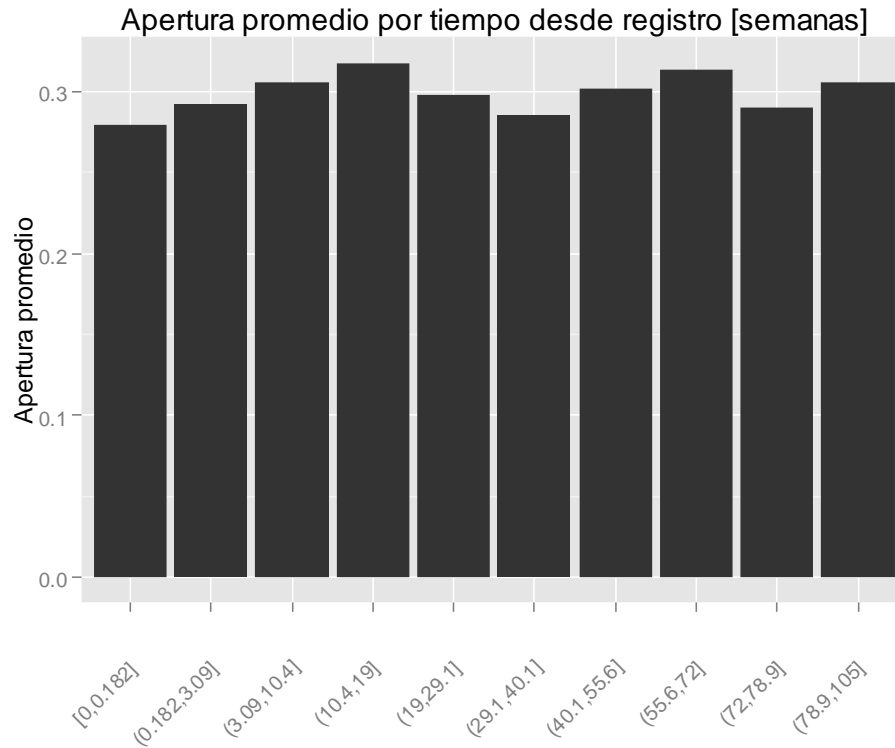


Ilustración 11 - Apertura promedio por tiempo desde registro

4.2.5. Envíos y aperturas históricas

Una vez más se hace presente el *trade-off* entre alcance y efectividad de promociones, pues al aumentar el número de envíos a los usuarios de la aplicación se aprecia una disminución clara de las tasas de apertura, siendo el valor máximo de dicha tasa un 50% y el valor mínimo 7% aproximadamente. Lo anteriormente comentado se ve reflejado en la Ilustración 42 de anexos.

Respecto a la cantidad de aperturas, la situación es similar a lo comentado sobre el stock de aperturas dinámico. A medida que los usuarios abren más de manera histórica, se evidencian tasas de apertura mayores (ver gráfica en anexos)

4.3. Promociones

La información sobre promociones es posible visualizarla a través de múltiples enfoques: *whispers*, campañas, marcas o categorías. El set de datos para estimar comprende 65 *whispers* distintos, 56 campañas, 23 marcas y 9 categorías.

Respecto a las marcas, la Ilustración 12 muestra que, abarcando el 37,7% del total de envíos, la marca que tiene mayor cantidad de envíos es *Club 57*, centro de evento/fiestas/restaurant ubicado en Patio Bellavista (por consiguiente, asociado al *fence* Patio Bellavista). La campaña asociada a estos envíos corresponde a un intento por fomentar el programa de fidelidad de Izit y

abarca 3 *whispers*: Uno asociado al canje de un bebestible por 2.000 *bananas*, otro para informar respecto al programa de *bananas* y finalmente un combo dos por uno en coctelería y bebestibles.

Las marcas que siguen a *Club 57* en cuanto a sus altas tasas de apertura son *Izit Regalos* y *Republicano* (restaurant ubicado en Patio Bellavista).

La marca que posee menor cantidad de envíos es Lolapalooza (21 envíos equivalente al 0,007% del total), debido a que se realizó una campaña de tiempo reducido para concursar en entradas para tal evento.

Es interesante notar que, al observar la Ilustración 13 es posible notar que, precisamente las marcas que poseen una cantidad pequeña de envíos son las que poseen mayor promedio de aperturas. Algo similar ocurre con las marcas que poseen poca cantidad de envíos.

Lo anterior indica que existe un *trade-off* entre el alcance que puede tener una marca (medido en la cantidad de envíos asociados a esta) y la efectividad de dicha marca en cuanto a tasas de apertura. En caso que un usuario este recibiendo continuamente la misma promoción o promociones similares, debiera incrementarse el grado de desmotivación y/o desagrado de este frente a los estímulos enviados. Sin embargo, existe la posibilidad que el usuario valore el envío continuo de promociones a su teléfono móvil, incrementando así la posibilidad de abrir y potencialmente aceptar los *whispers*. Es por esto que resulta interesante analizar la manera en que los usuarios acumulan tanto envíos como aperturas.

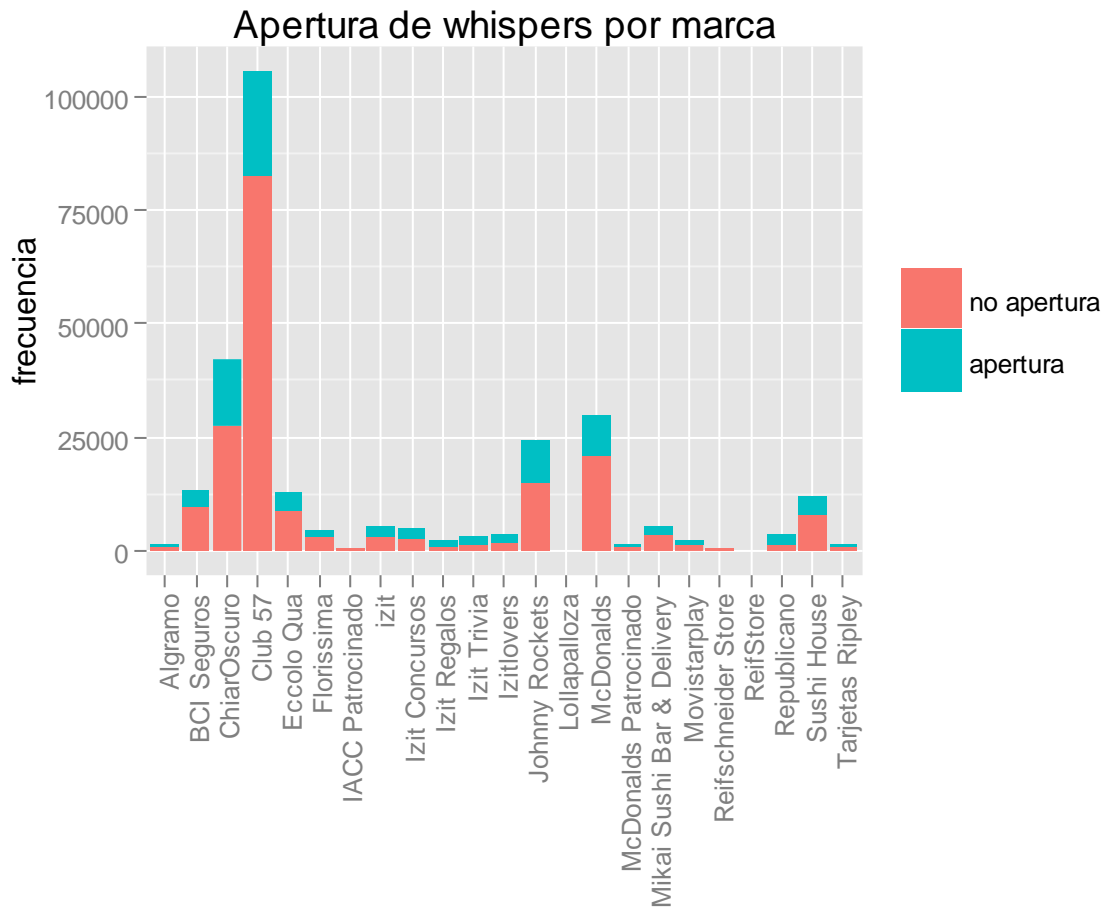


Ilustración 12 - Apertura y envíos por marca

Finalmente, respecto a las categorías ocurre un hecho similar al fenómeno visualizado con las marcas. La categoría que tiene mayor cantidad de envíos es la asociada a promociones dos por uno (Ilustración 14), sin embargo también es la que posee menor promedio de aperturas (Ilustración 15).

En estudios anteriores [5], quedó de manifiesto que la categoría o el contenido de la promoción *per se* no influyen en las tasas de apertura de estas, por lo que para efectos de este estudio no se controlará por dichos factores, sin embargo, resulta interesante hacerse cargo del *trade-off* entre alcance y efectividad.

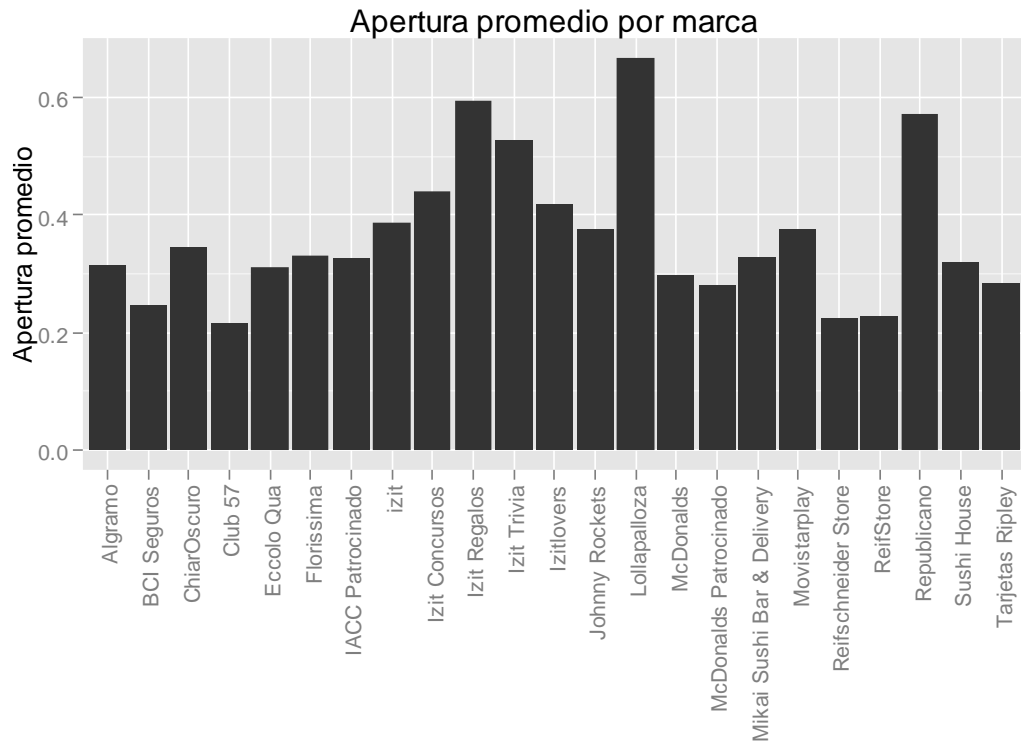


Ilustración 13 - Apertura promedio por marca

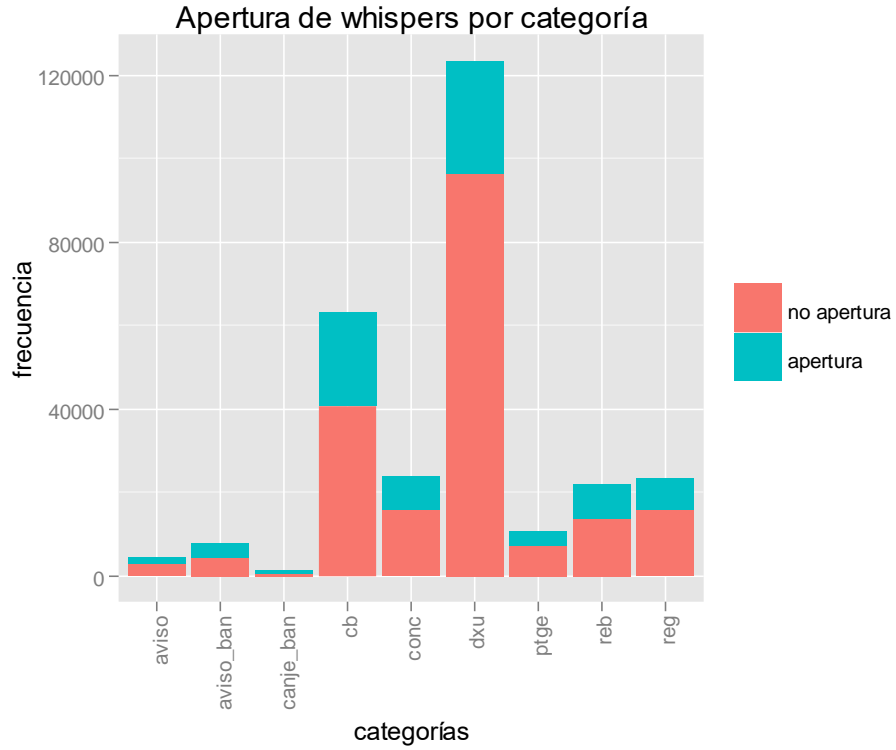


Ilustración 14 - Apertura y envíos por categoría

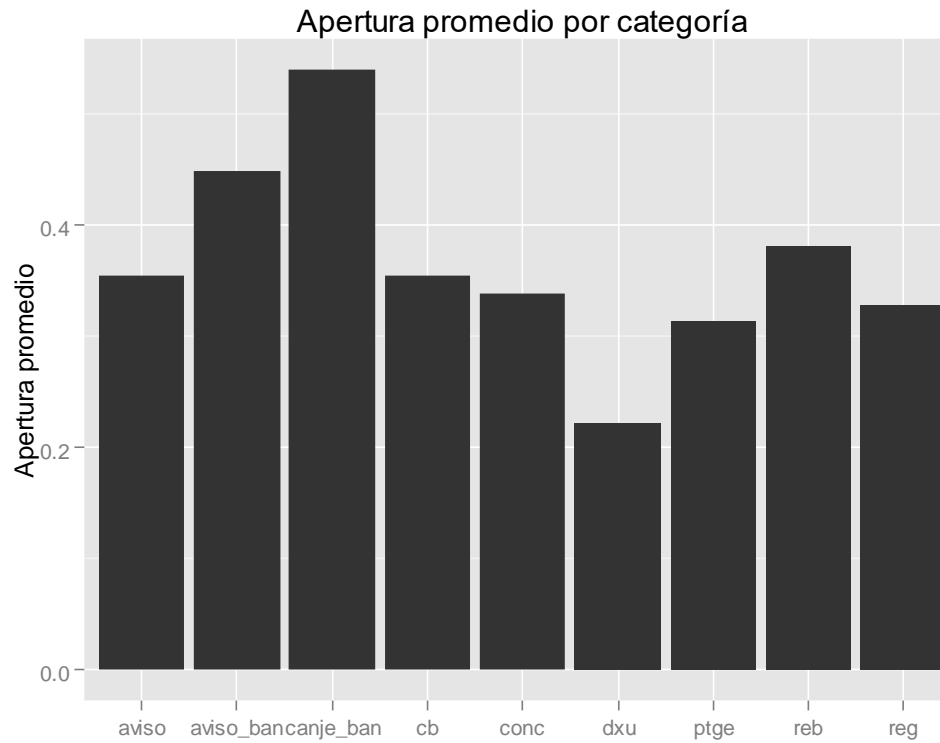


Ilustración 15 - Apertura promedio por categoría

En particular, uno de los factores que se desea medir es cómo el club de fidelización de la compañía, reflejado en promociones asociadas a *bananas*, modera las tasas de apertura de los usuarios en su interacción con la aplicación. Así, la Ilustración 16 muestra que en promedio, las tasas de apertura son mayores cuando la promoción enviada está asociada al canje de *bananas*. En lo concreto, existe una diferencia significativa de 0.24 puntos porcentuales entre ambas modalidades (p -valor < 0.001).

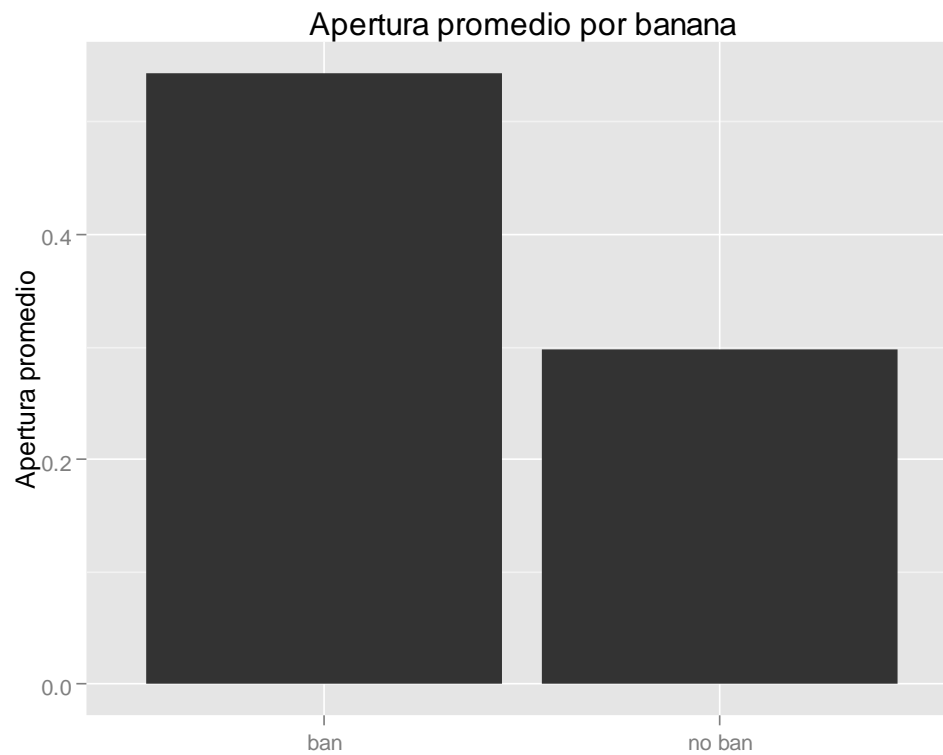


Ilustración 16 - Apertura promedio por categoría banana

4.4. Fences

Como se explicó en la sección 2.4, los *fences* a considerar en el estudio son los asociados al Mall Costanera Center (borde rojo), Metro Tobalaba (borde verde), Metro Baquedano (borde azul), Mall Plaza Vespucio y Patio Bellavista.

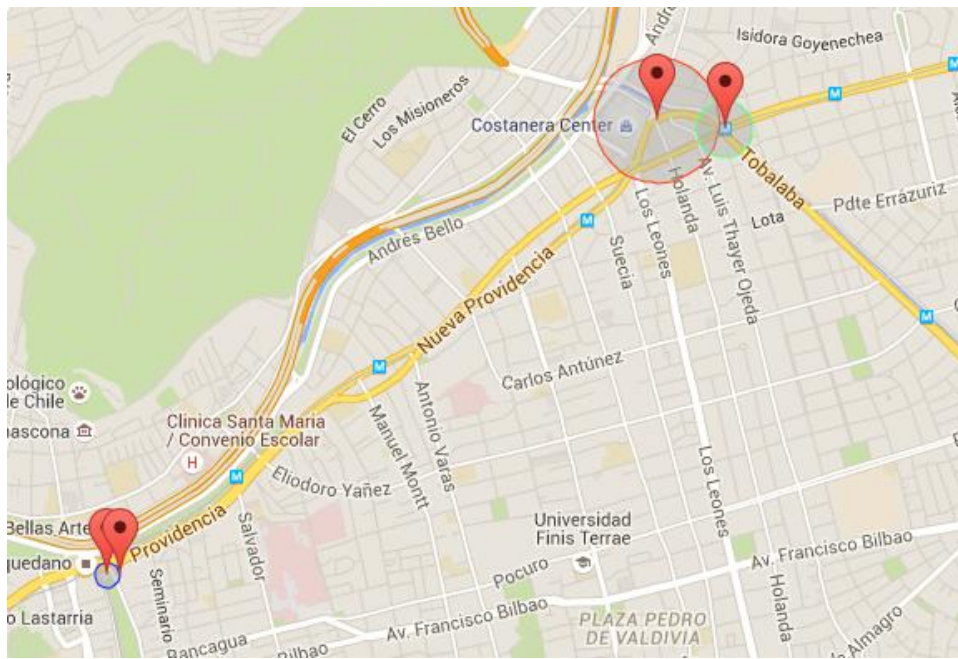


Ilustración 17 - Muestra de *fences* considerados

En la Ilustración 17 [13] es posible ver la ubicación y el radio de alcance de cada uno de estos, donde se observa que el *fence* asociado al Mall Costanera Center es el que posee mayor extensión de entre sus pares. La Tabla 6 muestra los *fences* considerados en el estudio y el radio medido en metros de cada uno de estos.

Fence	Radio [m]	Nro. Envíos
Costanera Center	300	86855
Mall Plaza Vespucio	220	30793
Metro Tobalaba	135	44292
Patio Bellavista	100	81609
Metro Baquedano	61	36109

Tabla 6 - Radio de *fences*

Agregando por *fences* también ocurre que existe un *trade-off* entre alcance y efectividad en cuanto a la apertura. El *fence* asociado al Mall Plaza Vespucio es el que posee mayor tasa de apertura, sin embargo es el que también posee menor cantidad de envíos respecto a sus pares. Por otro lado, si bien el *fence* asociado al Mall Costanera Center es el que posee mayor cantidad de envíos, no presenta la menor tasa de apertura.

