



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

**ESTUDIO PARA EL AUMENTO EN LA EFECTIVIDAD DE PROMOCIONES
PERSONALIZADAS EN UN SUPERMERCADO MAYORISTA**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

CARLA FRANCISCA NOVA ALARCÓN

**PROFESOR GUÍA:
LUIS ABURTO LAFOURCADE**

**MIEMBROS DE LA COMISIÓN:
RICARDO ESTEBAN MONTOYA MOREIRA
ANDRÉS IGNACIO MUSALEM SAID**

**SANTIAGO DE CHILE
ENERO 2014**

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL
POR: CARLA NOVA ALARCÓN
FECHA: 15/01/2014
PROF. GUIA: SR. LUIS ABURTO

Estudio para el aumento en la efectividad de promociones personalizadas en un supermercado mayorista

La industria del Retail es altamente compleja, ya que tiene gran cantidad tanto de clientes como de productos, y estos poseen una alta rotación, por lo que se requiere una estrategia para administrar las distintas áreas del negocio. Muchos esfuerzos se han puesto para diferenciar cada servicio prestado al cliente para mejorar su experiencia de compra. Uno de esos esfuerzos corresponde a las promociones personalizadas.

El objetivo principal consiste en estudiar potenciales mejoras en la efectividad de promociones del sistema kiosco de un mayorista, específicamente se estudiará la entrega de promociones al cliente correcto en el momento oportuno. Para cumplir con este objetivo se utilizarán datos transaccionales de clientes pertenecientes al club convenio del supermercado mayorista y se analizará la categoría de “postres refrigerados”. La base de datos contiene la información de 944 clientes considerando la historia de sus transacciones durante los años 2011 y 2012. Cabe destacar que el supermercado donde se realizará el estudio ya ha puesto esfuerzos por realizar promociones personalizadas, ya que éste consta de un sistema kiosco que brinda promociones a los clientes socios, quienes se inscriben con un rubro de negocio para recibir descuentos.

La metodología se basa en ofrecer un producto en el momento oportuno al cliente. Es por esto que en las variables para predecir la oferta de un producto por medio de una promoción del kiosco, se usan variables basadas en la técnica RFM.

Para estudiar el efecto de la promoción, se utiliza una regresión logística binaria que modela la decisión de compra de un cliente como función de la promoción recibida, la variable dependiente es la compra o no en la categoría.

Entre los principales resultados se observa que los atributos en los productos que influyen en la compra de los clientes son: para visitas al mayorista; los días entre compra, la frecuencia mensual, días acumulados entre compras móvil, días entre compra sobre frecuencia y, para la categoría; días promedio móvil entre compras y frecuencia móvil.

Se concluye que es posible aumentar la efectividad de las promociones mediante un cambio en la oferta de las promociones, brindando promociones al cliente correcto, del producto correcto, en el tiempo correcto, la cual se traduce en una disminución de costos de \$ 1.537.604 por cada 100.000 transacciones equivalente a MM\$106 al año.

Agradecimientos

Primero que todo quiero agradecer a Dios, por darme la perseverancia y apoyo incondicional en toda mi vida y en especial para realizar este trabajo, acompañándome en momentos difíciles y también en momentos gratos.

A mi familia, por su amor y consejos. Mis padres, por su cariño y apoyo, mi hermana mayor Valeska y su marido Roberto, que a pesar de estar lejos en estos momentos me han brindado todo su apoyo en momentos difíciles. A mi hermana Natalia y su marido Alexis, por los consejos de vida que me han dado, por alojarme en su casa para terminar este trabajo y por la tremenda alegría de la bendición que viene en camino.

A mi compañero de vida, Esteban, gracias por estar conmigo en los momentos difíciles cuando necesitaba un apoyo, o simplemente un abrazo, te amo.

A la familia de Esteban, por darme alegrías y por quererme como soy.

A mis amigos Nicolás Garcias, Enzo Solari y Hysan Kuo por acompañarme en todo mi periodo universitario.

A mi amiga Ana Castillo por todas sus buenas vibras a diario y conversaciones.

A una persona que llegó justo en el momento que la necesitaba, Margarita Marambio, muchas gracias por escucharme y ayudarme a convertirme en lo que soy actualmente, muchas gracias por todas nuestras conversaciones.

A mi profesor guía, Luis Aburto, gracias por su apoyo y comprensión.

A mis compañeros de Penta Analytics, Andrés Gormaz, Iván Castillo, Cristian Nuñez, Diego Saavedra y Yal Fuentes, gracias por el apoyo a diario que me brindaron para la realización de este trabajo, son unas excelentes personas, sigan así para seguir aumentando la productividad del área

A Patricio Fuentealba, gracias por tu paciencia y ayuda que me brindaste y por lo que me tuviste que entender para lograr un buen trabajo.

A mis amigos especiales e incondicionales, que siempre me esperan con un saludo caluroso, Arturo, Lukas, Bobby y Bruce, por su cariño y apoyo en las noches de frío y trabajo.

Tabla de Contenido

1.	Introducción.....	1
2.	Justificación.....	3
3.	Objetivos	6
	Objetivo General	6
	Objetivos Específicos	6
4.	Alcances.....	6
5.	Resultados Esperados	7
6.	Marco Conceptual	7
	6.1.Promociones.....	7
	6.2.Técnica KDD	8
	6.2.1. Modelo entre compra	9
	6.3.Técnica RFM	12
7.	Marco Metodológico.....	12
	7.1.Definición del problema y revisión bibliográfica	12
	7.2.Selección y procesamiento de datos.....	13
	7.2.1. Selección de clientes	13
	7.2.2. Selección de categoría.....	13
	7.3.Análisis descriptivo de los datos.....	14
	7.4.Desarrollo de modelo	14
	7.4.1. Modelo entre-compras	14
	7.5.Evaluación modelo.....	15
8.	Desarrollo.....	16
	8.1.Selección y análisis descriptivo de los datos.....	16
	8.1.1. Clientes.....	16
	8.1.2. Categoría a estudiar	20
	8.1.3. Variables.....	23
	8.2.Pre-procesamiento y transformación de datos	24
	8.2.1. Limpieza datos.....	24
	8.2.2. Transformación datos	24
	8.3.Desarrollo de Modelo	25
	8.3.1. Modelo para predicción de compra	25
9.	Conclusiones	31
10.	Trabajos Futuros	32

11. Bibliografía.....	33
12. Anexos	34
12.1. <i>Voucher</i> ejemplo de promociones por medio del sistema kiosco .	34
12.2. Detalle análisis categorías de rutina	35
12.3. Cohortes	35

Índice de Tablas

Tabla 1: Características rol de las categorías.....	21
Tabla 2: Caracterización categorías rutina.. ..	23
Tabla 3: Datos promedios.....	26
Tabla 4: Estadísticos categorías de destino.....	35

Índice de Ilustraciones

Ilustración 1: Participación de mercado.....	1
Ilustración 2: Kiosco supermercado mayorista. Sucursal Recoleta.....	4
Ilustración 3: Proceso KDD. Fuente: Lycos	8
Ilustración 4: Matriz de confusión..	10
Ilustración 5: ROC. Elaboración propia.....	12
Ilustración 6: Caracterización de la información.....	17
Ilustración 7: Rubro clientes estudio.	17
Ilustración 8: Caracterización de impresión y canje para clientes en estudio para Diciembre de 2012.	18
Ilustración 9: Caracterización de impresión y canje para transacciones en estudio para Diciembre de 2012.....	19
Ilustración 10: Mapa de roles de las categorías..	20
Ilustración 11: Mapping categoría destino.	22
Ilustración 12: Curva ROC.	26
Ilustración 13: Curva de costos..	26
Ilustración 14: Matriz confusión.....	27
Ilustración 15: Resultados.....	28
Ilustración 16: Simulación datos 2013. Enero -Agosto Predicción modelo. Elaboración propia.....	29
Ilustración 17: Simulación datos 2013. Enero - Agosto. Datos reales kiosco..	29
Ilustración 18: Voucher ejemplo promociones.....	34

1. Introducción

La industria del retail corresponde a un sector económico que comercializa masivamente productos y servicios en grandes cantidades y que se encarga de entregar bienes a su consumidor final, o en su defecto, a niveles correspondientes a los últimos eslabones de la cadena de distribución.

Es una industria bastante compleja, ya que se encarga de vender a un público numeroso, lo cual requiere una gran coordinación con sus proveedores, distribución y gestión de sus productos para que las personas obtengan lo que necesitan al tiempo correcto y cantidad que desean. También corresponde a una industria que está en constante crecimiento y que así se mantendrá, ya que las cifras indican que las ventas del sector en Chile aumentarán un 14,5% entre el 2011 y el 2015 [1], por lo que se vuelve un mercado atractivo para analizar, aprender y contribuir en su crecimiento y estrategias para lograrlo.

Unos de los formatos pertenecientes a la industria del retail son los supermercados, los cuales se estima un crecimiento del 26,6%, aumentando sus ventas desde US\$2,4 MM en 2011 a US\$3,5 MM en 2015 [1]. Dentro de este formato de supermercados, se pueden identificar dos, mayorista y minorista. El primero corresponde a ventas, compra, o contratos al por mayor y puede, o no, tener contacto directo con el consumidor final, habitualmente está ubicado entre el productor y un vendedor minorista. El segundo corresponde a actividades realizadas al por menor, donde normalmente es el vendedor al consumidor final.

El mercado de los mayoristas representan cerca del 30% del mercado de venta al por mayor en Chile, lo que se traduce en US\$700 Millones en ingresos [2]. Este mercado tiene como principales actores en el 2011 a Rabié, Adelco, La Caserita, Dimak, La Calera y al corporativo de SMU, los cuales tienen a su cargo mayoristas como Alvi, Mayorista 10, Maxi Ahorro, Vegamercado, Dipac y Comer. La participación de mercado para el año 2011 se puede apreciar en la siguiente imagen [3]:



Ilustración 1: Participación de mercado

Dentro del mercado mayorista se pueden identificar tres formatos [2]:

- ✓ Mayorista de Mesón: Corresponde al formato por el cual el cliente solicita los productos que quiere por medio de un mesón. Existen alianzas con proveedores de los productos para que el mayorista de mesón comercialice sus bienes ofreciéndoselos a minoristas. Una vez hecho el pedido por el cliente, los trabajadores emban los productos solicitados, donde se tiene la opción de retirarlos de inmediato, o en su defecto, solicitar un despacho, ya que generalmente poseen camiones para distribuir los pedidos. Algunos mayoristas correspondientes a este formato son los puntos Rabié, La Caserita y Dipac.
- ✓ Mayorista de autoservicio: Corresponde al formato por el cual el cliente recorre los pasillos, tal como en un supermercado minorista, donde va escogiendo sus productos a medida que transita por las góndolas del supermercado. Acá no existe la opción de despacho por medio de camiones, sino que el cliente debe llevarse sus productos adquiridos de forma particular. Algunos mayoristas correspondientes a este formato son Alvi, Mayorista 10, Vegamercado y Maxi Ahorro.
- ✓ Mayorista operador logístico: Corresponde al formato por el cual el cliente no tiene contacto físico con el mayorista, sino que realiza su pedido por medio de sistemas telefónicos o por plataformas web, para luego ser distribuidos al lugar donde indique el cliente a través de camiones ofrecidos por el supermercado. Algunos mayoristas correspondientes a este formato son Rabié, Adelco y Dimak.

