



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

**SEGMENTACIÓN DE CLIENTES DE UNA CADENA DE
SUPERMERCADOS EN BASE A ESTILOS DE VIDA**

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

SANTIAGO PATRICIO NÚÑEZ FUENTES

**PROFESOR GUIA:
LUIS ABURTO**

**MIEMBROS DE LA COMISIÓN:
MÁXIMO BOSCH
CLAUDIO PIZARRO**

**SANTIAGO DE CHILE
AGOSTO 2010**

Resumen Ejecutivo

El presente trabajo consiste en una segmentación de clientes, basado en el trabajo realizado por Dunnhumby para la cadena de supermercados Tesco. Uno de los factores claves para el éxito de los programas de lealtad tiene que ver con el marketing focalizado, para lo cual, es fundamental contar con una adecuada segmentación. Expertos concluyen que no basta con saber la frecuencia de compra, monto promedio o las características socio-demográficas del cliente, lo que realmente aporta valor a la empresa, es el conocimiento de sus gustos, preferencias y necesidades [4]. Debido a lo enunciado anteriormente, esta memoria tiene como objetivo la segmentación de clientes por estilos de vida, para la realización de marketing focalizado.

Para la realización de tal objetivo, la metodología propuesta consiste en: Definir una serie de atributos con los cuales se caracterizarán los productos. Luego, a partir de los productos, caracterizar al cliente vía la presencia de cada atributo en sus boletas. Finalmente realizar un análisis de conglomerados, utilizando los métodos ACP y K-Means, y la descripción de los segmentos en cuanto a frecuencia, ticket promedio, lealtad (usando análisis RFM) y las características demográficas.

Aplicando la metodología para un local de la cadena en estudio y con datos del primer semestre del año 2010, se tiene como resultado los siguientes segmentos: Tradicional (28%), Express (10%), Familiar (18%), Parrillero (5%), Gourmet (8%) y Saludables (31%). De los cuales, los Saludables son lo que promedian mayor edad (50 años, promedio general 45 años) y poseen mayor porcentaje de mujeres (80%, porcentaje general 70%). Con respecto a la lealtad, se obtuvo que el segmento Express es el que tiene los menores índices (64% de sus clientes leales) y el segmento Tradicional los mayores, con un 81% de sus clientes catalogados como leales. Este último, además es el segmento de mayor aporte en términos de ingresos al local (43% del total). Finalmente se observa que la frecuencia de compra es semanal para la mayoría de los clientes, los cuales poseen un ticket promedio en torno a los \$23mil. Por otro lado, con respecto de la metodología, para la automatización de la asignación de los atributos a los productos se propusieron 2 modelos, de los cuales el mejor asigna con buenas tasas de efectividad pero alcanza una reducida parte de los productos, por lo que la asignación manual es la mejor opción hasta el momento.

En conclusión se tiene que es posible segmentar los clientes en base a estilos de vida a partir de las boletas. Y los segmentos encontrados pueden ser alcanzados vía marketing focalizado relacionado a las características propias de cada estilo de vida.

Contenido

1.- Introducción	6
1.1 - La Industria del Retail	6
1.2 - El Retail en Chile	7
1.3 - Proyecto FONDEF	10
2.- Descripción del Proyecto	11
2.1 - Caso Tesco	11
2.2 - Programa de Lealtad	12
2.3 - Segmentación	14
2.4 - Marketing Focalizado.....	14
2.5 - Estilos de Vida	15
3.- Justificación	16
4.- Objetivos	17
5.- Alcances.....	18
6.- Resultados Esperados.....	19
7.- Marco Conceptual.....	20
7.1 - Gestión de Retail	20
7.2 - Clubes de Lealtad	23
7.3 - Tipos de Segmentación.....	25
7.4 - Marketing Masivo vs Marketing Focalizado.....	29
7.5 - Tipos de Promociones	31
7.6 - Métodos a Utilizar.....	32
8.- Descripción Tienda en Estudio	36
9.- Metodología	37
9.1 - Revisión Bibliográfica.....	38
9.2 - Selección de Atributos	38
9.3 - Realización de Muestra	39
9.4 - Limpieza del Maestro.....	39
9.5 - Marcación del Maestro de Productos.....	39
9.5.1 - Marcación Manual del Maestro de Productos.....	40
9.5.2 - Método Producto-Producto-Cliente	41
9.5.3 - Método Producto-Cliente-Productos	45
9.6 - Marcación de los Clientes	48

9.7 - Análisis de Conglomerados	49
9.8 - Descripción de cada Conglomerado	49
9.9 - Aplicar método RFM	50
10.- Resultados.....	51
10.1 - Revisión Bibliográfica	51
10.2 - Atributos.....	52
10.3 - Realización Muestra	53
10.4 -Limpieza Maestro	53
10.5 - Marcación de los Productos de Maestro	53
10.5.1 - Marcación Manual.....	53
10.5.2 - Resultados Método Producto-Producto-Cliente	55
10.5.3 - Resultados vía Producto-Cliente-Producto	58
10.5.4 - Conclusiones Marcación Productos	61
10.6 - Marcación Clientes.....	62
10.7 - Análisis Conglomerados	63
10.7.1 - Análisis de Componentes Principales.....	63
10.7.2 – Segmentación.....	66
10.7.3 - Descripción Segmentos	71
10.8 - Análisis RFM.....	74
11.- Conclusiones.....	77
12.- Trabajos Futuros.....	78
13.- Bibliografía.....	79
14.- ANEXOS	80
ANEXO 1: Definiciones	80
ANEXO 2: Extracto Cuestionario VALS	83
ANEXO 3: Limpieza Maestro	84
ANEXO 4: Diagrama CHAID	85
ANEXO 5: Matrices de Confusión del modelo PPC	86
ANEXO 6: Ejemplo Cuadro Análisis PCP.....	88
ANEXO 7: Comentarios Pruebas Modelo PCP	89
ANEXO 8: Varianza Explicada por Factores.....	90
ANEXO 9: Promedio Dimensiones por Clusters.....	91
ANEXO 10: Distribución Gastos de Clientes en Atributos	92
ANEXO 11: Información Demográfica de los Segmentos.....	92

1.- Introducción

1.1 - La Industria del Retail

Se entiende como retail al sector económico que engloba a las empresas especializadas en la comercialización masiva, de productos o servicios, a grandes cantidades de clientes.

Existen varios formatos o canales de retail, los principales son: Supermercados, Tiendas por Departamento, Tiendas de Mejoramiento del Hogar y Tiendas de Especialidad (farmacias, ferreterías, etc.). En el desarrollo del presente trabajo se pondrá foco en los supermercados.

La profesionalización de sus gestiones, la utilización de TI en sus operaciones y la expansión internacional de las principales cadenas, ha causado que la industria del retail haya tenido un gran crecimiento. El desarrollo de los supermercados mundialmente, ha ocurrido en tres fases de expansión:

- La primera tuvo lugar a inicios de los 90's en Sudamérica, Asia del este (a parte de China) y Europa central. Entonces, la participación de los supermercados en el retail de alimentos pasó de un 10 a 20% en 1990, a un 50 a 60% hacia el 2000.
- La segunda oleada ocurrió a mediados de los 90's y comprendió a Asia del sudeste, México y Centroamérica. La participación de los supermercados en el retail alimenticio pasó de un 5 a 10% en 1990, a un 30 a 50% hacia el 2000.
- La tercera oleada ocurrida recientemente tuvo réplicas aún más profundas y expansivas y sus epicentros fueron China, India y Rusia; donde la participación de los supermercados pasó de casi un 0% en 1990, a un 20% hacia el 2007, siendo el mayor crecimiento histórico registrado por el canal mayorista.¹

Tal ha sido el crecimiento del retail en el mundo, que la cadena de supermercados Walmart, figura tercera en el Ranking Global 500 de la revista Forbes del año 2009, detrás de las petroleras Royal Dutch Shell y Exxon Mobil, como las empresas más grandes del mundo.

El escenario mundial de los supermercados a finales del 2009, consiste en 4 cadenas que lideran los ingresos por ventas. En primer lugar la cadena Walmart, la cual registra ingresos por sobre los 400 mil millones de dólares para el 2009. En segundo lugar le sigue el retailer francés Carrefour, con ingresos cercanos a los 100 mil millones de dólares. En tercer y cuarto lugar,

¹ Fuente: Thomas Reardon, "Supermercados: Oportunidades y Desafíos para la Fruta Chilena en Economías Emergentes"

con ventas menores, les siguen el retailer alemán Metro AG y el retailer inglés Tesco respectivamente. ²

1.2 - El Retail en Chile

A inicios de la década de los 90's emerge el retail en Chile para convertirse en uno de los mercados más dinámicos de la economía nacional. Entre los años 2003 y 2007 la industria creció a una tasa promedio del 11% lo cual se traduce que para el 2007 el retail representó el 21,3% del PIB. Para el año 2009 se calcula que la industria generó el 22% del PIB.³ Lo anterior se ve reflejado en aumento sostenido en el número de tiendas de retail a nivel nacional entre los años 2002 y 2008, como indica la figura.



Figura 1: Crecimiento Tiendas; Fuente: Censo Nielsen 2008

En particular los supermercados tradicionales aumentaron de 650 a 745 locales, mientras que los hipermercados aumentaron de 56 a 58. Esto se traduce en una variación del 15% para el canal supermercados.

A pesar del significativo aumento de tiendas de supermercados, la penetración del canal en Chile es cercana al 60%, lo cual es mayor que algunos vecinos latinoamericanos que no alcanzan el 40%, pero menor que los países desarrollados que presentan del orden de 90% de penetración.⁴ En términos de

² Fuente: Wal-Mart 2009 Annual Report ; Carrefour Annual Results 2009 ; Metro AG Annual Report 2009; Tesco Annual Report 2009

³ Fuente: Banco Central, INE, Nielsen.

⁴ Fuente: Revista Capital, "El retail del Futuro"

ventas, los supermercados han presentado un crecimiento por sobre el crecimiento del PIB nacional, observando un 6,3% de crecimiento en supermercados vs un 4,5% en el PIB, para el periodo entre el 2001 y el 2007. La siguiente figura muestra el crecimiento en ventas del rubro.

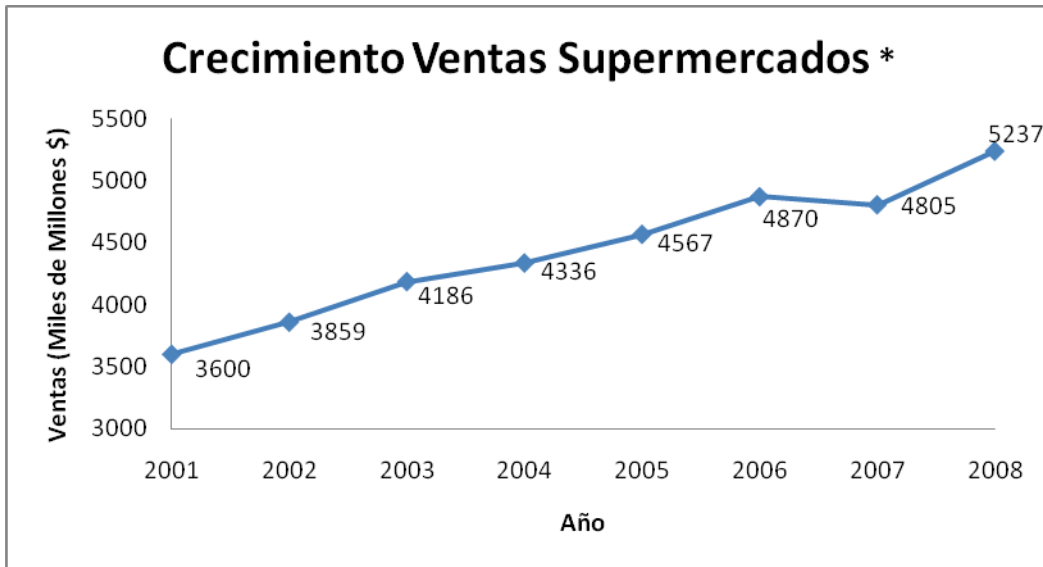


Figura 2: Crecimiento Supermercados; Fuente: Ac Nielsen, ASACH, INE, D&S, PLANET RETAIL, VIEWSIDE.

Los supermercados con respecto de los otros formatos son los de mayor participación de la industria del retail, la siguiente figura muestra como se distribuyen las ventas del retail con respecto a los formatos.

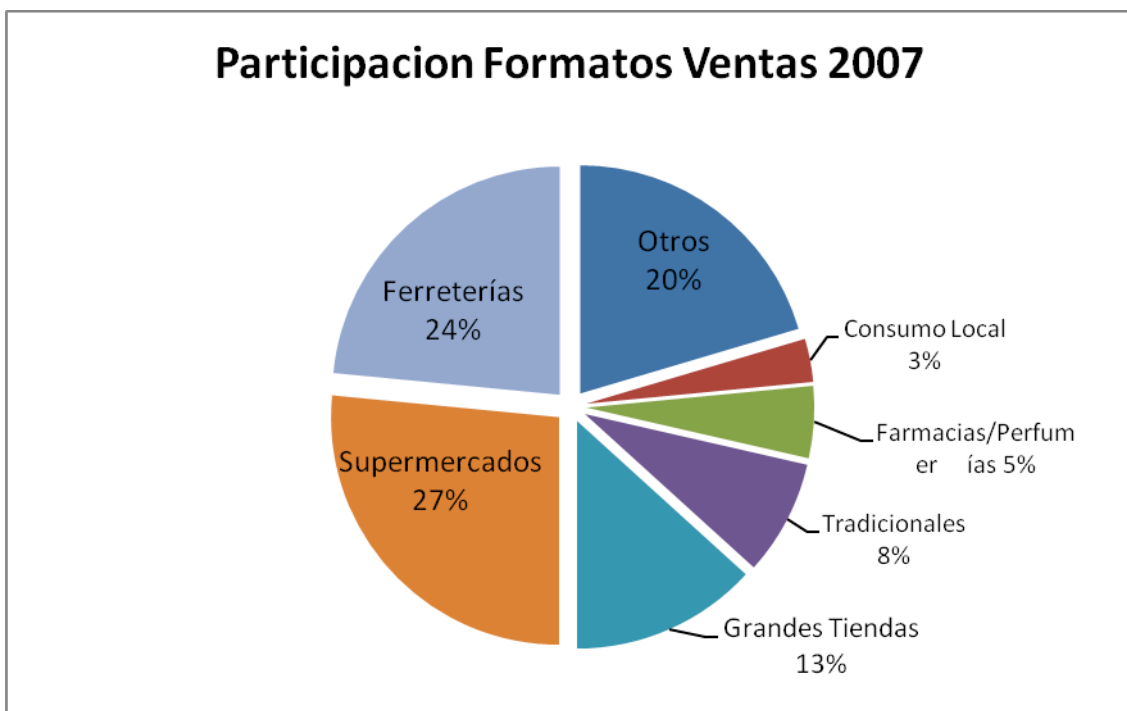


Figura 3: "Participación Formatos"; Fuente: Banco Central, INE, Fecusy Nielsen

Con respecto a la competencia en los formatos se tienen las siguientes marcas para los 2 formatos de estudio más importantes:

- En Supermercados:
 - Jumbo y Santa Isabel (Grupo Cencosud)
 - Líder y Ekono (Grupo D&S)
 - Tottus (Falabella)
 - Unimarc, Bryc y Deca (Grupo Saieh Rendic)
- En Tiendas por Departamento:
 - Falabella
 - Ripley
 - París
 - La Polar

Las participaciones de mercado de las marcas antes mencionadas para el año 2008, se muestran en las siguientes figuras.