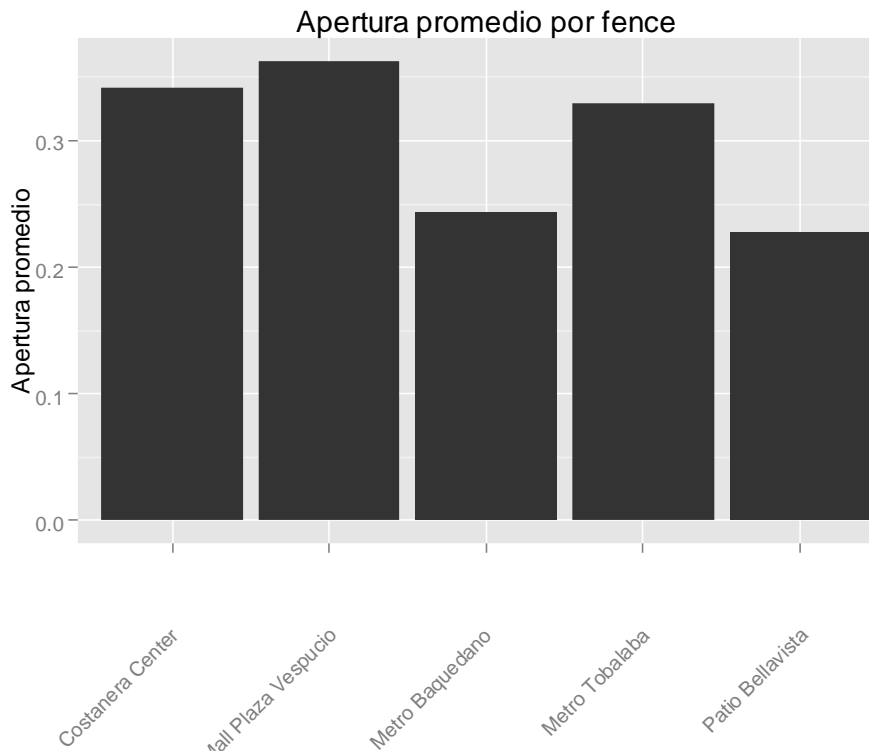


Ilustración 18 - Apertura promedio por fence

Es posible apreciar que existen diferencias pequeñas en cuanto a las tasas de apertura promedio entre los *fences* Costanera Center – Metro Tobalaba y Metro Baquedano – Patio Bellavista. Estas diferencias pequeñas pueden estar explicadas simplemente por la ubicación geográfica de cada *fence*. En muchos casos ocurre que, Metro Baquedano es un camino necesario para ir a los locales comerciales ubicados en Barrio Bellavista (donde se ubica el *fence* Patio Bellavista). Por otro lado, para ir al Mall Costanera Center, una de las visas naturales es mediante transporte subterráneo, donde precisamente la estación más cercana es Metro Tobalaba.

Respecto a las marcas asociadas a los envíos por *fences*, la Tabla 7 muestra que, en general, existe homogeneidad en cuanto a estas, excluyendo el *fence* Patio Bellavista, donde dado el contexto bohemio, se concentran marcas relacionadas a lugares nocturnos (a excepción de la marca Ripley).

Metro Tobalaba	Costanera Center	Plaza Vespucio	Metro Baquedano	Bellavista
Algramo	Algramo	Algramo	Algramo	Club 57
BCI Seguros	BCI Seguros	BCI Seguros	BCI Seguros	Izit Regalos
Eccolo Qua	ChiarOscuro	Florissima	Club 57	Izit Trivia
Florissima	Izit	IACC	Eccolo Qua	McDonalds
Izit	Izit Concursos	Izit	Florissima	Republicano
Izit Concursos	Izit Regalos	Izit Concursos	Izit	Ripley
Izit Regalos	Izit Trivia	Izit Regalos	Izit Concursos	
Izit Trivia	Izitlovers	Izit Trivia	Izit Regalos	
Izitlovers	JohnnyRockets	JohnnyRockets	Izit Trivia	
Johnny Rockets	Lollapalloza	Lollapalloza	JohnnyRockets	
Lollapalloza	McDonalds	McDonalds	McDonalds	
McDonalds	Movistarplay	Reifschneider	Republicano	
Mikai Sushi	Reifschneider	Ripley	Ripley	
Republicano	Ripley			
Sushi House				
Ripley				

Tabla 7 - Marcas por fences

4.5. Acumulación de envíos y aperturas

En la sección 4.2.5 se evidenció la relación negativa de la acumulación de envíos histórica en las tasas de apertura, como también la relación positiva de las aperturas históricas. Sin embargo, se tiene como hipótesis que el comportamiento de envíos y apertura podría explicarse de manera dinámica en la medida que los usuarios aprenden a interactuar con la aplicación.

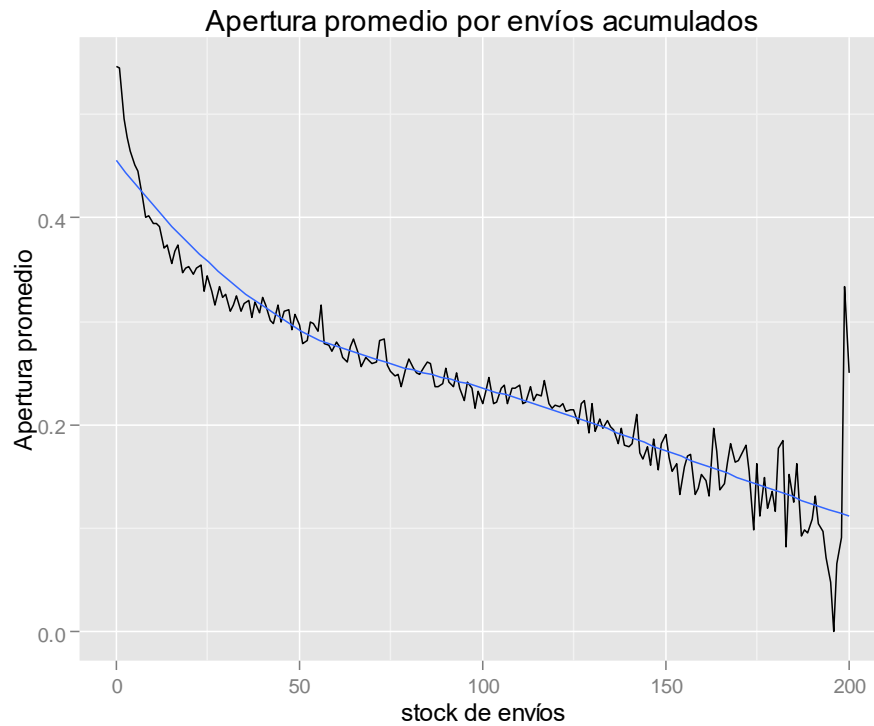


Ilustración 19 - Apertura promedio por stock de envíos

La Ilustración 19 muestra una relación negativa entre el stock de envíos y las tasas de apertura de los usuarios de la aplicación. Mientras que la Ilustración 20 muestra una relación totalmente contraria respecto al comportamiento inducido por la acumulación de aperturas.

Una posible explicación de lo anterior es la familiaridad que el usuario desarrolla al momento de interactuar de manera reiterativa con la aplicación. A medida que recibe más envíos, la sensación de invasividad puede manifestarse, por lo que las tasas de apertura tenderían a disminuir (resultaría interesante desagregar entre envíos geo localizados y no geo localizados).

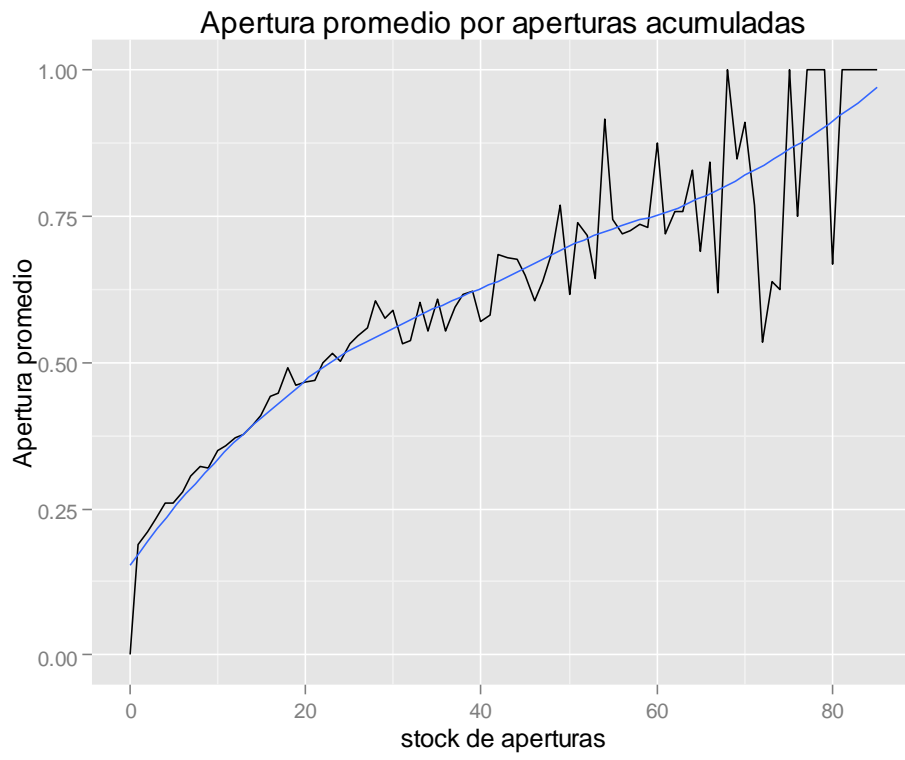


Ilustración 20 – Apertura promedio por stock de aperturas

5. Marco Teórico

A continuación se presentan las bases metodológicas y teóricas de los modelos a usar para el desarrollo del trabajo. Se partirá definiendo las generalidades de todos los modelos de elección discreta, para luego ahondar en detalle en las especificaciones de los modelos logit, logit con clases latentes, mixed logit y el modelo mixed logit jerárquico bayesiano.

5.1. Modelo logit

El modelo logit nace de asumir que los clientes n reportan una utilidad latente U_{nj} al seleccionar una alternativa j de un set de opciones. Dicha utilidad latente es posible descomponerla tanto en una porción determinística como una porción aleatoria, obteniendo que $U_{nj} = V_{nj} + \varepsilon_{nj}$, donde V_{nj} se parametriza en función de variables observables y coeficientes β asociados y ε_{nj} corresponde a una realización de una variable aleatoria valor extremo tipo I.

Una de las ventajas del modelo logit simple es la facilidad de cálculo respecto a las probabilidades de elección. Al especificar dicha distribución para el error, dicha probabilidad posee una forma cerrada para ser calculada:

$$P_{nj} = \frac{\exp V_{nj}}{\sum_k \exp V_{nk}}$$

Ecuación 2: Probabilidad de elección logit simple

5.2. Modelo logit con heterogeneidad discreta

Uno de los temas centrales en marketing es la heterogeneidad del comportamiento [15]. La heterogeneidad en el comportamiento de los agentes puede ser tanto observable (existen datos que evidencian comportamiento distinto entre agentes), como no observable (diferencias intrínsecas en el comportamiento de los agentes).

El enfoque logit con heterogeneidad discreta asume que los parámetros del modelo (β 's) posee una distribución discreta en la población. Esto es, existen segmentos discretos en la población caracterizados por parámetros distintos que internalizan el comportamiento diferenciados de los individuos inter segmentos y homogéneo dentro de él.

Así, definiendo $s_m \in \{1, \dots, S\}$ como la proporción de la población que pertenece al segmento m y β_m el vector de parámetros asociado al segmento m , se tendrá que la probabilidad de elección del individuo n ante la alternativa j es:

$$P_{nj} = \sum_{m=1}^S P_{nj|s_m} \cdot s_m$$

$$P_{nj} = \sum_{m=1}^S s_m \frac{\exp V_{nj}(\beta_m)}{\sum_k \exp V_{nk}(\beta_m)}$$

Ecuación 3: Probabilidad de elección modelo clases latentes

5.3. Modelo mixed logit

El modelo mixed logit corresponde a una generalización de lo expuesto en 5.2 permitiendo que las preferencias de los tomadores de decisiones varíen para cada uno de éstos. Así, se especifica la función de utilidad latente del agente n al elegir la alternativa j como $U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj}$. Donde x_{nj} corresponde a variables observables, β_n corresponde al gusto (preferencia) de cada individuo n respecto del atributo x_{nj} y ε_{nj} la porción no observable de la utilidad distribuyendo valor extremo tipo I (i.i.d). El modelo asume que los coeficientes varían a lo largo de los tomadores de decisiones con densidad $f(\beta|\theta)$, donde los parámetros a estimar serán los que caracterizan a la densidad (θ). A modo de ejemplo, si se especifica $\beta_n \sim N(\mu, \sigma^2)$, los parámetros a estimar serán $\theta = (\mu, \sigma^2)$.

La probabilidad de elección en este caso, calculada a partir de la ley de probabilidades totales es:

$$P_{nj} = \int \left(\frac{\exp \beta'_n x_{nj}}{\sum_k \exp \beta'_n x_{nk}} \right) f(\beta) d\beta$$

Ecuación 4: Probabilidad de elección - mixed logit

La especificación de $f(\beta)$ queda a criterio del modelador de la situación, sin embargo, en muchas aplicaciones se utiliza una especificación normal o log-normal.

5.4. Modelo mixed logit jerárquico bayesiano

Dados los avances en cuanto a materia computacional en las últimas décadas, el enfoque jerárquico bayesiano surge como una alternativa plausible de modelación en cuanto a modelos de elección se refiere. Una de las virtudes de métodos jerárquicos bayesianos es que mantienen la heterogeneidad a nivel individual usando data que describe el comportamiento específico de cada cliente, así como reconoce que el comportamiento de los otros individuos puede ser informativo. Además, para realizar inferencia no es necesario recurrir a argumentos asintóticos (como lo es el caso del método de máximo verosimilitud), esto pues se conoce la distribución posterior.

Bajo un enfoque jerárquico bayesiano, se asumirá que el modelo está descrito por la probabilidad de ocurrencia.

$$P_{nj}|\beta_n = \frac{\exp \beta'_n x_{nj}}{\sum_k \exp \beta'_n x_{nk}}$$

Ecuación 5: Probabilidad de elección condicional

Introduciendo un esquema jerárquico en donde la media de la distribución de los coeficientes aleatorios de cada usuario es una combinación lineal de variables demográficas asociadas [16].

$$\boldsymbol{\beta}_n \sim N(\Delta' z_n, V_\beta)$$

Ecuación 6: Distribución de coeficientes aleatorios modelo jerárquico

Para completar el modelo, es necesario agregar *priors*⁴ a los parámetros poblacionales Δ' y V_β .

$$\text{vec}(\Delta|V_\beta) \sim N(\text{vec}(\bar{\Delta}), A^{-1} \otimes V_\beta)$$

$$V_\beta \sim IW(v, V_0)$$

Ecuación 7: Distribución posterior de hiperparámetros

Donde \otimes corresponde al producto matricial de Kronecker y *IW* es la distribución Wishart inversa.

⁴ Distribución que refleja las creencias del investigador respecto a los posibles valores que puede tomar el parámetro en cuestión.

6. Desarrollo metodológico

Según el análisis descriptivo enunciado en la sección 4.5, los envíos y aperturas acumulados (dinámicos) tienen una influencia clara sobre las tasas de apertura de los usuarios de la aplicación, mientras que de igual forma se aprecia una diferencia significativa entre envíos asociados al club de fidelización de la compañía (4.3). Además, se incluirán como variables demográficas tanto el sexo como la edad los usuarios, por lo que los modelos presentados en esta sección tendrán la siguiente función de utilidad determinística del tipo:

$$v_{ijt} = f(\beta, stock_{env_{it}}, stock_{aper_{it}}, fence_{ijt}, ban_j, edad_i, genero_i)$$

Ecuación 8 - Especificación general utilidad modelo logit

Donde β es un vector de parámetros y:

- $stock_{env_{it}}$ corresponde al stock de envíos acumulados por el usuario i en el tiempo t .
- $stock_{aper_{it}}$ corresponde al stock de aperturas acumuladas por el usuario i en el tiempo t .
- env_i corresponde a una variable demográfica que contiene el número de envíos históricos entre las fechas 01/01/2015 y 29/09/2015 del usuario i .
- $aper_i$ corresponde a una variable demográfica que contiene el número de aperturas históricas entre las fechas 01/01/2015 y 29/09/2015 del usuario i .
- $fence_{ijt}$ es una variable categórica que indica el $fence$ por el cual el usuario i recibió la promoción j en el tiempo t .
- ban_j corresponde a una variable binaria que indica 1 si la promoción j está asociada al club de fidelización de la compañía.
- $edad_i$ corresponde a la edad del usuario i .
- $genero_i$ indica 1 si el usuario i es hombre y 0 si es mujer.

Finalmente, la variable objetivo del estudio es y_{ijt} que indica 1 si el usuario abrió la promoción que se envió y 0 si no.

6.1. Modelo logit simple

6.1.1. Modelo logit considerando envíos y aperturas acumuladas

En primera instancia, se decidió especificar un modelo en el que se excluye la acumulación dinámica tanto de aperturas como de envíos. Esto pues, en caso que dichas variables estáticas presenten altos niveles de significancia, motivaría el análisis dinámico de éstas.

Así, la especificación del modelo inicial es:

$$v_{ijt} = \beta_1 stock_{env_{it}} + \beta_2 stock_{aper_{it}} + \sum_{k \in \{fences\} \setminus \{Cost\}} \beta_k fence_{ijt}^k + \beta_7 ban_j + \beta_8 edad_i + \beta_9 gen_i + \varepsilon_{ijt}$$

Ecuación 9 - Especificación de utilidad modelo logit considerando comportamiento acumulado

Con i indexando a los usuarios de la aplicación, j indicando una promoción y t el tiempo.

Donde $fence_{ijt}^k$ indica el $fence$ considerado en la estimación (para efectos de identificación, se excluye el $fence$ Costanera Center).

Luego, la probabilidad de apertura del modelo logit simple será:

$$\Pr(y_{ijt} = 1) = \frac{\exp v_{ijt}}{1 + \exp v_{ijt}}$$

Ecuación 10 - Probabilidad de apertura modelo logit considerando comportamiento acumulado

Los resultados de la estimación mediante máxima verosimilitud se adjuntan en la Tabla 8, en donde se logra apreciar que todas las variables consideradas son significativas a un nivel $\alpha = 0.1\%$ de significancia, además los envíos y aperturas acumuladas tienen el efecto deseado, esto es, se aprecia la existencia de un *trade-off* entre ambas métricas. El coeficiente asociado a la edad indica que a medida que los usuarios son más longevos, la probabilidad de apertura disminuye.

Variables	Coficiente	Error estándar	Significancia
Envíos acumulados	-0,0187	0,0001	***
Apertura acumulada	0,1092	0,0007	***
Edad	-0,0209	0,0004	***
Plaza Vespucio	-0,0593	0,0144	***
Metro Baquedano	-0,3898	0,0158	***
Metro Tobalaba	-0,0736	0,0132	***
Patio Bellavista	-0,5074	0,0124	***
Género (hombre)	-0,1447	0,0091	***
Bananas	0,7690	0,0622	***
LL		-144.000	

Tabla 8 - Resultados modelo logit con envíos y apertura acumulada

Respecto a las variables categóricas, el factor de más importancia en cuanto al efecto sobre la probabilidad de apertura es si el envío corresponde a una promoción asociada al club de fidelización de la compañía, en donde el modelo sugiere que cuando un *whisper* es de esa naturaleza, la probabilidad de apertura aumenta respecto a los que no lo son. Por otro lado, respecto a los *fences*, el modelo sugiere que todos éstos disminuyen la probabilidad de apertura respecto a Costanera Center.

Cabe destacar la gran diferencia en cuanto a las magnitudes de los *fences* Metro Baquedano-Patio Bellavista y Metro Tobalaba-Plaza Vespucio. Se logra apreciar un efecto diferenciado entre estos grupos de *fences* lo que podría sugerir que existen ciertos tipos de clientes que frecuentan/prefieren ciertos geográficos en particular o que la proximidad de dichos *fences* genera en los mismos clientes tasas de aperturas diferentes.

Finalmente, el género indica que en promedio, las mujeres son más susceptibles a la apertura de promociones respecto a los hombres.

En anexos se puede observar una tabla similar a la anterior con las variables continuas normalizadas. Al realizar dicho procedimiento se logra observar que dentro de este grupo, las aperturas y envíos acumulados resultan ser los factores continuos más importantes en la estimación (seguidos por la edad), mientras que los efectos encontrados anteriormente en las variables binarias se mantienen.

6.2. Modelo logit con heterogeneidad discreta

6.2.1. Modelo logit con dos clases latentes

En un principio se calibrará un modelo logit asumiendo que existen dos clases latentes en la población. Dichas clases están caracterizadas por parámetros para cada una de ellas, los cuales intentan representar las preferencias heterogéneas entre clases y homogéneas dentro de cada una. Así, la especificación de la utilidad para cada clase será:

$$v_{ijt}^1 = \beta_1^1 stock_{env_{it}} + \beta_2^1 stock_{aper_{it}} + \sum_{k \in \{fences\} \setminus \{Cost\}} \beta_k^1 fence_{ijt}^k + \beta_7^1 ban_j + \beta_8^1 edad_i + \beta_9^1 gen_i + \varepsilon_{ijt}^1$$

$$v_{ijt}^2 = \beta_1^2 stock_{env_{it}} + \beta_2^2 stock_{aper_{it}} + \sum_{k \in \{fences\} \setminus \{Cost\}} \beta_k^2 fence_{ijt}^k + \beta_7^2 ban_j + \beta_8^2 edad_i + \beta_9^2 gen_i + \varepsilon_{ijt}^2$$

Ecuación 11 - Especificación utilidad clases latentes

Planteando la probabilidad de elección como en la **¡Error! No se encuentra el origen de a referencia.**, definiendo la proporción de los segmentos como en la **¡Error! No se encuentra el origen de la referencia.** con $S = 2$ y maximizando los parámetros por el método de máxima verosimilitud, se obtienen los resultados desplegados en la Tabla 9.

El modelo reconoce dos clases latentes; una compuesta por el 58% de la población y otra compuesta por el 42% (coeficiente asociado a la parametrización del tamaño de los segmentos es significativo al 0,1% de significancia).

Al analizar el comportamiento de los coeficientes por segmentos es posible concluir patrones que caracterizan de manera muy clara a ambas clases. En cuanto al coeficiente asociado a la acumulación de envíos y apertura, se puede apreciar que el segmento 1 valora menos, en comparación al segmento 2, la acumulación de aperturas. Esto es, dicho segmento deriva menor utilidad en la apertura de promociones y el aumentar su stock de apertura favorece en menor medida la probabilidad que abra un próximo envío respecto a la población del segmento 2. Además, el segmento 1 no es tan sensible a la acumulación de envíos en comparación con el segmento 2.

Variable	Segmento 1			Segmento 2		
	Coefficiente	Error estándar	Significancia	Coefficiente	Error estándar	Significancia
Apertura acumulada	0,2343	0,0030	***	1,0833	0,0390	***
Envíos acumulados	-0,0208	0,0002	***	-0,8237	0,0306	***
Edad	-0,0151	0,0007	***	0,0340	0,0030	***
Plaza Vespucio	-0,0343	0,0257		0,0639	0,0988	
Metro Baquedano	-0,5442	0,0258	***	0,6643	0,1629	***
Metro Tobalaba	-0,0532	0,0227	***	-0,0031	0,0987	
Patio Bellavista	-0,7417	0,0206	***	0,6348	0,1206	***
Género (hombre)	-0,1690	0,0154	***	0,0148	0,0723	
Bananas	1,3838	0,1498	***	0,6547	0,4922	
Tamaño segmento	58%			42%		
LL	-138.475					

Tabla 9 - Resultados modelo 2 clases latentes

Así, solo observando ambos coeficientes es posible caracterizar al segmento 1 como un grupo de usuarios que, quizá, mantiene abierta la aplicación por algún motivo no observable y no le importa en demasía el recibir gran cantidad de envíos, además, corresponde a un grupo de usuarios que si bien valora positivamente el acumular aperturas previas, no es del todo sensible a ellas en comparación con el otro grupo. El segmento 2 es un grupo de usuarios que es altamente sensible tanto a la acumulación de aperturas como a la acumulación de envíos.