Los consumidores cada vez están más sofisticados, y buscan un mayor valor agregado al adquirir los productos que consumen. Es por esto que cada mercado intenta diferenciarse, en especial el *retail*, donde esta diferenciación se basa en el diseño y la experiencia de compra para el cliente.

Un factor relevante para el crecimiento de este mercado es el factor gestión, específicamente la ingeniería de marketing [4] por medio de promociones personalizadas, ventas de servicios complementarios, acumulación de puntos para ser canjeados, entre otros.

Esta gestión de promociones ha obtenido fuerza durante estos últimos períodos. Bien se puede ver por el sistema “kiosco” que posee el supermercado mayorista donde se realizará el estudio, el cual consiste en una máquina que brinda descuentos y/o promociones en productos. Los principales actores que poseen “kioscos” son el “Club La Caserita” que posee la distribuidora del mismo nombre, que ofrece *vouchers* con descuentos en la próxima compra, y el “Club Mayorista” que posee el supermercado mayorista, que ofrece *vouchers* con descuentos previo a la compra.

Las promociones brindadas por el sistema kiosco del supermercado mayorista se realizan mediante un análisis del comportamiento de compra y canjes del cliente, para así ofrecer promociones personalizadas por medio de descuentos en productos que habitúan llevar y/o canjear en promociones pasadas, por lo cual se trata de un tema de fidelización de clientes para el mayorista. Tal metodología se puede ver en el trabajo realizado por Claudia Sotomayor [5] donde se diseñan promociones tanto de productos frecuentes como asociados.

Luego de la determinación de las reglas de negocios para la asignación de promociones, se realizó una evaluación de éstas [6] donde se refleja que aún quedan cosas por hacer para mejorar la eficiencia de las promociones brindadas, ya que en este trabajo sólo se midió qué tipos de promociones eran preferidas por los clientes, tales como el tipo de mensaje que se brinda en la promoción, la duración de éstas, la existencia de sensibilidad en los clientes a los descuentos, entre otros, con lo cual se llega a la conclusión de que las promociones brindadas no son eficientes, ya que no presentan cambios en el comportamiento de compra de los clientes. Por otro lado, en el 2012 se realizó una metodología para pronosticar la efectividad de promociones sobre los clientes basándose en sus respuestas pasadas [7]. Esta metodología mide la efectividad de las promociones actuales, lo que concluye que existe un 72% de los clientes que no muestran cambios en el comportamiento de compra del producto ofrecido, por lo que es evidente que existe campo donde mejorar la eficiencia de las promociones actuales. Ambos trabajos muestran que aún queda por aumentar la efectividad de las promociones brindadas por el kiosco del mayorista.

Si bien el tipo de promociones que brinda actualmente el kiosco permite aumentar la canasta del cliente, no permite aprovechar las promociones para aumentar la efectividad de estas reflejadas en los canjes de los *vouchers*. Es aquí en donde se enmarca la presente memoria, en identificar a aquellos clientes que se les pueda brindar promociones aumentando el “ítem” de los productos que frecuenta comprar y prediciendo el momento en que comprará el producto para así lograr una mayor eficiencia en las promociones brindadas.

2. Justificación

La presente memoria se realiza bajo el alero de la empresa Penta Analytics que se encarga de realizar proyectos relacionados con la inteligencia de negocios, donde se analizan datos, utilizando técnicas de Data Mining, para aportar con información valiosa a las distintas empresas con el fin de aumentar su rentabilidad. Se realizan labores tanto en temas de eficiencia operacional (gestión de personas, detección de fraude, logística e inventario, entre otros) como en gestión de *retail* (gestión de *shopper*, gestión de ventas, gestión de promociones, entre otros)

Entre los clientes más antiguos, está el supermercado mayorista del cual se obtendrán los datos a ser analizados. Penta Analytics maneja todos los datos

transaccionales del supermercado mayorista, gestionándolos y brindándoles reportes periódicos según lo establecido por ambas partes.

Con respecto a las promociones brindadas por la empresa, corresponden a 300 al año, donde el 96% corresponde a promociones en descuento del precio final del producto ofrecido al cliente.

Particularmente, el supermercado mayorista donde se trabajarán los datos para su gestión, tiene a su disposición promociones que brindan los proveedores, los cuales deben ser gestionados por el *retail*. Es por esto que se dispone de un sistema “kiosco”, el cual consta de un dispensador, como se muestra en la ilustración 2, que brinda promociones tipo descuento en precios por medio de la impresión de un *voucher* por cada cliente (Anexo 12.1: Ejemplo *voucher*). Estas promociones contienen descuentos en productos frecuentes y asociados. Sin embargo, la tasa de canje no es la misma para cada una. Vale la pena destacar, que la disposición para obtener el *voucher* de descuentos, se tiene que estar inscrito en el supermercado mayorista como socio, es decir, acreditar algún rubro de negocio.



Ilustración 2: Kiosco supermercado mayorista. Sucursal Recoleta

Los datos utilizados serán del tipo transaccionales y sociodemográficas tales como sucursal de compra, rubro del negocio del socio, frecuencia de compra, entre otras.

Lo que se quiere lograr con la presente memoria es poder identificar qué tipo de clientes pueden cambiar su comportamiento de compra a un producto de una gama superior por medio de promociones personalizadas tipo *up-selling*¹ identificando además cuando es más probable que adquieran un producto específico para así aumentar la efectividad de las promociones brindadas por el kiosco. Es importante aquí el valor del cliente, ya que es éste quien decide si aceptar o no ciertas promociones en la compra de un producto de una gama superior, por lo que ofrecerle lo que está dispuesto a aceptar es fundamental para lograr el objetivo propuesto, pero a su vez es de igual importancia el margen obtenido por el mayorista.

El aprovechar de mejor forma estos descuentos brindados por los proveedores es beneficioso para el mayorista, ya que utilizando los mismos recursos (promociones brindadas por el proveedor) pueden gestionarse de mejor forma para obtener mayores márgenes debido a la identificación de compra de los clientes, a los cuales se le ofrecerán descuentos a su medida, es decir, personalizadas a partir de comportamientos de compras pasadas.

Actualmente no se tiene eficiencia en las promociones brindadas por el kiosco en temas de ofrecer un producto de mayor valor para el cliente (*up-selling*) ni la identificación de qué producto es más probable que se compre en cada ocasión de compra. Por lo tanto, lo que se quiere lograr con el presente trabajo es brindar promociones del producto correcto, al cliente correcto y en el tiempo correcto [8].

Para realizar los análisis correspondientes en este trabajo, se escogerá una categoría en la cual gestionar los datos, con el objetivo que sirva como metodología para ser replicado en las otras categorías.

¹ La promociones *up-selling* consisten en ofrecer al cliente un producto de una gama superior al que frecuenta comprar, por ejemplo, ofrecerle un producto de una marca mejora a la que lleva. Más adelante se brindará una mejor descripción de la promoción.

3. Objetivos

Objetivo General

Estudiar un rediseño de promociones focalizadas para mejorar la efectividad de éstas en el sistema kiosco de un supermercado mayorista.

Objetivos Específicos

- ✓ Identificar los perfiles de los clientes a los cuales enfocar las mejoras en las promociones personalizadas.
- ✓ Valuar variables relevantes para estimar el momento oportuno en el cual ofrecer una promoción personalizada.
- ✓ Estimar el momento oportuno en el cual ofrecer una promoción personalizada.
- ✓ Evaluar el impacto en los canjes dada la propuesta de mejora.

4. Alcances

- ✓ Sólo se estudiarán clientes activos pertenecientes al club convenio del supermercado mayorista entre los años 2011 y 2012, los cuales corresponden al 25% de los clientes socios que compraron al menos una vez durante el periodo de tiempo descrito. Determinar y validar las variables relevantes para estimar el momento oportuno en el cual ofrecer una promoción personalizada.
- ✓ El estudio se realizará sólo para la categoría de “postres refrigerados”, siendo replicable a otras.
- ✓ No se brindarán recomendaciones de promociones tipo *up-selling*, sin embargo, una vez estimado la compra o no en la categoría, sería directo realizar una promoción a nivel de marca por ejemplo, con el objetivo de aumentar el margen promedio por producto canjeado.
- ✓ Con las recomendaciones se remite evaluar potenciales mejoras en la efectividad de las promociones, es decir, cambiar el comportamiento en sala, no el comportamiento de compra futuro.
- ✓ No se brindarán recomendaciones por parte del porcentaje de descuento eficiente que se debe ofrecer.

5. Resultados Esperados

- ✓ Perfiles de clientes a quienes enfocar las promociones.
- ✓ Cálculo de variables posibles que predigan la compra en cierta categoría.
- ✓ Estimación de la probabilidad de compra en la categoría de los clientes identificados, dada la promoción.
- ✓ Evaluación de la aceptación de promociones.

6. Marco Conceptual

6.1. Promociones

Existen muchas definiciones de lo que son las promociones, pero en el sentido más amplio y referente al presente trabajo se podría definir como la cuarta herramienta del marketing-mix, la cual incluye las distintas actividades que desarrollan las empresas para comunicar los méritos de sus productos y persuadir a su público objetivo para que compren **[9]**.

Cabe destacar la existencia de promociones masivas y promociones focalizadas. Las primeras van dirigidas a todos los clientes, sin previo estudio, ya que su objetivo principal es básicamente aumentar el volumen de ventas, por lo tanto cualquier tipo de cliente puede acceder a ellas. Las segundas corresponden a promociones dirigidas a uno (o más) segmento objetivo, ya que se tiene un fin estratégico, por lo cual se tiene un interés por detrás ya sea para realizar técnicas de retención, fidelización de clientes, adquisición de nuevos clientes, entre otras acciones. En el presente informe se realizarán promociones personalizadas.

Para realizar promociones, existen distintas técnicas, como lo son descuento directo, packs, cupones, muestra gratis, entre otros **[10]**. Con las que trabaja en el kiosco del mayorista son de descuento directo, la cual consiste en la reducción del precio final del producto. Según estudios anteriores **[6]** corresponde a una de las técnicas con mayor aceptación, ya que la mayoría de los clientes conocen el precio lista de los productos, por lo que un descuento lo perciben de inmediato.

Unas de las técnicas que se usan este tipo de promociones de descuento directo son:

- ✓ *Cross-selling*: Consiste en ofrecer descuento en el precio de productos complementarios a los que compra frecuentemente el cliente. No son productos que estén en la canasta habitual de compra, pero existe una gran probabilidad que se compren juntos porque corresponden a productos complementarios, ya que por medio de transacciones conjuntas se realiza

una relación entre ellos. Uno de los objetivos de esta técnica es aumentar el número de productos en la canasta del cliente.

- ✓ *Up-selling*: Consiste en ofrecer descuento en el precio de productos de un formato o margen mayor que el que regularmente compra, es decir, en productos que responden a una misma necesidad pero que requerirá un mayor gasto por parte del cliente. Uno de los objetivos de esta técnica es aumentar el ticket promedio de compra del cliente, lo cual se traduce en mayores márgenes para la empresa. También tiene como objetivo la fidelización del cliente, ya que se le está ofreciendo descuento en un producto más “sofisticado” del que actualmente compra.