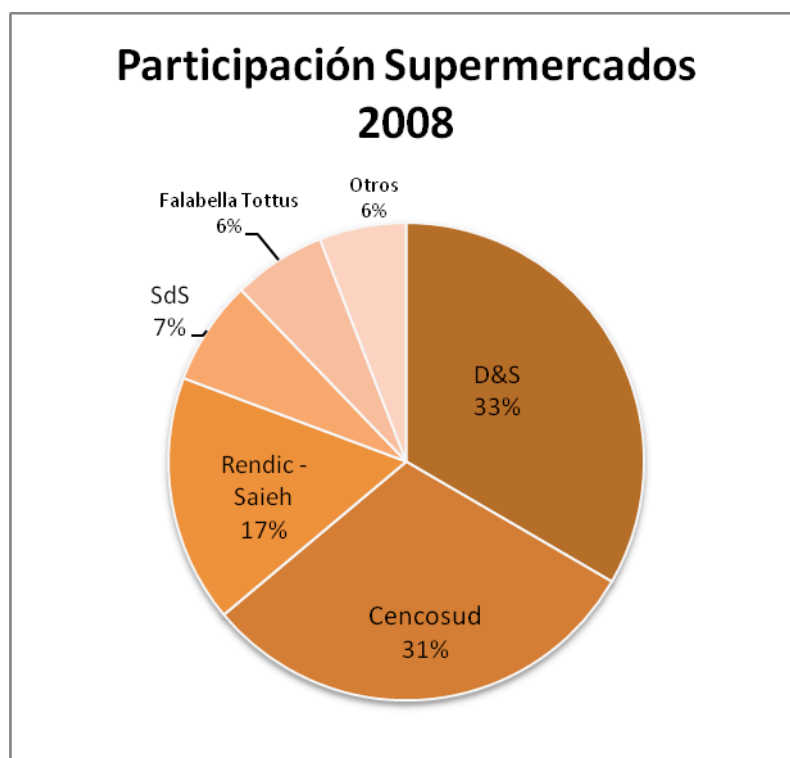


Figura 4: "Participación de Mercado Supermercados 2008";
Fuente: Memoria Anual y Estados Financieros D&S (2008), Memoria Anual y Estados Financieros Cencosud (2008)

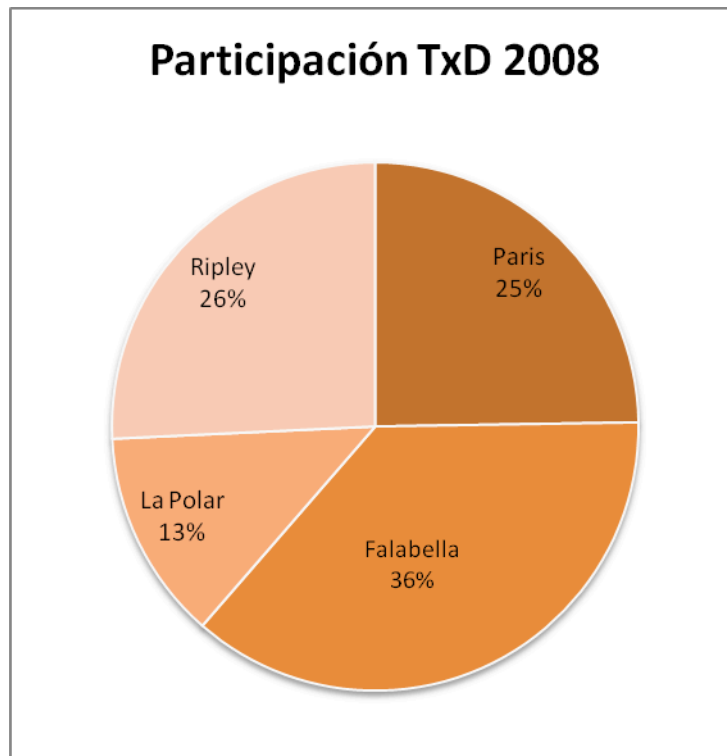


Figura 5: “Participación de Mercado Tiendas por Departamento 2008”;
Fuente: Memoria Anual y Estados Financieros D&S (2008), Memoria Anual y Estados Financieros Cencosud (2008)

Durante el año 2009, sin dudas, el suceso más importante ocurrido en la industria del retail ha sido la adquisición del 75% del grupo D&S por parte de la cadena Walmart. Se cree que tal acontecimiento, traerá consigo una serie de nuevas prácticas en la integración con los proveedores, redes logísticas y políticas de precio.⁵

1.3 - Proyecto FONDEF

Este trabajo forma parte del proyecto Fondef número D0611015, el cual se ha adjudicado la Escuela de Ingeniería de la Universidad de Chile y tiene como finalidad contribuir al desarrollo de la ingeniería del retail en Chile.

⁵ Fuente: Memoria Anual y Estados Financieros D&S (2009)

2.- Descripción del Proyecto

El presente trabajo consiste en la realización de una segmentación por estilos de vida para una cadena de supermercados. Para realizar la segmentación, se basará la experiencia realizada por el retailer inglés Tesco, quién junto a la empresa consultora Dunnhumby, llevaron a cabo la primera segmentación por estilos de vida, a partir de la información contenida en las boletas de los clientes.

2.1 - Caso Tesco

En 1994 la cadena de supermercados Tesco poseía millones de ítems de información transaccional y aun no les había sido posible obtener *insights* de aquello, es decir, transformar tal información en conocimiento útil para la empresa. Fue entonces cuando se asociaron con la empresa Dunnhumby, consultora que había ganado cierta reputación tras sus trabajos con BWM y Mercury (Empresa de Telecomunicaciones).

Al principio, Tesco entregó a Dunnhumby cierta cantidad de datos para que se realizasen los primeros estudios y pruebas. Estos datos contenían información de 12 supermercados y correspondía aproximadamente a un 0,5% de los tickets o boletas de Tesco (Aprox. 500.000).

El primer gran resultado fue darse cuenta que los clientes más leales, una pequeña porción del total, eran los responsables de gran parte de los ingresos. Hoy en día casi el 60% de los ingresos de Tesco proviene del 20% de sus clientes. Con esta información algunos gerentes de local rápidamente tomaron iniciativas, se sabe que uno invitó a los 300 mejores clientes a un desayuno para conocerlos personalmente, otro gerente mandó cartas de disculpas a sus clientes por la remodelación que se estaba llevando a cabo en el local.

El segundo gran resultado fue la elaboración la segmentación por estilos de vida para ser aplicada mediante la *Clubcard*, el programa de lealtad. Basándose el supuesto "*you are what you eat*", se estimó que a partir de información en las boletas era posible encontrar patrones de compras, que revelen distintos estilos de vida que tienen los clientes. Lo primero fue investigar en las boletas, que productos eran predictores de una necesidad o estilo de vida, como por ejemplo los clientes preocupados de cuidar su peso. El resultado, fue una serie de 25 "dimensiones de compras" o tipologías, las cuales describirían a los productos, como por ejemplo: verde, orientado a familia, calorías-consciente, etc. Luego, a un maestro de cerca de 40.000

productos se les dio una puntuación “si” o “no” en cada una de las dimensiones. Realizado lo anterior, se procedió a realizar una segmentación a partir del comportamiento de los clientes en el tiempo con respecto de las tipologías. Finalmente, se realizó una aplicación de la segmentación, generando catálogos de descuentos en productos afines a cada segmento.

En la mañana del 22 de Noviembre de 1994 Clive Humby (Cofundador de Dunhumby) y el equipo de marketing de Tesco, presentaron los resultados del estudio de prueba, realizado con el *Clubcard*, a la junta directiva de Tesco. El argumento fue simple, si se replica el aumento en ventas observado con la prueba de estudio, Tesco puede convertirse en el retailer N°1 de Inglaterra por primera vez. La junta aprobó los resultados y se marco el inicio de la masificación de la *Clubcard*.

En febrero de 1995 se realizó el lanzamiento oficial de la *Clubcard*, entregando más de 7 millones de tarjetas, con las cuales los clientes accedían a descuentos en sus compras por su uso. Al cabo de unos días cerca del 70% de las ventas de Tesco, habían sido realizadas por usuarios de la tarjeta *Clubcard*.

En los primeros días Dunhumby era capaz de analizar solo el 1% de las transacciones y el uso de la tarjeta *Clubcard* estaba creciendo enormemente, por lo cual fue necesario elaborar un sistema más sofisticado para la interpretación de la información. Fue entonces cuando desarrolladores de computación empezaron a elaborar un algoritmo que permitiría darle puntajes a cada producto, con respecto de todas las dimensiones, como resultado se obtuvo el llamado *Rolling Ball Algorithm*. Tomó 2 años en HH de análisis, para lograr una base de datos en que cada producto estuviese con un puntaje en cada dimensión [3].

El resultado del Clubcard es claro, hoy en día Tesco, es el retailer N°1 en Inglaterra y la cuarta cadena de supermercados más grande del mundo.

2.2 - Programa de Lealtad

Motivado por el caso Tesco, lo que se busca en este trabajo es realizar una segmentación similar, para luego ser utilizada por las estrategias de marketing del club de lealtad.

La industria del retail, en particular los supermercados, posee ciertas características que hacen de los programas de lealtad una interesante estrategia para competir. Estas características son:

- **Productos *Comodities*:** La productos ofertados en las distintas cadenas de supermercados se por lo general son los mismos en las cadenas.

- Bajo Costo de Cambio: Para un cliente no representa mayor costo cambiarse de marca y comprar en otro local de la competencia.
- Mínima información del Cliente: Sin un programa de lealtad los retailers no son capaces de identificar ni tener información sobre los consumidores.
- Servicio como factor diferenciados: El servicio otorgado por el retailer es un factor clave en la experiencia de compra y satisfacción del cliente. Los retailers pueden diferenciarse entre ellos por medio de los distintos niveles de servicio.

Adicionalmente a las dificultades antes mencionadas. Las tendencias en el negocio, indicarían la necesidad de contar con un programa de lealtad, debido a las siguientes razones:

- Disminución de la Lealtad de los Clientes: Los hipermercados han condicionado al consumidor a comprar basándose casi exclusivamente en el precio, por lo mismo los clientes acudirán al retailer que ofrezca el menor precio. Se estima que el 70% de los clientes de los supermercados están siempre en búsqueda de alternativas a su retailer usual. Además, estudios han concluido que los clientes generalmente dividen sus compras entre varios retailers. 1
- Extrema Competencia de Precios: Cadenas de supermercados, como Walmart, se mantienen competitivos por medio de la oferta de los precios más bajos. Otros retailers, que no pueden competir por bajos precios, necesitan competir de otra manera que no sea por precios.
- Sinergia de la múltiples marcas: Aplicable preferentemente en tiendas por departamento, los retailers incrementan el *share of wallet* de los clientes aumentando las compras de los clientes en múltiples marcas.

Los programas de lealtad ayudan a los retailers por medio de los siguientes factores [7]:

- ✓ Disminución de la competencia por precios.
- ✓ Incremento del costo de cambio de retailer.
- ✓ Disminución en los costos de marketing.
- ✓ Construyendo una relación positiva y de confianza con el cliente.
- ✓ Fortaleciendo la marca.
- ✓ Incrementando el *share of wallet*

Como se ha visto, los clubes de lealtad, son una buena estrategia que tienen los retailers para mantenerse competitivos en la industria, sin tener que caer en la “guerra de precios”. Pero por otro lado, se sabe que lograr el éxito de éstos, es tremendamente difícil y costoso. A modo de ejemplo, se tienen los casos de ASDA (Principal competencia de Tesco), que gastó US\$50 millones en 5 años

y Safeway UK, que gastó US\$375 millones, antes de que descontinuaran sus respectivos programas de lealtad y adoptaran la estrategia de precios bajos [1].

2.3 - Segmentación

Una adecuada segmentación de clientes es esencial para el éxito de los programas de lealtad, ya que le permite al retailer, la realización de marketing focalizado. Los retailers, normalmente, cometen el error de dar el mismo trato a todos sus clientes.

A continuación se nombran las principales razones por las cuales la segmentación es necesaria [7].

- Identifica los clientes más y menos rentables.
- Permite focalizar el marketing en los consumidores, de manera de obtener mayor eficiencia en las campañas.
- Evita los mercados que no son rentables para la empresa.
- Construye una relación de lealtad con el consumidor por medio de ofertar productos y servicios que estos quieren.
- Permite mejorar el servicio al cliente.
- Permite adelantarse a la competencia en mercados específicos.
- Permite usar los recursos en forma sabia.
- Permite mejorar la oferta de los locales en función de las necesidades del cliente.
- Permite incrementar las ganancias por medio de la disminución de costos y el aumento en precio que se le pueden dar a algunos productos o servicios.

2.4 - Marketing Focalizado

El objetivo puntual de la segmentación propuesta, es permitir la realización de marketing focalizado.

El marketing focalizado es uno de los componentes más importantes en los programas de lealtad, ya que permite aumentar las ventas y las ganancias, por medio de un efectivo *up-selling* y *cross-selling* de los clientes.