También existen diferencias en cuanto al coeficiente de edad de ambos segmentos, evidenciando que en el segmento 1, a medida que la edad aumenta, la probabilidad de apertura disminuye, mientras que el segmento 2 tiene un comportamiento contrario.

Respecto al impacto del *fence* en cada clase, se puede observar que los usuarios pertenecientes al segmento 1 aumentan su probabilidad de apertura cuando reciben envíos provenientes de los *fences* asociados a Mall Costanera Center y Mall Plaza Vespucio respecto a los *fences* restantes. En consecuencia, dicho segmento parece estar asociado a usuarios que frecuentan centros comerciales, y es en ese contexto que abren la aplicación para comenzar a recibir promociones. Por otro lado, los usuarios del segmento 2 aumentan su probabilidad de apertura cuando reciben envíos provenientes de los *fences* asociados a Patio Bellavista y Metro Baquedano.

Así, al parecer este segmento está compuesto por usuarios que gustan de lugares de esparcimiento más que centros comerciales.

En cuanto al coeficiente asociado al género, el segmento 2 es indiferente, mientras que en el segmento 1 los envíos a hombres disminuyen la probabilidad de apertura.

Finalmente, se logra apreciar que el segmento 1 valora en demasía el club de fidelización de la compañía, aumentando la probabilidad de apertura al recibir mensajes alusivos a *bananas*, mientras que el segmento 2 es indiferente a éstos.

Respecto a un modelo con 3 clases latentes, en anexos es posible ver los resultados de su calibración. En dicho modelo, los resultados indican que el parámetro asociado a la proporción de una de las clases es no significativo al 95% de confianza, por lo que no se sustenta estadísticamente la inclusión de 3 clases en la especificación.

6.3. Modelo mixed logit jerárquico bayesiano

Bajo el enfoque jerárquico bayesiano, la especificación de la utilidad de los usuarios queda descrita como:

$$v_{ijt} = \beta_i' x_{ijt}$$

Ecuación 12 - Especificación de utilidad modelo jerárquico

Donde el vector x_{ijt} comprenderá todas las variables consideradas en la especificación de la Ecuación 8, **salvo las variables demográficas asociadas a la edad y al género**. Lo anterior debido a que, el vector de coeficientes aleatorios β_i será definido precisamente en función de variables demográficas que inducirán la estructura jerárquica. Así, se asumirá que el vector de coeficientes posee la siguiente distribución:

$$\beta_i \sim N(\Delta' z_i, V_\beta)$$

$$z_i = (gender_i, bananas_{count_i}, age_i, timestamp_i, envios_i, aper_i, acep_i)$$

Ecuación 13 - Jerarquía en modelo mixed logit

Donde las variables demográficas consideradas corresponden al sexo, contador de *bananas* históricas, edad, tiempo entre el registro y el primer envío recibido, envíos históricos⁵, número de aperturas históricas y aceptaciones históricas. Además se especifican la distribución posterior de Δ y V_β al igual que en la **¡Error! No se encuentra el origen de la referencia.**

El modelo se ejecutó a lo largo de $R = 20.000$ iteraciones, almacenándolas cada 20 y con un período de *Burn-in* de 8.000 iteraciones. Al estimar el modelo mediante el algoritmo Metropolis-Hasting (Random Walk), se obtiene las distribuciones presentadas en la Ilustración 21 alcanzando un valor de log verosimilitud de -122.550 (en anexos se presenta la evolución de la log verosimilitud a medida que avanzan las iteraciones).

⁵ Variable estática que intenta reflejar la historial de envíos del usuario. Variable de stock de envíos es dinámica, por lo que cambia a lo largo del tiempo.

Al observar las distribuciones, es posible apreciar la heterogeneidad existente en la mayoría de los parámetros. Al igual que lo concluido en secciones anteriores, la acumulación de envíos presenta una distribución que en general presenta valores negativos, sin embargo es posible apreciar una cola derecha que indica que existen usuarios que valoran positivamente el envío de promociones a lo largo del tiempo. De igual forma, la distribución de las preferencias por la acumulación de aperturas es en su mayor cantidad positivas, existiendo una pequeña cola izquierda que indica a usuarios que no gustan de acumular aperturas.

Respecto a la valoración de los usuarios por los distintos *fences* es posible observar que en promedio esta es negativa, lo cual indica que el *fence* asociado a Costanera Center se encuentra por sobre sus pares en cuanto a preferencias. Sin embargo, aún se aprecia suficiente heterogeneidad como para identificar a usuarios que valoran otros *fences* por sobre Costanera.

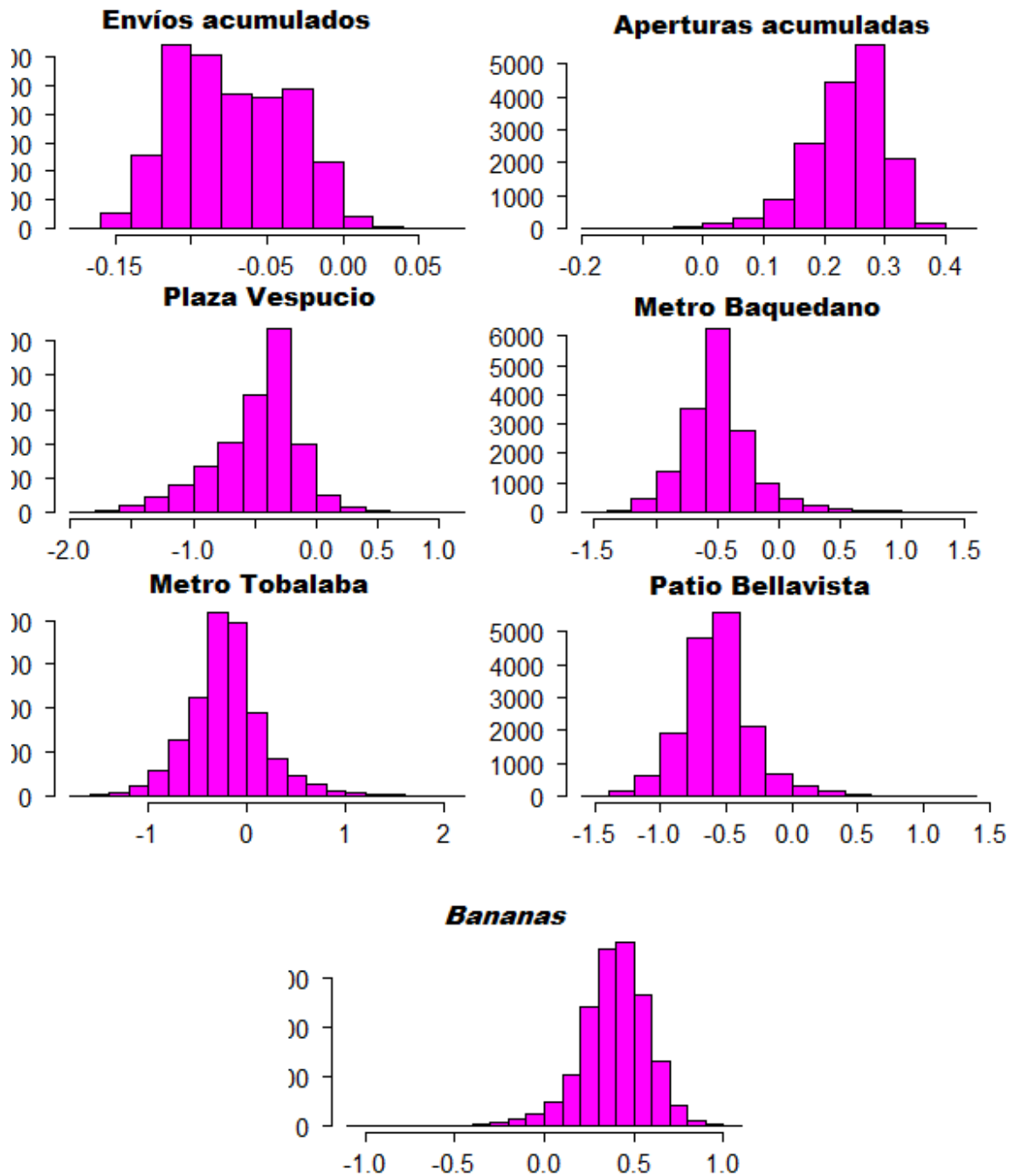


Ilustración 21 - Promedio de posterior de coeficientes

Finalmente, respecto a las preferencias por los envíos asociados a *bananas* se tiene que posee una cola débil a la izquierda, lo cual va acorde con lo concluido anteriormente respecto a la gran importancia de dicho atributo en cuanto a las variables categóricas. En la sección de anexos se puede consultar los valores promedio de cada distribución.

Una de las ventajas de incluir estructura jerárquica en el modelo es que, permite explicar las preferencias de los usuarios por los atributos mediante variables demográficas. La Tabla 10 indica los valores promedio obtenidos para el vector de coeficientes Δ que caracteriza a la distribución de

las preferencias individuales de cada usuario (en rojo se muestran los parámetros no significativos a un 95% de confianza y en anexos se detallan las desviaciones estándar de cada uno).

Coeficiente	Constante	Género (H)	Contador bananas	Edad	Tiempo	Envíos	Apertura	Aceptación
Envíos acumulados	-0,0000003	0,0044	-0,0007	-0,0088	-0,0055	-0,0064	-0,0001	-0,1313
Apertura acumulada	-0,0011	0,0000	-0,0143	0,0171	-0,0050	-0,0015	-0,0059	-0,0003
Plaza Vespucio	0,0000	-0,0067	0,0000	0,0467	0,0178	-0,0067	-0,0057	-0,0263
Metro Baquedano	-0,0006	0,0000	0,0391	0,0000	-0,0108	0,0225	-0,0049	0,0070
Metro Tobalaba	-0,0009	0,0017	-0,0001	0,0437	0,0000	0,0136	0,0153	0,0034
Patio Bellavista	0,0005	0,0033	-0,0113	-0,0001	-0,0517	0,0000	-0,0394	0,0012
Bananas	0,0003	-0,0006	-0,0087	-0,0036	-0,0002	0,0910	0,0000	0,1289

Tabla 10 - Valores promedio Δ

Así, es posible observar que, intrínsecamente la acumulación de envíos posee una valoración negativa por parte de los usuarios (intercepto negativo), mientras que el ser de género masculino favorece las preferencias por recibir envíos. Por otro lado, tanto la edad, el tiempo desde el registro hasta el primer envío y el contador de *bananas* disminuyen las preferencias por acumular promociones recibidas.

Respecto a las preferencias por la acumulación de aperturas, el género parece no ser del todo importante en cuanto a magnitud (sin embargo, es significativo al 95% de confianza), mientras que nuevamente el contador de *bananas* histórico posee un efecto negativo en el coeficiente de stock de aperturas para los usuarios. Una diferencia respecto al análisis de los envíos es que, para el stock de aperturas la edad actúa positivamente, aumentando la sensibilidad de los usuarios al abrir promociones previas.

En cuanto a las preferencias por los distintos *fences*, es posible ver que tanto Plaza Vespucio como Patio Bellavista poseen valoraciones intrínsecamente positivas, mientras que Metro Baquedano y Metro Tobalaba negativas. Esto va en línea con la concepción que estos últimos corresponden a *fences de paso*, en donde los usuarios no necesariamente se encuentran susceptibles a realizar compras. Es probable que estos se encuentren realizando actividades no relacionadas (transporte hacia sus hogares, por ejemplo) o se dirijan hacia un destino en el cual podrían ser más propensos a responder positivamente a una promoción enviada (Metro Baquedano precisamente se ubica a pasos de Patio Bellavista, mientras que Metro Tobalaba se ubica cercano a Costanera Center (*fence* de referencia)). Lo anterior se alinea con el coeficiente negativo de la edad sobre las preferencias de los usuarios respecto al *fence* Patio Bellavista: dado que es un lugar bohemio y frecuentado por adolescentes/adultos jóvenes, al aumentar los años del usuario, el coeficiente de dicho *fence* disminuye.

Respecto al coeficiente de *bananas* de los usuarios, es posible observar que posee intercepto positivo, esto es, los envíos asociados a esta modalidad son intrínsecamente bien valorados por los usuarios. Por otro lado, es el único coeficiente que posee signo negativo en cuanto al género, por los estos envíos debieran ser enviados con mayor prioridad a mujeres de cualquier edad (coeficiente de edad no significativo).

Hechos importantes a destacar es que, el coeficiente asociado al tiempo (en semanas) desde el registro hasta el primer envío recibido en la base de datos es en general negativo, lo cual podría indicar que a medida que pasa el tiempo la presencia de la aplicación en los usuarios se va degradando. Por otro lado, la intuición sobre las medidas de actividad de los usuarios con la compañía (en particular, aperturas y aceptaciones históricas) era que estas influían positivamente en las preferencias de los usuarios en todas sus dimensiones, sin embargo este efecto solo se observa en el coeficiente individual para *bananas* y Metro Tobaralaba.

7. Desarrollo metodológico – Inclusión de medida de distancia

Uno de los factores más importantes y diferenciadores de la aplicación Izit es la geo localización, y en ese sentido, la distancia de cada usuario a las dependencias físicas del avisador asociado a la promoción enviada es una medida interesante de analizar respecto al cómo esta modera las tasas de apertura de los usuarios.

Sin embargo, como es posible apreciar en la descripción detallada de la base de datos en la sección 3.1, no es posible realizar el cruce directo de información entre las tablas *inbox* y *User Position*. Un primer enfoque consistió en crear una llave *usuario-fecha* en ambas bases de datos, de modo de evaluar la cantidad de registros de distancias exactas que se rescatan.

Al realizar lo anterior, se logran rescatar 100 registros de la tabla *inbox* (es decir, 100 registros de la tabla de interés poseen *usuario-fecha* exacta en la tabla de posiciones), lo cual corresponde al **0,002%** de la base total aproximadamente. Más aún, se pierde la estructura de panel debido a que la base resultante posee 100 registros de 100 usuarios únicos.

Dada la reducción de la base de datos, se procedió a aproximar las distancias de los usuarios, por medio de la creación de una llave *usuario-fecha* (sin hora), de modo de encontrar la fecha más cercana a cada registro de la tabla *inbox*. La idea principal del algoritmo se muestra en la Ilustración 22.

user_id	viewStatus	inboxdate
163	1	2015-02-25 22:13:21
163	1	2015-03-02 16:05:37
163	0	2015-03-02 16:09:45
163	0	2015-03-02 16:13:21
163	0	2015-03-02 16:16:45

userSID	lon	alt	timeTrack	distance
163	-70.65033516	564.0474243	2015-02-25 22:06:24	175354
163	-70.65033516	564.0474243	2015-02-25 22:06:24	132522
163	-70.65033516	564.0474243	2015-02-25 22:06:24	236456
163	-70.64761360	566.2254639	2015-02-25 22:19:22	176087
163	-70.04790669	551.6760864	2015-02-26 02:13:10	97972

Ilustración 22 - Ejemplo de aproximación de distancia

Se logra apreciar que el usuario en cuestión recibió y aceptó una promoción enviada en la fecha 25/02/2015 a las 22:13, mientras que la tabla de posición registra su ubicación el mismo día, pero 3 minutos después. Por lo que se asigna una distancia de 176 metros al *fence* en cuestión (la tabla posee la información de distancia medida en milímetros). Cabe destacar que también se cuenta con información de posición del usuario a las 22:06, sin embargo no es la hora más cercana al usuario.

La programación de lo anterior, resulta ser computacionalmente costosa debido al tamaño de la tabla de posiciones, sin embargo, se logró realizar mediante el siguiente esquema de programación:

- Ordenar las tablas *User Position* e *inbox* según usuario y fecha.
- Recorrer la tabla *inbox* secuencialmente y la tabla *User Position* en bloques de submatrices de clientes.
- Para cada registro de *inbox* recorrer la submatriz correspondiente al usuario en cuestión. Calcular la diferencia de fechas y seleccionar el valor absoluto del mínimo entre éstas.
- Si el valor mínimo es menor a una hora, recuperar la distancia y el nivel de batería asociado a dicha posición. En caso contrario, pasar al otro registro de *inbox* y repetir el proceso.

El algoritmo anterior recupera 209.574 registros de la tabla *inbox* original entre las fechas 25/02/2015 y 31/03/2015 con 29.881 usuarios únicos de la aplicación y equivale al 6% de la base de datos original.

Mediante el mismo procedimiento utilizado en la sección 3.3, se procedió a realizar la separación del conjunto de estimación con los respectivos filtros y creación de variables. Así, al finalizar el proceso se obtuvo un set de datos con 7.527 registros a lo largo de 2.270 usuarios de la aplicación.

El principal filtro adicional que se realizó en los datos fue la eliminación de distancias sin sentido, esto es, distancias que fuesen mayores al radio de alcance de cada *fence* considerado (Tabla 6) y que no correspondan a envíos no geo localizados.

Respecto a cómo se comporta la variable distancia, la Ilustración 23 muestra que las distancias más grandes están concentradas en el *fence* Costanera Center, mientras que las más pequeñas están asociadas al *fence* ubicado en Metro Baquedano. Lo anterior tiene sentido, pues se alinea perfectamente con las medidas de los radios de los *fences* considerados.

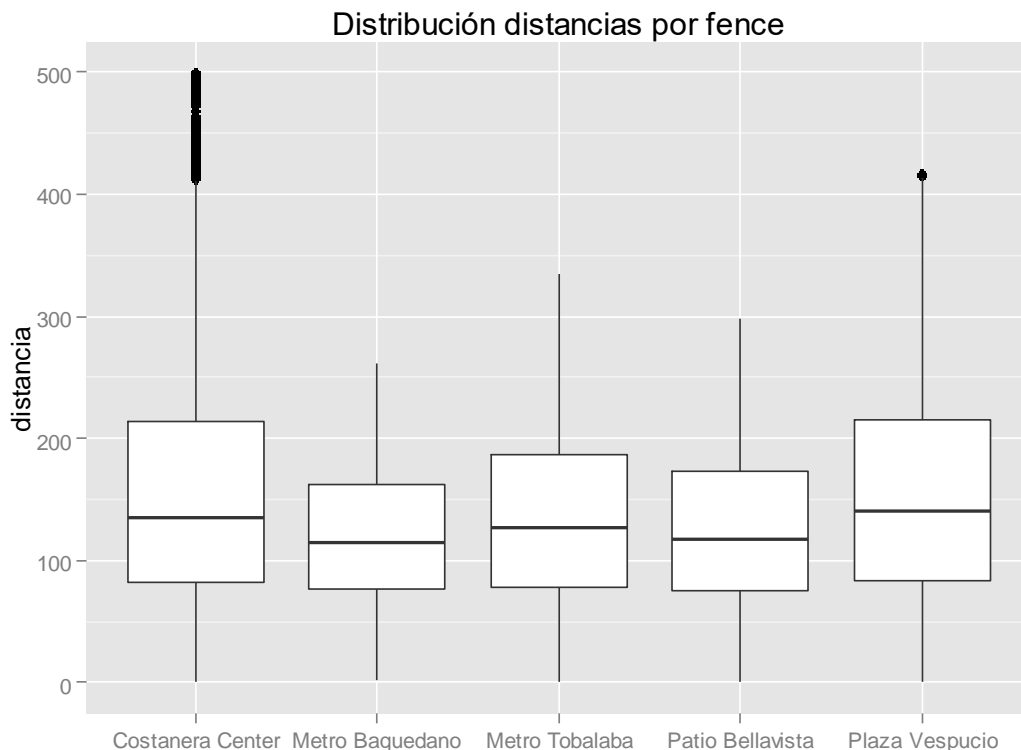


Ilustración 23 - Distribución de distancias por fence

Por otro lado, la Ilustración 24 muestra el promedio de aperturas para los cuartiles de la variable distancia. Es posible observar que, al parecer, la distancia posee un comportamiento irregular frente a las tasas de apertura, lo cual sugiere que existe alta heterogeneidad en cuanto a las preferencias por la distancia de los usuarios. Es probable que existan individuos con alta tolerancia a los envíos distantes del *fence* en cuestión, mientras otros que solo abren las promociones en caso de encontrarse lo suficientemente cerca de la tienda asociada.

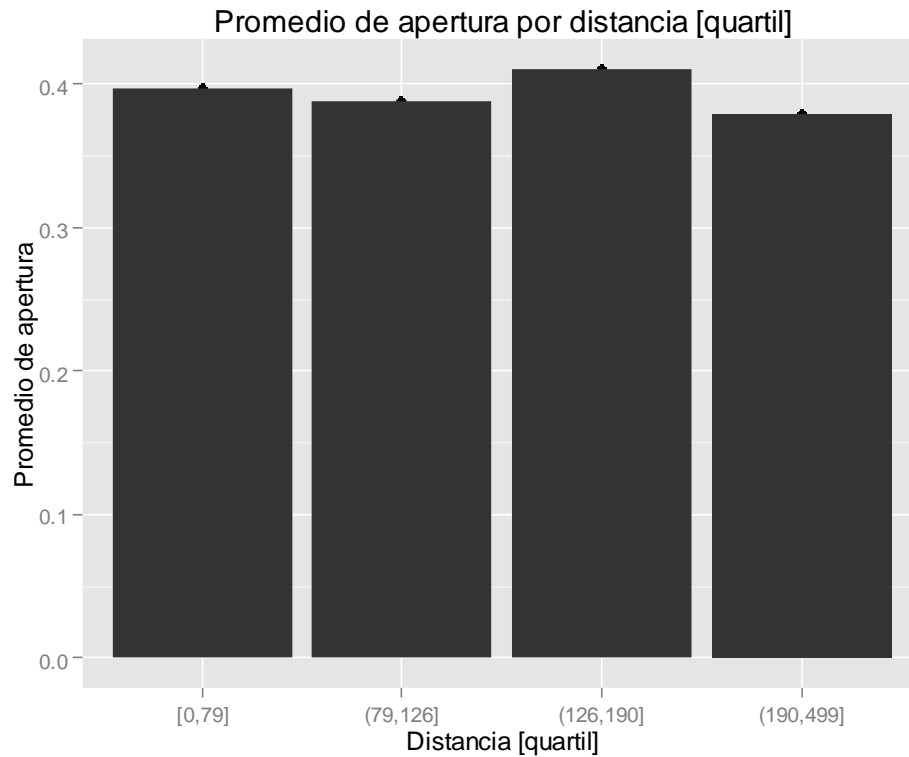


Ilustración 24 - Promedio de aperturas por distancia

7.1. Modelo logit simple

Para incluir la variable distancia en el modelo, se usará en primera instancia la siguiente especificación:

$$v_{ijt} = \beta_1 distance_{ijt} + \beta_2 stock_{env_{it}} + \beta_3 stock_{aper_{it}} + \sum_{k \in \{fences\} \setminus \{Cost\}} \beta_k fence_{ijt}^k + \beta_8 ban_j + \beta_9 edad_i + \beta_{10} gen_i + \varepsilon_{ijt}$$

Ecuación 14 - Especificación modelo logit (distancia)

Los resultados del modelo anteriormente especificado se pueden observar en la Tabla 11, donde se aprecia que todos los *fences* salvo el asociado a Patio Bellavista son no significativos. Además, la variable asociada a la distancia posee el signo esperado, siendo significativa a un nivel de significancia $\alpha = 5\%$.