Específicamente en este trabajo se dará énfasis a las promociones tipo *up-selling*, ya que actualmente los *vouchers* ofertados a los clientes poseen promociones en productos frecuentes y los asociados a éstos, lo que corresponde a promociones *cross-selling*, ya que los últimos son ofrecidos en base a lo que habitúa el cliente a comprar.

6.2. Técnica KDD

KDD (Knowledge Discovery in Databases) es una técnica de Data Mining que consiste en un proceso no trivial de identificación de patrones en los datos que a simple vista no se pueden obtener.

Data Mining Model

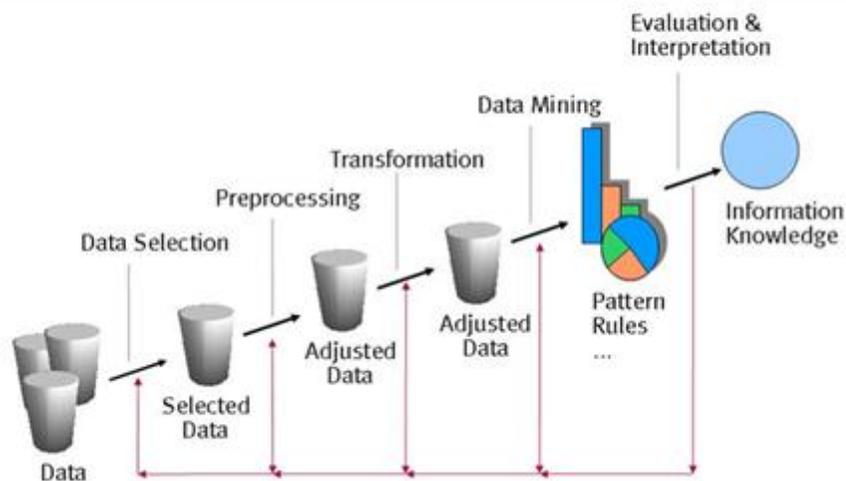


Ilustración 3: Proceso KDD. Fuente: Lycos

Los principales pasos del proceso corresponden a:

- ✓ Selección de datos: Corresponde a la extracción y selección de datos a utilizar desde una base de datos. La selección de datos a utilizar dependerá de cada tipo de problemática que se quiere resolver.
- ✓ Pre-procesamiento de datos: Consiste en la limpieza de los datos. Básicamente en detectar valores fuera de rango, faltantes o nulos.
- ✓ Transformación de datos: Consiste en la modificación de los datos a ser ingresados al modelo, ya que generalmente los datos no poseen la forma correcta para extraer conocimiento a partir de éstos. Básicamente consiste en normalizar a aquellos datos que lo requieran, extracción de algunos atributos, discretización, cambio de escala, entre otros.
- ✓ Data Mining o aplicación de modelos: Esta etapa consiste en la aplicación del modelo en sí. También dependerá de la problemática a tratar. Algunas técnicas usadas con árboles de decisiones, redes neuronales, support vector machine, entre otros.
- ✓ Evaluación e Interpretación: Es la etapa más importante del proceso, ya que es aquí donde se obtiene la información oculta en los datos y se debe saber interpretar los resultados obtenidos a partir de los modelos con la problemática real de negocio que se quiere resolver.

Para este trabajo es fundamental seguir los pasos de esta metodología, ya que la información se encuentra almacenada pero no se puede tomar decisiones sin antes de utilizarla de buena forma. Es por esto que es fundamental escoger los datos necesarios que servirán para el análisis de esto con el objetivo de sacar información relevante para el negocio.

Por lo tanto, es esencial encontrar las variables correctas para tener resultados adecuados al negocio.

Específicamente, en este trabajo el modelo utilizado para uno de los objetivos que consiste en la predicción de compra en una categoría, es el siguiente:

6.2.1. Modelo entre compra

Dentro del paso de aplicación de modelos del proceso KDD, se pueden encontrar muchas metodologías para realizar la predicción, sin embargo en este trabajo se utilizará:

6.2.1.1. Regresión logística binaria

Consiste en una técnica multivariante donde la variable dependiente es categórica y las variables independientes pueden ser de cualquier naturaleza, es

decir, cualitativas o cuantitativas. El objetivo es, por medio de probabilidades, predecir a qué grupo de la variable objetivo pertenece un sujeto dada las características del mismo (variables independientes).

El objetivo primordial de esta técnica es modelar cómo influye en la probabilidad de aparición de un suceso, dicotómico, la presencia o no de diversos factores y el nivel de estos.

Se requiere determinar la probabilidad de que a alguien le ocurra un determinado suceso. En el caso de la presente memoria es identificar si un cliente comprará un producto en una categoría y ocasión de compra dada, para el caso de predicción de compra y la compra de una marca dada para la realización de up-selling.

La probabilidad de aparición de un suceso está determinada por:

$$P(x) = \frac{1}{1 + e^{-\beta_0 - \beta_i x_i}}$$

Donde β_0 corresponde al intercepto, β_i al coeficiente estimado de la variable x_i y x_i son las distintas variables independientes.

Una vez estimada la probabilidad, se asigna la pertenencia a un suceso u otro. Una de las metodologías para fijar el cohorte de pertenencia, es el costo de cometer un error, los cuales se ven reflejados en la matriz de clasificación que brinda los resultados del modelo:

		Valor Observado	
		Clase 1	Clase 2
Predicción	Clase 1	TP Verdadero Positivo	FP Falso Positivo Error Tipo I
	Clase 2	FN Falso Negativo Error Tipo II	TN Verdadero Negativo

Ilustración 4: Matriz de confusión. Elaboración propia.

Donde se define:

$$\text{Sensibilidad (TPR)} = \frac{TP}{TP + FN}$$

$$\text{Especificidad (TNR)} = \frac{TN}{TN + FP}$$

$$FPR = \frac{FP}{TN + FP}$$

$$\text{Accuracy} = \frac{(TP + TN)}{N + P}$$

Dónde:

TPR: Ratio verdaderos positivo (True positive rate), Sensibilidad

TNR: Ratio verdaderos negativos (True negative rate). Especificidad

FPR: Ratio falsos negativos (False negative rate)

P= TP +FN

N=FP+TN

TP: Verdadero positivo (True positive). Dónde la clase del valor predicho corresponde a la misma clase que el valor observado (Clase 1)

FP: Falso positivo (False positive). Dónde la clase del valor predicho (Clase 1), es distinta a la clase del valor observado (Clase 2) Error Tipo 1.

FN: Falso negativo (False negative). Dónde la clase del valor predicho (Clase 2), es distinta a la clase del valor observado (Clase 2) Error Tipo 2.

TN: Verdadero negativo (True negative). Dónde la clase del valor predicho corresponde a la misma clase que el valor observado (Clase 2)

Otra forma de ver los puntos posibles de cohorte es el área bajo la curva ROC:

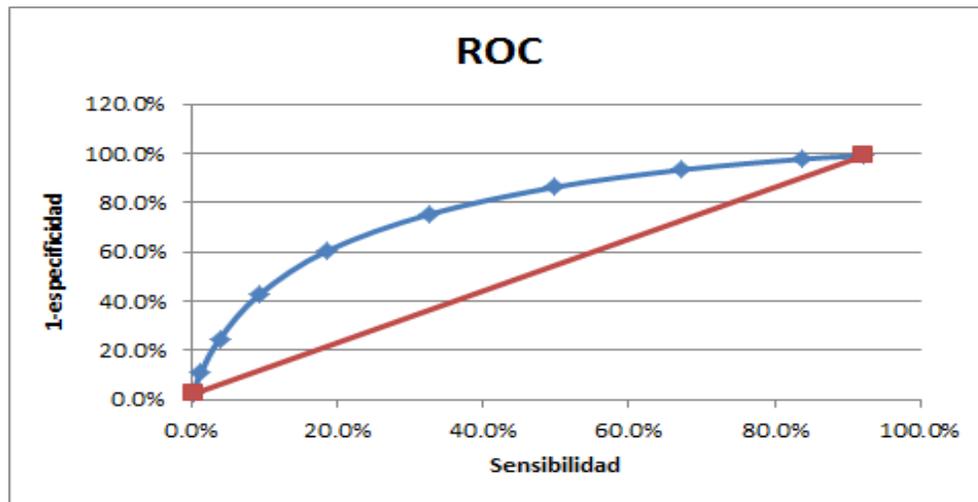


Ilustración 5: ROC. Elaboración propia.

6.3. Técnica RFM

Consiste en una técnica de segmentación resumida, la cual se basa en el análisis de datos del pasado para identificar con alguna certeza el futuro.

Generalmente, los clientes que compran frecuentemente, que han comprado en el último periodo y los que gastan una mayor cantidad de dinero, son los que responden mejor a las técnicas de marketing que se les ofrece.

Básicamente se requiere identificar la “Recency” (tiempo desde la última compra), “Frequency” (frecuencia de compra) y “Monetary value” (monto de compra) para identificar a aquellos clientes que están comprando y que responderán en forma asertiva a las distintas técnicas de marketing que se les pueda ofrecer.

Para este trabajo, esta técnica es fundamental para la creación de las variables a utilizar en el modelo de predicción entre-compras.

7. Marco Metodológico

7.1. Definición del problema y revisión bibliográfica

El primer paso consiste en conocer la industria y revisar bibliografía acerca de ésta, con el objetivo de entender el mercado en donde está inmersa la empresa de la cual se analizará. Luego de conocer la industria, es necesario comprender la situación actual que tiene la empresa en temas de promociones personalizadas, ya que el presente trabajo analizará la efectividad de éstas, comprender trabajos pasados realizados anteriormente, con el fin de complementarlos.

Por otro lado, es necesario realizar revisión bibliográfica acerca de los modelos que se necesitarán para cumplir los objetivos expuestos, es decir, investigar métodos de predicción de compra.

El objetivo principal de esta primera etapa es comprender la situación actual y determinar el problema u oportunidad de negocio que sustentará el presente trabajo.

7.2. Selección y procesamiento de datos

Esta etapa consiste primeramente en escoger a los clientes a ser estudiados, con el objetivo de obtener datos relevantes para llegar a conclusiones representativas. Luego es necesario escoger la categoría estudio para finalizar eligiendo las variables relevantes que puedan servir en los modelos para explicar la variable dependiente, que en un caso corresponde a si un determinado cliente compra o no en una determinada categoría en una ocasión de compra asociada; y en el otro caso corresponde a estimar la probabilidad de compra de un producto de una gama superior al que el cliente habitúa comprar.

Por otro lado, se eliminar aquellos datos inconsistentes, tales como los montos negativos, fechas fuera de rango, entre otros.

7.2.1. Selección de clientes

Se escogen aquellos clientes que van constantemente al mayorista, es decir, aquellos que son fieles al supermercado. En base a esto, las variables relevantes a considerar si un cliente es fiel o no son las siguientes:

- Temporalidad de datos: Se obtendrán los datos de un periodo de dos años (2011-2012), con el objetivo de tener datos transaccionales suficientes para estimar los modelos, especialmente el predictivo.
- Escoger aquellos clientes que son fieles, es decir, que tengan cierta periodicidad mensual y que no falten más de un tiempo determinado, ya que se considera como clientes fugados.