Con el objetivo de vislumbrar los alcances del *up-selling* y *cross-selling* de los clientes, se tiene el caso de Kroger y Dunhumby. En el año 2002 la cadena de supermercados Kroger se asoció con la consultora Dunhumby con el objetivo implementar un club de lealtad como el *Clubcard* de Tesco. En las primeras conversaciones, los gerentes de Kroger señalaron que sus tiendas operaban muy bien y lo que necesitaban era tener más clientes. Dunhumby analizando

la información disponible, no tardó mucho en darse cuenta que Kroger no necesitaba más clientes, sino ponerle más atención a los que ya tenía. Dunnhumby estimó que los clientes de Kroger, incluso los mejores, gastaban menos de la mitad su presupuesto de compras en la tienda. Y en el caso en que los clientes gastaran la totalidad de su presupuesto en la cadena, las ventas de la marca podrían triplicarse [3].

En la actualidad se observa que: Si los supermercados hacen marketing focalizado, por lo general, es sólo los clientes de mayor valor (comúnmente definido por el gasto que realizan), a pesar de que estudios señalan, que las campañas de fidelización a menudo son más efectivas para cambiar el comportamiento de los clientes de valor medio y bajo. Los clubes de lealtad líderes hacen un trabajo bastante mejor, por ejemplo se sabe que Tesco, hace 80.000 variaciones de cartas y revistas para los miembros de su club de lealtad [7].

2.5 - Estilos de Vida

Tal como el modelo de Tesco, lo que se propone en este trabajo es desarrollar una segmentación por estilos de vida a partir de las transacciones. La razón que hace a la segmentación por estilos de vida conveniente, es porque ésta responde las 4 preguntas fundamentales, para el éxito de las campañas de marketing, estas son:

- i. - ¿Quiénes son los mejores clientes y que los hace diferentes con el resto?
- ii.- ¿Dónde se encuentran otros clientes tal como ellos?
- iii.- ¿Cuáles son sus hábitos y preferencias, de manera de confeccionar una campaña a ellos?
- iv.- ¿Qué medio utilizan frecuentemente, de manera de saber cómo poner la campaña?

Otro punto interesante es la ventaja de la segmentación de Tesco por sobre las otras. Cuando a Edwina Dunn, cofundadora de Dunnhumby, le preguntaron: ¿Cuál es la diferencia entre la segmentación por estilos de vida de Tesco y los modelos clásicos como ACORN y MOSAIC? Su respuesta fue: La diferencia está en que los estilos de vida de Tesco, son un indicador mucho más sensible, éstos cambian a medida que el comportamiento del cliente cambia.⁶

⁶ Fuente: Noelle McElhatton "Case Study Tesco"

3.- Justificación

El retail, en particular los supermercados, es un sector empresarial muy competitivo, por lo mismo, para que una marca sobreviva y sea exitosa es fundamental que esta se posicione por sobre el resto, en la mente de los consumidores. Para lograr tal posicionamiento, es necesario que la marca gane la fidelidad y preferencia por parte de los clientes, lo cual es posible sólo si se cuenta con un acabado conocimiento de los clientes. No basta con saber su frecuencia de compra, su monto promedio o sus características socio-demográficas, lo que realmente aporta valor a la empresa, es el conocimiento de los gustos, las preferencias y las necesidades del cliente [4]. Como solución a este desafío, el presente trabajo propone una segmentación por estilos de vida.

La segmentación permite focalizar las acciones de marketing, pero la segmentación por estilos de vida, permite realizar marketing focalizado, a partir de las preferencias y necesidades de cada segmento, lo cual es fundamental para crear relaciones de confianza y de largo plazo con el cliente. En la medida que el supermercado se sincronice con el cliente, se podrá llevar a cabo un club de lealtad de valor para los clientes. Estudios concluyen que los retailers que puedan desarrollar un buen programa de lealtad, que cambie el comportamiento de los clientes, tendrían una verdadera ventaja competitiva [2].

Los problemas con las segmentaciones por estilos de vida, son el alto costo y el tiempo que demoran en llevarse a cabo, pensar que el método VALS se construye a partir de cuestionarios y es aplicable para una sala en particular. La segmentación por estilos de vida propuesta, disminuye los costos, demora menos tiempo en implementarse, es replicable para toda la cadena y tiene la ventaja de ser dinámica, es decir, la segmentación es capaz de percibir los cambios de hábitos que va teniendo el cliente en el tiempo. Los beneficios mencionados anteriormente son posibles, ya que la segmentación propuesta, a diferencia del resto de las segmentaciones, se elabora a partir de la información transaccional de cada cliente.

En conclusión, las posibilidades y beneficios que se pueden obtener a partir de la segmentación propuesta son especialmente atractivos, ya que pueden marcar la diferencia con la competencia, en la relación con los clientes. Y una mejor relación con los clientes, tiene como consecuencia mayores ganancias y mayor valor de la marca.

4.- Objetivos

Objetivo General

Segmentar los clientes de una cadena de supermercados en base a estilos de vida para la realización de marketing focalizado.

Objetivos Específicos

- 1.- Determinar atributos de productos que permitan encontrar estilos de vida en los clientes.
- 2.- Desarrollar metodología para marcar los productos del maestro con respecto a atributos.
- 3.- Definir metodología para marcar clientes en base a atributos presentes en los productos de sus transacciones.
- 4.- Determinar conglomerados de clientes representables por estilos de vida.
- 5.- Caracterizar segmentos en base a las ocasiones de consumo.
- 6.- Generar recomendaciones para cada uno de los segmentos determinados.

5.- Alcances

Como principal supuesto para el desarrollo de este trabajo, se considera que la canasta de compra permite predecir el estilo de vida del cliente, en particular, los alimentos que el cliente compra, por lo que se trabajará sólo con los productos comestibles ofrecidos.

Otro supuesto importante a considerar es que la compra en el local identifica totalmente al cliente, es decir, que el cliente realiza la totalidad de sus compras dentro del local. Por ejemplo, se calificará a un cliente de saludable si compra en su mayoría productos catalogados como tal, asumiendo que no va a comprar productos no saludables en otros lugares.

Como se mencionó anteriormente, para el estudio se consideraron sólo productos comestibles presentes en boletas del año 2010, desde el 1 de Enero hasta el 30 de Junio de un local de la comuna de Vitacura. Estos datos serán entregados por el supermercado en estudio, por lo que no se considera la recopilación de éstos dentro de las tareas a realizar.

Por último, no se considera la implementación de acciones (por ejemplo, promociones) a base de los resultados entregados.

6.- Resultados Esperados

El resultado del presente trabajo corresponde a una segmentación por estilos de vida de los clientes de un local de la cadena en estudio.

Los segmentos finales deberán tener las siguientes condiciones:

- Suficientemente grande: Los segmentos deben tener un número de clientes tal que sean económicamente viables.
- Relevantes: Los segmentos deben ser abordables en función de los productos y servicios ofrecidos por la cadena.
- Alcanzables: Los segmentos deben ser alcanzables por en términos de precio, promociones, distribución, etc.
- Diferenciable: Deben existir diferencias claras entre los segmentos tal que permitan elaborar marketing focalizado.
- Interpretables: Los segmentos deberán ser interpretables en base a comportamientos de compras o estilos de vida.

Como referencia, se presentan los segmentos finales encontrados por la consultora Dunnhumby para el retailer Tesco.

Segmentos por Estilos de Vida	% Clientes	Características
Lujosos (<i>Finer Foods</i>)	17%	Conscientes del tiempo, disfrutan productos de lujo y son ávidos de experimentar.
Saludables (<i>Healthy</i>)	9%	Interesados en productos orgánicos, beneficios ambientales, bajos en azúcar/grasas y conscientes de las calorías
Cómodos (<i>Convenience</i>)	22%	Consideran los alimentos como combustibles, son ocupados y cuentan fuertemente con el microondas.
Precio Sensible (<i>Price Sensitive</i>)	16%	Miran en primer lugar con la conveniencia y compran productos genéricos.
Familiares (<i>Mainstream</i>)	26%	Gustan de las marcas establecidas, tienen gustos amplios y son muy influenciados por las necesidades de los hijos.
Tradicionales (<i>Traditional</i>)	10%	Gusta de cocinar pero sus compras son estándar por lo que no compran impulsivamente.

Figura 6: “Segmentos Dunnhumby”; Fuente: Bibliografía [8]

Se observa que cada segmento es suficientemente grande para realizar acciones comerciales focalizadas, además son representativos de un comportamiento o estilo de vida lo cual permite elaborar promociones en función de sus necesidades.

7.- Marco Conceptual

A continuación se contextualizará el área de estudio del presente trabajo. Para ello se ordenaron los temas de general a específico. Se comienza por una descripción global de las estrategias de retail, en seguida se centrará en clubes de lealtad, para luego dar paso a las técnicas de segmentación. Después el tema segmentación, se verán ventajas y desventajas del marketing focalizado. Consecutivamente se describirán las principales promociones que se pueden ofrecer vía marketing focalizado y finalmente se describirán los métodos y algoritmos a utilizar.

7.1 - Gestión de Retail

Corresponde a las distintas áreas de acción que los retailers tienen con el fin de mejorar las relaciones con los clientes y rentabilizar los recursos disponibles. Se tienen cuatro principales áreas de estudio como indica el siguiente cuadro.

Gestión Clientes	Gestión Tienda
Gestión Categoría	Gestión Disponibilidad

Figura 7: "Gestión de Retail"; Fuente: Elaboración Propia

A continuación una descripción de cada una de las gestiones [9].

Gestión de Clientes:

Para los retailers, existen tres temas que es necesario conocer de sus clientes y poder hacer una mejor gestión de ellos:

- ¿Qué es lo que lleva a los clientes a la tienda?
- ¿Cómo compran?
- Conocer el significado de "su" marca para los clientes.

Para responder tales preguntas el retailer debe ser capaz de:

- Clasificar o Segmentar a sus clientes.
- Determinar la Ocasión de Consumo preferente de la tienda.
- Estimar la Experiencia de Compra y Valor Percibido del cliente.

Gestión de Tienda

A cargo de esta gestión se tiene 4 principales campos de acción.

1. Cantidad y Tipo de Personal.
2. Configuración e Infraestructura (*layout* de los productos y disponibilidad de m²)
3. Ambientación y Merchandising.
4. Atención al Comprador – Informaciones

Típicamente la forma de evaluar esta gestión es por medio de la *rentabilidad por m²*.

Gestión de Categoría

Con alto involucramiento de los proveedores, en este campo tiene como finalidad administrar categorías como negocios independientes dentro del local, produciendo mejores resultados tanto para el retailer como para el proveedor. Las áreas de acción se denominan decisiones tácticas del retail, las cuales son:

- Precio.
- Surtido.
- Promoción
- Disponibilidad
- Ubicación
- Exhibición
- Atención (Sólo para Tienda por Departamentos)

Gestión de Disponibilidad

Esta gestión tiene que ver con la eficiencia de la cadena de suministro y la disponibilidad del producto para el cliente final. El objetivo es La idea es evitar los *stock out*, ya que no solo perjudican la venta spot, sino también la imagen del retail y las futuras transacciones.

Algunos indicadores para esta gestión son:

- Quiebres de Stock.
- FMG: Faltante de Mercadería en Góndola.
- OSA: On-Shelf Availability.

El presente trabajo se enmarca dentro la gestión de clientes para la cadena en estudio,

7.2 - Clubes de Lealtad

Un Club de Lealtad es una herramienta de marketing que permite crear, organizar, administrar e implementar de programas de fidelización. Esta aplicación tiene como finalidad la construcción de vínculos emocionales y duraderos entre las personas y la marca.

Un adecuado programa de lealtad provee beneficios intangibles, como mejora en la reputación y aumento de valor de marca, así como también proveen beneficios concretos. A continuación se explicarán que tipos de beneficios concretos es posible obtener como consecuencia de los programas de lealtad [7]:

1.- Decrecimiento en la Competencia de Precios

Uno de los desafíos más grandes de los retailers es la competencia por precios. En los supermercados esta competencia se presenta a nivel estratégico de las compañías, como WalMart, que utiliza su eficiencia en la cadena de suministro y grandes compras para batir el precio de la competencia. Este tipo de competencia puede ocurrir cuando la publicidad del retailer es genérica, es decir, cualquier persona puede acceder a sus descuentos. Las promociones de lealtad focalizadas pueden hacer frente a la competencia por precios, cuando un retailer provee descuentos focalizados a sus clientes en forma individual (vía correo, kiosco, POS, etc.), la competencia no tiene acceso a tal información por lo que no se le puede ganar en forma efectiva.

2.- Mayor retención de Clientes

Un programa efectivo conlleva a un cliente a ser leal a un retailer ahora y en el futuro, lo que provee beneficios basados en las compras y en el valor del cliente en el tiempo. La retención de los clientes aumentará con tiempo debido a que si el cliente cambia de retailer, lo hará por uno que sabe muy poco sobre él, por lo que sufrirá una disminución en su nivel de servicio y a menudo mayores precios en sus productos que son importantes.

3. - Disminución en los Costos de Marketing

Un adecuado programa de lealtad permite realizar marketing focalizado, lo cual permite aumentar la efectividad de las promociones y disminuir los costos, ya que se evita que clientes no rentables hagan uso de la promoción y que clientes que estaban dispuestos a pagar precio completo lo hagan.

4.- Aumento del *Share of Wallet*

Los retailers pueden aumentar significativamente sus ganancias por medio de un aumento de la rentabilidad por venta. Los programas ayudan a generar más ganancias por venta porque mejoran el cross-selling de los clientes, esto debido a que las promociones pueden ser focalizadas a que miembros que

compran sólo una cierta porción de un set de productos, compren otros productos del set, por ejemplo clientes que compran comida para perros se les puede dar beneficios para comprar juguetes para perros. Este tipo de promociones son razones para los miembros del programa para unificar sus ventas en el retailer y que no compren los artículos por separado e varias tiendas.

5.- Mejoras la Gestión de Inventarios y Rotación

Los programas de lealtad permiten al retailer mejorar su surtido y servicio. Por ejemplo en estados unidos, la cadena de farmacias CVS rediseño su sección de cosméticos guiado por su aumento en ventas. LA información de los programas también permiten al retailer disminuir sus costos en servicio al cliente, usando una adecuada cantidad de personal para servir a los clientes basándose en su valor, en vez de ofrecer altos servicios a todos los clientes, incluso los menos valiosos.