Nuevamente el stock tanto de envíos como aperturas resultan ser significativos a un nivel de significancia $\alpha = 0,1\%$, mientras que la edad y la variable que asocia envíos relativos a *bananas* poseen un comportamiento similar a los modelos expuestos con anterioridad. Esto es, el que un

envío pertenezca al club de fidelización de Izit resulta en un aumento de la probabilidad de apertura mayor en magnitud respecto sus variables pares binarias.

Variabales	Coefficiente	Error Estándar	Significancia
Distancia	-0,0006	0,0003	*
Envíos acumulados	-0,1131	0,0055	***
Aperturas acumuladas	0,5487	0,0180	***
Metro Baquedano	-0,0979	0,0782	
Metro Tobalaba	0,0309	0,0702	
Patio Bellavista	-0,2993	0,0710	***
Plaza Vespucio	-0,1749	0,0993	.
Género (hombre)	0,0085	0,0520	
Edad	-0,0200	0,0026	***
Bananas	0,8170	0,2469	***
LL		-4.390	

Tabla 11 - Resultados modelo logit con distancia

7.2. Modelo con heterogeneidad discreta

Al igual que en la sección 6.2.1 se busca calibrar un modelo que incluya la variable distancia y que incorpore clases latentes. Si bien, al modelar la situación mediante dos clases latentes se obtienen coeficientes significativos en ambos segmentos, el parámetro que caracteriza a cada clase resultó ser no significativo (ie, el tamaño de una de las clases se aproxima a 0), por lo que los datos no sugieren que exista evidencia estadística para separar en dos segmentos a la población según la especificación

7.3. Modelo mixed logit

Una de las falencias del modelo logit simple es que no permite acomodar variaciones aleatorias en las preferencias de los usuarios y que no permite correlación entre las componentes no observables de la utilidad a lo largo del tiempo. Precisamente, es muy probable que la decisión de abrir o no una promoción por parte de un usuario se encuentre ligada a las decisiones o experiencias pasadas del mismo con la aplicación. Factores como la factibilidad del canje (en caso de aceptar), experiencia de atención con el personal de la tienda asociada a la promoción o expectativas del usuario con el producto/servicio ofrecido son factores por los cuales no se controla y podrían influir en gran medida el cómo los clientes se relacionan con la compañía.

La especificación del modelo es idéntica a la propuesta anteriormente, sin embargo se asumirá que el vector de parámetros β asociado al modelo sigue una distribución normal $N(\mu, \Sigma)$ con $\dim \mu = 10$ y $\dim \Sigma = 10 \times 10$ diagonal⁶. La diferencia entre el modelo aquí propuesto es que se asume una distribución continua a nivel de parámetros, no así en el caso abordado en 7.2, cuya distribución es discreta.

⁶ Sin término fuera de la diagonal de la matriz

Los resultados del modelo se adjuntan en el anexo Tabla 22 de anexos, en donde se logra apreciar que existe una gran diferencia en cuanto la estimación del coeficiente asociado a la distancia, pues el modelo mixed logit estima un coeficiente 66,7% mayor que el obtenido mediante el modelo logit sin heterogeneidad.

Respecto al resto de los coeficientes, el modelo mixed logit reporta parámetros similares en cuanto a magnitud (no así en cuanto a errores estándar).

7.4. Modelo mixed logit jerárquico bayesiano

Para la inclusión de heterogeneidad mediante un enfoque jerárquico bayesiano, se especificaron parámetros idénticos a los enunciados en la sección 6.3, sin embargo, se agregó la variable relacionada con la distancia de los usuarios al *fence*.

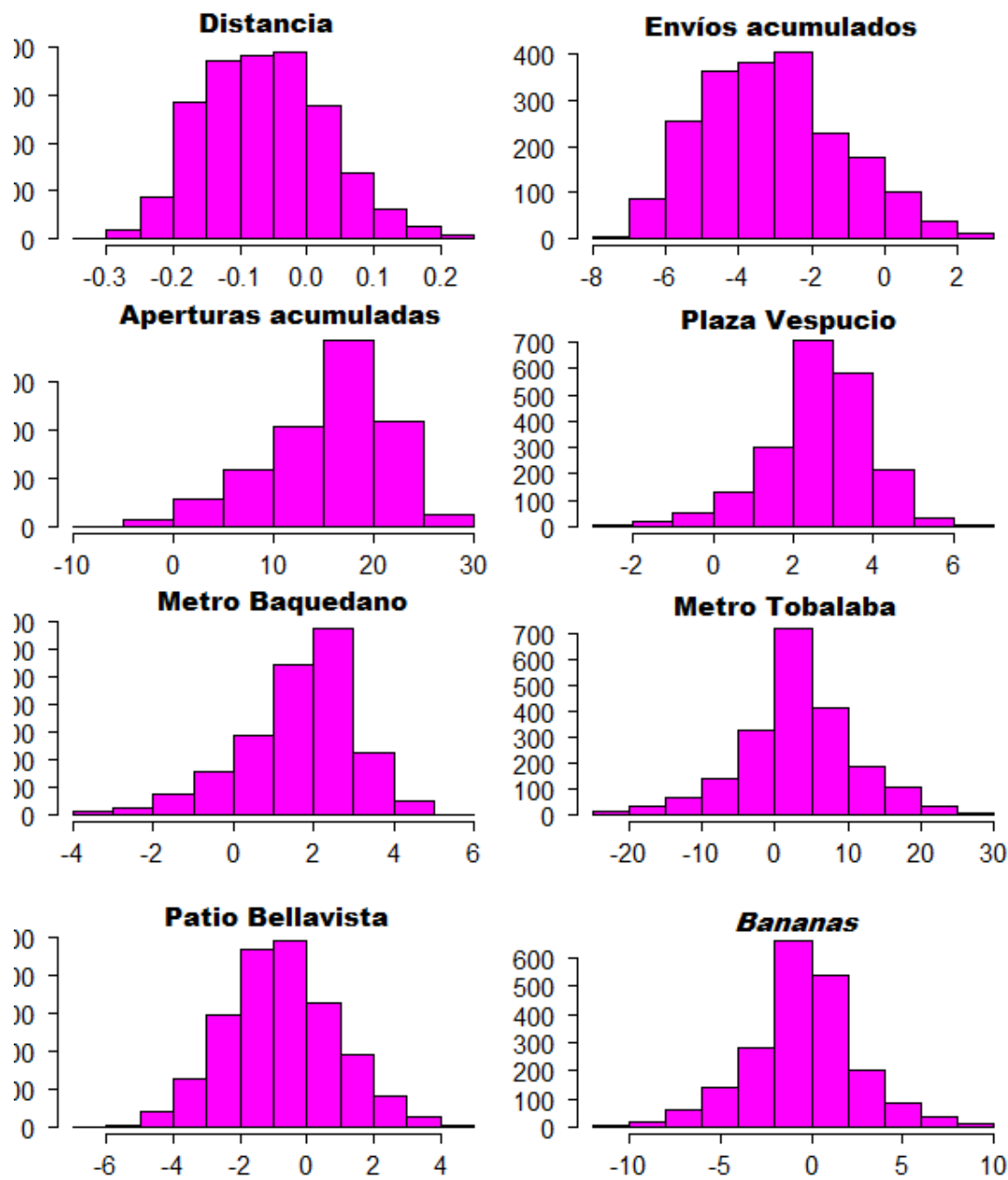


Ilustración 25 - Promedio de posterior de coeficientes (distancia)

Obteniendo un valor de log verosimilitud de -712.8, es posible observar que existe gran heterogeneidad en cuanto a las preferencias por la distancia de los usuarios. La intuición indica que la distancia, en general, tiene un efecto negativo en cuanto a cómo los usuarios responden a los estimulados entregados por las compañías de avisaje, sin embargo, la distribución del parámetro de distancia desplegado en la Ilustración 25 muestra que existe una porción de la población que no valora de manera negativa dicha variable. No así el parámetro de acumulación de aperturas, donde se logra apreciar que la distribución es asimétrica y altamente inclinada hacia coeficientes positivos, esto es, a medida que los usuarios abren promociones, su disposición para abrir próximas aumenta (en promedio). De igual forma, se aprecia que, y en línea con lo expuesto a lo largo del estudio, la preferencia por la acumulación de envíos en los usuarios es principalmente negativa. En

la Tabla 23 de anexos se despliegan los promedios para cada uno de los coeficientes a lo largo de todos los usuarios.

La Tabla 12 muestra los valores promedio obtenidos para el vector de coeficientes Δ que caracteriza a la distribución de las preferencias individuales de cada usuario (en rojo se muestran los parámetros no significativos a un 95% de confianza y en anexos se detallan las desviaciones estándar de cada uno).

VARIABLES	CONSTANTE	GÉNERO (H)	CONTADOR BANANAS	EDAD	TIEMPO	ENVÍOS	APERTURA	ACEPTACIÓN
Distancia	-0,0000	-0,0000	0,0001	0,0000	0,0000	0,0000	-0,0000	0,0000
Envíos acumulados	0,0029	-0,0335	0,2672	0,6766	-0,2111	1,4631	-0,2514	-1,1439
Aperturas acumuladas	0,0000	0,0003	-0,0009	0,0012	-0,0001	0,0011	0,0005	0,0002
Bananas	-0,0018	-0,0529	0,4015	0,0968	0,0104	-0,1958	-0,0900	0,0137
Plaza Vespucio	-0,0017	-0,0802	0,4100	0,0733	0,0567	0,0609	-0,0220	-0,0084
Metro Baquedano	-0,0003	0,0166	0,0822	0,0069	0,0064	-0,0365	-0,0175	-0,0240
Metro Tobalaba	0,0012	-0,0049	-0,2126	-0,0386	-0,0450	-0,0053	0,0147	0,0189
Patio Bellavista	-0,0173	0,0921	0,1081	0,3678	0,1681	3,6683	-0,7912	-1,2357

Tabla 12 - Valores promedio Δ (distancia)

Esta vez, los envíos y aperturas acumuladas parecen no estar moderados en gran medida por variables demográficas de los usuarios. Respecto a esta última, la tabla indica que el único factor demográfico que influye de manera significativa es la edad, manteniendo la misma tendencia de lo ocurrido para el caso sin distancia: A medida que aumenta la edad de los usuarios, estos valoran en mayor medida la acumulación de aperturas como moderador de su comportamiento.

Respecto al coeficiente de distancia, se logra apreciar que intrínsecamente su valor es negativo (constante menor que cero y significativa), mientras que si un usuario es de género masculino el valor del coeficiente disminuye aún más, El contador de *bananas* histórico favorece la sensibilidad del usuario hacia la variable distancia, por lo que (*ceteris paribus*), al aumentar de edad los usuarios valoran en mayor medida la distancia como variable moderadora de la apertura. De igual forma, el coeficiente asociado a las métricas restantes de actividad de los usuarios (tiempo desde el registro y envíos históricos) revelan que (*ceteris paribus*) a medida que un cliente amplía su relación con la aplicación, su sensibilidad por la distancia aumenta.

La interpretación del resto de los coeficientes demográficos asociados se torna difusa, debido a la no significancia de los parámetros.

8. Discusión y recomendaciones

8.1. Modelo logit con clases latentes (sin distancia)

A lo largo del estudio se ha desarrollado una serie de modelos que tienen como objetivo principal el entender qué variables moderan la apertura de promociones en un contexto geo localizado.

Del modelo con 2 clases latentes propuesto en 6.2.1 es posible caracterizar a los usuarios de la aplicación, asignándolos a un segmento en particular y analizando el comportamiento de cada clase como un conjunto. Lo anterior es un requerimiento explícito por parte de la empresa, esto es, Izit busca clusterizar a sus usuarios en base a atributos de modo de realizar acciones de marketing focalizadas.

Observando la Tabla 9 es posible caracterizar en primera instancia a los usuarios pertenecientes a cada clase:

Segmento 1	No es del todo sensible a la acumulación de aperturas.
	Tolera en mayor medida la acumulación de envíos.
	Valora los <i>fences de paso</i> negativamente respecto a Mall Costanera Center. Plaza Vespucio y Costanera los considera de igual manera.
	Hombres disminuyen la probabilidad de apertura.
	A mayor edad, disminuye la probabilidad de apertura.
	Valora en gran medida los envíos asociados a <i>bananas</i> .
	Segmento 2
Más sensible a la acumulación de aperturas.	
No tolera la acumulación de envíos.	
Valora los <i>fences de paso</i> positivamente. El resto al mismo nivel que Costanera Center.	
Indiferente respecto al género.	
A mayor edad, es más probable que abran.	
Indiferente frente a promociones asociadas a <i>bananas</i> .	

Tabla 13 - Caracterización de clases latentes

La Tabla 13 permite caracterizar al segmento 1 como un grupo de usuarios que tolera en gran medida los estímulos enviados por la compañía. Si bien el acumular envíos no es un problema para dichos usuarios en comparación al otro segmento, tampoco el acumular aperturas hace que abran en gran medida promociones actuales

Otro aspecto diferenciador de dicho segmento es su grado de valoración respecto a los envíos asociados con el club de fidelización de la compañía, lo cual habla de la estrecha relación que podrían tener los usuarios de este segmento con Izit.

Finalmente, respecto a los *fences*, el segmento 1 parece valorar de manera negativa los *fences de paso* (Metro Tobalaba y Metro Baquedano), además de valorar negativamente el *fence* asociado a Patio Bellavista. Lo anterior permite concluir que los usuarios relativos a esta clase valoran en mayor medida Centros Comerciales (Mall Costanera Center y Mall Plaza Vespucio).

Por otro lado, el segmento 2 parece ser mucho más sensible tanto a la acumulación de envíos como a la acumulación de aperturas, además de conglomerar a usuarios que a medida que aumentan de edad se hace más probable que abran promociones. Los usuarios pertenecientes a este segmento valoran positivamente los *fences* asociados a Metro Baquedano y Patio Bellavista (en comparación con *fence* Costanera), mientras que los coeficientes del resto parecen ser valorados de igual forma que el *fence* base.

Finalmente, el segmento 2 parece ser indiferente respecto a los envíos asociados al club de fidelización de la compañía, lo cual podría ser indicio de un bajo nivel de involucramiento de estos con la compañía.

Usando la expresión de la **¡Error! No se encuentra el origen de la referencia.**, es posible alcular la probabilidad de pertenencia a un segmento en particular para cada usuario de la base de datos, de modo de poder definir un umbral de corte y así asignar a un usuario a una clase si y solo si supera dicho umbral. Para efectos del estudio, dicho umbral fue 0.7, esto es, en caso que la probabilidad de pertenencia de un usuario supere dicho valor, se asignó a la clase 1.

Así, se obtuvo una participación del 79% para el segmento 1 y de un 21% para el segmento 2. Además, analizando las variables demográficas de los usuarios para cada segmento es posible apreciar un comportamiento acorde a lo analizado en la Tabla 13, pues el segmento uno posee métricas de actividad superiores a los usuarios pertenecientes al segmento 2.

Segmento	Contador de bananas	Tiempo	Envíos	Apertura	Aceptación	Género (hombre)
1	131,16	35,67	73,63	22,02	0,27	50,6%
2	54,50	25,36	32,28	9,21	0,17	47,3%

Tabla 14 - Variables demográficas promedio por segmento

La Tabla 14 indica que el segmento 1 efectivamente conglojera a un grupo de usuarios que posee un nivel de relación superior con la compañía respecto al segmento 2 en todas las métricas promedio consideradas. Así, el segmento 1 recibe más del doble de promociones que el segmento 2, abre una cantidad de promociones mucho mayor, posee (en promedio) una mayor cantidad de *bananas* históricas, etc. Sin embargo, los valores promedio no son del todo informativo para caracterizar a los usuarios (podría existir mucha varianza).

Efectivamente, una forma más clara de visualizar lo anterior es mediante los histogramas de las métricas demográficas a lo largo de todo el segmento. Así, es posible visualizar en la Ilustración 26 que el segmento 1 posee clientes que en promedio reciben mucho más en relación al segmento 2, mientras que de igual forma parece ser mucho más intenso en relación a las aperturas históricas que los usuarios realizan.

En anexos se encuentran un gráfico similar para el tiempo hasta el primer envío de cada usuario (Ilustración 46 de anexos), mostrando que el segmento 1 acumula una cantidad mucho mayor de usuarios que llevan una gran cantidad de semanas con la aplicación instalada.

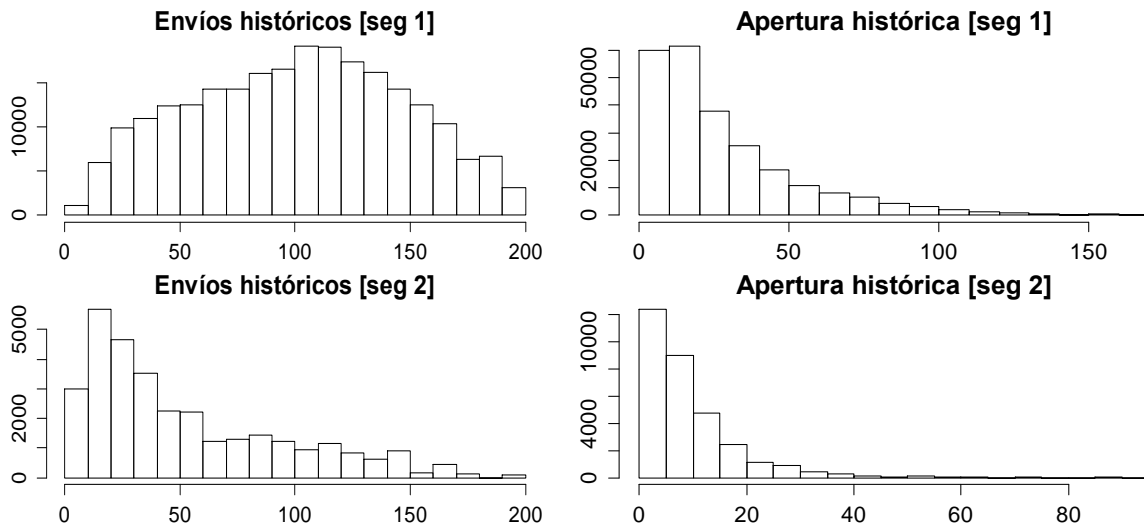


Ilustración 26 - Comparación envíos/aperturas históricas por segmentos

Respecto a las categorías que cada segmento abre, la Ilustración 28 muestra que, si bien ambos segmentos suelen abrir promociones asociadas a *compras con beneficio* (si consume un producto accede a un beneficio en otro producto/servicio asociado) y *dos por uno* (lleva dos productos por el precio de uno), las tasas de apertura⁷ entre segmentos difieren considerablemente respecto al resto. Así, sumado a que el segmento 1 prefiere en demasía las promociones asociadas a *bananas*, una estrategia por parte de la compañía podría ser enviar mayor cantidad de promociones de las categorías mencionadas al segmento 1 agregando un estímulo con *bananas* para aumentar las tasas de apertura.

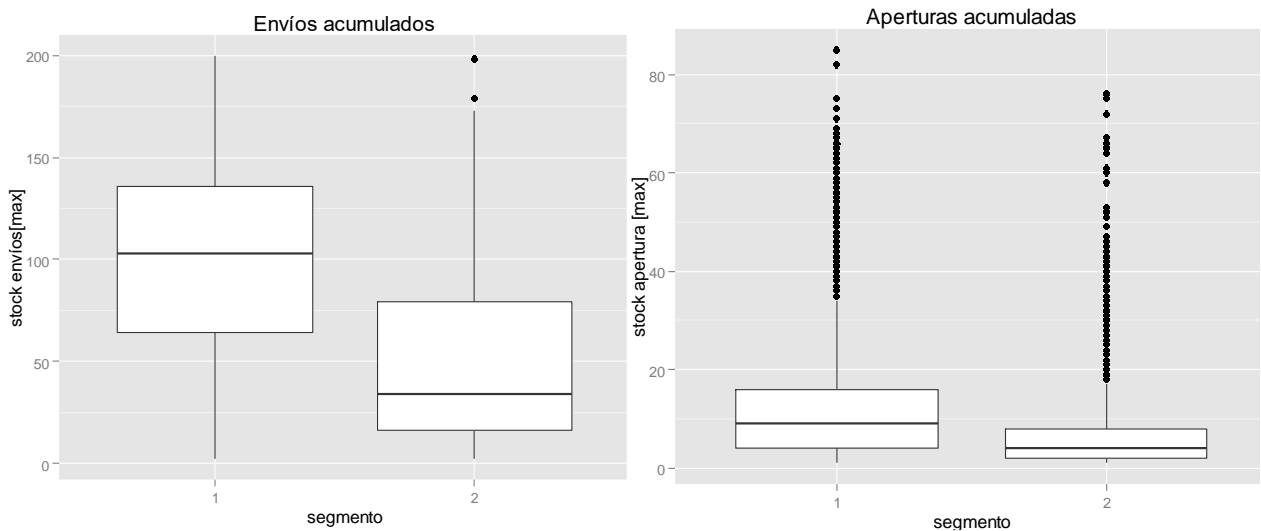


Ilustración 27 - Comparación envíos/aperturas acumuladas por segmento (clases latentes)

⁷ Calculada para cada categoría, esto es, dividiendo la cantidad de aperturas de cada categoría sobre el total de promociones enviadas de esa categoría

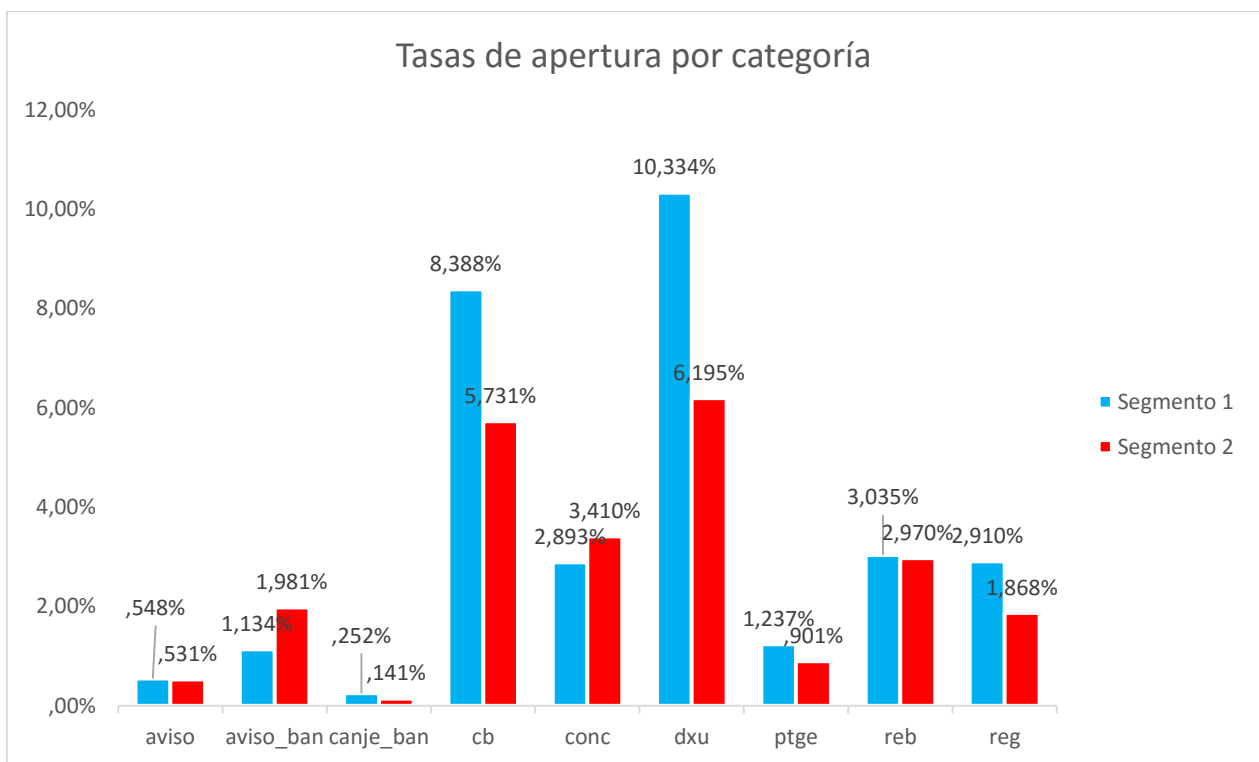


Ilustración 28 - Tasas de apertura por categoría y segmentos

Al observar la distribución de los valores máximos que alcanzan los envíos y aperturas acumulados para los usuarios de la muestra (Ilustración 27), se observa que el efectivamente el segmento 1 es más tolerante a recibir promociones que el segmento 2, mientras que respecto a la apertura acumulada el segmento 1 parece contener a clientes que acumulan mayor cantidad de aperturas. Si bien se esperaría que el segmento 2 tenga una mayor proporción de usuarios con alta acumulación de aperturas (debido a que dicho segmento valora en mayor medida el coeficiente asociado), lo anterior puede estar ocasionado debido a que la apertura es condicional al envío, por lo que proporcionalmente el segmento 1 posee mayor cantidad de posibles usuarios para abrir promociones.