7.2.2. Selección de categoría

Para la elección de la categoría, se analizarán varios indicadores de venta de las categorías existentes en los clientes identificados del punto anterior.

La elección se basa en temas estratégicos, tales como identificar el rol de cada categoría² para escoger aquella que sea representativa dentro del supermercado.

Los indicadores que se utilizan son los siguientes:

² Destino, habitual, ocasional y de conveniencia. Más adelante se detalla la estrategia de cada uno de los roles que cumplen las categorías.

- Margen: Para especificar el margen de la categoría, primero se tiene que ver la contribución de ésta que corresponde a:

$$\text{Contribución} = \text{Monto} - \text{Costo}$$

Ya obtenida la contribución por categoría se calcula el margen como:

$$\text{Margen} = \frac{\text{Contribución}}{\text{Monto}}$$

- Presencia: Corresponde al porcentaje de representatividad de que tiene la categoría en las boletas. Se calcula de la siguiente manera, en base a los clientes en estudio:

$$\text{Presencia} = \frac{\text{Distintas transacciones de la categoría}}{\text{Distinta tansacciones totales}}$$

- Monto de ventas: Corresponde a la suma de ticket total de los clientes en estudio.

7.3. Análisis descriptivo de los datos

Esta etapa tiene el objetivo de comenzar a familiarizarse con los datos, conocer a los clientes en estudio, como tratarlos, saber el comportamiento de éstos y analizar la situación actual de la efectividad de las promociones personalizadas.

7.4. Desarrollo de modelo

Para el desarrollo de los modelos es necesario usar el método KDD mencionado en el marco metodológico, ya que es necesario extraer las variables relevantes para que éstos no pierdan capacidad predictiva. Para ambos modelos se utiliza una regresión logística binaria, variando la variable target u objetivo y las variables independientes que describen a ésta.

A continuación se detalla la metodología para el modelo que se utiliza:

7.4.1. Modelo entre-compras

Las variables escogidas para este modelo se basan en la metodología RFM. Las variables escogidas son³:

- Recency: Corresponde a los días entre transacciones por cada cliente.
- Monto transacción: Corresponde al monto total transado por cliente, aunque no compre en la categoría escogida.

³ Las variables RFM, se repiten para la categoría, es decir, se calculan tanto para las transacciones totales, como para las transacciones donde se compra en la categoría escogida.

- Frecuencia mensual: Frecuencia mensual de asistencia al mayorista por cada cliente y ocasión de compra.
- Frecuencia semestral móvil: Corresponde a la frecuencia histórica de los últimos seis meses de historia por cliente.
- Frecuencia diaria: Corresponde al cálculo de la frecuencia semestral móvil pero expresada en días.
- Recency/Frecuencia: Indicador que muestra la relación entre la los días entre compra que lleva y la frecuencia con la cual habitúa asistir al supermercado. Se pueden dar los siguientes casos:
 - ✓ Menor a uno: Indica que los días que han transcurrido desde la última compra que realizó el cliente, es menor a la frecuencia que habitúa tener. Por lo tanto, es poco probable que compre en la próxima ocasión de compra.
 - ✓ Igual a uno: Indica que los días que han transcurrido desde la última compra que realizó el cliente, es igual a la frecuencia que habitúa tener.
 - ✓ Mayor que uno: Indica que los días que han transcurrido desde la última compra que realizó el cliente, es mayor a la frecuencia que habitúa tener. Por lo tanto, la probabilidad que compre en la próxima ocasión de compra es mayor.
- Indicador de días feriados y fines de semana: Corresponde a una variable dummy, donde toma el valor uno cuando es feriado o fin de semana y cero si no.
- Target: Variable objetivo dummy, donde toma el valor uno cuando el cliente compra en la categoría en estudio y cero si no.

Ya identificadas las variables, se procederá a evaluar la representatividad de cada una de éstas en el modelo con el objetivo de eliminar aquellas que no sean representativas o que no tengan capacidad de predicción en la variable objetivo.

La variable target u objetivo del modelo es la compra o no en la categoría escogida para el análisis. Por lo tanto, las variables antes descritas, deberán tener capacidad de predicción de la compra en cierta categoría.

7.5. Evaluación modelo

Para evaluar el modelo se considerará el estadístico pseudo R^2 , el cual puede tener un valor entre 0 y 1. A mayor predicción del modelo, mayor es el R^2 , sin embargo se considera como buena predicción de 0.5 hacia arriba.

Por otro lado, se realizará la evaluación del modelo con datos de Enero 2013 a Agosto 2013, de los clientes escogidos para la calibración del modelo. Se escoge este rango de fechas, ya que para registros anteriores (2011 y 2012) no se tiene información de la oferta en los *voucher*, al tener esta información, se contrastará lo que ofrecería el modelo para ofertar versus la oferta real brindada

por el kiosco, y por otro lado se ve finalmente si el cliente compra o no en la categoría.

8. Desarrollo

8.1. Selección y análisis descriptivo de los datos

8.1.1. Clientes

Para comenzar el análisis es necesario crear una base de datos completa con la información necesaria para realizar un análisis específico.

El mayorista posee una base de datos general de las transacciones de cada cliente registrado. Se comenzó identificando a aquellos clientes pertenecientes al Club Convenio, quienes se inscriben en el supermercado asociando un rut a un rubro de negocio con el cual podrán tener acceso a promociones personalizadas de descuento por medio de la impresión de *vouchers* del sistema kiosco. Para esto se realizaron consultas usando el lenguaje SQL hasta obtener a los clientes objetivos para obtener la data transaccional.

Aumentando la productividad de Data Mining

Para consolidar la base de datos de los clientes a estudiar, se realizaron los siguientes filtros:

1. Socios 2011-2012: Se analizaron las transacciones realizadas de los clientes pertenecientes al Club Convenio entre los años 2011 – 2012.
2. Frecuencia mensual menor a 60: Se eliminaron aquellos clientes cuya frecuencia mensual superaba las 60⁴ transacciones, ya que existían registros que tenían muchas transacciones. Esto se debe básicamente a que los clientes “prestan su rut” para tener mayor acceso a descuentos.
3. Frecuencia mensual, con 3 fallas: Se eliminaron aquellos clientes que no tienen una frecuencia mensual de asistencia al mayorista, permitiendo 3 meses de holgura, ya que se quiere tener la certeza de que el cliente realiza transacciones con el fin de tener la información necesaria para la estimación de tiempo entre compra en una categoría dada.
4. Fallas no seguidas: Finalmente, se eliminaron aquellos clientes que no tuvieron transacciones durante 3 meses seguidos, ya que otra regla de negocio establecida por la empresa, es que después de 3 meses sin transacciones se comienza a catalogar al cliente como fugado.

Explicado los filtros realizados, a continuación se muestra una descripción de cada regla aplicada:

⁴ Se considera un corte de 60 transacciones mensuales debido a reglas de negocio utilizadas por la empresa.

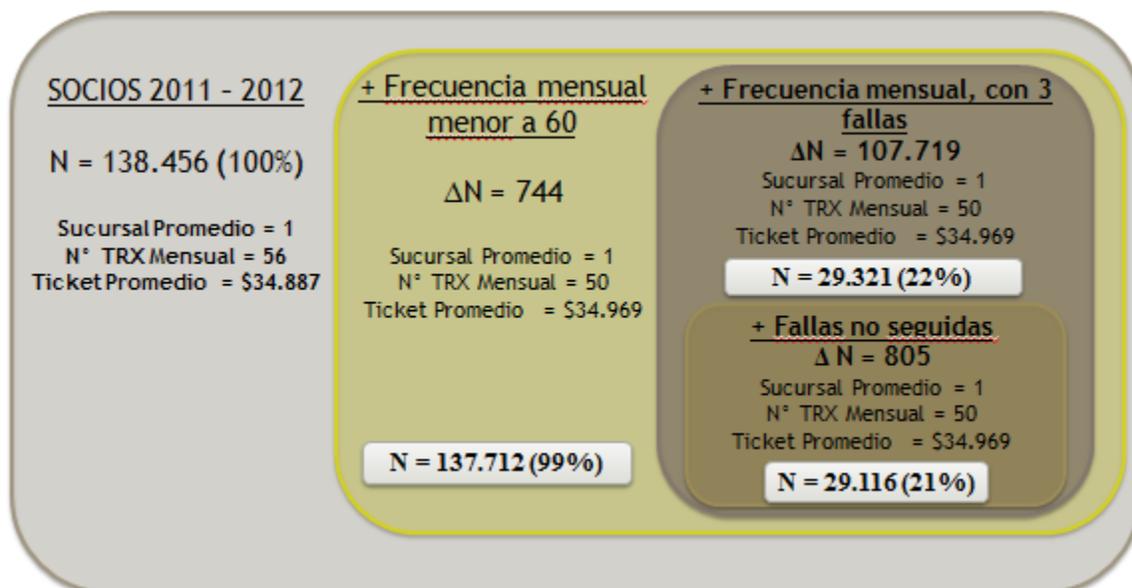


Ilustración 6: Caracterización de la información. Elaboración propia.

Dado que los clientes pertenecen al Club Convenio, se tiene la información de a cuál rubro pertenecen sus negocios. El detalle del giro del negocio de los clientes a estudiar se presenta a continuación:

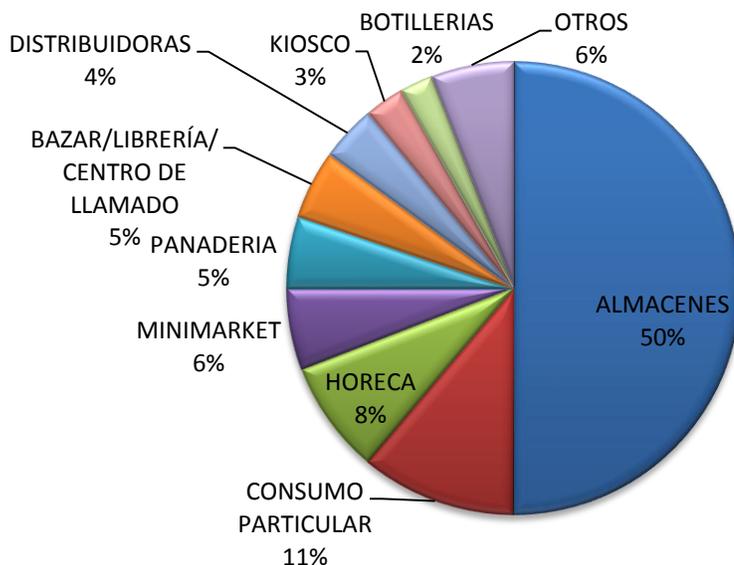


Ilustración 7: Rubro clientes estudio. Elaboración propia

Se puede observar que la mayoría de los clientes pertenece al rubro “Almacenes”. Esta información es importante, ya que son clientes que en sus transacciones tienen un gran mix de productos transados, gracias al alto mix que deben ofrecer ellos a su vez a sus clientes. Por lo tanto, es posible obtener valiosa información de ellos.