6.- Nuevas Formas de Obtener Ganancias

Un programa de lealtad puede también obtener 2 nuevas formas de obtener ganancias, la primera mediante la venta de información agregada de información a un grupo de proveedores y la segunda mediante alianzas con pequeños retailers. Por ejemplo un proveedor de vinos puede pagar por la información de todos los clientes compran vinos. Entonces se convierte en un escenario *win-win-win* dado que el proveedor gana por la rentabilidad de sus gastos en marketing, el retailer gana por las ventas y por la venta de la información y el cliente que recibe descuentos en productos que son importantes para él. Con respecto de la segunda fuente nueva de ganancias, si los clientes son muy leales al programa, otras compañías como pequeños retailers pueden asociarse al programa en vez de empezar uno ellos.

7.3 - Tipos de Segmentación

La segmentación de mercado es el proceso de dividir un mercado en grupos uniformes más pequeños que tengan características y necesidades semejantes.

Los segmentos son grupos homogéneos y que tienen como particularidad que dentro de cada grupo, es probable que respondan de modo similar a determinadas estrategias de marketing. De esta forma, los integrantes del segmento tendrán probablemente las mismas reacciones acerca del marketing mix de un determinado producto, vendido a un determinado precio, distribuido en un modo determinado y promocionado de una forma dada.

A continuación se muestra una tabla resumen con los principales tipos de segmentación y sus variables.

Tipos de Segmentación	Variables Utilizadas
Segmentación Demográfica	<ul style="list-style-type: none"> • Sexo. • Edad. • Raza. • Lugar de Residencia. • Características Físicas.
Segmentación Socio-Económica	<ul style="list-style-type: none"> • Niveles de Ingreso. • Nivel de Educación. • Profesión. • Clase Social.
Segmentación Socio-Económica	<ul style="list-style-type: none"> • Niveles de Ingreso. • Nivel de Educación. • Profesión. • Clase Social.
Segmentación por Uso	<ul style="list-style-type: none"> • Por cantidad de uso. • Por tipo de uso. • Por oportunidad de uso. • Por lealtad de marca.
Segmentación por Estilo de Vida	<ul style="list-style-type: none"> • Engloban diversos criterios a la vez. • Se basan en datos estadísticos reales que combinan las variables anteriores obteniendo grupos con individuos que piensan y consumen de manera similar y que comparten además ciertas variables socio-económicas y demográficas.

Figura 8: "Tipos de Segmentaciones";
Fuente: Emigdio Alfaro, "Segmentación de los mercados"

Existen una serie de segmentaciones que explicarían estilos de vida de una población a partir de variables geodemográficas, las destacadas son las siguientes:

Nielsen Claritas PRIZM: Es un set de clústeres usando variables geodemográficas desarrollada para segmentar la población de EEUU. Es sus inicios fue desarrollada por Claritas Inc, empresa que más tarde sería adquirida por Nielsen Company. Esta segmentación fue ampliamente usada para aplicaciones de marketing en los años 90 y aun continua. Los clústeres fueron desarrollados aplicando también análisis del censo de EEUU.

CAMEO System: Es un set de clasificaciones usadas por las empresas en sus planes de ventas, redes y marketing. CAMEO Reino Unido se elaboró dividir el mercado en 57 tipos de zonas y 10 claves de marketing. Este proyecto fue desarrollado por Eurodirect.

ACORN System (A Classification Of Residential Neighborhoods): Fue la primera herramienta en identificar y entender a la población del Reino Unido en relación a la demanda por los productos y servicios. ACORN categorizó 1.9 millones de casillas de correo usando cerca de 125 estadísticos demográficos y empleando 287 variables de estilos de vida. La clasificación ACORN consiste en 56 tipos de personas con respecto a 14 grupos y 5 categorías.

MOSAIC System: Este método clasifica la población del Reino Unido en 15 grupos socio-económicos. Posterior a esta segmentación se desarrolló Mosaic Global que se basa en que las ciudades de mundo responden a los mismos patrones de segregación, esta clasificación utiliza información de 29 países e identifica 10 tipos de barrios.

GeoSmart System: Segmentación desarrollada en Australia basada en que las personas con perfiles demográficos y estilos de vida similares tienden a cerca unos de otros.

En cuanto a las segmentaciones por estilos de vida propiamente tal, el método más estudiado se denomina VALS y se explica a continuación.

VALS (Values, Attitudes and Lifestyles)

Desarrollada en 1970 por Arnold Mitchell con base en los trabajos de Maslow (1954), esta metodología clasifica a las personas en 8 segmentos acorde a los resultados del cuestionario VALS. (Extracto cuestionario VALS en ANEXO 2)

De acuerdo a la estructura de análisis de VALS, los grupos de personas se definen en base a 2 dimensiones, las cuales son Recursos y Motivaciones. El eje de los recursos se construye a partir de los ingresos, educación, seguridad,

inteligencia, habilidades de liderazgo y energía de las personas. Por su parte el eje de las motivaciones tiene 3 niveles: El primero son los consumidores guiados por su conocimiento y principios, los cuales son motivados por las ideas; El segundo nivel son los consumidores guiados por demostrar éxito, estos son motivados por los logros.; Por último se tienen los consumidores guiados por actividades sociales, estos son motivados por la auto-expresión

Los 8 segmentos de la metodología VALS son los siguientes:⁷

1. **Innovadores:** Los innovadores tienen éxito, sofisticado, personas con alta autoestima. Son personas receptivas a nuevas ideas y tecnologías. Los innovadores son consumidores muy activos en sus compras y reflejan gustos de lujo en un nicho de productos y servicios.
2. **Sobrevivientes:** Son personas con pocos recursos con los que tratan de hacer frente. Se sienten cómodos con lo familiar y la seguridad es de vital importancia. Se centran en la satisfacción de necesidades, más que en el cumplimiento de deseos. Son consumidores prudentes y representan un mercado muy modesto para la mayoría de productos y servicios.
3. **Cumplidores:** Son personas maduras, satisfechas, valoran el orden, el conocimiento y la responsabilidad. Tienden a ser personas educadas y en el proceso de toma de decisiones busca activamente información. En el mercado del consumo a pesar de tener una alta disponibilidad de recursos, buscan la durabilidad, la funcionalidad y el valor de los productos que compran.
4. **Creyentes:** Se sienten motivados por ideales. Son conservadores, basadas en creencias tradicionales como la familia, la religión, la comunidad y la nación. Los creyentes son consumidores predecibles, ya que eligen productos y marcas establecidas. Son clientes fieles.
5. **Logradores:** Son personas bien educadas, adineradas, exitosos, muy trabajadores y felices, tienen mucha confianza en sí mismos. Han logrado objetivos orientados a su estilo de vida y profundo compromiso con el trabajo y la familia. En mercado del consumo caracterizan por sus muchos deseos y necesidades, su consumo en productos y servicios de prestigio.
6. **Competitivos:** Están preocupados por las opiniones y la aprobación de los demás. El dinero define el éxito de estos, que no tienen suficiente para satisfacer todos sus deseos. Buscan productos que reflejen alto estatus, para emular a personas con alto estatus. Son consumidores activos, porque es una actividad social que les da la oportunidad de demostrar a sus compañeros de su capacidad para comprar.

⁷ Fuente: Sonia Losada y Laura Lucas "Estudio de VALS"

7. **Experimentales:** Son jóvenes, entusiastas y consumidores impulsivos. Buscan variedad, emoción, les gusta probar lo nuevo, les gusta el deporte, la recreación al aire libre y las actividades sociales. En base al mercado del consumo, gran parte de sus ingresos van destinados a la moda, entretenimiento y socialización.
8. **Realizadores:** Viven dentro de un contexto familiar, del trabajo práctico de la recreación física. Tienen poco interés por lo que está fuera de este contexto. Son los encargados de proponer nuevas ideas y de trabajar en grandes empresas e instituciones. Se impresionan por las posesiones materiales que no buscan practicidad y funcionalidad.

7.4 - Marketing Masivo vs Marketing Focalizado

Una de los beneficios más importantes de la segmentación, es que permite realizar marketing focalizado. A continuación se muestran los pros y contras de la realización de marketing masivo y marketing focalizado.

Tipo se Marketing	Pros	Contras
<p>Marketing Masivo</p> <p>(Radio, Televisión, Diarios)</p>	<ul style="list-style-type: none"> ▪ Llega a mucha gente. ▪ Fácil de implementar. ▪ Se logra ver y oír la campaña en medios masivos como TV, radio y diarios. ▪ A menudo aumenta el reconocimiento de la marca. ▪ Puede atraer una amplia variedad de clientes. 	<ul style="list-style-type: none"> ▪ Difícil de medir los resultados. ▪ Llega a una gran audiencia que no desea el servicio. ▪ Mala rentabilidad con respecto a la inversión. ▪ Controlado por los proveedores. ▪ Se llega a clientes que no pueden acceder a los beneficios, por ejemplo clientes que viven muy lejos.
<p>Marketing Focalizado</p> <p>(Cartas, catálogos, Fliers)</p>	<ul style="list-style-type: none"> ▪ Menor gasto para hacer llegar la campaña a un segmento específico. ▪ Se pueden medir los resultados y hacer cambios rápida y fácilmente. ▪ Se puede decidir quién exactamente recibirá la campaña. ▪ Mayor rendimiento de la inversión. ▪ Rápido rendimiento sobre la inversión. ▪ Focalizado a la comunidad específica del(los) local(es). ▪ Reconocimiento de nombre en la comunidad. ▪ Permite lograr que una determinada área tenga ventajas sobre el resto de la competencia. ▪ Permite ser creativo y fresco con las promociones. 	<ul style="list-style-type: none"> ▪ Tiene que basarse en una estrategia de marketing. ▪ Hay que saber a QUIEN para focalizar. ▪ Hay que elaborar promociones afines para segmento.

7.5 - Tipos de Promociones

Los tipos de beneficios que se pueden otorgar a los miembros de un club de lealtad son variados, van desde descuentos hasta modificar el *layout* de la sala. La siguiente tabla resume las principales promociones o beneficios [7].

Beneficio/Premio	Descripción
Descuentos	<ul style="list-style-type: none"> ▪ Ofrecer descuentos simples (Ej: 50% descuento). ▪ Ofrecer descuentos de valor (Ej: Gasta \$100 y recibe un producto gratis). ▪ Ofrecer descuentos de frecuencia (Ej: Compra 5 y recibe el 6° gratis). ▪ Presupuesto de descuento por vendedor.
Promociones Focalizadas	<ul style="list-style-type: none"> ▪ Ganancia de puntos y cambio de los puntos por adquirir productos ▪ Ofrece promociones basados en las preferencias del cliente, historial de compra, <i>tier</i> y segmento.
Lista de compras online personalizada	<ul style="list-style-type: none"> ▪ Usar tarjetas de lealtad para fabricar listas de compras para ser usadas en la Web o en las tiendas. ▪ Actualizar lista de compras ▪ Automatizar las listas de compras basándose en lo que el cliente es propenso a comprar. ▪ Incitar up-selling y cross-selling agregando nuevos productos en las listas.
Lista de compras personalizada impresa en la tienda	<ul style="list-style-type: none"> ▪ Permite al cliente imprimir su lista de compras en la tienda. ▪ Proveer asistencia de compras (Ej: organizar los productos de la lista)
Promociones en Puerta	<ul style="list-style-type: none"> ▪ Ofrecer promociones focalizadas en la puerta de la tienda (Ej: Permitiendo que impriman en Kiosco) de esta manera los clientes son influenciados antes de empezar sus compras.
Cajas Preferenciales	<ul style="list-style-type: none"> ▪ Proveer cajas de pago preferenciales, esto con el fin de aumentar la satisfacción de los clientes del club.
Niveles de servicio focalizados	<ul style="list-style-type: none"> ▪ Distribuir las categorías acorde a las preferencias de los miembros de mayor valor.
Invitaciones exclusivas	<ul style="list-style-type: none"> ▪ Generar ventas solo para miembros (Ej: Abrir la tienda temprano solo para miembros en una promoción matutina) ▪ Ofrecer eventos afines sólo a miembros (Ej: degustaciones)
Manejo de la oferta	<ul style="list-style-type: none"> ▪ Productos ofertados en función de las preferencias de los clientes de mayor valor (Ej: Destacar los productos que gustan estos clientes y tenerlos siempre en stock, no remover ítems que son valiosos para estos clientes)

	<ul style="list-style-type: none"> Mantener áreas valorizadas por los clientes de mayor valor (Ej: Una sección de productos japoneses), incluso si la categoría no genera ganancias.
Comprador Personalizado (<i>Clienteling</i>)	<ul style="list-style-type: none"> Proveer de un comprador privado que escoja productos basándose en las preferencias de los miembros de mayor valor. (Aplicable fundamentalmente en tiendas por departamento)
Regalos Gratis	<ul style="list-style-type: none"> Otorgar regalos como por ejemplo un cumpleaños gratis.
Sorteos	<ul style="list-style-type: none"> Permite a los clientes usar sus puntos para adquirir un boleto de sorteo, el cual puede tener como premio un producto valioso o una cantidad grande de puntos.
Estacionamientos	<ul style="list-style-type: none"> Proveer estacionamientos gratis o exclusivos para los miembros más valiosos. Cobrar a los clientes no leales.
Consideración Personal	<ul style="list-style-type: none"> Hacer que los clientes se sientan valorados por medio de la inclusión de sus opiniones en la planificación de la tienda.
Despacho e Incentivos	<ul style="list-style-type: none"> Proveer despacho gratis a los clientes de mayor valor. Proveer despacho gratis u otro incentivo a los miembros que expandan las categorías compradas usualmente o si llevan productos de ciertas categorías.

Figura 9: “Tipos de Beneficios”; Fuente: [7]

7.6 - Métodos a Utilizar

A continuación se describen las principales herramientas que permitirán en análisis de los datos y posterior segmentación. Para realizar la segmentación de los clientes se utilizarán los métodos ACP y K-Means, mientras que para la descripción de los segmentos se utilizará el análisis RFM. Estos métodos se describen a continuación.

A.- Análisis de componente principales (ACP)⁸

El Análisis de Componentes Principales (ACP) es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será

⁸ Fuente: Vargas, María Fernanda. (2009), Rediseño de un modelo de incentivos para vendedores de una tienda por departamentos. Santiago, Chile.

reducirlas a un menor número perdiendo la menor cantidad de información posible.

Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serán independientes entre sí. Un aspecto clave en ACP es la interpretación de los factores, ya que ésta no viene dada a priori, sino que será deducida tras observar la relación de los factores con las variables iniciales (habrá, pues, que estudiar tanto el signo como la magnitud de las correlaciones).

Fases de un análisis de componentes principales:

1. Análisis de la matriz de correlaciones: Un análisis de componentes principales tiene sentido si existen altas correlaciones entre las variables, ya que esto es indicativo de que existe información redundante y, por tanto, pocos factores explicarán gran parte de la variabilidad total.
2. Selección de los factores: La elección de los factores se realiza de tal forma que el primero recoja la mayor proporción posible de la variabilidad original; el segundo factor debe recoger la máxima variabilidad posible no recogida por el primero, y así sucesivamente. Del total de factores se elegirán aquéllos que recojan el porcentaje de variabilidad que se considere suficiente. A éstos se les denominará componentes principales.
3. Análisis de la matriz factorial: Una vez seleccionados los componentes principales, se representan en forma de matriz. Cada elemento de ésta representa los coeficientes factoriales de las variables (las correlaciones entre las variables y los componentes principales). La matriz tendrá tantas columnas como componentes principales y tantas filas como variables.
4. Interpretación de los factores: Para que un factor sea fácilmente interpretable debe tener las siguientes características, que son difíciles de conseguir:
 - a. Los coeficientes factoriales deben ser próximos a 1.
 - b. Una variable debe tener coeficientes elevados sólo con un factor.
 - c. No deben existir factores con coeficientes similares.
5. Cálculo de las puntuaciones factoriales: Son las puntuaciones que tienen los componentes principales para cada caso, que permitirán su representación gráfica.

B.- Método K-Means

Es un método de análisis de conglomerados (clústeres), el cual particiona n observaciones en un número fijo K de clústeres, por medio de las cercanías entre las medias de la observación y del clúster.

El análisis de conglomerados de K medias es un método de agrupación de casos que se basa en las distancias existentes entre ellos en un conjunto de variables (este método de aglomeración no permite agrupar variables). Versiones anteriores del procedimiento comenzaban el análisis con la asignación de los primeros K casos a los centros de los K conglomerados (los centros multivariantes de los conglomerados se denominan *centroides*). En la versión actual se comienza seleccionando los K casos más distantes entre sí (el usuario debe determinar inicialmente el número K de conglomerados que se desea obtener). Y a continuación se inicia la lectura secuencial del archivo de datos asignando cada caso al centro más próximo y actualizando el valor de los centros a medida que se van incorporando nuevos casos. Una vez que todos los casos han sido asignados a uno de los K conglomerados, se inicia un proceso iterativo para calcular los centroides finales a esos K conglomerados.

El análisis de conglomerados de K medias es especialmente útil cuando se dispone de un gran número de casos. Existe la posibilidad de utilizar la técnica de manera exploratoria, clasificando los casos e iterando para encontrar la ubicación de los *centroides* o solo como técnica de clasificación, clasificando los casos a partir de *centroides* conocidos suministrados por el usuario. Cuando se utiliza como técnica exploratoria, es habitual que el usuario desconozca el número idóneo de conglomerados y compare las soluciones obtenidas; en estos casos también puede utilizarse el método de análisis de conglomerados jerárquicos con una submuestra de casos.

C.- Análisis RFM

Es un método utilizado para el análisis de comportamiento y para segmentación de clientes. Este método es particularmente útil para el retail ya que utiliza como input la información del POS. Entre las aplicaciones dadas se tiene la identificación de respuestas a las acciones de marketing, medir lealtad y determinar ocasiones de compra de los clientes. Se realiza a partir de la construcción de sus 3 parámetros, las cuales son:

i.- Recency (R): La cual dice relación con el tiempo que ha transcurrido desde la última compra de un individuo medida en días.

ii.- Frequency (F): La cual representa la frecuencia de compra de un cliente y es medida como días entre compra.

iii.- Monetary Value (M): La cual indica el monto de compra gastado por un cliente.

Usualmente se calcula la razón R/F y luego se definen intervalos, que definirán si el cliente es: Leal, En Fuga o Fugado. Con el parámetro M, se estima la

importancia, en términos de ingresos, que representan los clientes para la cadena retail en estudio.

8.- Descripción Tienda en Estudio

El supermercado en estudio pertenece a una cadena que a la fecha supera el 20% de la participación de mercado. Esta posee con 246 locales en todo Chile, cerca de 2.000 trabajadores y registra ingresos estimados en US\$2 mil anuales para inicios del año 2009.

Para el desarrollo del presente se trabajará con un solo local. Este queda ubicado en un sector residencial de la comuna de Vitacura, posee entre 1500 y 3000 m² y realiza ventas al detalle.

A continuación se presenta el mapa de roles del local en estudio.

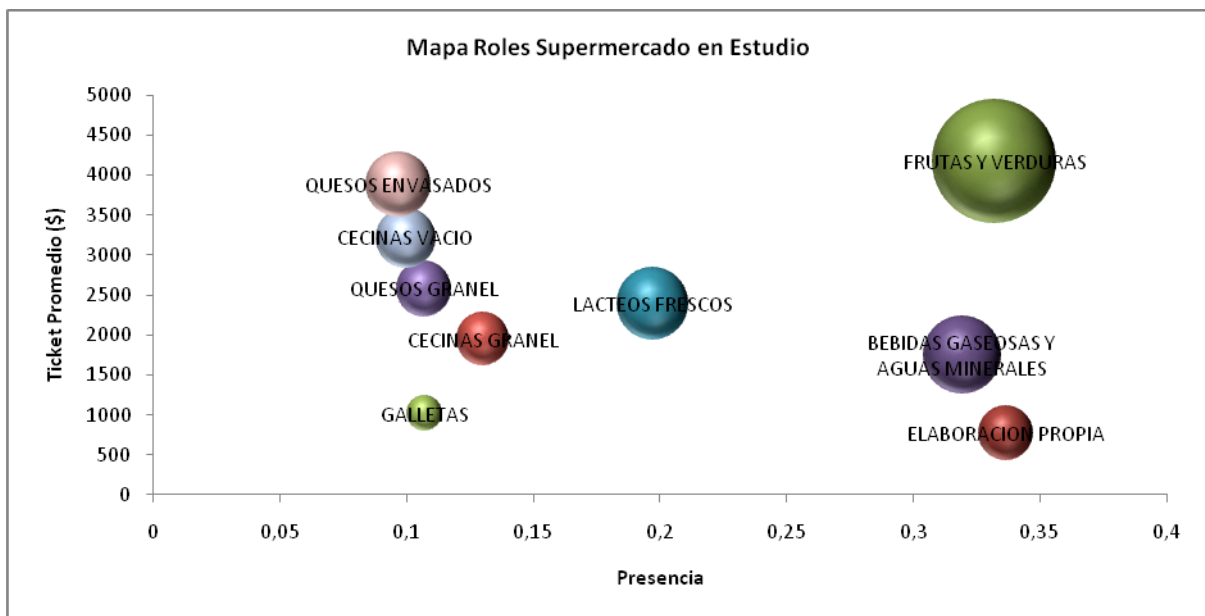


Figura 10: "Mapa de Roles"; Fuente: Elaboración Propia

Del mapa se pueden extraer los siguientes resultados⁹.

Categorías de Destino:

- Frutas y Verduras.
- Bebidas Gaseosas y Aguas Minerales.
- Elaboración Propia (Pan).

Categorías de Rutina:

- Lácteos Frescos.

⁹ Significados de los roles de categorías en Anexo 1

Categorías de Conveniencia:

- En general todos aquellos con Presencia menor a 15%

9.- Metodología

El presente trabajo se basa en el trabajo realizado por la empresa consultora Dunnhumby para la cadena de supermercados Tesco. Acorde a la bibliografía estudiada, la metodología se basa en la siguiente secuencia:

- 1- Se generan una serie de atributos (o tipologías) con los cuales se describen todos los productos de un supermercado.
- 2- Se caracteriza cada cliente del supermercado en base a las presencias de los atributos en los productos presentes en sus compras.
- 3- Se generan grupos de clientes con respecto a las caracterizaciones antes mencionadas.
- 4- Se estudia cada grupo y se identifican en relación a estilos de vidas.

En consideración de lo anterior, para la realización del presente trabajo se propone la siguiente metodología.

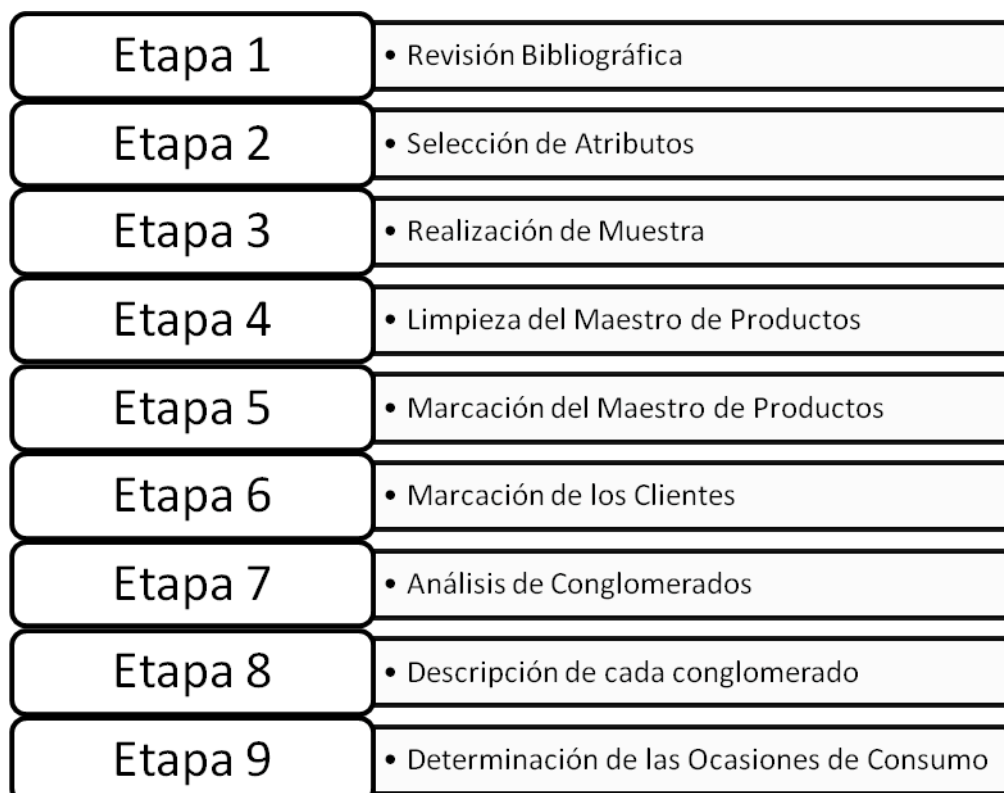


Figura 11: "Etapas Metodología". Fuente: Elaboración Propia

A continuación la descripción de cada una de las etapas enunciadas en la figura.

9.1 - Revisión Bibliográfica

Esta etapa tiene por objetivo obtener un acabado entendimiento del proyecto a realizar. Para ello se procede realizando una recopilación, selección y estudio del material bibliográfico afín para llevar a cabo el proyecto. Entre las principales actividades a realizar en esta etapa se tienen:

- Entrevistas con profesionales expertos en el tema.
- Estudio de libros y trabajos publicados con respecto del tema.
- Lectura de entrevistas y reportajes a las personalidades relacionadas con el tema.
- Estudio de las tendencias y avances al respecto.
- Investigación del material disponible en Internet.

9.2 - Selección de Atributos

Etapa crítica para el éxito del trabajo, ya que los atributos seleccionados tienen directa relación los segmentos finales. Por esta razón es importante generar atributos afines a los resultados esperados, es decir, que permitan reconocer estilos de vidas en los clientes, por ejemplo: Si para los productos se tiene el atributo *Aventurero*, que vendrían siendo productos particularmente extraños según la cultura, como hamburguesas de avestruz o carne de cocodrilo, tiene sentido que los clientes que compren productos con este atributo sean clientes *Lujosos (Finer Foods)*. Luego para generar los atributos es importante preguntarse ¿Cuáles Atributos? y ¿Para qué me serviría este atributo?

Para la selección de los atributos con los cuales se describirán los productos, se propone en primer lugar hacer una recopilación de todos los estilos de vida previsible, luego en función de estos generar atributos afines y finalmente hacer la selección, en base a evaluaciones y discusiones.

Una vez seleccionados los atributos, se tiene la discusión de las escalas y niveles que debe tener cada atributo, por ejemplo: Si se tiene el atributo *Saludable* este puede tener una escala *Si/No (Si el producto es saludable o no)*, puede tener varios niveles discretos o su extremo una escala continua. Si bien mientras más niveles por atributo es mejor, hay que considerar que la complejidad también va aumentando, especialmente para la etapa 5 que viene después.

Es importante tener en cuenta dos cosas: La primera es la constante validación de estos atributos con distintas personalidades con el fin de evitar sesgos; La segunda consideración, es seleccionar atributos y escalas que sean realizables con respecto de la información disponible, por ejemplo se podría pensar en el atributo “Calorías” pero si no se dispone de esta información por producto, no es posible llevarla a cabo.

9.3 - Realización de Muestra

Una vez realizada la etapa 2, se recomienda seleccionar una muestra de productos representativos por cada atributo, esto con dos fines: El primero es lograr un mayor entendimiento y tangibilizar lo que representa el atributo y en segundo lugar poder evaluar el atributo.

Este ejercicio permite por una parte apreciar la relación de las categorías con los atributos y afinar la definición de cada atributo, lo cual permite identificarlo mejor y hacerlo más objetivo.

Para la realización de esta etapa se recomienda realizar reiteradas visitas al supermercado, para observar el perfil de los clientes que compran los productos y tener la descripción de cada SKU seleccionado.

Una vez seleccionados los productos por atributo, es útil evaluar con terceros si cada producto representa el atributo en cuestión.

9.4 - Limpieza del Maestro

Esta etapa consiste en filtrar el maestro de producto con el cual se trabajará. Cabe decir que esta etapa tiene relación con los alcances definidos para el proyecto, en el presente se trabajará sólo con productos de tipo comestibles, por lo tanto se excluirán los llamados productos *Non-Foods*, luego no serán consideradas categorías como: Desodorantes, Artículos de Aseo, Artículos de Higiene Personal, etc.