Una conclusión importante de este modelo es que, es posible aumentar el alcance de promociones utilizando al segmento 1. Dado que dichos usuarios no son del todo sensibles a la acumulación en de envíos en cuanto al efecto que tienen dicha variable sobre las tasas de apertura, es posible enviar promociones no geo localizadas cuando este segmento tenga la aplicación abierta (en desmedro del otro segmento, al cual el acumular envíos disminuye su probabilidad de apertura).

8.2. Modelo jerárquico bayesiano

Como se vio en la sección 6.3, el enfoque jerárquico bayesiano permite estimar la distribución de las preferencias de los usuarios por los distintos atributos considerados en el modelo, incluyendo un nivel demográfico para explicar dichas preferencias. Así, una de las ventajas que supone este enfoque es la caracterización individual de cada usuario de la aplicación en cuanto a la heterogeneidad en su comportamiento.

A modo de ejemplo, la Ilustración 29 muestra la distribución del coeficiente de envíos acumulados para una muestra de usuarios de la aplicación. En dicha ilustración es posible apreciar la heterogeneidad en los usuarios en cuanto al coeficiente en cuestión, logrando identificar individuos que, a lo largo de las 500 muestras de coeficientes, poseen valores en general negativos. Sin embargo, de igual manera es posible identificar usuarios que no son del todo sensible a la acumulación de envíos, más aún, se logra apreciar individuos que en general poseen un coeficiente positivo, lo cual indica que dichos usuarios disfrutaron de recibir promociones y ven aumentada su probabilidad de apertura al recibir una mayor cantidad de estímulos por parte de la compañía.

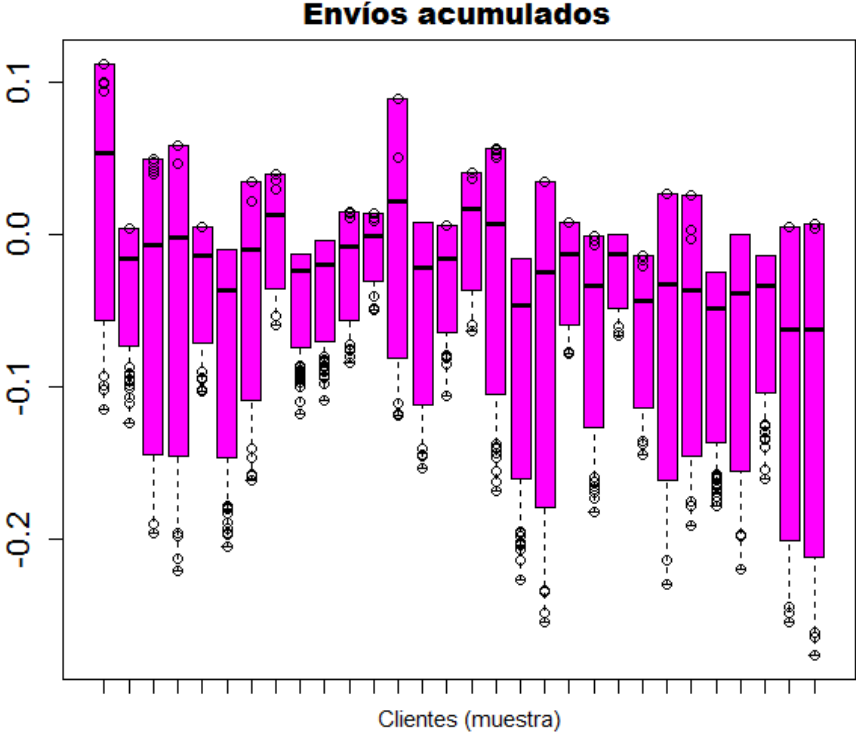


Ilustración 29 - Distribución de envíos acumulados (muestra clientes)

Gráficas similares para los coeficientes asociados a las aperturas acumuladas, *fences* y *bananas* se encuentran en la sección de anexos.

A modo de dejar de manifiesto la heterogeneidad capturada por el modelo, la Ilustración 30 muestra la distribución de los coeficientes asociados a la acumulación de envíos y a la acumulación de aperturas para un conjunto de 4 usuarios de la aplicación.

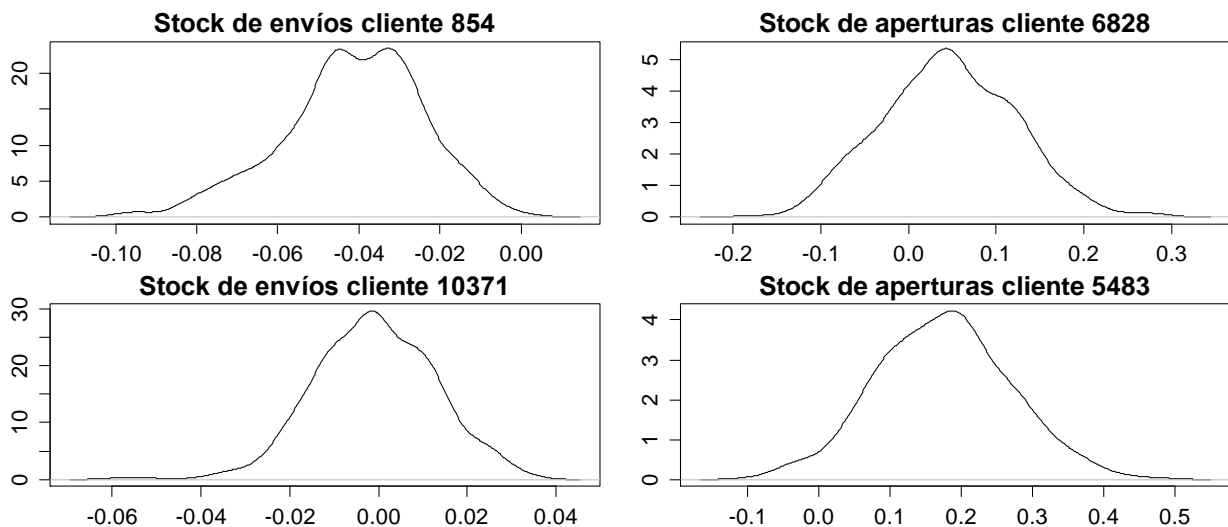


Ilustración 30 - Distribución acumulación de envío y aperturas (muestra usuarios)

En este gráfico, es posible ver que el usuario 854 penaliza en mayor proporción la acumulación de envíos en comparación con el usuario 10.371 (primera columna de la Ilustración). Así, la compañía ha de saber que el segundo usuario tolera en mayor medida la acumulación de envíos, por lo que una forma de aprovechar este hecho es mediante el envío de promociones no geo localizadas que permitan aumentar el alcance asociado a los clientes de Izit.

Por otro lado, la Ilustración 48 de anexos muestra que en general el coeficiente asociado a la acumulación de aperturas es positivo para todos los usuarios que interactúan con la aplicación en el horizonte temporal considerado, sin embargo se logra apreciar que aún existe heterogeneidad en cuanto a las preferencias por dicho indicador en los usuarios. Así, la segunda columna de la Ilustración 30 refleja que, si bien ambos clientes considerados poseen una distribución que toma en general valores positivos, el usuario 6.828 posee un grado de intolerancia a la acumulación de aperturas relativamente menor en comparación con el usuario 5.483.

Es posible realizar un análisis similar respecto a los *fences* y como estos afectan en la probabilidad de apertura de los usuarios. Desde la Ilustración 49 a la Ilustración 52 en anexos, es posible observar que de igual forma existen preferencias heterogéneas por recibir envíos en ciertos *fences*, por lo que se podría asignar un orden de cada uno de estos para todos los usuarios, de modo de poder entender de mejor manera el comportamiento del conjunto de clientes considerado.

Si bien con este enfoque es posible caracterizar a cada usuario respecto a sus preferencias por las distintas variables consideradas en el modelo, una alternativa que ayuda a focalizar de manera más operativa los esfuerzos de marketing por parte de la compañía es la segmentación. Una forma de clusterizar a los usuarios de la aplicación es mediante sus preferencias por cada uno de los atributos en cuestión.

Así, para cada usuario se obtuvo el valor promedio de los 500 coeficientes muestreados por el método jerárquico bayesiano y se aplicó el método k-medias para clusterizarlos a partir de dichos atributos. Para calcular el número de segmentos a calibrar se usó la regla del codo [16] obteniendo como resultado tres segmentos de tamaño 17%, 30% y 83%, mientras que la Ilustración 53 de anexos representa gráficamente las componentes discriminantes de la clusterización realizada.

Respecto a la caracterización por variables demográficas, en cuanto a los promedios, la Tabla 15 muestra que el segmento 2 posee mayor actividad en cuanto a cantidad de aperturas históricas, contador de *bananas* y tiempo entre el registro y el primer envío. Sin embargo, nuevamente resulta poco ilustrativo analizar solo los promedios sin tomar en cuenta la dispersión de los datos.

Segmento	Contador de bananas	Tiempo	Envíos	Apertura	Aceptación	Género (hombre)
1	79,51	33,24	108,58	12,84	0,16	52,6%
2	304,60	43,84	90,28	55,35	0,48	41,9%
3	89,93	31,53	44,18	14,45	0,23	52,3%

Tabla 15 - Variables demográficas promedio por segmento (cluster jerárquico)

En la Ilustración 31 se logra apreciar que a nivel de envíos, tanto el segmento 1 como el segmento 2 parecen ser bastante activos en comparación con el segmento 3, el cual acumula una mayor masa en la cola izquierda y menor en la cola derecha de la distribución.

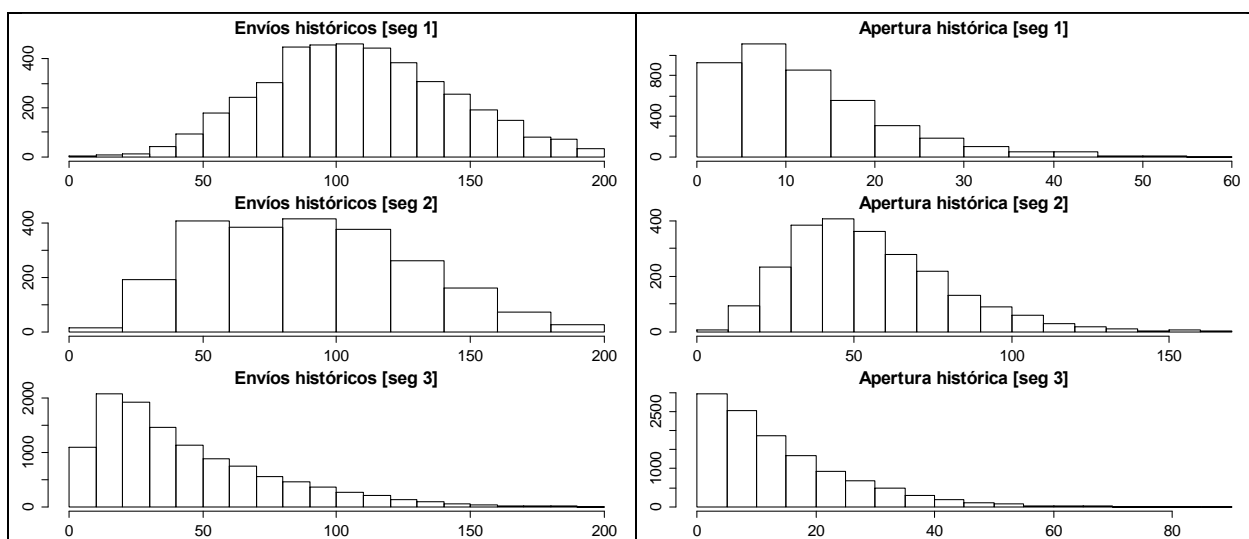


Ilustración 31 - Comparación envíos/aperturas históricas por segmentos (cluster jerárquico)

Por otro lado, respecto a la cantidad de aperturas históricas por parte de los usuarios, el segmento 2 parece ser el que posee usuarios con una mayor cantidad de aperturas a lo largo de su relación con la compañía en el horizonte temporal considerado, mientras los segmentos 1 y 3 poseen distribuciones más balanceadas hacia la izquierda, esto es, usuarios con menor cantidad de aperturas históricas.

El resto de las métricas demográficas parece no ser del todo concluyentes. Para ver más en detalle, consultar la sección de anexos.

La clasificación propuesta también permite observar diferencias en cuanto a los valores que toman las variables de interés, en particular, los envíos y aperturas acumuladas por los usuarios. La Ilustración 32 muestra una comparación entre segmentos tanto de los envíos acumulados como las aperturas acumuladas, donde para la construcción de los gráficos se calculó, para cada usuario, el máximo nivel de dichos atributos.

La Ilustración muestra que, el segmento 3 concentra en general a usuarios que poseen pocos envíos acumulados, lo cual habla sobre el bajo nivel de actividad de estos para con la aplicación

(pues, dicha métrica contiene información de todos los *fences* en el horizonte temporal impuesto); mientras que parecen no existir diferencias tan marcadas entre los segmento 1 y 2.

Por otro lado, en cuanto a la cantidad máxima de aperturas que acumulan los usuarios, el segmento 2 es el que posee una mayor cantidad, mientras que los segmentos 1 y 2 acumulan una cantidad muy por debajo.

Lo anterior indica que, al parecer el segmento 1 recibe muchas promociones (lo cual podría estar causado por usuarios que siempre tienen la aplicación abierta), sin embargo abren una cantidad por debajo en comparación a los envíos que reciben, mientras que el segmento 3 recibe una cantidad baja de promociones y abre muy poco, de igual manera. Esto permite a la compañía identificar un segmento de la población que tiene una disposición positiva a la acumulación de envíos, pero que no está abriendo promociones en la misma proporción. Así, es posible realizar acciones de marketing a este segmento en particular, de modo que aumente su tasa de apertura promedio y logre interactuar de mejor manera con la aplicación.

Sin duda que los incentivos que puede generar la compañía para tales efectos están ligados al contenido mismo de las promociones enviadas. Una manera plausible de enfocar dichos esfuerzos es simplemente enviando a cada uno de estos usuarios promociones parecidas a las que abrió con más frecuencia en el pasado o en alguna ventana temporal sensata.

Por ejemplo, una simple búsqueda de la actividad del usuario 386 (perteneciente al segmento 1) indica que el 60% de las promociones que abrió están asociadas a la marca ChiarOscuro (cafetería), por lo que una buena idea para incentivar las aperturas de dicho cliente sería usar dicha marca.

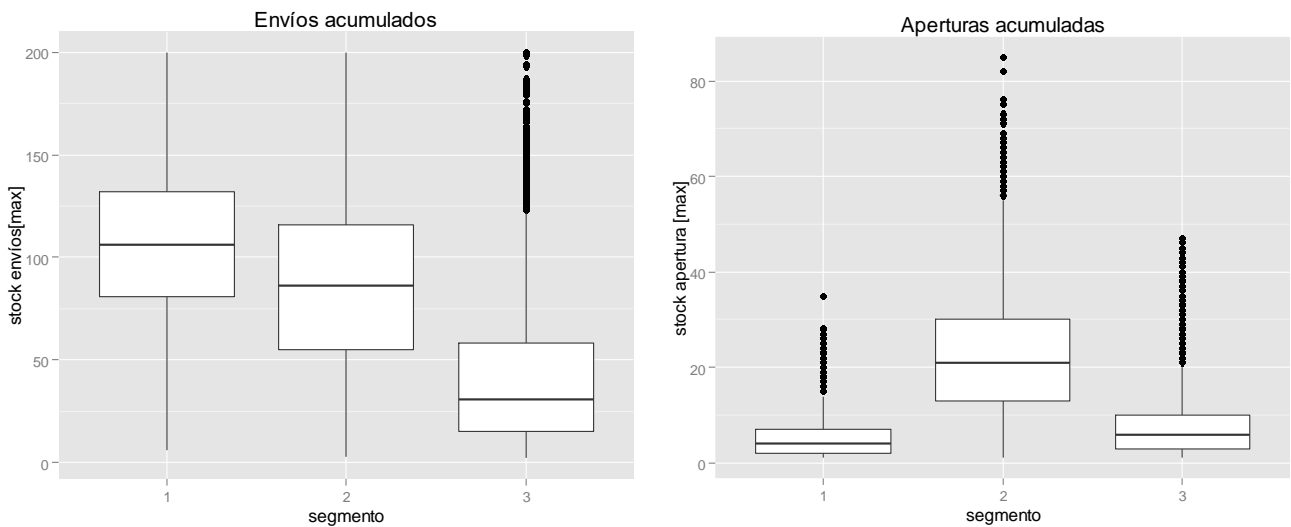


Ilustración 32 - Comparación envíos/aperturas acumuladas por segmentos (cluster jerárquico)

Si bien la caracterización anterior permite identificar parcialmente a grupos de usuarios, queda de manifiesto que las métricas resultantes no presentan gran diferenciación en cuanto a las métricas presentes. Así, y para comparar de mejor manera con el modelo de clases latentes de la sección 6.2.1, se decidió calibrar un modelo k-means con dos clases latentes.

El algoritmo tiene como resultado 2 segmentos de tamaño 67% y 33% (en anexos se presentan gráficos de sus componentes discriminantes) y logra capturar un comportamiento mucho más diferenciado en comparación con el modelo de tres segmentos. En la Ilustración 33 se logra apreciar que el segmento 1 (el de mayor tamaño) acumula una menor cantidad de envíos en comparación con el segmento 2, sin embargo el segmento 2 posee una menor disposición a acumular aperturas. Así, y en contraste con el modelo de clases latentes, se identifica un grupo de usuarios que gusta de recibir promociones, pero sin embargo no las abre activamente.

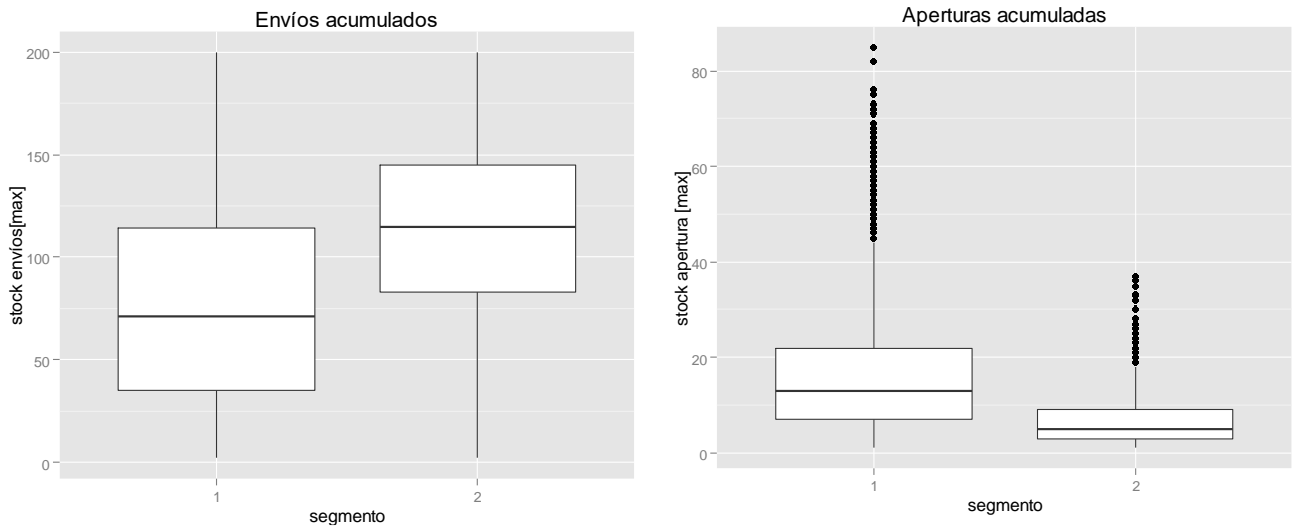


Ilustración 33 - Comparación envíos/aperturas acumuladas por segmentos - 2 clases (cluster jerárquico)

Dicha distinción es importante, pues el modelo detecta una oportunidad de aumentar las tasas de apertura para aquellos usuarios que están dispuestos a recibir una cantidad considerable de promociones. Una forma de realizar lo anterior podría ser personalizando más aún el mensaje que el usuario recibe en la bandeja de entrada o hacer uso de los mensajes con mayor tasa de apertura por el usuario en cuestión en el pasado para hacer del envío aún más relevante.