Luego de los filtros realizados, otra variable importante de observar dentro de los clientes identificados, es el comportamiento que tienen con respecto a la impresión y canje de *vouchers* que se les ofrece cada vez que se ven enfrentados a una compra. Esta información es relevante, ya que se puede observar a primera vista si las promociones son efectivas o simplemente se están gastando recursos innecesariamente que no contribuyen con el objetivo que tiene las promociones personalizadas: fidelización de clientes. Para observar este fenómeno, se mide la impresión y canje de los *vouchers* ofrecidos.

Para analizar las impresiones y canjes de los clientes, se observan los datos correspondientes al mes de Diciembre del año 2012. Dentro de este periodo de tiempo, sólo se tienen 27.426 clientes que presentan al menos una transacción en el mes, donde las impresiones y canjes de éstos se presentan a continuación⁵:

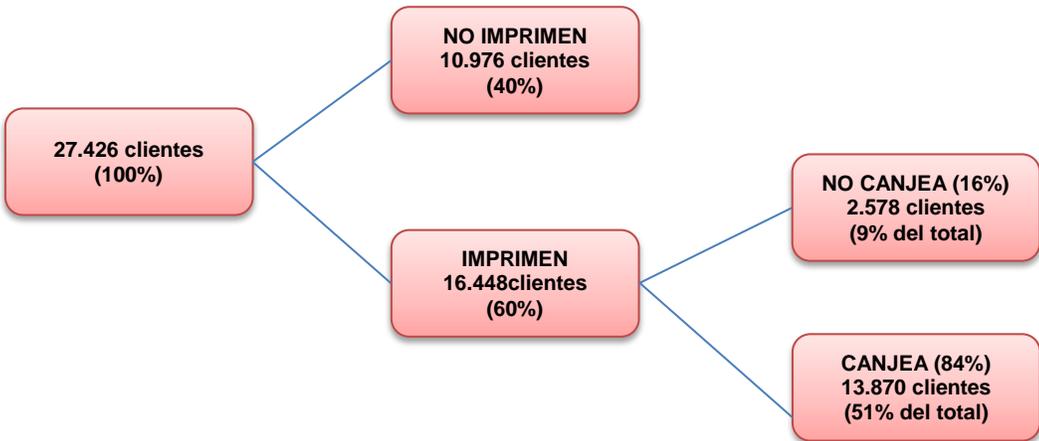


Ilustración 8: Caracterización de impresión y canje para clientes en estudio para Diciembre de 2012. Elaboración propia

Se puede observar que la tasa de impresión de los clientes observados corresponde al 60%, y la tasa de canje de los que imprimen en el periodo de diciembre de 2012 es de 84%, que representa el 51% de los clientes que tuvieron transacciones durante el periodo descrito. Cabe destacar además, que 3.296 clientes (24% de los que canjean) son los que canjean postres refrigerados (categoría que se analizará más adelante).

Al analizar estos indicadores, se puede ver que la tasa de canje e impresión es alta (sobre el 50%), sin embargo, hay que tener en cuenta que se considera como canje al menos un producto que es canjeado, por lo que si se analizara a nivel de producto o transacción por cliente, la tasa de canje disminuiría. Por otro lado, existe un 16% de clientes que imprimen y no canjean ninguno de los seis productos ofertados en el *voucher*, por lo que aún queda por mejorar.

⁵ Cabe mencionar que para el supermercado mayorista se considera exitoso si el cliente ha impreso al menos una vez en el periodo de Diciembre 2012 y, por otro lado, en los canjes, se considera exitoso si se canjea al menos un producto en el mes descrito.

Además, si se analizan estos indicadores a nivel de transacciones se obtiene lo siguiente:

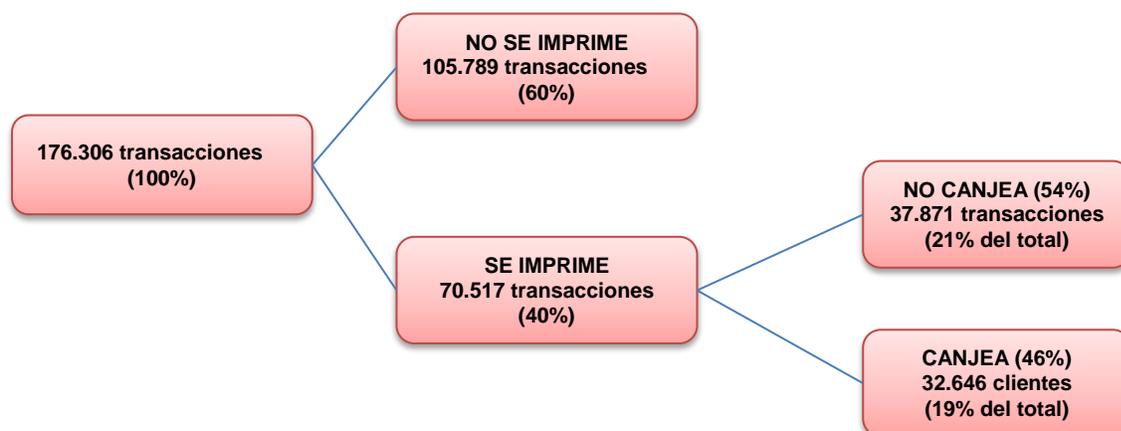


Ilustración 9: Caracterización de impresión y canje para transacciones en estudio para Diciembre de 2012. Elaboración propia

Se puede ver que la tasa de canje disminuye a un 46% (de un 84%) al analizarlo con respecto a las transacciones (en lugar de clientes) que contienen impresión y canje en el periodo de diciembre de 2012. Además, 5.442 transacciones corresponden a postres refrigerados (17% de las transacciones canjeadas).

Por lo tanto, el 51% de los clientes ha utilizado, al menos una vez, alguna promoción que aparece en el *voucher* durante diciembre del 2012. Pese a lo anterior, en términos de las transacciones, en un 54% de éstas se imprime el *voucher* pero no se canjea (no se utiliza). Es aquí donde se ve la oportunidad de aumentar la tasa de canje, ya que existe un 54% de las promociones que no se utilizan, es decir, que son promociones mal asignadas, de las cuales se debe mejorar. Lo anterior queda mejor representado mencionando que de cada 100 clientes, 51 canjea al menos una vez durante diciembre, y suponiendo que cada uno de esos clientes realiza 10 transacciones durante diciembre, solamente en el 46% de dichas transacciones (aproximadamente en 5 transacciones del mes) se canjea algún producto de los que se ofrecen en el *voucher*.

Los estudios que se realizarán, serán con respecto a los canjes de aquellos productos canjeados, condicionado a la impresión, ya sea a nivel de cliente o a nivel de transacción.

Finalmente, se puede apreciar que existe un espacio aún para mejorar la efectividad de los productos que aparecen en el *voucher*, de manera que sean productos que realmente le interesen al cliente tratando de maximizar el canje de éstos, y llegar a persuadir efectivamente en la compra. Es posible también que los clientes no utilicen las promociones durante todas las transacciones que realizan (54%) debido a que en el *voucher* aparecen productos que no son de su interés, así como también puede ser que los clientes no estén imprimiendo los *vouchers*

(60%) debido a que se han dado cuenta de que los productos que ahí aparecen están lejanos a sus requerimientos.

8.1.2. Categoría a estudiar

En base a los clientes identificados, se realizó un *mapping* de las distintas categorías presente en las transacciones. A continuación se muestran las categorías según la ocasión de compra que se tiene en cada una de éstas.

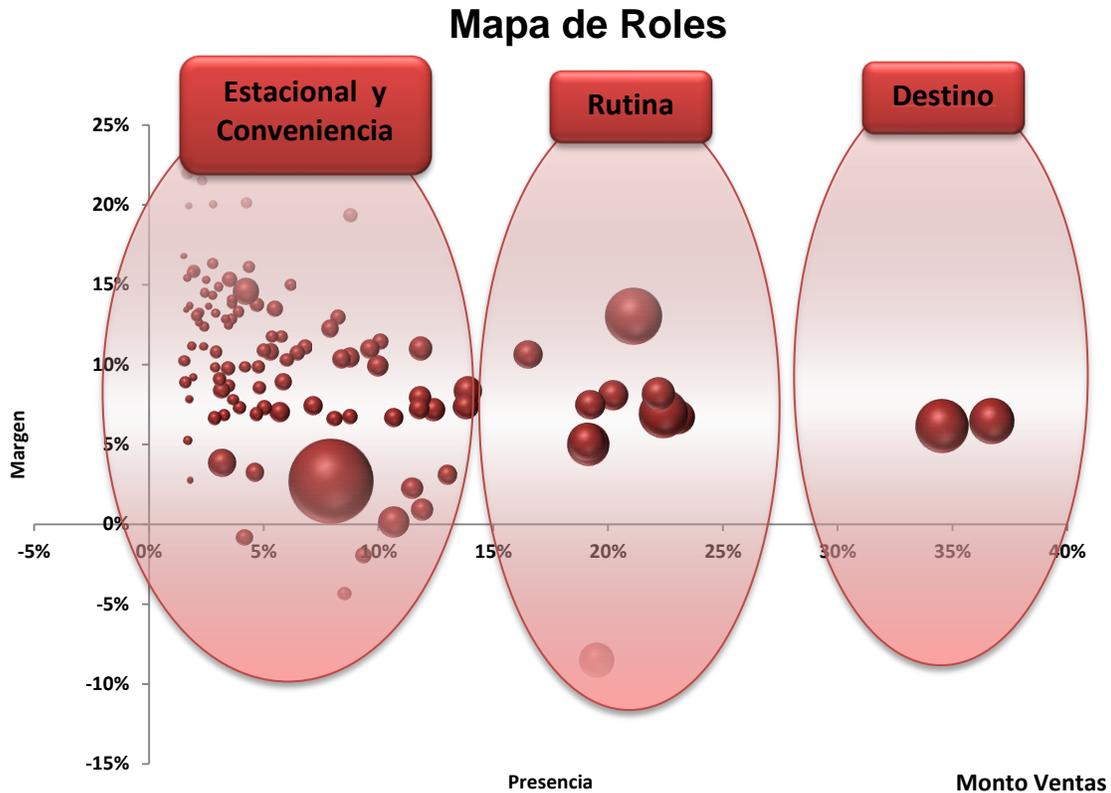


Ilustración 10: Mapa de roles de las categorías. Elaboración propia.

Se puede observar claramente tres grupos. Donde el grupo que posee mayor presencia corresponde a categorías de destino (“Yogurt” y “Leches líquidas”), ya que el cliente va en busca de éstas específicamente en una ocasión de compra dada. Los que poseen menor frecuencia corresponden a categorías de rutina, las cuales son habituales dentro de la canasta de compra de los clientes. Finalmente, los que poseen menor presencia, corresponden a la mayoría de las categorías, donde la compra de éstas se vuelve estacional y no se ven muchas transacciones en los clientes.

En base al papel que tienen las categorías según la identificación realizada, es que se analizan para escoger en cuál basar el estudio, del cual se espera realizar reglas que permitan mejorar los resultados para aumentar la efectividad de

las promociones. Las características de cada papel en las categorías se presenta a continuación [13]:

Tabla 1: Características rol de las categorías.