Por otro lado, es usual que los maestros de productos tengan productos sin descripción, en el presente trabajo tales productos también serán excluidos, ya que los productos necesitan ser identificables para su descripción en base a los atributos.

9.5 - Marcación del Maestro de Productos

Esta etapa es sin duda una de las etapas más complejas de todo el trabajo, ya que tiene por objetivo la asignación de cada atributo para cada producto del maestro. Si consideramos el caso simplificado en que un supermercado tenga 50mil SKU, se seleccionaron 10 atributos y cada atributo tiene dos valores posibles, se tiene entonces el desafío de asignar 1 millón de valores. Notar que un maestro de productos de un hipermercado tiene cerca de 100 mil SKU.

El desarrollo de esta etapa comprende 3 actividades principales, estas son: La realización de la marcación del maestro de productos en forma manual y la ejecución de los modelos Producto- Producto-Clientes y Producto-Cliente Producto. El detalle de cada una de las actividades se detalla a continuación.

9.5.1 - Marcación Manual del Maestro de Productos

Esta actividad consiste en realizar en forma personal y manual, la asignación de los valores de cada atributo sobre el maestro de productos de un local de supermercados.

Es necesario realizar la marcación manual por los siguientes motivos:

- Permite seguir avanzando en la metodología.
- Es la forma que se tiene para evaluar los resultados de los modelos predictivos.
- Permite concluir, después de la segmentación, si los atributos seleccionados sirven.

Para la realización de esta actividad, se sugiere la siguiente secuencia de búsqueda:

1. Utilizar las categorías disponibles y marcarlas según un atributo cuando sea posible. Por ejemplo: La categoría *Frutas y Verduras* para marcar si se tiene el atributo *Verde*.
2. Buscar por productos que representen cada atributo. Si el atributo es *Saludable* se podrían seleccionar todos los productos *pan pita, avena, etc.*
3. Buscar por etiquetas transversales a los productos, como podrían ser *Light o 0%Calorias* si se tiene el atributo *Saludable*.
4. Proceder reconociendo las marcas comerciales de los productos que sirvan para marcar cada atributo, así se tiene por ejemplo la marca *Livean* para marcar el atributo *Saludable*.
5. Finalmente, buscar por SKU directamente.

9.5.2 - Método Producto-Producto-Cliente

Este método recibe su nombre debido a que busca asociar productos con productos reiteradas veces y después pasar a los clientes.

El proceso que sigue este método se muestra en la siguiente figura:

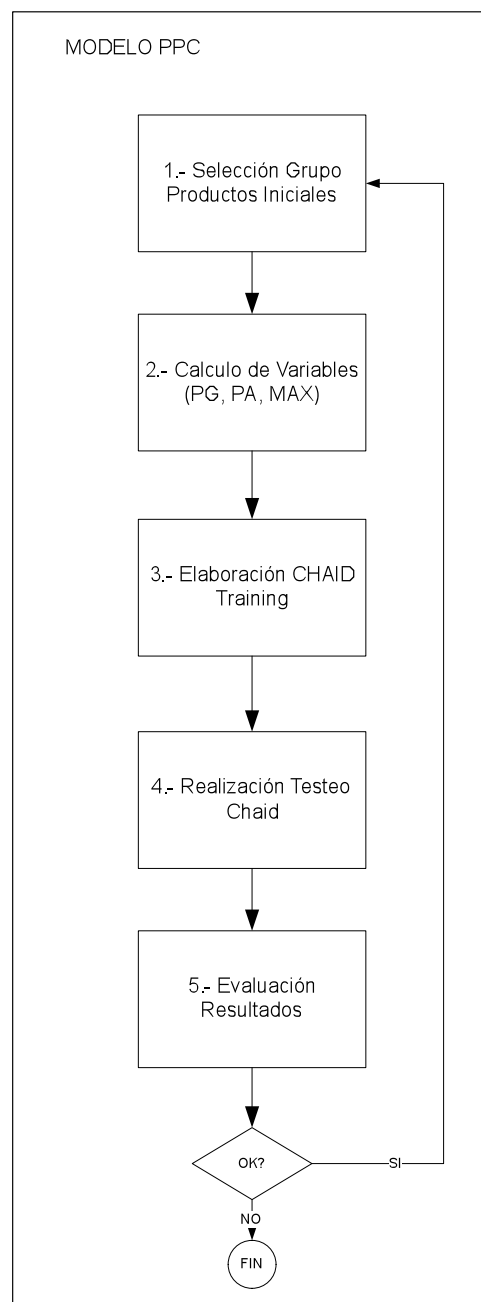


Figura 12: "Proceso PPC"; Fuente: Elaboración Propia

A continuación se describe cada una de las etapas del proceso.

1.- Selección de Grupo de Productos Iniciales

De acuerdo al atributo en cuestión, se selecciona un grupo de productos en que todos sean representativos del atributo.

2.- Cálculo de Variables

Utilizando el grupo de productos iniciales, se proceden a calcular las variables, las cuales asocian un producto desconocido con el grupo de productos iniciales. Estas son:

- Promedio de Ganancias (PG): Corresponde al promedio de las *Ganancias*¹⁰ del grupo de productos iniciales para cada producto desconocido. Considerar que a mayor ganancia entre dos productos, mayor es su probabilidad de compra conjunta.
- Productos Asociados (PA): Corresponde al número de productos iniciales con los cuales el producto desconocido ha tenido transacciones conjuntas. Esta variable se fundamenta en el supuesto de que si un producto que lleva con productos que tienen cierto atributo, entonces el producto también tiene el atributo.
- Máximo de Ganancias (MAX): Corresponde a la máxima *Ganancia* de un producto desconocido con un producto del grupo inicial.

Para entender mejor la procedencia de las variables, se tiene el siguiente diagrama:

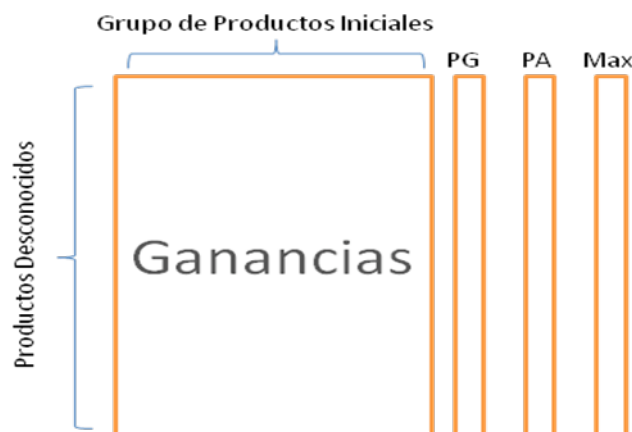


Figura 13: "Diagrama Variables PPC"; Fuente: Elaboración Propia

¹⁰ Ver Anexo 1

Se tiene el cálculo de la *Ganancia*, entre cada producto desconocido y cada producto del grupo de productos iniciales, generando una tabla representada por el cuadro “Ganancia” de la figura, donde las columnas son los productos iniciales y las filas son los productos desconocidos. Y luego por cada producto desconocido, se calculan las variables PG, PA y MAX.

3.- Elaboración CHAID Training

Calculadas las variables de la etapa anterior, se selecciona un grupo de productos que tiene la siguiente característica; El 50% de sus productos son el grupo inicial y el otro 50% son productos desconocidos pero que no poseen el atributo en estudio. A este grupo se le denominará *grupo de training*.

El siguiente paso consiste en la creación de nodos por medio de la aplicación del algoritmo CHAID al grupo de training, la idea es que el algoritmo genere reglas (o perfiles de nodos) que discriminen los productos del grupo inicial con los productos desconocidos, para más tarde aplicar tales reglas a los demás productos desconocidos.

Con respecto al *input* del algoritmo CHAID, se tiene:

- Variable a explicar:
 - ✓ Vector que tendrá valor 1 para los productos del grupo inicial y 0 para los productos desconocidos.

- Variables explicativas:
 - ✓ Promedio de Ganancias (PG)
 - ✓ Productos Asociados (PA)
 - ✓ Máximo de Ganancia (MAX)
 - ✓ Categorías de los productos.

4.- Realización Testeo CHAID

Obtenidas las reglas o perfiles de nodos, se procede a la aplicación de tales reglas sobre el grupo de testeo, el cual corresponde a todos los productos del maestro excluyendo el grupo de training.

5.- Evaluación Resultados

Se evalúan los resultados utilizando los indicadores Precisión y Soporte, que se definen como sigue:

- Precisión: Corresponde a la efectividad de la predicción, que viene siendo el porcentaje de los productos predichos que fueron predichos correctamente.
- Soporte: Corresponde al porcentaje de productos predichos correctamente del total de productos.

Si los resultados son satisfactorios, se vuelve al primer paso del proceso pero considerando los nuevos productos descubiertos. Si los resultados no son satisfactorios, se termina el proceso.

9.5.3 - Método Producto-Cliente-Productos

Este método recibe su nombre debido a que relaciona productos con otros productos por medio de los clientes. Esto se basa el supuesto de que *Cientes saludables compran productos saludables*, a diferencia del modelo anterior que se basa en que *Productos saludables se llevan con productos saludables*.

El proceso que sigue este método se muestra en la siguiente figura:



Figura 14: "Proceso PPC"; Fuente: Elaboración Propia

A continuación se describe cada una de las etapas del proceso.

1.- Selección Productos Iniciales

Se comienza el proceso con la selección de un grupo de productos que sea altamente representativo del atributo en cuestión.

2.- Identificación de Clientes

Se realiza una búsqueda de todos los clientes que hayan comprado al menos una vez cualquiera de los productos del grupo inicial.

3.- Elaboración de Variables Discriminatorias de Clientes

Utilizando el grupo de productos iniciales, se calculan las variables que permitirán ordenar los clientes para luego seleccionar lo más convenientes. Estas variables son:

- Presencia Transacciones: Corresponde al porcentaje de transacciones (compras) del cliente en que hay al menos un producto del grupo inicial.
- Presencia Productos: Porcentaje que corresponde a los productos comprados del grupo inicial con respecto a todos los productos distintos comprados por el cliente.
- Presencia Unidades: Corresponde al porcentaje que representan las unidades de los productos iniciales sobre el total de unidades compradas por el cliente.
- Presencia Monto: Corresponde a la penetración del grupo de productos iniciales sobre el monto total gastado por el cliente.

Una vez calculadas estas variables para cada cliente, se procede a la siguiente etapa.

4.- Selección de Clientes

Obtenidas las variables para cada cliente, se procede a seleccionar los clientes utilizando algún criterio, entre ellos:

- Criterio de Corte: Escoger cortes en todas o algunas de las variables antes mencionadas.
- Número Fijo: Ordenar los clientes y escoger los primeros 500 o 1000 clientes.
- Porcentaje: Ordenar los clientes y escoger un porcentaje fijo, tal como el primer 5% o 10% de los clientes.

5.- Identificación de Productos

Del grupo de clientes seleccionados, se buscan sus productos comprados al menos 1 vez por al menos 1 cliente.

6.- Elaboración de Variables Discriminatorias de Productos

Con el objetivo de hacer un *ranking* con los productos, de manera de poder elegir los productos que tengan el atributo en cuestión. Para ello se generan dos variables para cada producto, que se construyen en base a selección clientes. Estas variables son:

- Lift Presencias: Razón entre la presencia del producto en los Clientes Saludables y la presencia general del producto en el supermercado.
- Presencia en Clientes: Porcentaje que corresponde al número de clientes seleccionados distintos que llevaron el producto, sobre el número total de clientes seleccionados.

7.- Selección de Productos

Utilizando las variables discriminatorias de clientes antes descritas, se seleccionan los productos representativos del atributo en estudio. Para esto se proponen los siguientes criterios:

- Criterio de Corte: Escoger cortes en las variables mencionadas anteriormente.
- Número Fijo: Ordenar los productos y escoger un número fijo de productos.
- Porcentaje: Ordenar los productos y escoger un porcentaje fijo.

8.- Evaluación de Resultados

Se evalúan los resultados comparando los resultados del modelo con los resultados de la marcación manual del maestro. Si se está conforme se repite el proceso incluyendo los nuevos productos encontrados, por el contrario, si no se está conforme se termina el proceso.

9.6 - Marcación de los Clientes

Una vez marcado el maestro de productos con respecto de los atributos se procede a la marcación de los clientes. El objetivo de esta etapa es traspasar la información de los atributos de los productos a los clientes a partir de sus compras. Se propone calcular la presencia promedio del atributo en el total de las transacciones de cada cliente. La realización de esta etapa será llevada a cabo por medio de la siguiente secuencia:

1.- Preparación Atributos

Se separa cada atributo en una serie de variables *dummy*, por ejemplo: si se tiene el atributo "Precio" que toma los valores 1,2 o 3, según el precio del producto, se separa las variables "Precio1" que toma los valores 1 o 0, "Precio2" que toma los valores 1 o 0 y "Precio3" que toma los valores 1 o 0.

2.- Calculo de Presencias por Atributo

Contar cuantas veces se encuentra la variable *dummy* entre las transacciones de cada cliente y se determina el número de transacciones totales del mismo cliente. De esta forma se obtendrá una tabla que contenga el identificador del cliente, el número de transacciones totales y el número de veces que el cliente ha comprado cada atributo. Realizado esto se limpiará el maestro de los clientes que presentan anomalías como por ejemplo demasiadas o muy pocas transacciones.

3.- Porcentaje Atributo por Cliente

Por cada cliente, obtener el porcentaje que representa cada atributo con respecto a todas las compras realizadas, esto se hace simplemente dividiendo cada atributo por el número total de productos llevados.

4.- Calculo del Lift

Calcular las diferencias porcentuales por atributo entre todos los clientes, es decir, el lift por atributo. Esto con el objetivo de hacer comparables los atributos y tener referencia de magnitud para cada uno de ellos. El resultado será una tabla con el identificador de cada cliente y las diferencias porcentuales de cada atributo.

9.7 - Análisis de Conglomerados

Para esta instancia ya se tiene cada cliente caracterizado por cada uno de los atributos, los atributos que antes caracterizarían a los productos, ahora lo hacen con los clientes.

En términos generales esta etapa se divide en dos actividades principales, estas son:

1.- Análisis de Componentes Principales

Con el fin de mejorar la segmentación se aplica un ACP sobre los atributos, de esta manera se disminuyen el número de variables para la segmentación, generando nuevas dimensiones (o constructos), los cuales tienen correlación cero entre ellos.