8.3. Modelo jerárquico bayesiano (incluida variable distancia)

En esta sección se realizará un análisis similar al de la sección 8.2, pero con la salvedad que se incluye la variable asociada a la distancia de un usuario al *fence* por el cual la promoción se gatilla. Si bien se obtienen distribuciones para todos los coeficientes asociados a la especificación planteada en 7.4, se pondrá foco solo en las distribuciones asociadas al coeficiente de distancia.

Como se logra apreciar en Ilustración 34, existe heterogeneidad en cuanto a las preferencias de los usuarios por el atributo distancia. Si bien es de esperar que en general el coeficiente posea un signo negativo, existen algunos usuarios de la aplicación que en promedio poseen una valoración positiva o cercana a cero respecto a la distancia. Esto es, dichos usuarios son indiferentes a la cercanía al momento decidir si abrir o no una promoción enviada por Izit.

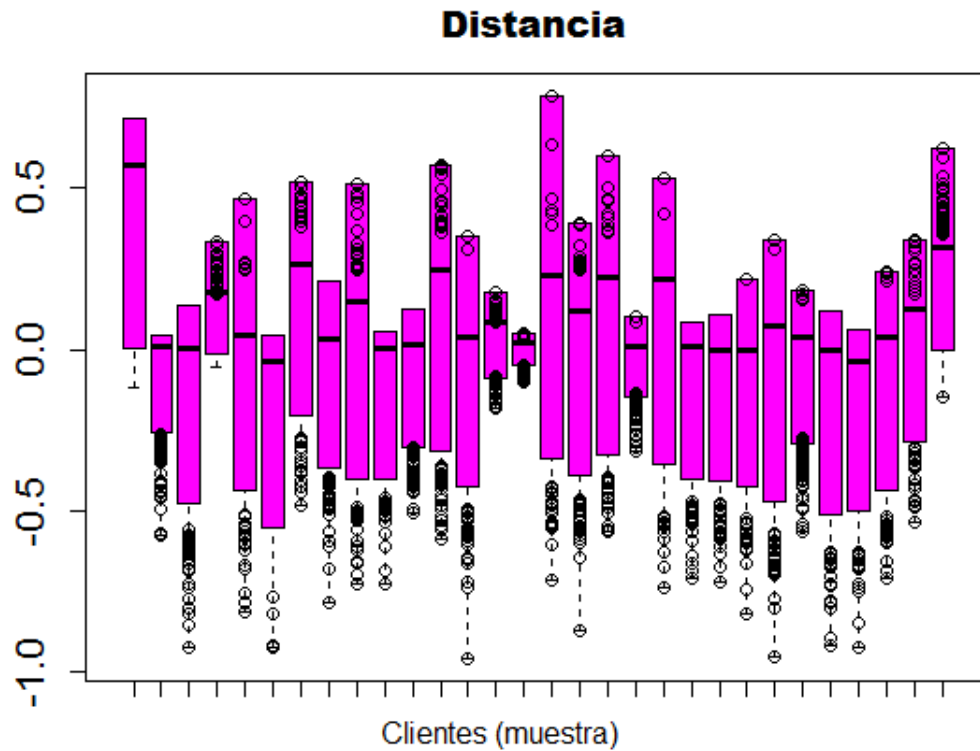


Ilustración 34 - Distribución de distancia (jerárquico) (muestra clientes)

Para expresar gráficamente la heterogeneidad presente en la modelación incluyendo la cercanía del usuario al *fence*, la Ilustración 35 despliega la distribución de los coeficientes asociados a distancia, acumulación de envíos y acumulación de aperturas.

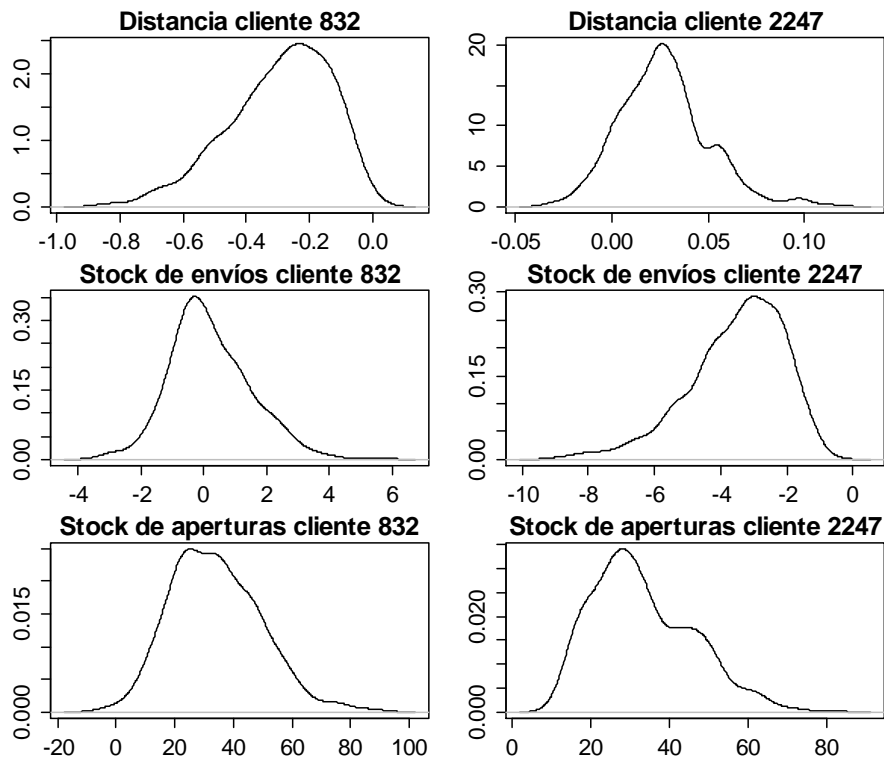


Ilustración 35 - Distribución de variables modelo con distancia (jerárquico) (muestra usuarios)

Al observar la gráfica se puede evidenciar la diferencia que existe en cuanto a las preferencias por la distancia de los usuarios en cuestión. El cliente 832 valora de forma negativa la excedida distancia entre él y el *fence* (en la mayoría de las muestras generadas), mientras que el cliente 2247 parece bordear, en promedio, la indiferencia en cuanto a dicha métrica.

Si bien el cliente 2247 considera que la distancia no es un atributo del todo relevante en su utilidad, ocurre totalmente lo contrario al observar el coeficiente de acumulación de envíos, pues la distribución de dicho coeficiente se concentra principalmente en valores negativos, y en consecuencia, el usuario en general tiene una valoración negativa en cuanto al efecto de la acumulación de envíos en su utilidad. Mientras que por otro lado, el cliente 832 tolera en mayor medida la acumulación de envíos, distribuyendo sus preferencias en torno al valor nulo.

Así, una manera de usar dichos resultados para aumentar el alcance de las promociones enviadas por la compañía es identificar a los usuarios que poseen una sensibilidad menor a la distancia y además toleran la acumulación de envíos (como es el caso del usuario 832). De esta forma, es posible gatillar promociones no geo localizadas a dichos usuarios, aumentar el alcance de las marcas asociadas a la compañía y satisfacer en mayor medida a los clientes de Izit.

9. Conclusiones y trabajos futuros

9.1. Conclusiones

A lo largo del estudio se han planteado una serie de modelos que intentan explicar la manera en que se comportan las tasas de apertura de los usuarios de la aplicación en el horizonte temporal considerado.

Respecto al modelo logit sin heterogeneidad, es posible concluir que de las variables continuas por las cuales se controla, la acumulación de aperturas pasadas es el factor que más importa en cuanto a las chances de abrir una promoción en el presente, aumentando dicha probabilidad. Además, la acumulación de envíos resulta disminuir, en promedio, la probabilidad de apertura de una promoción, en línea con la idea que un exceso de envíos hacia los usuarios puede tener un efecto *spam* en el comportamiento de estos. En cuanto a los *fences*, en promedio el asociado a Costanera Center es el que promueve en mayor medida la apertura de promociones, mientras que, de igual forma, los envíos asociados a *bananas* aumentan la probabilidad de apertura.

En cuanto a los modelos de clases latentes y jerárquico bayesiano, si bien con ambos es posible segmentar a los usuarios de modo de analizar comportamiento conglomerado, se considera que el modelo con heterogeneidad discreta describe de mejor manera la situación en función de su parsimonia y la interpretabilidad de sus resultados. Sin embargo ofrece un abanico de posibilidades de análisis individual, pudiendo analizar cuáles son los factores que determinan en mayor medida las preferencias de los usuarios.

El modelo de clases latentes permite identificar dos tipos de usuarios en el set de datos disponible: un segmento que tolera en mayor medida los envíos excesivos y no es del todo sensible a la acumulación de aperturas, además de valorar en mayor medida los *fences* asociados a centros comerciales. En ese segmento los hombres disminuyen la probabilidad de apertura respecto a las mujeres, la edad actúa de forma negativa en dicha probabilidad y los envíos asociados a *bananas* parecen ser valorados en gran medida.

Una forma de caracterizar a este segmento es como usuarios que frecuentan centros comerciales y abren la aplicación en esas ubicaciones esperando obtener ofertas convenientes, además de ser un segmento que con un gran apego y conciencia de marca respecto a los envíos asociados a *bananas*.

El segundo segmento es mucho más sensible a la acumulación tanto de envíos como de aperturas, esto es, la excesiva recepción de envíos disminuye la probabilidad de abrir promociones, mientras que el acumular aperturas aumenta en demasía dicha probabilidad comparativamente con el segmento 1. Además, este segmento valora positivamente los *fences de paso*, es indiferente tanto al género como a los envíos asociados a *bananas* en cuanto a las tasas de apertura y a medida que aumenta la edad de sus usuarios, favorece la apertura de promociones.

Una forma de caracterizar a este segmento es como usuarios que frecuentan lugares de esparcimiento (*fences* asociados a Barrio Bellavista) y abren promociones precisamente asociadas al tipo de establecimientos ahí ubicados. Aparentemente es un segmento que no está interesado en el club de fidelización de Izit.

Dado que el segmento 1 no es del todo sensible a la acumulación de envíos, una posible estrategia para aumentar el alcance de los clientes de la compañía es enviar a dicho segmento promociones no geo localizadas que incluyan contenido relativo al club de fidelización.

El modelo jerárquico bayesiano, si bien no permite establecer una segmentación del todo clara basándose en los promedios de los coeficientes para cada usuario, sí permite entender qué variables demográficas moderan las preferencias de cada usuario al momento de enfrentarse a una promoción. El coeficiente asociado a la acumulación de envíos es inherente negativo, sin embargo cuando el usuario es hombre tiende a aumentar su valor, *ceteris paribus*. Así, los hombres en promedio parecen tener mayor tolerancia a la acumulación de envíos en relación con las mujeres, mientras que al aumentar la edad de los usuarios dicha tolerancia disminuye.

Respecto a coeficiente de acumulación de aperturas, es inherentemente nulo en promedio, sin embargo aumentan en la medida que el género del usuario es hombre. Al contrario del coeficiente de acumulación de envíos, la apertura está moderada positivamente por la edad, esto es, a medida que los años del usuario en cuestión aumentan su valoración por la acumulación de aperturas también lo hace, *ceteris paribus*.

Finalmente, respecto a la inclusión de la variable distancia, se logra apreciar que con la muestra considerada, en el modelo logit simple dicha métrica posee un efecto promedio negativo sobre la probabilidad de apertura. Esto es, a medida que la distancia entre el *fence* y el usuario aumenta, se hace menos probable que dicho usuario abra la promoción enviada. Sin embargo, al incluir heterogeneidad mediante un enfoque bayesiano, se logra apreciar que, si bien en promedio los usuarios valoran negativamente la distancia, existen un grupo de usuarios que hace caso omiso de esta métrica al momento de decidir abrir o no un envío. Lo anterior puede ser debido a la imposición de normalidad de los coeficientes, por lo que una posible alternativa para verificar robustez consiste en aplicar otro tipo de distribuciones a los parámetros (tipo log-normal, por ejemplo). Así, el modelo bayesiano logra identificar aquellos usuarios y la compañía podría ver en ellos un oportunidad para aumentar el alcance de las campañas de sus clientes.

9.2. Trabajos futuros

El trabajo desarrollado deja una serie de posibles líneas de investigación en cuanto al comportamiento de los usuarios se trata.

En primera instancia, queda de manifiesto que la acumulación de envíos y aperturas modera el aprendizaje de los usuarios a lo largo de las interacciones que tiene este con la aplicación, sin embargo dichos inventarios son *mentales*, en el sentido que los individuos podrían olvidar dicho inventario y comenzar a *acumular* nuevamente. Así, una posible extensión del modelo propuesto es considerar inventarios con umbrales, de modo que luego de un tiempo dicha acumulación se

actualice. Otra posible extensión al modelo es considerar inventarios desagregados por categorías, marcas o campañas.

Por otro lado, respecto a la inclusión de la medida de distancia en el modelo, es posible sensibilizar aún más la aproximación de ésta considerando un nuevo umbral de corte. Más aún, si no se posee información de localización de un usuario luego de un umbral, podría definirse un estado latente de inactividad para poder incorporar dichos efectos al modelo.

Si bien es posible sensibilizar y mejorar en cuanto a eficiencia computacional el cálculo aproximado de las distancia, es vital que por parte de la empresa se generen mediciones de distancia en la medida que se envían promociones a los usuarios. De esta forma, se contaría con las coordenadas exactas de los individuos al momento de recibir una promoción y no sería necesario su cálculo aproximado. Otra extensión en cuanto a distancia es adoptar un enfoque de regresión geográfica ponderada [18], de modo explorar cómo influye la ubicación en cuanto a latitud y longitud de los usuarios en las tasas de apertura.

Otro aspecto a considerar es el tiempo en el que se registran las mediciones de la compañía. Esto pues, la fecha con la que se cuenta es el instante en el que se envió la promoción al individuo y, por simplicidad, durante el presente trabajo se asumió que dicha fecha es un buen *proxy* del instante en que el usuario efectivamente abrió la promoción. Una extensión podría ser el controlar por el tiempo que transcurre desde que el estímulo fue enviado hasta que el usuario abre (en caso de abrir). Intuitivamente, se esperaría que en promedio el tiempo transcurrido sea inversamente proporcional a las tasas de apertura, sin embargo ha de existir bastante heterogeneidad a lo largo de los usuarios.

Respecto a los *fences*, dado que se manifestaron comportamientos diferenciados dependiendo de la ubicación de éstos, un posible estudio futuro podría ser el estudio de la efectividad de cada uno de dichos *fences* como función de la actividad comercial a su alrededor.

Finalmente, sin duda uno de los desafíos próximos con los datos de la aplicación es intentar controlar durante todas las fases de decisión de los usuarios: envío, apertura, recomendación y aceptación. En particular, es posible explotar los datos respecto a envíos recomendados y como se expanden éstos a través de las redes sociales del usuario en cuestión. La compañía posee datos de todas las acciones del usuario en cuanto a *rewhis*, por lo que se podría incorporar este efecto *social* en la modelación de tasas de apertura/aceptación/recomendación.

10. Bibliografía

- [1] «Pew Research Center,» Febrero 2014. [En línea]. Available: <http://goo.gl/1qEXdb>. [Último acceso: 2015].
- [2] «EMarketer,» Abril 2015. [En línea]. Available: <http://goo.gl/G9btJi>. [Último acceso: Noviembre 2015].
- [3] «Izit,» [En línea]. Available: <http://www.izitapp.com/negocios.html>. [Último acceso: Noviembre 2015].
- [4] «Izit,» [En línea]. Available: <http://www.izitapp.com/quesomos.html>. [Último acceso: Noviembre 2015].
- [5] F. Toledo, «Análisis del comportamiento de canje de promociones geo localizadas a través de smartphones,» 2015.
- [6] Banerjee y Dholakia, «Mobile Advertising: Does location based advertising work?,» *International Journal of Mobile Marketing*, 2008.
- [7] Spiekermann, Rothensee y Klafft, «Street Marketing: How proximity and context drive coupon redemption,» *Journal of Consumer Marketing*, pp. 280-289, 2011.
- [8] Molitor, Reichhart, Spann y Ghose, «Measuring the effectiveness of location-based advertising: A randomized field experiment (Doctoral dissertation, Working paper, Ludwig Maximilian University of Munich, Munich, Germany),» 2013.
- [9] Andrews, Luo, Z. Fang y Ghose, «Mobile crowdsensing,» 2014.
- [10] Tversky y Kahneman, «The framing of decisions and the psychology of choice,» *Science*, pp. 453-458, 1981.
- [11] «R Core Team,» [En línea]. Available: www.r-project.org.
- [12] «Smart Insights,» 2015. [En línea]. Available: <http://www.smartinsights.com/email-marketing/email-communications-strategy/statistics-sources-for-email-marketing/>.
- [13] «Map Developers,» [En línea]. Available: www.mapdevelopers.com/.
- [14] K. Train, *Discrete choice methods with simulation*, Cambridge university press, 2009.
- [15] M. Goic, Apunte curso Marketing II - Heterogeneidad discreta.
- [16] Rossi, Allenby y McCulloch, *Bayesian statistics and marketing*, John Wiley & Sons, 2012.

- [17] A. K. Jain, «Data clustering: 50 years beyond K-means,» *Pattern recognition letters*, pp. 651-666, 2010.
- [18] C. Brundon, «Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity,» *Geographical Analysis*, 1996.
- [19] A. K. Jain, «Data clustering: 50 years beyond K-means,» *Pattern recognition letters*, vol. 31, n° 8, pp. 651-666, 2010.

11. Anexos

A continuación se presenta la descripción detallada de las tablas obtenidas por medio de la empresa subyacente. En paréntesis se muestra la naturaleza de los identificadores de cada tabla, donde *P* indica una llave primaria y *F* una llave foránea.

- **Inbox [3.465.152 registros]:** Tabla que contiene información relacionada al mensaje en la bandeja de entrada de la aplicación de los usuarios Izit. Los registros de esta tabla son:
 - **messageId:** Corresponde al id del mensaje enviado. (*P*)
 - **whisperId:** Indica el id del whisper. (*F*)
 - **user_id:** Corresponde al id del usuario. (*F*)
 - **fenceid:** Indica el id del fence asociado al envío. (*F*)
 - **viewStatus:** Variable dummy que indica 1 si usuario abrió la promoción y 0 si no.
 - **inboxdate:** Fecha del envío.
 - **burned:** 1 si el envío fue aceptado y 0 si no.
 - **given:** 1 si el envío fue rewhisperado y 0 si no.
 - **received:** 1 si el envío fue recibido mediante rewhisper y 0 si no.
 - **downloaded:** 1 si el envío fue descargado en la aplicación y 0 si no.
- **Users [198. 285 registros]:** Tabla que entrega información asociada al usuario⁸. Los registros de esta tabla son:
 - **user_id:** Corresponde al id del usuario. (*P*)
 - **userBirthday:** Fecha de nacimiento del usuario.
 - **gender:** Género del usuario (M/F)
 - **device_brand:** Marca de Smartphone del usuario.
 - **user_categories:** Categorías marcadas por el usuario como preferenciales.*
 - **last_seen:** Fecha y hora de la última visita a la aplicación.*
 - **bananas_count:** Contador de bananas a la fecha.*
 - **current_latitude:** Latitud actual del cliente.*
 - **current_longitude:** Longitud actual del cliente.*
 - **near_fence:** Fence más cercano.
 - **distance_to_fence:** Distancia del usuario al fence más cercano.
 - **signup_date:** Fecha de primer inicio de sesión de usuario.
- **Whispers [188 registros]:** Tabla que contiene información sobre los mensajes enviados a cada usuario. Los registros de esta tabla son:
 - **whisper_id:** Id del whisper. (*P*)
 - **whisperCampaign:** Id de la campaña. (*F*)
 - **whisperContent:** Contenido del whisper.
 - **whisperContentShort:** Contenido resumido del whisper.
 - **whisperContentLong:** Contenido extendido del whisper.
 - **whisperTitle:** Título del whisper.

⁸ Asterisco indica que fechas fueron medidas con referencia el día de la consulta.

- **Campaigns [160 registros]:** Información relativa a las campañas. La utilidad de esta tabla es poder realizar el cruce con la información de **brand**. Los registros de la tabla son:
 - **whisperCampaign:** Id de la campaña. (*P*)
 - **campaignName:** Nombre de la campaña.
 - **ownerId:** Owner asociado.

- **Fences [1.960 registros]:** Información relacionada con los fences (Lugar geográfico con un radio definido que hace factible el envío de promociones cuando un usuario toma contacto con él). Los registros de la tabla son:
 - **fenceid:** Indica el id del fence asociado al envío. (*P*)
 - **lat:** Latitud del fence.
 - **lon:** Longitud del fence.
 - **radius:** Radio de alcance del fence.
 - **fenceName:** Nombre del fence.

- **UserPosition [50.602.198 registros]:** Tabla que indica información relativa a la posición instantánea de los usuarios.
 - **user_id:** Corresponde al id del usuario. (*F*)
 - **lat:** Latitud del usuario.
 - **lon:** Longitud del usuario.
 - **timeTrack:** Fecha.
 - **fence_lat:** Latitud del fence asociado.
 - **fence_lon:** Longitud del fence asociado.
 - **distance:** Distancia entre usuario y fence.
 - **bat:** Nivel de batería del teléfono.

- **Brand [45 registros]:** Tabla que contiene los nombres de las marcas asociadas a las campañas. Sus registros son:
 - **ownerId:** Owner asociado. (*P*)
 - **brand:** Nombre de la marca.

- **Burned_whispers [18.686 registros]:** Tabla con información relativa al canje de promociones
 - **messageId:** Corresponde al id del mensaje enviado. (*P*)
 - **storeuid:** Identificador de la tienda.
 - **trans_date:** Fecha de transacción.

- **User_whisper_gift [165.752 registros]:** Contiene información sobre rewhispers de usuarios
 - **whisper_id:** Id del whisper.
 - **fenceid:** Indica el id del fence asociado al envío.
 - **messageId_env:** Id del mensaje que envía rewhisper.
 - **messageId_rec:** Id del mensaje que recibe rewhisper.
 - **from_user:** Id del usuario que envía.