Papel de la categoría	Surtido	Precios	Presentación en el lineal	Promoción
Destino	Surtido completo	Liderazgo	Ubicación óptima de la tienda	Alto nivel de actividad promocional
Habitual	Muchas variedades	Competitivo de forma consistente	Ubicación intermedia	Nivel medio de actividad promocional
Ocasional (Estacional)	Variedades en el período estacional	Competitivo de forma estacional	Buena ubicación en el período estacional	Actividad promocional estacional
Conveniencia	Variedades elegidas selectivamente	No inflar	Ubicación que quede disponible	Bajo nivel de actividad promocional

Fuente: Elaboración propia.

Al ver las categorías de destino, se observa que deben tener un alto nivel de promoción, ya que corresponden a categorías que fomentan la fidelidad hacia el supermercado mayorista por parte de los clientes. Si bien es una buena opción para estimar la compra en la categoría, ya que se tiene la suficiente información para el modelo, no lo es para ofrecer promociones *up-selling*, pues los clientes que van a comprar en las categorías de destino, tienen una marca y formato específico del cual compran, por lo que será muy difícil que acepten una promoción que cambie su decisión de compra.

El otro extremo corresponde a las categorías estacionales y de conveniencia. Estas deben tener una baja actividad promocional, por lo que no representan una categoría para promociones y además se tiene baja presencia en los tickets de los clientes, por lo que se tiene poca información para alimentar al modelo con el objetivo de estimar la compra en la categoría.

Finalmente, las categorías de rutina, poseen mayor presencia para estudiar el comportamiento de compra en los clientes y además se podrían realizar promociones tipo *up-selling* con el objetivo que cambien el producto que habitúan comprar los clientes por medio de promociones ofreciéndoles otro más sofisticado, el cual es más probable que acepten la promoción comparándolo con las categorías de destino.

A continuación se muestra un *mapping* de las categorías de rutina:

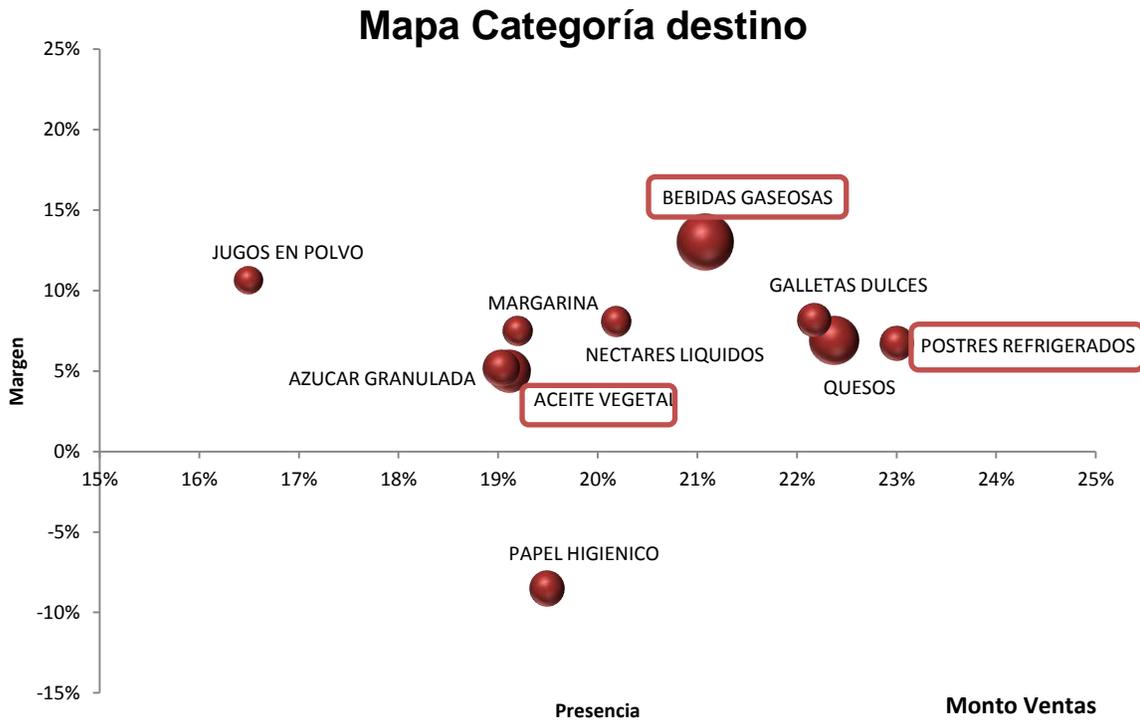


Ilustración 11: Mapping categoría destino. Elaboración propia

El volumen de las burbujas representan las ventas totales durante el periodo de estudio. Se puede observar que la categoría “PAPEL HIGIENICO” muestra que tiene un margen negativo. Sin embargo, en la práctica no es así, ya que es una categoría que frecuenta estar en promoción, por lo que se observa que margina negativamente. El detalle de cada una de las categorías se presenta en el Anexo 2.

Dentro de las categorías que poseen un mayor margen y mayor presencia en las boletas de las transacciones de los clientes en estudio, están las categorías de “galletas dulces”, “bebidas gaseosas”, “quesos” y “postres refrigerados”. Al conversar con expertos en el negocio, es que se vuelve interesante observar las categorías “postres refrigerados”, “bebidas gaseosas” y “aceite vegetal”, ya que son representativas en el monto de ventas.

Se analizaron las 3 categorías antes mencionadas, para las cuales se calcularon las correlaciones entre 3 variables (días entre transacciones, transacciones promedio por cada cliente en la categoría y monto transado) para cada una de las categorías. Además, para cada categoría, se analizaron cuatro fuentes de información distintas: los datos totales de aquellos clientes que han comprado alguna vez en la categoría (1), los datos totales de aquellos clientes que han comprado alguna vez en la categoría pero normalizado con el monto

promedio transado por cada cliente (2), los datos de aquellos clientes que tienen una frecuencia entre 40 y 300 transacciones en el periodo de estudio (3) y los datos de aquellos clientes que tienen una frecuencia entre 40 y 300 transacciones en el periodo de estudio pero normalizado con el monto promedio transado (4).

El detalle de clientes y número de transacciones realizadas en el periodo de estudio por categoría se presenta a continuación:

Tabla 2: Caracterización categorías rutina. Elaboración propia.

Categoría	# Clientes	# TRX
Bebidas Gaseosas	27.200	3.681.352
Postres Refrigerados	26.477	3.513.348
Aceite Vegetal	27.561	1.022.694

Fuente: Elaboración propia.

Los resultados de los signos de las correlaciones del análisis en las categorías antes mencionadas se presentan en Anexo 3.

Al observar el comportamiento en cada una de las categorías, se puede ver que aquella categoría que tiene un comportamiento más uniforme es la categoría “postres refrigerados”, ya que para el análisis de los días entre compra y el monto transado se ve que a mayor días entre compras, menor es el monto transado; para el análisis entre los días entre compras y el número de transacciones, se tiene que a mayor número de transacciones, menos días transcurren entre compras y finalmente la relación entre las transacciones y el monto es positiva lo cual indica que a mayor número de transacciones, mayor es el monto.

En base a lo anterior, es que la categoría a analizar corresponde a “postres refrigerados”, ya que corresponde a una categoría que tiene un comportamiento estable en el análisis realizado, se tienen muchas transacciones por parte de los clientes observados y además corresponde a una categoría de alimento perecible, por lo cual tiene una rotación alta. Esta categoría posee 126 SKU distintos, los cuales pertenecen a 6 marcas y a 15 tipos de postres.

8.1.3. Variables

Las variables que se utilizarán para estimar la compra en la categoría son basadas en la técnica RFM, se tiene la siguiente dinámica tanto para la *recency*, frecuencia y monto:

Se calculan cinco variables con respecto a la *recency*, frecuencia y monto, tanto para las distintas transacciones (independiente si se compra o no en la categoría) como para las compras en la categoría. Por lo tanto existen diez variables con respecto a cada ítem de la técnica RFM: cinco para transacciones

generales y cinco para compras en la categoría, por lo tanto, se calculan 30 variables. Cabe destacar que no todas se incluirán en el modelo, ya que algunas corresponden a cálculos previos de las variables finales.

La dinámica de cálculo es la siguiente⁶:

- i. La primera corresponde a los días transcurridos desde la última compra por cada cliente. Para el caso de compra en la categoría, corresponde a los días transcurridos entre compras de la categoría por cliente.
- ii. La segunda corresponde a los días acumulados desde la primera compra, tanto a nivel general como en la categoría.
- iii. La tercera corresponde al *recency* promedio, tanto para la categoría como a nivel general.
- iv. La cuarta corresponde al *recency* acumulado móvil calculado desde el día de la transacción a seis meses móviles anteriores.
- v. La quinta corresponde

Además se incluyen el monto sobre los días transcurridos y los días transcurridos sobre la frecuencia, tanto para transacciones en general como para transacciones que contienen la categoría escogida.

8.2. Pre-procesamiento y transformación de datos

8.2.1. Limpieza datos

Una vez calculadas las variables, la única limpieza de éstos fue eliminar aquellos registros en el cual el monto era negativo, ya que correspondían a evaluaciones.

Los datos se agruparon por fecha de transacción, es decir, si un cliente asistió al mayorista más de una vez al día, estas transacciones se ven reflejadas en un dato, el cual contempla la suma de todas las transacciones diarias.

8.2.2. Transformación datos

En la transformación de datos se fue muy minucioso en el cálculo de cada una de las variables, ya que se realizó un cálculo especial en aquellos días donde las variables de la categoría no existían, ya que el cliente no compraba en la categoría, por lo que se trató de crear variables lo más representativo de la realidad, de lo que está sucediendo como inventario de la categoría estudiada y que tenga capacidad de predicción para la variable target, es decir, si comprará o no en la categoría de estudio en la próxima ocasión de compra.

⁶ Se ejemplifica para el cálculo de la *recency*.

8.3. Desarrollo de Modelo

8.3.1. Modelo para predicción de compra

8.3.1.1. Desarrollo modelo

El modelo utilizado en una regresión logística binaria, se utiliza un software para su desarrollo, donde se utiliza la técnica hacia delante de Wald. Esta técnica comienza con el intercepto y va agregando variables representativas al modelo hasta llegar a introducir todas las variables.

Para realizar la calibración del modelo, se iteró diez veces el modelo, donde se fue variando el cohorte para la estimación de la probabilidad, por ejemplo, si el cohorte es al 50%, todas aquellas probabilidades mayores a éste porcentaje, el modelo lo clasificará como 1, es decir, que ofrecerá al cliente la promoción en la categoría, por otro lado si el cohorte es 80%, todas probabilidades menores a ésta, la clasificará como 0, es decir, no ofrecerá promoción en la categoría.

Para escoger el cohorte óptimo y obtener los coeficientes, se analizaron los errores de cada una de las iteraciones en base a los errores típicos que brinda la matriz de confusión.

Específicamente para este análisis, el error tipo 1 como se muestra en la ilustración 4, corresponde a no ofrecer postre, cuando debería haber ofrecido, ya que el cliente compra en la categoría, lo cual se traduce a perder un canje en postre, que corresponde a la contribución promedio de un postre por la probabilidad que se compre un postre por medio de un canje. Vale mencionar que esta probabilidad corresponde a nivel de producto, es decir, dado la oferta, cual es la probabilidad de que sea canjeada, y no la probabilidad de éxito de canje como se mencionaba en el análisis descriptivo (dado seis productos que se ofertan en el *voucher*, si se canjea al menos uno, se considera exitoso)

Por otro lado, el error tipo 2, corresponde a ofrecer postre cuando no debería haber ofrecido, es decir, pierdo que el cliente canjee algún otro producto que se ponga en lugar de postre, lo cual se traduce a costos por medio de la contribución de un postre promedio por la probabilidad de canje de un producto promedio. La probabilidad acá sigue la misma intuición que la probabilidad de canje de un producto en la categoría postre refrigerado.