Notar que el resultado del ACP son una serie de variables llamadas factores, que van desde 1 hasta el número total de atributos. De los factores se escogerán aquellos que tengan autovalor mayor o igual a 1, y serán estos los que pasarán a ser las nuevas dimensiones con las que se segmentará.

2.- Segmentación vía K-Means

Realizada la actividad anterior se procede a generar los clusters de clientes, para esto se utilizará el algoritmo K-Means.

Se escogerá el número de clusters utilizando la llamada "Regla del Codo" (Anexo 1), y luego se identificará cada cluster por medio de los valores de los promedios de cada dimensión en cada cluster.

Notar que para la identificación de cada cluster se busca asignar un estilo de vida particular, tal que este reúna la información entregada por las dimensiones.

9.8 - Descripción de cada Conglomerado

Se realizará un análisis y descripción de cada segmento encontrado, en función de su frecuencia de compra, ticket promedio, promedio de edad y porcentaje de hombres y mujeres.

Adicionalmente a la caracterización de cada cluster se sugerirán recomendaciones en función de la realización de marketing focalizado.

9.9 - Aplicar método RFM

Mediante la aplicación del método *Recency, Frecuency and Value* (RFM), será posible evaluar la lealtad e importancia que tiene cada segmento para el supermercado.

10.- Resultados

En esta sección se detallarán los resultados obtenidos para cada etapa planteada en la metodología.

10.1 - Revisión Bibliográfica

Se comenzó estudiando los trabajos que se han hecho con respecto al tema, destacando el trabajo de memoria de Sebastián Mena [6]. Luego, con el fin de conocer bien y entender el caso Tesco, se hizo lectura de los libros *Scoring Points:How Tesco is winning customer loyalty* [4] y el libro *Any colour you like as long as it's any colour you like* [3], ambos mencionados en la bibliografía. Finalmente se hizo una acabada búsqueda y estudio de las presentaciones y publicaciones con respecto a la segmentación por estilos de vida realizada por Dunnhumby para Tesco. Adicionalmente se tuvieron varias entrevistas con profesores expertos en el tema.

Por otro lado, se estudiaron las distintas segmentaciones por estilos de vida existentes, el concepto de los clubes de lealtad y las estrategias que existen en el retail para competir. Esto a partir del material presente en la bibliografía.

10.2 - Atributos

Una vez revisado el material bibliográfico disponible, se procedió al estudio de la información para generar los atributos que describirán los productos. El resultado final de tal estudio se presenta en la siguiente tabla.

Nº	Atributo	Escala	Descripción
1	Precio	-2,-1,1,2	Corresponde al rango de precio, en que se ubica el producto c/r a su categoría. Siendo -2 en nivel de precios más bajo.
2	Saludabilidad	1,0,-1	Grado en que el producto es reducido en grasas y beneficioso para la salud. 1 es saludable, 0 es indiferente y -1 es no saludable.
3	Perecibilidad	1,0,-1	Indica el periodo de vencimiento que tiene el producto. 1 indica que el producto vence rápidamente y -1 si el producto es no perecible.
4	Preparación	1,0,-1	Nivel de preparación que debe tener los productos antes de ser consumidos. 1 si el producto es instantáneo o requiere poca preparación, 0 si el producto es de consumo directo y -1 si requiere preparación.
5	Verde	1,0	De procedencia vegetal y no haya sufrido transformaciones físicas importantes. Por ejemplo: La harina no califica como Verde.
6	Carnes	1,0	De procedencia animal.
7	Chileno	1,0	Si el producto es típicamente Chileno.
8	Fiesta	1,0	Si el producto es de consumo en celebraciones, juntas de amigos, etc.
9	Extranjero	1,0	Si el producto es de marca extranjera y no común. Ej: Cerveza Gulden Draak.
10	Exótico	1,0	Si el producto no pertenece a la cultura chilena común. Ej: Guayaba.
11	Niño	1,0	Si el producto es de consumo típico de niños y bebés.
12	Congelado	1,0	Si el producto viene congelado.

Figura 15: "Atributos"; Fuente: Elaboración Propia

10.3 - Realización Muestra

Para la realización de muestras de producto por cada atributo, se visitó el supermercado en estudio, registrando 100 productos por cada atributo, para luego evaluar con distintas personas. Este ejercicio fue de gran ayuda para clarificar el concepto de cada atributo y ser consecuente para marcar el maestro en forma manual.

10.4 -Limpieza Maestro

Como se mencionó en la metodología se descartaron todos los productos no comestibles, esto se hizo eliminando las categorías de productos que no cumplieran el requisito. Adicionalmente se excluyeron todos los productos que no poseían descripción, ya que sin descripción no son identificables. En Anexo 3 se encuentran todas las categorías descartadas.

Finalmente se obtuvieron cerca de 17 mil productos de 24 mil iniciales.

10.5 - Marcación de los Productos de Maestro

10.5.1 - Marcación Manual

Como lo indica la metodología, se realizó una marcación manual de todos los productos del maestro con respecto de todos los atributos seleccionados. A continuación se menciona en términos generales como se realizó tal marcación.

Atributos Verde, Carnes, Fiesta, Congelado

Para la marcación de estos atributos se utilizaron las clasificaciones de los productos con que cuenta el maestro, en particular la clasificación “familia”, que viene siendo la categoría del producto. Para el atributo Verde fueron seleccionadas las familias “Frutas y Verduras”, “Verduras Congeladas” y “Conserva Frutas”. Para el atributo Carnes se seleccionaron todas las familias relacionadas con productos de origen animal, por ejemplo: fiambrería, vacuno, cerdo, pavo, pollo, pescados, etc. Para el atributo Fiesta se seleccionaron las familias de alcoholes y snacks. Por último para el atributo congelado, se contaba con la familia de los productos congelados.

Atributos Percibilidad y Preparación.

Para el atributo preparación, se seleccionaron como 1 las familias de productos que requieran cocción, entre estas están todas las carnes, arroz, legumbres, entre otros. Con -1 los productos que requirieran mínima preparación como platos preparados, sopas enlatadas, etc. Se marcaron con 0 los productos que vienen solo en un formato, por ejemplo: Azúcar, aceite, chocolates, etc. Notar que el objetivo de este atributo es descubrir aquellos clientes que por comodidad o falta de tiempo no se dan el tiempo de preparar comidas.

Con respecto al atributo Percibilidad, se marcaron con 1 los productos que fueran no perecibles tales como los cereales, legumbres, arroz, conservas, etc. Con -1 los productos que tuvieran un pronto vencimiento como por ejemplo las carnes frescas, productos refrigerados, huevos, etc. Por último con 0 los productos que no se caractericen por ningún extremo, por ejemplo los snacks.

Atributos Extranjero, Exótico y Chileno

Los atributos Extranjero, Exótico y Chileno se construyeron a partir de la búsqueda de marcas y palabras claves. Para el caso de Exótico, se disponía de la familia de productos "Étnico" la cual contenía productos no comunes de distintas partes del mundo y además se buscó por productos asociados a palabras claves como por ejemplo: avestruces, caviar, paté de foie, etc. Para el caso del atributo Extranjero se buscaron los productos que fueran de ciertas marcas, tales como: Kraft, Parade, etc. Ídem procedimiento para el caso del atributo Chileno, en las cuales se buscó por quínoa, las legumbres en general, zapallo, etc. Para encontrar estas palabras se realizaron una serie de visitas al supermercado, en las cuales se fueron registrando los nombres de las marcas y productos

Atributo Saludabilidad

Se marcaron con 1 los productos que sean bajos calorías y los productos que tengan propiedades beneficiosas para la salud, esto se hizo mediante la búsqueda de los productos Diet, Light, los productos Livean, Svelty, productos integrales, etc. Con -1 se marcaron los productos altos en grasas o que no sean recomendables para la salud como por ejemplo paté, hamburguesas, nuggets, alcoholes, etc. Con 0 se marcaron los productos que son indiferentes en cuanto a su contribución vitamínica o en grasas.

Atributo Precio

El atributo precio se construyó de la siguiente manera: Lo primero fue calcular el promedio de los precios por familia de productos, luego se marcaron los productos que están por sobre y bajo el promedio. Consecutivamente, por cada partición, se volvió a calcular el promedio y división de los productos, obteniendo finalmente cuatro niveles de precio, estos son: Alto- alto, alto-bajo, bajo-alto y bajo-bajo.

Atributo Fiesta

El atributo fiesta se construyó tomando las familias de productos Alcoholes y Snacks.

10.5.2 - Resultados Método Producto-Producto-Cliente

La aplicación del modelo se realizó exclusivamente para predecir el atributo Muy Saludable. Siguiendo el procedimiento planteado en la metodología y variando el número de productos iniciales, se llegó a resultados expuestos en la tabla.

Prods. Iniciales	Modelo	Precisión	Soporte
70	Training	97,0%	94,2%
	Test	40,8%	7,7%
150	Training	98,6%	96,0%
	Test	33,7%	6,1%
300	Training	93,5%	95,3%
	Test	37,2%	22,9%
500	Training	93,3%	94,8%
	Test	33,2%	14,4%
1000	Training	89,1%	96,0%
	Test	27,7%	86,4%

Figura 16: "Resultados PPC"; Fuente: Elaboración Propia

Nota: El número de productos iniciales indica el número de productos saludables iniciales, por lo tanto en cada prueba el número de productos del grupo de training es dos veces el número de productos iniciales.

Como se aprecia en la tabla, ocurre que para el grupo de training tiene buenos indicadores de precisión y Soporte, pero en la aplicación de las reglas sobre el grupo de training no ocurre lo mismo.

Con respecto del indicador precisión, el cual indica que porcentaje de los productos predichos saludables eran realmente saludables, se observa que para el caso de 70 productos saludables iniciales, se logra la mayor efectividad en el grupo de test. En un principio, al ser esta la primera prueba, se pensó que 70 productos no eran suficientes para que el algoritmo aprenda, pero resultó ser, que a mayor cantidad de productos de training, no mejoró el indicador.

Con respecto al indicador soporte, el cual indica el porcentaje de los productos saludables que fueron predichos correctamente, se observa que para el grupo de training no baja del 94%, mientras que para el grupo de test, no supera el 14%, exceptuando el caso de 1000 productos iniciales. Para el caso de los 1000 productos iniciales, si bien el soporte es alto, la efectividad es la más baja entre todas las pruebas.

En Anexo 5 están disponibles las matrices de confusión de cada prueba y el diagrama del CHAID, Anexo 4, para el caso de la prueba con 150 productos iniciales.

Una razón que explicaría los resultados obtenidos es que la utilización del indicador Ganancia (Anexo 1) no adecuada, ya que no logra diferenciar un atributo del resto de los productos. Se tiene por ejemplo el SKU *Leche descremada Soprole 1 litro*, el cual ha tenido transacciones con cerca de 8500 SKU distintos (en 2 meses), de los cuales el 24% de esos productos fueron marcados manualmente como saludables y que al graficar el comportamiento de los productos, no se observan diferencias entre el comportamiento de los productos saludables y los productos en general, para todos los niveles de Ganancias.

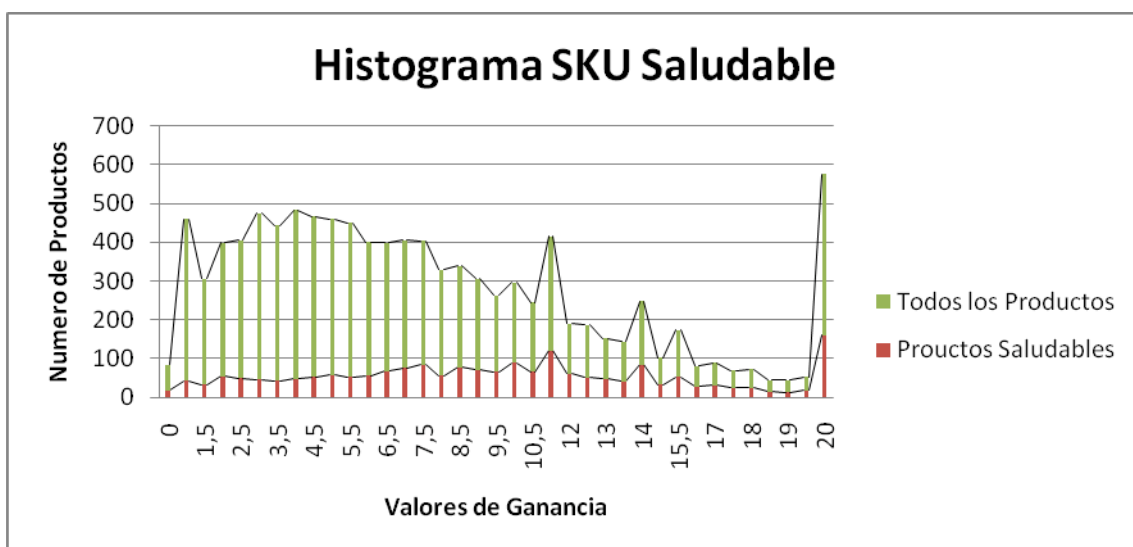


Figura 17: "Histograma Ganancias SKU"; Fuente: Elaboración Propia

Un resultado útil sería aquel que se pudiera apreciar un aumento de la proporción de productos saludables a medida que es mayor la ganancia, lo cual por lo visto no ocurre.

10.5.3 - Resultados vía Producto-Cliente-Producto

Tal como en la aplicación del modelo anterior, en este también se trabajó exclusivamente con el atributo Muy Saludable. Siguiendo el procedimiento descrito en la metodología se encontraron los resultados que se presentan a continuación.

Para comenzar el proceso se procedió a la selección del grupo de productos saludables iniciales, para ellos se consideraron 90 productos. Siguiendo con el proceso, se encontraron 6669 clientes.

Nota: Los datos obtenidos son de 6 meses (1 Enero 2010 al 30 Junio 2010) y se considera cliente a las personas que hayan tenido al menos una transacción mensual y máxima una diaria.

A continuación se muestra una tabla resumen, la cual es una matriz de dos entradas que por un lado tiene criterios de corte de acuerdo a la variable Presencia Producto (Presencia de los productos iniciales en el total de productos distintos comprados por el cliente) y por otro lado se tienen criterios de corte respecto de la variable Presencia Transacciones (Presencia de los productos iniciales en el número de boletas totales de cada cliente) y que en cada celda se indica el número de clientes que cumplen tales criterios.