- **to_user:** Id del usuario que recibe.
- **date:** Fecha.

```
> pairwise.t.test(inbox$viewStatus,inbox$gender, pool.sd = FALSE)

Pairwise comparisons using t tests with non-pooled SD
data:  inbox$viewStatus and inbox$gender

      F
M <2e-16

P value adjustment method: holm
```

Ilustración 36 - Test medias indicador de apertura y género [anexos]

```
> pairwise.t.test(inbox$viewStatus,inbox$fenceName, pool.sd = FALSE)

Pairwise comparisons using t tests with non-pooled SD
data:  inbox$viewStatus and inbox$fenceName

Costanera Center Mall Plaza Vespucio Metro Baquedano Metro Tobalaba
Mall Plaza Vespucio 2e-10 - - -
Metro Baquedano <2e-16 <2e-16 - -
Metro Tobalaba 2e-06 <2e-16 <2e-16 -
Patio Bellavista <2e-16 <2e-16 2e-08 <2e-16

P value adjustment method: holm
```

Ilustración 37 - Test medias indicador de apertura y fences [anexos]

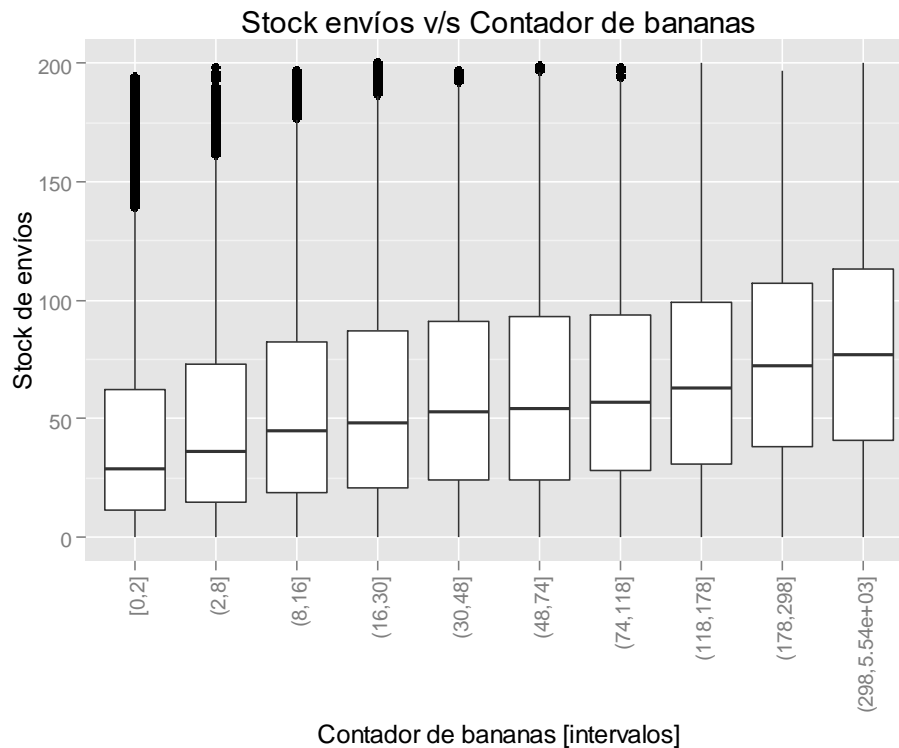


Ilustración 38 - Box plot envíos acumulados vs contador de bananas [anexos]

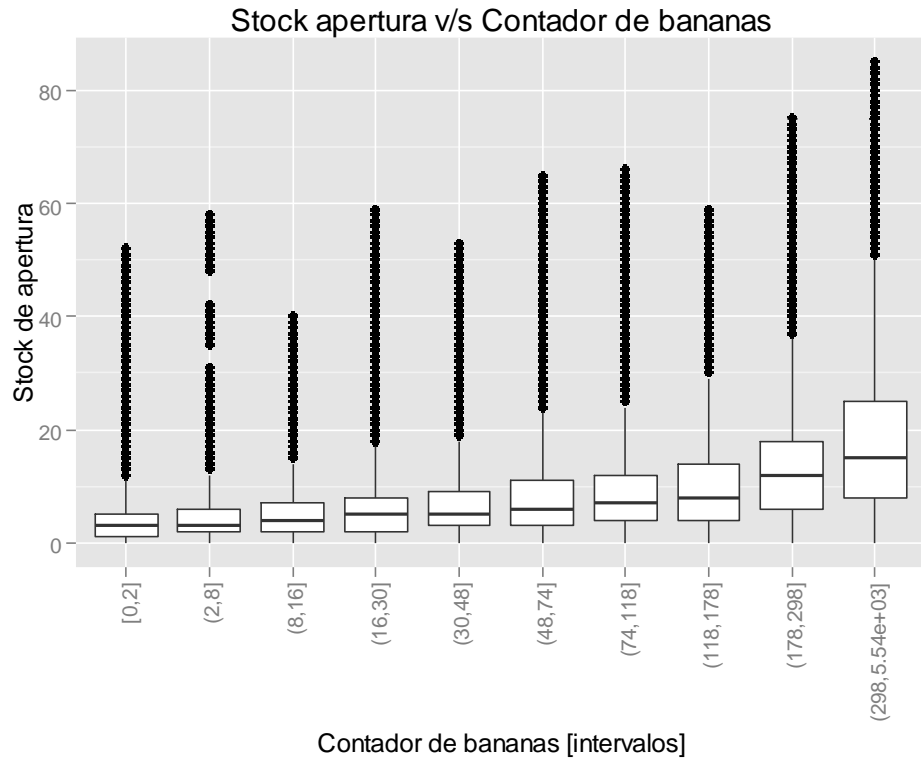


Ilustración 39 - Box plot aperturas acumulados vs contador de bananas [anexos]

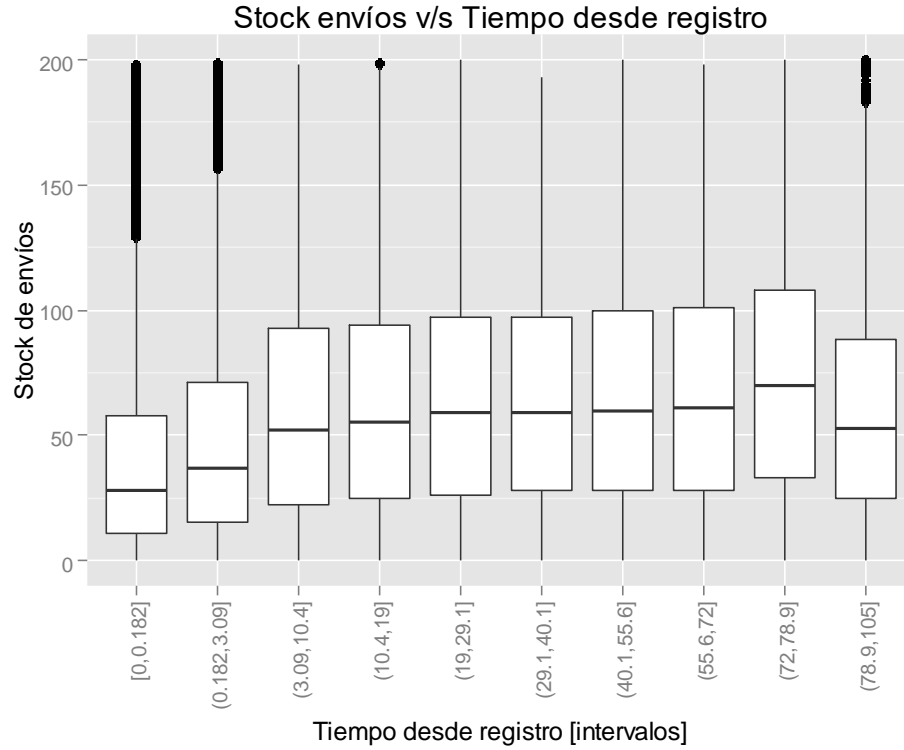


Ilustración 40 - Box plot envíos acumulados vs tiempo desde registro [anexos]

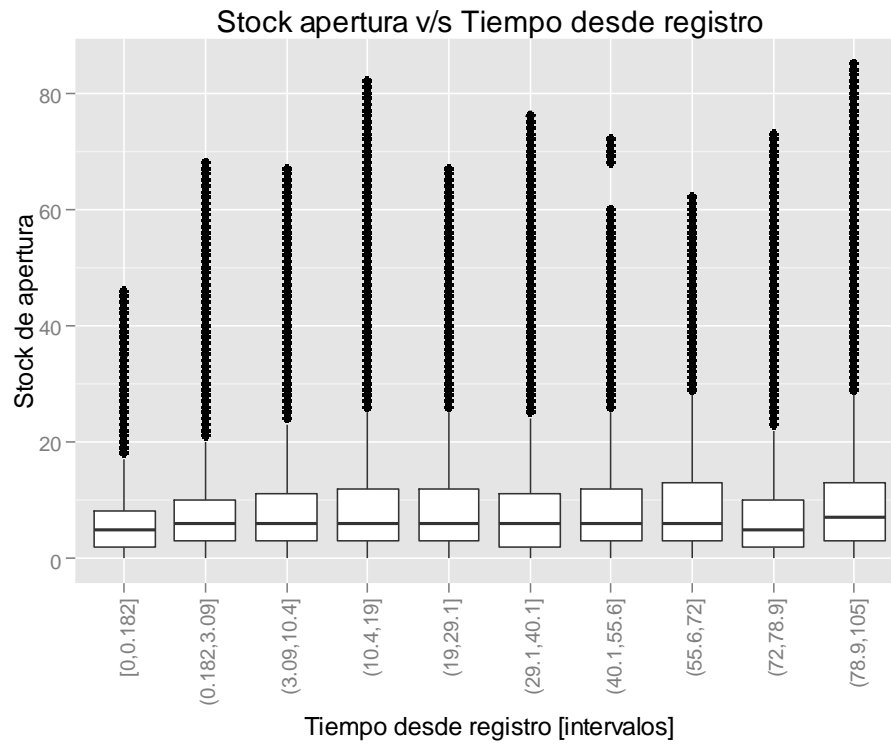


Ilustración 41 - Box plot aperturas acumuladas vs tiempo desde registro [anexos]

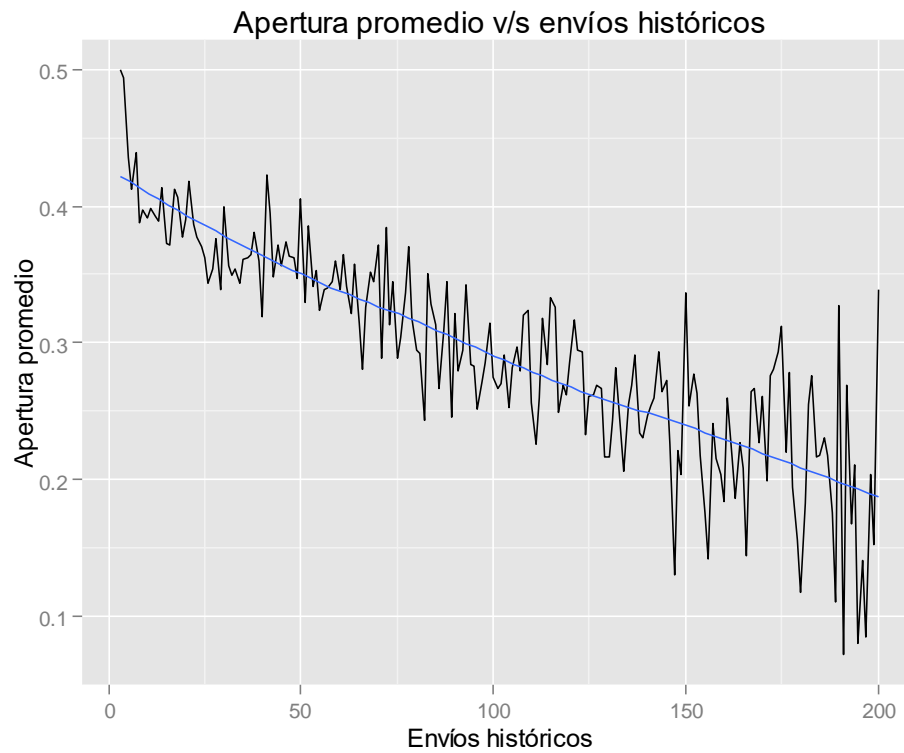


Ilustración 42 - Apertura promedio vs envíos históricos [anexos]

```
> pairwise.t.test(inbox$viewStatus,inbox$banana, pool.sd = FALSE)
      Pairwise comparisons using t tests with non-pooled SD
data:  inbox$viewStatus and inbox$banana
      ban
no ban <2e-16
P value adjustment method:holm
```

Ilustración 43 - Test medias indicador de apertura y bananas [anexos]

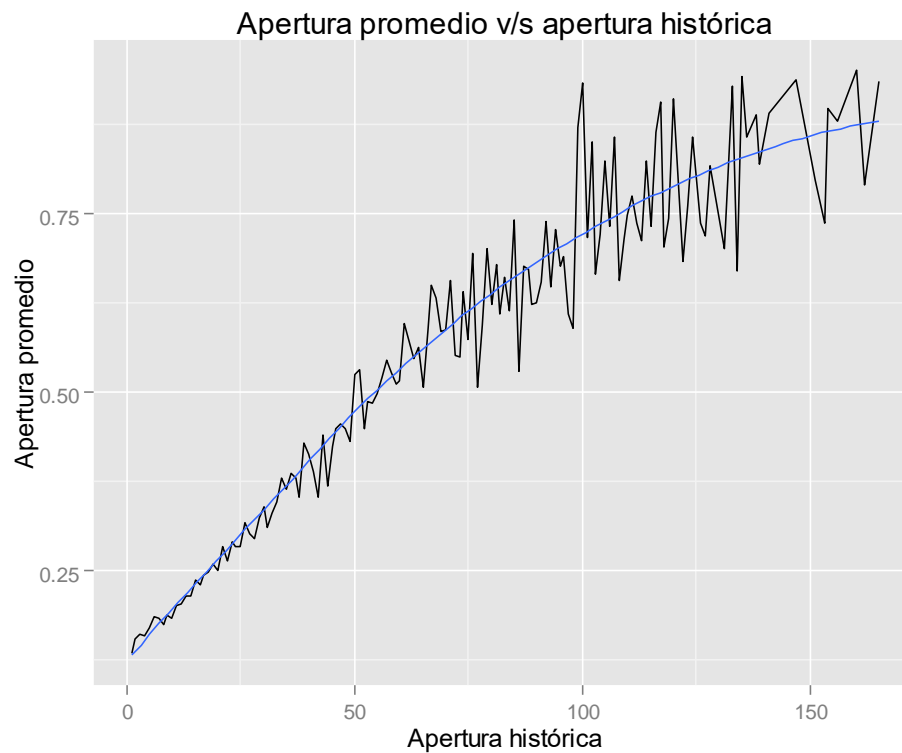


Ilustración 44 - Apertura promedio vs apertura histórica

Variables	Coficiente	Error estándar	Significancia
Envíos	-0,8713	0,0062	***
Apertura	1,1861	0,0062	***
Edad	-0,0270	0,0004	***
Plaza Vespucio	-0,2087	0,0151	***
Metro Baquedano	-0,6175	0,0159	***
Metro Tobalaba	-0,1226	0,0136	***
Patio Bellavista	-0,7959	0,0121	***
Género (hombre)	-0,0598	0,0094	***
Bananas	0,8694	0,0653	***
LL	-137.000		

Tabla 16 - Resultados logit considerando indicadores históricos escalados [anexos]

Variables	Coficiente	Error estándar	Significancia
Envíos acumulados	-0,793864	0,006376	***
Apertura acumulada	0,936804	0,005806	***
Edad	-0,027965	0,000351	***
Plaza Vespucio	-0,077497	0,014438	***
Metro Baquedano	-0,430986	0,015611	***
Metro Tobalaba	-0,096946	0,013157	***
Patio Bellavista	-0,554856	0,012106	***
Género (hombre)	-0,15067	0,009101	***

Bananas	0,754578	0,062226	***
LL	-144.000		

Tabla 17 - Resultados modelo logit con indicadores acumulados escalados [anexos]

Coficiente	Constante	Género (H)	Contador bananas	Edad	Tiempo	Envíos	Apertura	Aceptación
Envíos acumulados	0,0000	0,0010	0,0001	0,0004	0,0009	0,0034	0,0001	0,1104
Apertura acumulada	0,0012	0,0000	0,0037	0,0012	0,0005	0,0005	0,0031	0,0002
Plaza Vespucio	0,0000	0,0050	0,0000	0,0292	0,0009	0,0004	0,0005	0,0049
Metro Baquedano	0,0001	0,0000	0,0346	0,0000	0,0321	0,0008	0,0004	0,0012
Metro Tobalaba	0,0000	0,0005	0,0000	0,0417	0,0000	0,0353	0,0009	0,0012
Patio Bellavista	0,0000	0,0001	0,0045	0,0001	0,0335	0,0000	0,0228	0,0013
Bananas	0,0000	0,0001	0,0006	0,0028	0,0001	0,0365	0,0000	0,0403

Tabla 18 - Desviación estándar Δ [anexos]

Coficiente	Promedio
Envíos acumulados	-0,0712
Apertura acumulada	0,2344
Plaza Vespucio	-0,5148
Metro Baquedano	-0,5187
Metro Tobalaba	-0,2173
Patio Bellavista	-0,5760
Bananas	0,5342

Tabla 19 - Promedio preferencias de usuarios (jerárquico) [anexos]

Variable	Segmento 1			Segmento 2			Segmento 3		
	Coficiente	Error estándar	Significancia	Coficiente	Error estándar	Significancia	Coficiente	Error estándar	Significancia
Envíos acumulados	3,207	0,332	***	1,479	0,057	***	0,270	0,005	***
Apertura acumulada	-0,735	0,077	***	-1,209	0,049	***	-0,018	0,000	***
Plaza Vespucio	0,338	0,228		-0,161	0,119		-0,043	0,035	
Metro Baquedano	-0,891	0,397		0,448	0,218		-0,588	0,033	***
Metro Tobalaba	-0,304	0,258		-0,118	0,126		-0,041	0,030	
Patio Bellavista	-1,584	0,420	***	0,663	0,165	***	-0,806	0,028	***
Género (hombre)	-0,131	0,185		-0,035	0,090		-0,177	0,020	***
Edad	-0,061	0,008	***	0,024	0,004	***	-0,010	0,001	***
Bananas	62,801	7,026	***	1,753	0,670	***	0,429	0,182	***
α_1	-0,804	0,032	***	α_2	-0,020	0,017	α_3	1,000	Base

Tabla 20 - Resultados modelo logit 3 clases latentes [anexos]

Variable	Segmento 1			Segmento 2		
	Coficiente	Error estándar	Significancia	Coficiente	Error estándar	Significancia
Distancia	-0,004	0,001	***	0,002	0,004	***
Envíos acumulados	6,567	0,457	***	10,520	3,630	
Apertura acumulada	-0,173	0,018	***	-9,680	3,625	***
Metro Baquedano	-1,203	0,397	***	1,381	0,939	***
Metro Tobalaba	-1,173	0,348	***	0,952	0,845	
Patio Bellavista	-1,923	0,352	***	-0,595	1,150	
Plaza Vespucio	-0,709	0,540		-0,031	1,561	
Género (hombre)	-0,218	0,234		0,307	0,633	
Edad	-0,089	0,011	***	-0,159	0,061	***
Bananas	-6,389	2,022	***	53,178	18,819	***
α	-0,008	0,035				

Tabla 21 - Resultados modelo logit 2 clases latentes (distancia) [anexos]

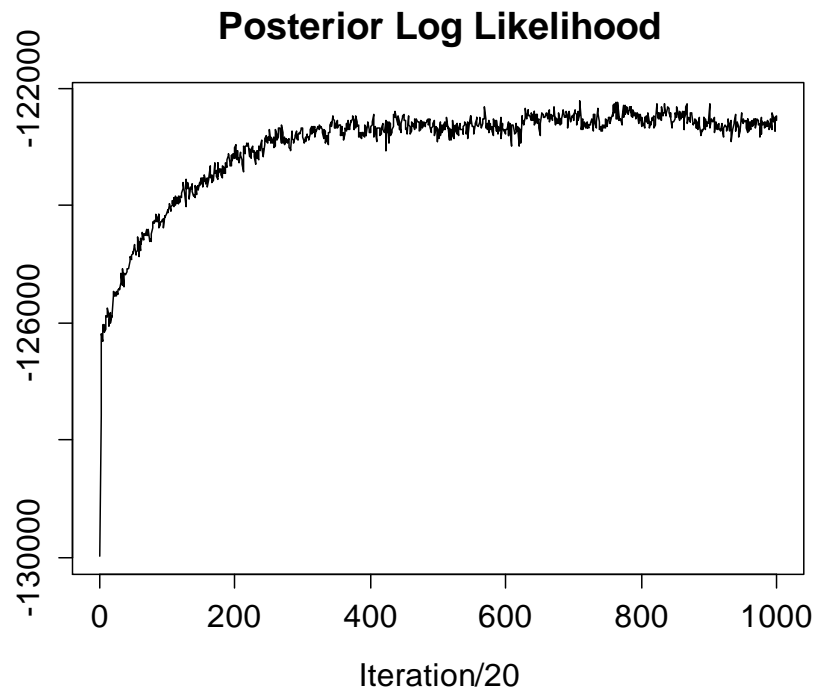
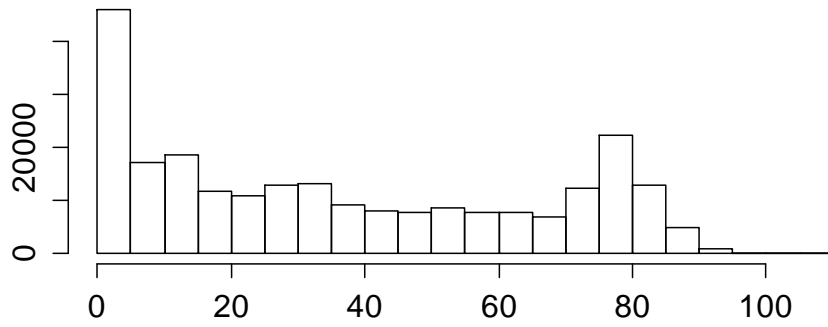


Ilustración 45 - Evolución log verosimilitud modelo jerárquico (sin distancia) [anexos]

Variables	Coficiente	Error estándar	Significancia
Distancia	-0,001	0,000	*
Envíos acumulados	-0,147	0,007	***
Apertura acumulada	0,767	0,028	***
Metro Baquedano	-0,096	0,086	
Metro Tobalaba	0,032	0,076	
Patio Bellavista	-0,349	0,078	***
Plaza Vespucio	-0,205	0,105	.
Género (hombre)	-0,008	0,057	
Edad	-0,025	0,003	***
Bananas	0,844	0,283	**
Sd Distancia	0,000	0,006	
Sd envíos acumulados	0,056	0,009	***
Sd apertura acumulada	0,214	0,022	***
Sd metro Baquedano	-0,039	1,570	
Sd metro Tobalaba	-0,010	1,550	
Sd patio Bellavista	0,023	0,975	
Sd plaza Vespucio	-0,037	1,370	
Sd género (hombre)	0,002	0,568	
Sd edad	0,000	0,014	
Sd bananas	0,113	2,870	
LL		-4350,000	