Los valores con lo cual se calculan los errores estándar para cada uno de las iteraciones se muestran a continuación:

Tabla 3: Datos promedios. Elaboración propia.

Probabilidad de canje de un postre	Probabilidad de canje de un producto x	Contribución promedio de un postre	Contribución promedio de un producto x
25%	20%	\$69 pesos	\$168 pesos

Fuente: Elaboración propia.

La curva ROC se ve representada de la siguiente forma:

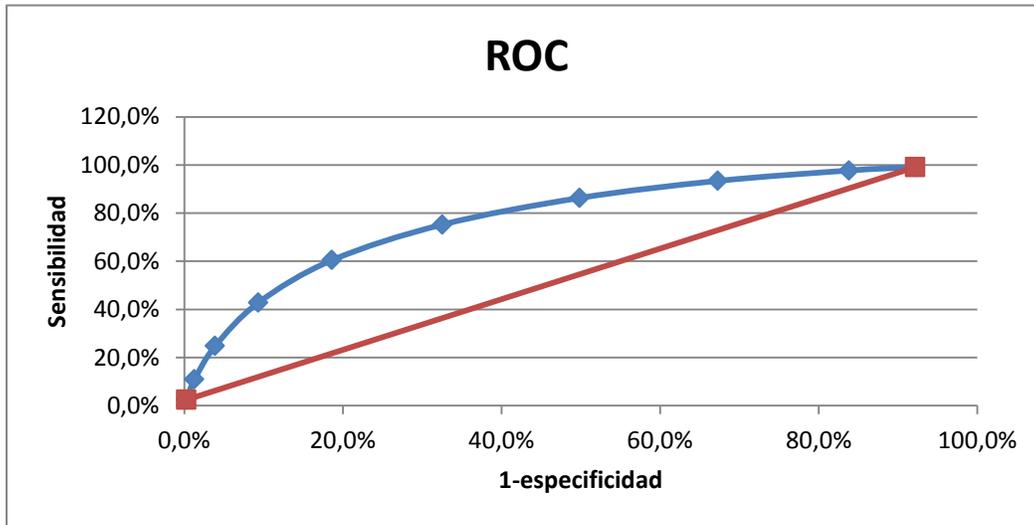


Ilustración 12: Curva ROC. Elaboración propia.

Por otro lado, la curva de beneficios con respecto a cada cohorte se ve de la siguiente manera:

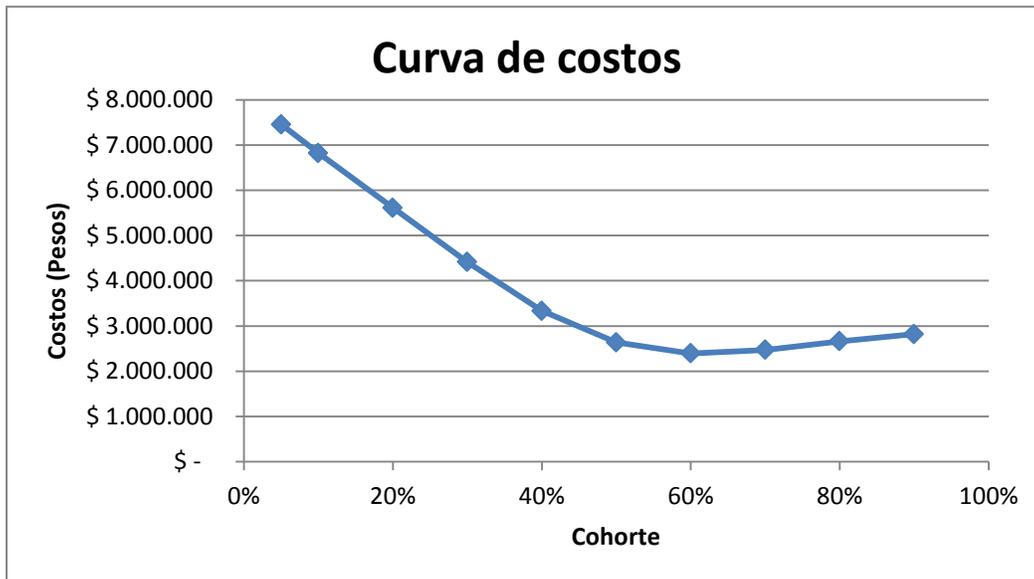


Ilustración 13: Curva de costos. Elaboración propia.

Al observar la curva de costos, se puede ver que el mínimo costo se obtiene al 60%, es decir, todas aquellas probabilidades mayores al 60%, estimado ya los coeficientes, el modelo dirá que el cliente tiene alta probabilidad de comprar en la categoría, por lo que le ofrecerá una promoción.

Una vez escogido el cohorte, se puede observar la siguiente matriz de confusión⁷:

		REAL	
		Corte 60%	0
PREDICCIÓN	0	217.369	22.461
	1	95.076	71.268

Ilustración 14: Matriz confusión. Elaboración propia.

Al calcular los parámetros de la matriz se obtiene lo siguiente:

Sensibilidad=69,6%

Especificidad=76,0%

Accuracy=71,1%

Se puede observar que el modelo acierta correctamente al 71% del total de los casos, es decir, acierta al 69,6% de los casos cuando el cliente no compra en la categoría, por lo que no se le ofrecerá promoción y, por otro lado, acierta en un 76% a aquellos clientes que sí compran en la categoría, por lo que se le ofrecerá promoción.

Los betas correspondientes a las variables significativas (al 95%) son las siguientes:

- R: Corresponde a los días entre visitas al mayorista, independiente o no si compra en la categoría poste.
- R acumulado móvil: Corresponde a los días promedio entre cada visita al mayorista durante los seis últimos meses, independiente si compra o no en la categoría.
- Frecuencia mensual: Corresponde a la frecuencia con la cual habitúa ir el cliente en estudio, independiente si compra o no en la categoría.
- R/F: Corresponde a la relación entre la los días entre compra que lleva y la frecuencia con la cual habitúa asistir al supermercado.

⁷ Los resultados del resto de los cohortes se puede ver en anexo 12.3.

- R promedio móvil en la categoría: días promedios entre compras, cuando se ha comprado al menos un producto perteneciente a la categoría.
- Frecuencia móvil en la categoría: frecuencia promedio mensual en la categoría calculada hasta seis meses atrás.

Variables	Betas	Significancia
Recency	0,073	0,033
Recency Móvil	-0,006	0
Frecuency Mensual	-0,069	0
Recency/Frecuency	0,001	0,013
Recency Prom en la categoría	-0,018	0,017
Frecuency móvil en la categoría	0,042	0,001
Constante	0,128	0

Ilustración 15: Resultados. Elaboración propia.

Se puede concluir que al aumentar los días entre compra a nivel del mayorista, aumenta la probabilidad de compra en la categoría, pero si los días entre compra en los últimos seis meses, es menor, la probabilidad de compra en la categoría aumenta, además los días entre compras promedio en la categoría disminuye, la probabilidad de compra en la categoría aumenta. Por otro lado, si la frecuencia mensual de compra en el mayorista disminuye, la probabilidad de compra en la categoría aumenta, es decir, a medida que pasan menos días entre compras, es más probable que se compre en la categoría, compra más seguido en la categoría.

Además, el coeficiente del indicador que muestra la relación entre los días entre compra y la frecuencia, es positivo, por lo que a medida que aumenta el valor del factor, aumenta aún más la probabilidad de compra en la próxima visita al mayorista. Recordar que si el factor es mayor a uno quiere decir que los días que transcurren desde la última compra, es mayor que la frecuencia que habitúa tener, por lo que mayor es la probabilidad de compra en la próxima visita.

Finalmente, al aumentar la frecuencia de compra en la categoría, aumenta la probabilidad de compra de ésta en la próxima visita.

8.3.1.2. Evaluación e Interpretación

Para la evaluación del modelo, se calcularon las mismas variables con las que se calibró el modelo, pero con datos pertenecientes a Enero 2013 hasta Agosto 2013, ya que en este período se tiene la información de la oferta en las promociones brindadas en los *voucher*.

Por lo tanto, se contrasta la oferta real ofrecida por el kiosco (observar ilustración 17) con lo que el modelo predijo (ilustración 16), es decir, si es oportuno

ofrecer una promoción en la categoría postres refrigerados, dada las variables significativas que arrojó el modelo.

Los resultados de la evaluación final del modelo con datos del 2013 se puede ver en la ilustración 16 y 17.

		REAL	
		Corte 60%	0
PREDICCIÓN	0	1.780	170
	1	776	590

Ilustración 16: Simulación datos 2013. Enero -Agosto Predicción modelo. Elaboración propia.

		REAL (Compra)	
		Corte 60%	0
KIOSCO	0	350	2.029
	1	137	800

Ilustración 17: Simulación datos 2013. Enero - Agosto. Datos reales kiosco. Elaboración propia.

La sensibilidad para la oferta que brindaría el kiosco con el modelo antes descrito es de un 69,9%, por otro lado, la asignación real que tuvo el kiosco es de un 71,9%. Si bien la asignación actual predice más cuando el cliente no compra en la categoría (0), la especificidad del modelo corresponde a un 77,6% contrastado con lo que ofreció el kiosco en el mismo periodo, que corresponde a un 28,3%, es decir, si se realizara la asignación de promociones con el modelo antes descrito, se acertaría mucho más a aquellos clientes que comprarán en la categoría (1).

A nivel general, el modelo tiene una *accuracy* de un 71,1% contrastado con lo que ofertó el kiosco en el mismo periodo, que corresponde a un 34,7%. Se puede ver que existe una gran mejora, ya que se acierta mucho más aquellos clientes que no comprarán y que comprarán en la categoría estudiada.

Se puede ver que el error tipo II es muy alto, es decir, no ofrecer cuando debería haber ofrecido. Es por esto que el modelo intenta reducir este costo de cometer este error.

Por otro lado, el modelo aumenta la oferta en un 46%, ya que de ofertar 937 postres para a ofertar 1366 postres de los 3316 registros utilizados para la evaluación.

Además se puede ver que el error tipo I de los datos del modelo corresponden a \$ 403.783 y los del error tipo II a \$ 185.805, los cuales suman un total de \$ 589.587. Por otro lado, el costo del error tipo I del actual kiosco es de \$ 71.268 y el del tipo II es de \$ 2.055.923, los cuales suman un total de \$ 2.127.191.

A nivel general se puede ver que existe un ahorro en costos de \$ 1.537.604 pesos en el periodo considerado (enero – agosto 2013).

9. Conclusiones

Al realizar este trabajo tanto a nivel cualitativo como cuantitativo, se puede observar la dificultad de asignar promociones a los clientes, con el objetivo de no mal gastar recursos. Además se requiere que los clientes recepcionen de buena manera y acepten dichas promociones para lograr así la efectividad en las asignaciones de éstas.

Es posible mencionar que a lo largo del desarrollo del presente trabajo de memoria, a través de la metodología planteada y llevada a cabo, se lograron concretar los objetivos planteados en un comienzo.

Se pudo evaluar una posible mejora en las promociones focalizadas que actualmente se tiene en el kiosco del mayorista. Al analizar cuando será el momento oportuno de ofrecer un producto al cliente correcto, se puede obtener una mejora de la eficiencia en la asignación utilizando los mismos recursos disponibles y actuales usados en el sistema kiosco.

Si bien, este estudio no está completo, ya que existe una variable clave que no se pudo incluir por falta de recursos e información disponible para el período de tiempo considerado (la oferta en el *voucher*), se puede apreciar que es factible disminuir los costos del actual kiosco, lo cual se ve reflejado en los resultados, que de cada cien mil productos transados, se puede ahorrar cerca de \$1.537.604 de pesos en el periodo considerado. Sabiendo que en el año 2012 se transaron 69.222.781 productos, el mayorista estaría ahorrando en costos cerca de MM\$106 al año.

Si bien el estudio realizado fue para la categoría específica de postres refrigerados, es replicable para cualquier otra categoría, siendo riguroso en que existan los datos suficientes para llegar a conclusiones concretas y lo más cercanas a la realidad.

Al evaluar posibles mejoras en la asignación de promociones focalizadas, el enfoque realizado en el presente trabajo, no es la única manera de abordarlo. Otra aplicación es la continuación de la asignación del momento oportuno, ya que si se sabe que con alta probabilidad el cliente comprará en una cierta categoría se podría aprovechar esta oportunidad para ofrecer un producto perteneciente a una gamma superior del que actualmente lleva (*up-selling*). Ya sea ofreciéndole un producto de mejor marca, un pack, entre otros, con el objetivo que el cliente cambie su comportamiento de compra y comience a llevar productos mejores que los que habitúa llevar, que otorguen una mejor contribución.

Finalmente, se puede observar que existen oportunidades de mejoras en diversos ámbitos dentro del tema de las promociones y retail, ya que el tema y los recursos son amplios. Existen muchos estudios que mejoran la asignación de promociones que se pueden utilizar para obtener beneficios de éstas y además, que es lo más importante para los supermercados en estos tiempos, lograr la fidelización del cliente y mantenerlo satisfecho en cada ocasión de compra.

10. Trabajos Futuros

Con el objetivo de seguir aumentando la efectividad de las promociones, incluyendo un beneficio extra para el mayorista (a nivel de margen por producto canjeado por ejemplo), se pueden brindar promociones tipo *up-selling*, ya que con el presente trabajo se ha obtenido el momento oportuno en el cual ofrecer un producto. Al ofrecer promoción del tipo *up-selling*, sería tentador para el cliente cambiarse de marca por ejemplo, dado que ya va a comprar en cierta categoría.

Otro trabajo futuro tiene que ver con las variables consideradas dentro de la estimación de compra en una categoría, ya que no son las únicas, pudiendo influir una variable que entre el 2011 y 2012 no se tenía registro de ella, correspondiente a la oferta de productos que se da en el *voucher* promocional al cliente. **Dicha** variable puede ser muy influyente en la decisión de compra en sala para el cliente, ya que al ver un producto en promoción, del cual tiene bajo stock.

Con estas recomendaciones se complementarían el trabajo actual de aumentar la efectividad de las promociones personalizadas que ofrece el kiosco del mayorista., así como detectar oportunidades de mejoras en los ingresos del retail mediante la implementación de promociones de tipo *up-selling*.

11. Bibliografía

- [1] Tapia, M.J. Ventas de retail crecerán 14,5% en Chile a 2012 y bajarán peso en la región a 3,1%. [en línea] La Tercera en Internet. 21 de febrero, 2012 < <http://diario.latercera.com/2012/02/21/01/contenido/negocios/10-101277-9-ventas-del-retail-creceran-145-en-chile-a-2015-y-bajaran-peso-en-la-region-a-31.shtml>> [consulta: 23 de noviembre 2012]
- [2] Jorquera, A. Mercado Mayorista 10 en Chile Factura US\$2.000 Millones al año. [en línea] Diario la estrategia online. 19 octubre, 2009. <http://www.estrategia.cl/detalle_noticia.php?cod=24085> [consulta: 23 noviembre 2012]
- [3] Jorquera, A. Mayorista 10 triplica participación de mercado en dos años. [en línea] Diario la estrategia online. 3 de febrero, 2011. < http://www.estrategia.cl/detalle_noticia.php?cod=37262> [consulta: 23 de noviembre 2012]
- [4] Bosch, M. 2011 El crecimiento de la Industria Chilena del Retail, Y su por qué. [diapositivas] Feria ANTAD 2011, Guadalajara, México. CERET Universidad de Chile.
- [5] SOTOMAYOR, C. 2008. Definición de un sistema de recomendación para una empresa de retail. Memoria de Título. Universidad de Chile.
- [6] ARAYA, M. 2011 Estimación de la efectividad promocional en un supermercado mayorista. Memoria de Título. Universidad de Chile.
- [7] FUENZALIDA, B. 2012 Pronóstico de efectividad de promociones sobre clientes, dadas sus características y respuestas pasadas. Memoria de Título. Universidad de Chile.
- [8] Duran Alejandro. MasterBase. 2012 [en línea]. <http://info.masterbase.com/blog/bid/80899/C%C3%B3mo-crear-el-mensaje-correcto-en-el-tiempo-correcto-a-trav%C3%A9s-de-la-conducta>. [consulta: 20 febrero 2013].
- [9] Kotler. P. 2000. Dirección de Marketing. Prentice Hall Iberia. S.R.L. 98p.
- [10] Bosch, M. y Goic, M. 2006. «Promociones: Más que rebajas de precios. Trent Management Edición Especial, Vol. 8, pp. 132-142.
- [11] Gupta, S. 1988 Impact of Sales Promotions on When, What, and How Much to Buy. Journal of Marketing, Vol. 25, nº 4, pp. 432 - 355.
- [12] Gupta, S. Stochastic Models of Interpurchase Time with Time-Dependent Covariates. Journal of Marketing Research, Vol. 28, nº 1, pp. 1-15.
- [13] del Castillo, Á.M. 2004. Gestión por categorías: Una integración eficiente entre fabricantes y distribuidores. Madrid, Gesbiblo S.L. p. 95.

12. Anexos

12.1. *Voucher* ejemplo de promociones por medio del sistema kiosco

Cupón de Promociones: Juan Pérez	
Promociones del día	
YOGHURT SOPROLE YOGHITO DAMASCO 36X125GR -10%	
BEBIDA ORANGE CRUSH VIDRIO 1L - 15%	
YOGHURT SOPROLE BATIDO FRUTILLA 175G - 7%	
FIAMBRE DE JAMON SAN JORGE - 15%	
SHAMPOO BALLERINA BOLSA 5X10UN - 5%	
YOGHURT SOPROLE YOGHITO FRUTILLA 36X125G - 10%	
Promociones válidas sólo por el día	
Club del Supermercado Mayorista	

Ilustración 18: Voucher ejemplo promociones.

12.2. Detalle análisis categorías de rutina

Tabla 4: Estadísticos categorías de destino. Elaboración propia.

CATEGORÍA	Cantidad Clientes	Cantidad TRX (MM)	Monto (\$MM)	Costo (\$MM)	Contribución (\$MM)	Margen	Penetración	Presencia	Ticket Promedio
POSTRES REFRIGERADOS	26,477	1.0	\$ 3,506	\$ 3,269	\$ 236	7%	91%	23%	\$ 3,335
QUESOS	27,374	1.0	\$ 7,278	\$ 6,772	\$ 506	7%	94%	22%	\$ 7,118
GALLETAS DULCES	28,429	1.0	\$ 3,262	\$ 2,994	\$ 267	8%	98%	22%	\$ 3,218
BEBIDAS GASEOSAS	27,200	0.9	\$ 9,755	\$ 8,480	\$ 1,275	13%	93%	21%	\$ 10,129
NECTARES LIQUIDOS	27,556	0.9	\$ 2,711	\$ 2,491	\$ 220	8%	95%	20%	\$ 2,939
PAPEL HIGIENICO	27,843	0.8	\$ 3,643	\$ 3,952	-\$ 308	-8%	96%	19%	\$ 4,090
MARGARINA	27,152	0.8	\$ 2,574	\$ 2,380	\$ 193	8%	93%	19%	\$ 2,934
ACEITE VEGETAL	27,561	0.8	\$ 5,464	\$ 5,188	\$ 275	5%	95%	19%	\$ 6,254
AZUCAR GRANULADA	27,923	0.8	\$ 4,066	\$ 3,855	\$ 211	5%	96%	19%	\$ 4,674
JUGOS EN POLVO	26,517	0.7	\$ 2,374	\$ 2,121	\$ 253	11%	91%	16%	\$ 3,150

Fuente: Elaboración propia.

12.3. Cohortes

		REAL		
		Corte 90%		
PREDICCIÓN	0	239.159	671	99,7%
	1	162.272	4.072	2,4%
		59,6%	85,9%	59,9%

Sensibilidad Especificidad Accuracy

		REAL		
		Corte 80%		
PREDICCIÓN	0	236.723	3.107	98,7%
	1	148.239	18.105	10,9%
		61,5%	85,4%	62,7%

Accuracy

		REAL		
		0	1	
PREDICCIÓN	Corte 70%			
	0	230.448	9.382	96,1%
	1	125.051	41.293	24,8%
		64,8%	81,5%	66,9% Accuracy

Sensibilidad Especificidad

		REAL		
		0	1	
PREDICCIÓN	Corte 60%			
	0	217.369	22.461	90,6%
	1	95.076	71.268	42,8%
		69,6%	76,0%	71,1% Accuracy

Sensibilidad Especificidad

		REAL		
		0	1	
PREDICCIÓN	Corte 50%			
	0	195.131	44.699	81,4%
	1	65.770	100.574	60,5%
		74,8%	69,2%	72,8% Accuracy

Sensibilidad Especificidad

PREDICIÓN		REAL		
		Corte 40%	0	1
0		161.737	78.093	67,4%
1		41.161	125.183	75,3%
		79,7%	61,6%	70,6%

Sensibilidad Especificidad Accuracy

PREDICIÓN		REAL		
		Corte 30%	0	1
0		96.227	95.562	50,2%
1		18.190	114.511	86,3%
		84,1%	54,5%	64,9%

Sensibilidad Especificidad Accuracy

PREDICIÓN		REAL		
		Corte 20%	0	1
0		78.421	161.409	32,7%
1		10.969	155.375	93,4%
		87,7%	49,0%	57,6%

Sensibilidad Especificidad Accuracy

		REAL		
		Corte 10%	0	1
PREDICCIÓN	0	38.744	201.086	16,2%
	1	3.795	162.549	97,7%
		91,1%	44,7%	49,6%

Sensibilidad Especificidad Accuracy

		REAL		
		Corte 5%	0	1
PREDICCIÓN	0	18.717	221.113	7,8%
	1	1.444	164.900	99,1%
		92,8%	42,7%	45,2%

Sensibilidad Especificidad Accuracy