Tabla 22 - Resultados modelo mixed logit (distancia) [anexos]

Tiempo desde registro [seg 1]



Tiempo desde registro [seg 2]

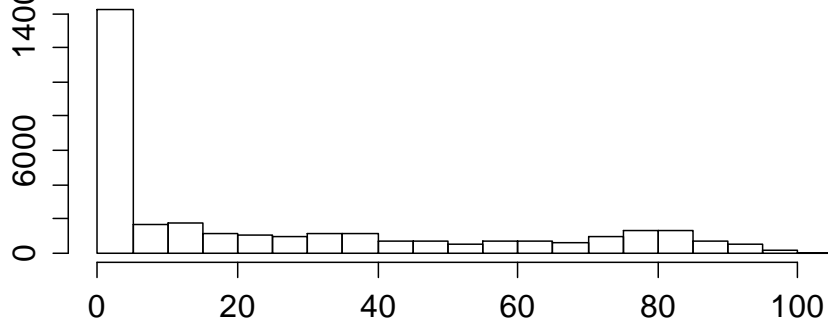
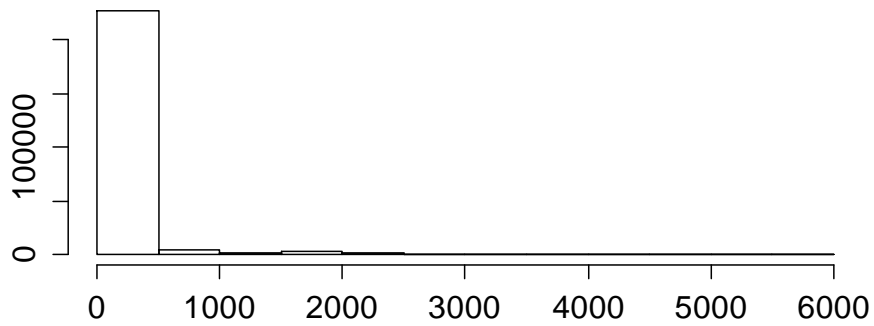


Ilustración 46 - Comparación tiempo desde registro entre segmentos[anexos]

Contador de bananas [seg 1]



Contador de bananas [seg 1]

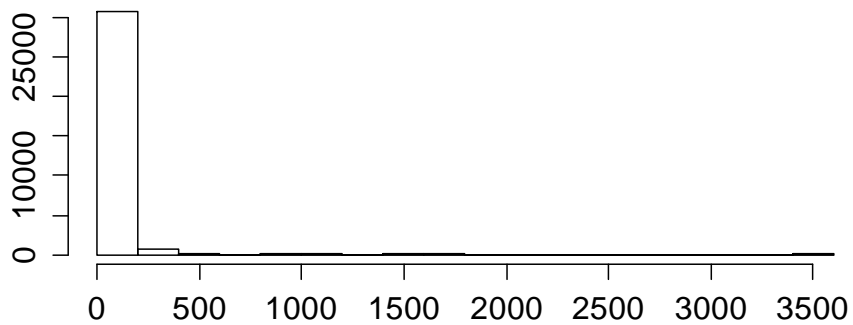


Ilustración 47 - Comparación contador de bananas entre segmentos[anexos]

Coefficiente	Promedio
Distancia	-0,087
Envíos acumulados	-4,488
Apertura acumulada	22,958
Bananas	3,939
Plaza Vespucio	2,953
Metro Baquedano	4,383
Metro Tobalaba	-1,046
Patio Bellavista	-0,364

Tabla 23 - Promedio preferencias usuarios (jerárquico-distancia) [anexos]

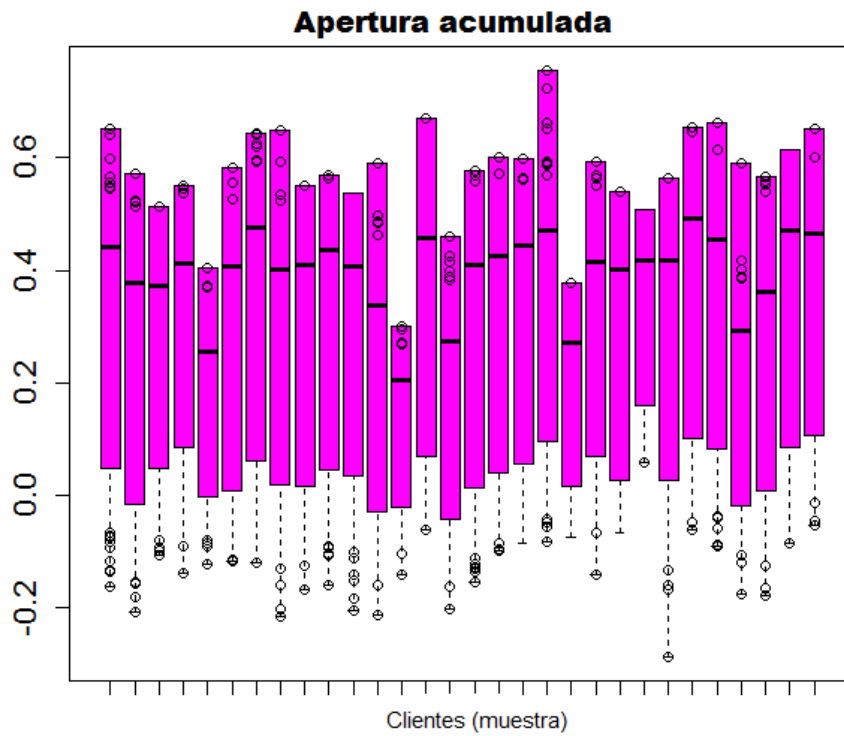


Ilustración 48 - Distribución de apertura acumulada (muestra clientes) [anexos]

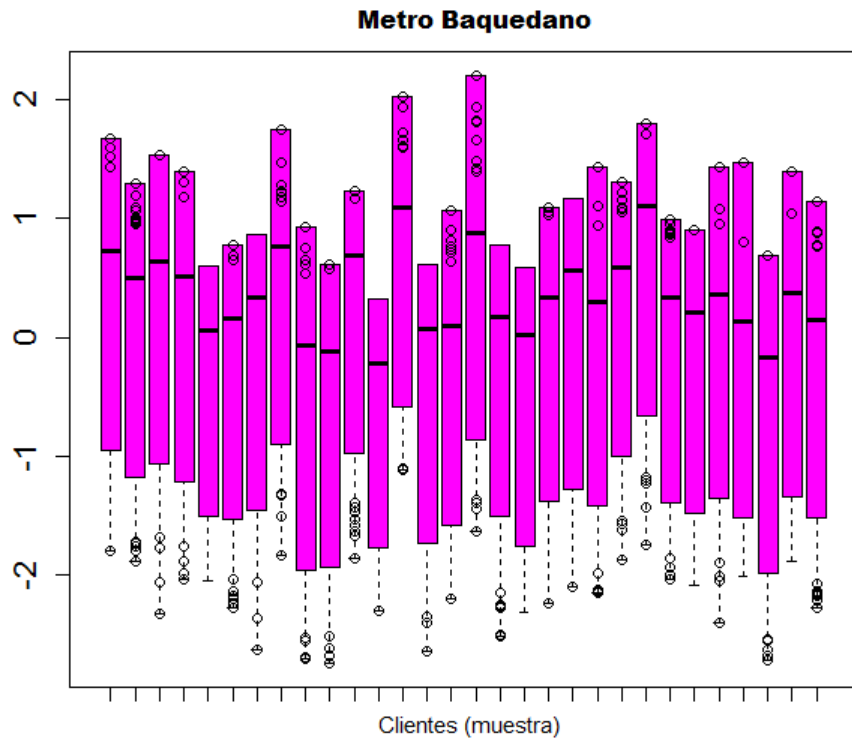


Ilustración 49 - Distribución fence Metro Baquedano (muestra clientes) [anexos]

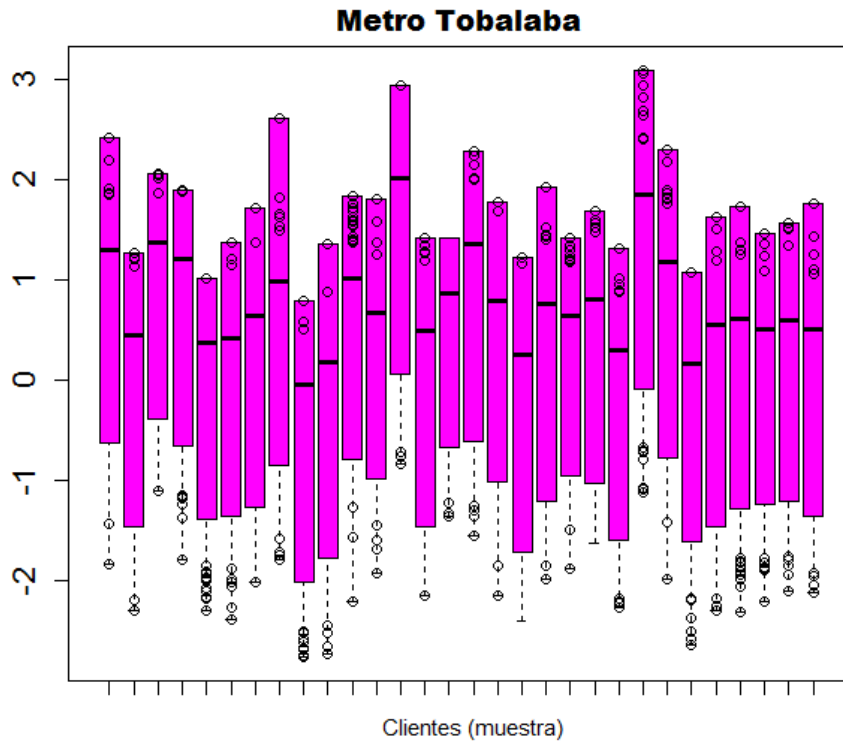


Ilustración 50 - Distribución fence Metro Tobalaba (muestra clientes) [anexos]

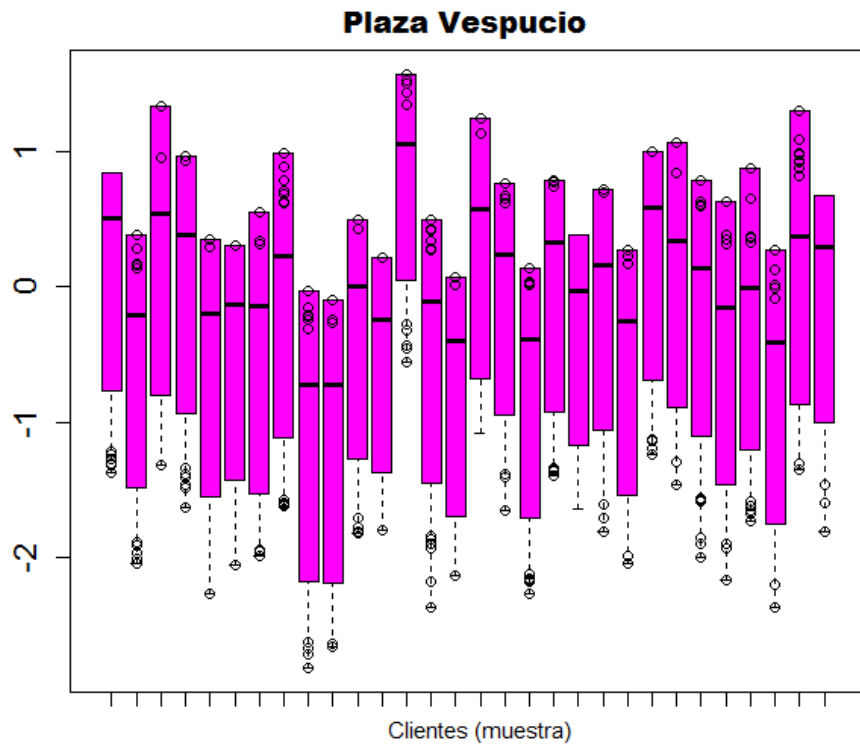


Ilustración 51 - Distribución fence Plaza Vespucio (muestra clientes) [anexos]

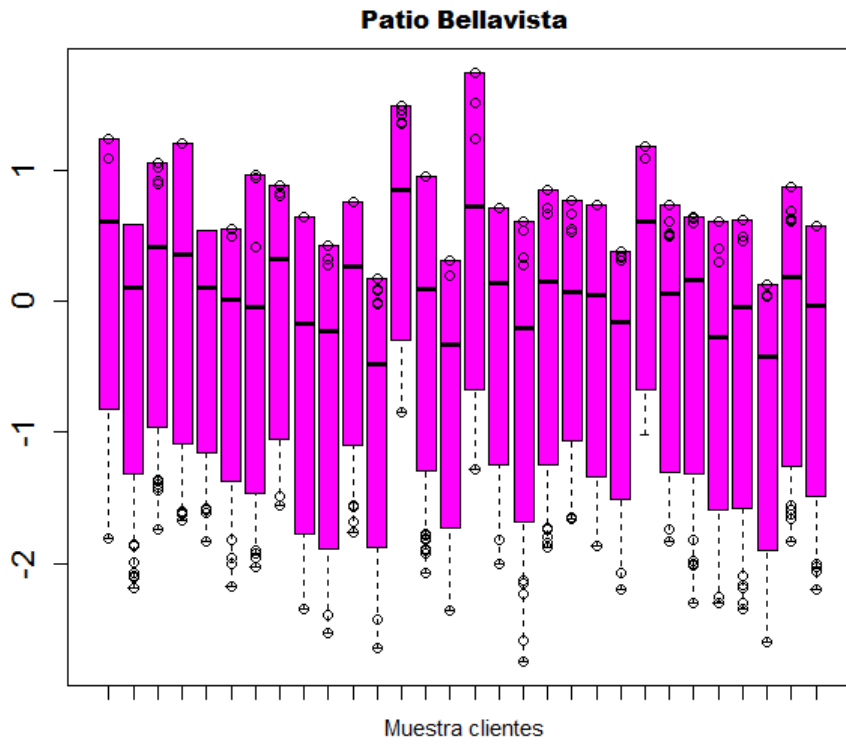


Ilustración 52 - Distribución fence Patio Bellavista (muestra clientes) [anexos]

Representación gráfica de clusters

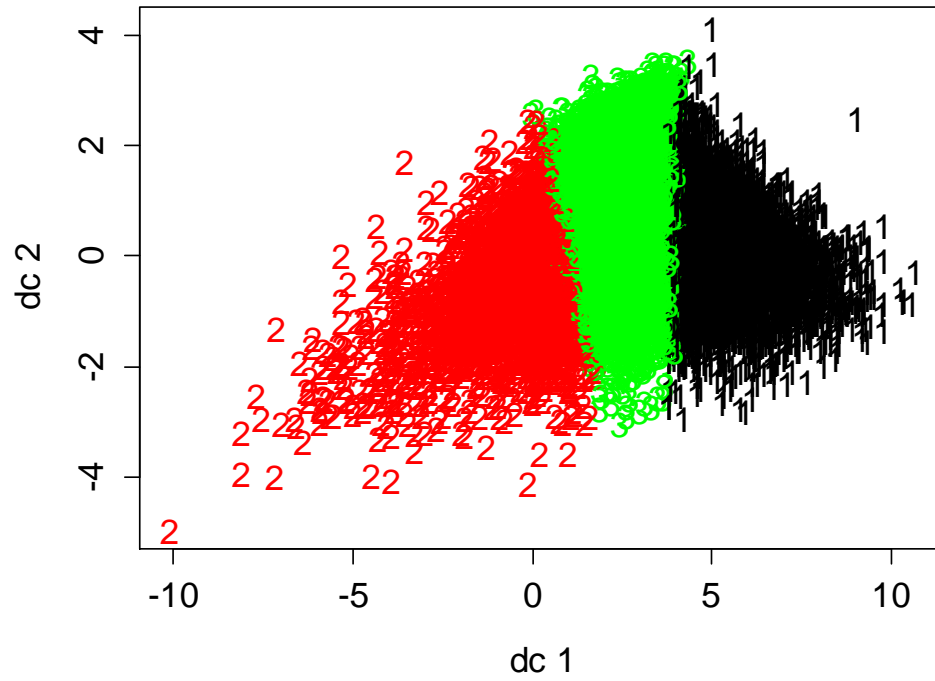


Ilustración 53 - Representación gráfica de clusters (jerárquico) [anexos]

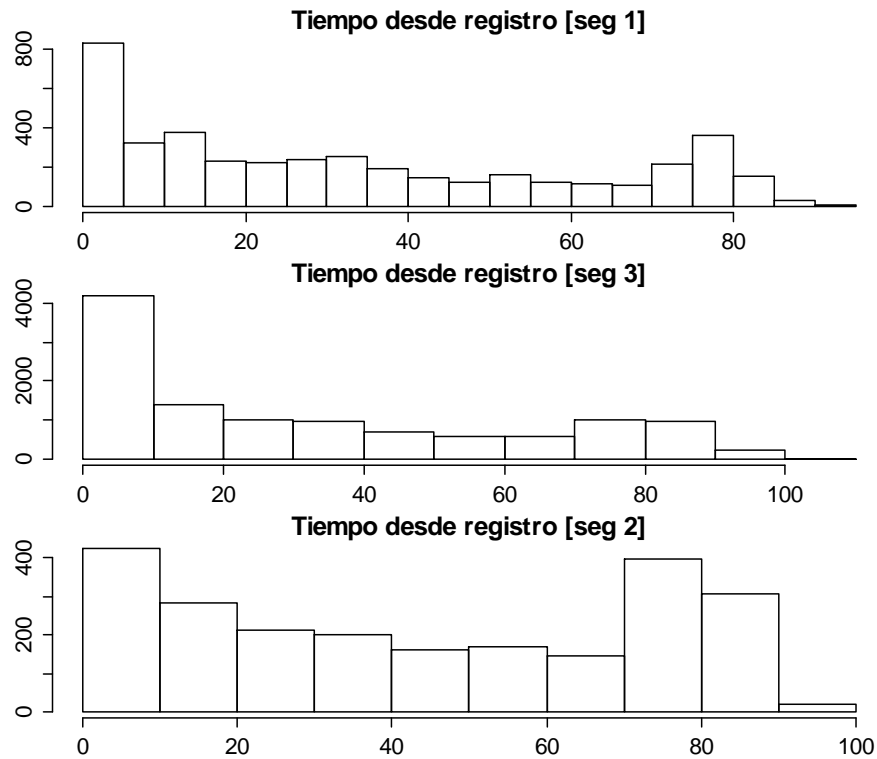


Ilustración 54 - Comparación tiempo desde registro entre segmentos (jerárquico)[anexos]

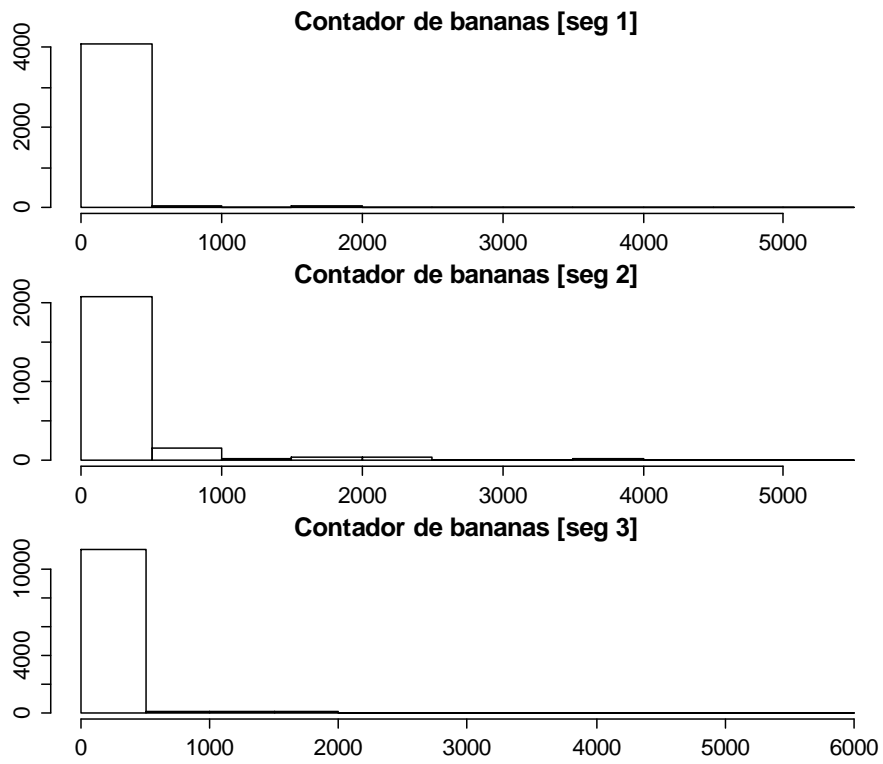


Ilustración 55 - Comparación contador de bananas entre segmentos (jerárquico) [anexos]

Representación gráfica de clusters

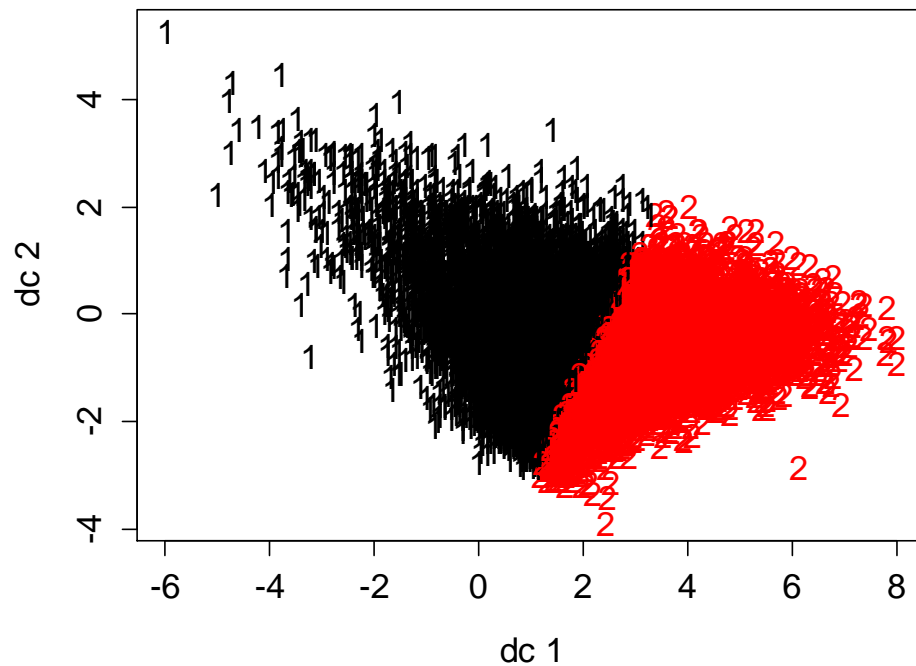


Ilustración 56 - Representación gráfica de clusters 2 clases (jerárquico) [anexos]