		Pres- Trxs					
		Corte 40%	Corte 50%	Corte 60%	Corte 70%	Corte 80%	Corte 90%
Pres- Prods	1%	2617	1849	1033	520	265	77
	2%	2586	1835	1026	517	263	77
	3%	2346	1695	962	488	250	73
	4%	1797	1324	783	402	207	62
	5%	1221	924	563	281	149	44
	6%	740	571	350	185	101	33
	7%	467	362	226	124	69	23
	8%	279	218	126	63	38	12
	9%	171	135	74	38	23	9
	10%	111	86	44	20	12	5
	11%	73	56	28	14	9	4
	12%	50	39	21	11	7	2

Figura 18: "Cuadro Criterios para Clientes"; Fuente: Elaboración Propia

Para comprender la lectura de la tabla se tiene el siguiente ejemplo: La primera celda contiene 2617 clientes, estos clientes son aquellos que tienen una Presencia-Producto mayor o igual a un 1% y una Presencia-Transacciones mayor o igual al 40%.

Las celdas destacadas en amarillo son aquellas que fueron utilizadas durante el estudio. Se observa del estudio realizado que los clientes que tienen mayor Presencia-Producto son mejores para discriminar los productos saludables que los clientes de mayor Presencia-Transacciones. Esto debido a que los últimos, al no discriminar bien, dispersan los productos lo que obliga aplicar criterios muy ajustados para lograr buenos índices de efectividad. (Ver Anexo 6)

Los resultados de todas las pruebas realizadas se resumen en la siguiente tabla.

PRUEBA	CRITERIO CLIENTES	CRITERIO PRODUCTO	BUSQUEDA	10	11	12	13	14	15	18	17
A	Pres. Trxs >= 80% Pres.Prod>=3%	-	-	-	-	-	-	-	-	-	-
B	Pres. Trxs >= 60% Pres.Prod>=5%	Lift Presencias >= 3 y Pres. Clientes >= 6%	Todos Prods	90	-	-	-	-	-	-	-
C	Pres. Trxs >= 60% Pres.Prod>=5%	Lift Presencias >= 4 y Pres. Clientes >= 6%	Todos Prods	90	9 (100%)	-	-	-	-	-	-
D	Pres. Trxs >= 60% Pres.Prod>=5%	Lift Presencias >= 5 y Pres. Clientes >= 3%	Todos Prods	90	48 (92%)	-	-	-	-	-	-
E	Pres. Trxs >= 50% Pres.Prod>=8%	Lift Presencias >= 4 y Pres. Clientes >= 2%	Todos Prods	90	86 (78%)	-	-	-	-	-	-
F	Pres. Trxs >= 50% Pres.Prod>=8%	Lift Presencias >= 3 y Pres. Clientes >= 3%	Todos Prods	90	115 (85%)	85 (70%)	-	-	-	-	-
G	Pres. Trxs >= 50% Pres.Prod>=8%	Lift Presencias >= 2 y Pres. Clientes >= 5%	Todos Prods	90	-	-	-	-	-	-	-
H	Primeros 500	Lift Presencias >= 2 y Pres. Clientes >= 2%	Todos Prods	90	77 (87%)	74 (72%)	-	-	-	-	-
I	Primeros 1000	Lift Presencias >= 2 y Pres. Clientes >= 2%	Todos Prods	90	125 (86%)	97 (85%)	42 (67%)	-	-	-	-
J	Primeros 1000	Lift Presencias >= 3 y Pres. Clientes >= 1%	Todos Prods	90	47 (91%)	42 (79%)	-	-	-	-	-
K	Primeros 1000	Primeros 100	Todos Prods	90	100 (84%)	100 (76%)	100 (74%)	100 (72%)	100 (60%)	100 (58%)	100 (53%)
L	Primeros 1000	Primeros 100	Nuevos Última Iteración	90	100 (84%)	100 (84%)	100 (84%)	100 (68%)	100 (41%)	-	-

Figura 19: "Resultados PCP"; Fuente: Elaboración Propia

Nota: En esta sección se pondrá foco en las pruebas K y L, información sobre las demás pruebas en Anexo 7.

De la tabla:

La primera contiene los nombres asignados para cada prueba, la segunda columna contiene los criterios que se utilizaron para seleccionar los clientes

encontrados a partir de los productos iniciales, los cuales se obtienen de la siguiente manera:

- Presencia-Producto: Numero de Productos Iniciales distintos sobre el número total de productos distintos comprados por el cliente.
- Presencia-Transacciones: Numero de Transacciones en los cuales hay presente al menos un producto de grupo inicial sobre el numero de transacciones totales realizadas por el cliente.
- Primeros 500 y Primeros 1000: Equivale a escoger los primeros 500 o 1000 clientes, ordenados según alguna variable, en este caso Presencia-Producto.

La tercera columna contiene los criterios utilizados para la selección de los productos, los cuales se obtienen de la siguiente manera:

- Lift Presencias: razón entre la presencia de un producto en los clientes seleccionados sobre la presencia general del producto.
- Presencia-Clientes: Número de clientes del grupo seleccionado que compraron el producto sobre el número total de clientes del grupo seleccionado.

En la cuarta columna se tiene el tipo de búsqueda con la cual se buscan los clientes en la siguiente iteración, esta puede ser utilizando todos los productos saludables conocidos (Acumulativo) o utilizando sólo los productos nuevos encontrados en la última iteración. Por último se tienen las columnas que siguen, las cuales presentan los resultados de las iteraciones, entregando el número de productos nuevos encontrados y entre paréntesis el porcentaje de los productos que habían sido etiquetados como saludables, a lo cual se llamará efectividad.

El primer resultado interesante resulta el hecho de que la selección del grupo de clientes usando criterios de corte no es adecuada, esto debido a que ocurre que en cada iteración, en la medida que se van a agregando más productos el grupo de clientes seleccionado también crece, lo cual perjudica la posterior selección de productos. Por ejemplo, para el caso de la prueba F, en la tercera iteración el grupo de clientes seleccionados es de 3000 clientes, que corresponde al 45% de los clientes totales, lo cual dificulta la posterior selección de productos. Es por esta razón que la selección de clientes resulta más conveniente hacerla de manera determinística, en particular se vió que el grupo de los primeros 1000 clientes entrega mejores resultados que escoger los primeros 500, según los resultados de las pruebas H e I.

El segundo aprendizaje tiene que ver con los desempeños de los criterios de corte en los productos. Resulta ser que el criterio Lift-Presencias tiene un mejor desempeño que el criterio Presencia-Clientes, esto debido a que el último tiende a seleccionar productos denominados *Traffic Carrier*, que son aquellos

productos *de destino*, es decir, que la gran parte de los clientes los incluye en sus compras, lo cual representa un problema cuando se haga la selección de clientes posterior ya que se incluyen la mayoría de los clientes totales.

Con respecto de las pruebas con criterios de clientes fijos, lo que se hizo fue ordenar los clientes según Presencia-Producto, de mayor a menor, y luego seleccionar los 500 o 1000 primeros. Con respecto de las pruebas K y L, las cuales tienen criterios de productos fijos, lo que se hizo fue exigir un 1% de Presencia-Clientes y luego ordenar los productos según Lift-Presencias, de mayor a menor, para finalmente escoger los 100 primeros. La diferencia entre K y L, es la búsqueda de los nuevos productos para la iteración sucesiva, por una parte en K se hace una búsqueda acumulada, es decir, incluye todos los productos nuevos para buscar los clientes en la segunda iteración. Por su parte en L, se realiza la búsqueda de los clientes utilizando los productos nuevos encontrados en la iteración anterior.

Como conclusión, se tiene que para predecir o identificar productos saludables a partir de un grupo inicial, los mejores resultados se obtuvieron con las pruebas K y L, de las cuales parece funcionar mejor K se según las efectividades hasta la iteración 4.

En resumen, escogiendo la prueba K, se tiene que este modelo logra identificar 306 nuevos productos saludables, en 4 iteraciones, a partir de 90 productos iniciales conocidos, con una efectividad del 77%. En relación a los 3900 productos que fueron identificados como saludables vía la marcación manual, se tiene que el modelo tiene un soporte del 12%.

10.5.4 - Conclusiones Marcación Productos

Las principales conclusiones a partir de los resultados encontrados en esta etapa son:

1. El modelo Producto-Cliente -Producto logra predecir de mejor manera los productos saludables desconocidos que el modelo Producto-Producto-Cliente. Si bien sólo logra identificar el 12% de los productos saludables totales, ese es un problema que tiene solución (Dunnhumby lo resuelve).
2. Los resultados indicarían que el supuesto predominante es "*clientes saludables compran productos saludables*", por sobre el supuesto "*productos saludables se llevan con productos saludables*".

3. La marcación de productos del maestro en forma manual es crítica para la medición de los modelos.
4. Estos primeros resultados sirven por un lado para descartar opciones en la elaboración de modelos de predicción y por otro lado dan luces para guiar trabajos futuros.

10.6 - Marcación Clientes

Utilizando el maestro de productos marcado en forma manual, se procedió la marcación de los clientes como se indica en la metodología.

Un resultado interesante es el que presenta la siguiente tabla.

Atributo	Demanda	Oferta
Extranjero	0,8%	2%
Exótico	1,9%	3%
Instantáneo	20,1%	18%
Indif. Preparación	68,3%	66%
Lenta Preparación	11,6%	16%
Niño	15,4%	17%
Chileno	1,8%	1%
No Saludable	30,1%	43%
Indif. Saludable	27,0%	34%
Saludable	42,8%	23%
Vencimiento Rápido	64,2%	40%
Indif. Perecibilidad	16,0%	20%
No Perecible	19,8%	40%
Fiesta	6,9%	15%
Congelado	1,9%	4%
Verde	24,4%	10%
Carnes	10,9%	11%
Nivel Precios Alto-Alto	9,3%	
Nivel Precios Alto-Bajo	30,1%	
Nivel Precios Bajo-Alto	27,6%	
Nivel Precios Bajo-Bajo	32,9%	

Figura 20: "Oferta y Demanda"; Fuente: Elaboración Propia

En la tabla se presenta la demanda, determinada a partir de número total de productos comprados por cada atributo sobre el total de productos comprados, y la oferta, determinada como el porcentaje de productos disponibles por cada atributo.

Es interesante ver el caso de los atributos Saludable y No Saludable, se observa que la demanda es mayor a la oferta para el caso de los productos Saludables, mientras que para el atributo No Saludable la oferta es mayor a la demanda.

10.7 - Análisis Conglomerados

10.7.1 - Análisis de Componentes Principales

Con el objetivo de eliminar la correlación entre las dimensiones y mejorar la segmentación se realizó un análisis de factores. El resultado fueron 7 factores, los cuales tenían autovalores (*eigenvalues*) mayores que 1, que explican un 71% de la varianza (Anexo 8).

Con el fin de optimizar el proceso se le aplicó la rotación de ejes *Varimax*. La siguiente tabla muestra la construcción de los 7 factores a partir de las dimensiones originales.

	Componente (Factores)						
	1	2	3	4	5	6	7
Extranjero	,02	,00	,18	-,04	,05	-,05	,61
Exótico	-,09	,01	-,13	,04	-,08	,11	,71
Poca Preparación	-,25	-,69	-,35	-,04	-,28	-,19	,04
Indiferente Preparación	,11	,76	,18	-,49	,16	,14	-,01
Lenta Preparación	,18	-,20	,21	,83	,16	,05	-,03
Niño	-,47	,00	-,03	-,29	,54	,25	-,14
Chileno	,54	,13	,06	-,01	-,08	,10	-,05
No Saludable	-,78	,14	,28	-,09	-,08	-,06	,14
Indiferente Saludable	-,05	-,81	,07	,04	,06	,03	-,18
Muy Saludable	,74	,46	-,31	,05	,03	,03	,00
Muy Perecible	,54	-,24	-,68	-,02	-,10	-,01	,11
Indiferente Perecible	-,66	,50	-,05	,03	-,09	-,10	-,07
No Perecible	-,10	-,13	,87	,00	,20	,10	-,08
Fiesta	-,24	,26	,66	-,10	-,31	-,12	,22
Congelado	-,03	,26	-,10	,59	,00	,09	,08
Verde	,80	,33	-,22	-,02	-,05	,16	,01
Carnes	-,04	-,26	-,10	,74	-,13	-,11	,05
Precio Alto-Alto	,00	,11	-,07	,10	-,04	-,16	,64
Precio Alto-Bajo	-,19	-,15	-,05	-,06	-,46	-,75	-,12
Precio Bajo-Alto	,00	,09	,09	,04	,88	-,10	-,03
Precio Bajo-Bajo	,19	,01	,01	-,03	-,28	,90	-,23

Figura 21: "Matriz de Componentes Rotados"; Fuente: Elaboración Propia

Factor 1

De la figura 21, se observa que el factor 1 se construye utilizando las dimensiones Verde y Muy Perecible como máximos ponderadores, con 0.8 y 0.74 respectivamente. En el otro extremo, se tiene que la dimensión No Saludable es el menor con -0.78. En consideración de lo anterior, se le denominó a este factor como Saludable.

Factor 2

Se observa de la tabla, que en la construcción de este factor las dimensiones Muy Saludable y Verde tiene valores 0.46 y 0.33, que vienen siendo de los más altos. A su vez, la dimensión Carnes tiene un valor de -0.26 que es una de las menores. En consideración de esta información a este factor se le denominó Vegetariano.

Factor 3

Este factor destaca por tener un alto valor en la dimensión Fiesta, el cual es de 0.66. Por esta razón a este factor se le denominó Fiestero.

Factor 4

Se observa que este factor tiene sus mayores ponderadores en las dimensiones Lenta Preparación y Carnes, con valores de 0.83 y 0.74. Por el contrario, la dimensión Poca Preparación, que vienen siendo productos instantáneos, es una de las menores con un valor de -0.04. En consideración de lo anterior, a este factor se le denominó Cocinable.

Factor 5

Este factor pondera la dimensión Niño con 0.54, que viene siendo la mayor ponderación. Por el contrario atributos que representa productos para adultos como Fiesta y Exótico, tienen bajos valores. Utilizando esta información, a este factor se le denominó Infantil.

Factor 6

Este factor destaca por tener su valor más alto en los productos de más bajo precio, el cual es de 0.9. Por esta razón se le denominó Económico.

Factor 7

Este factor tiene la particularidad de que sus mayores ponderadores son de las dimensiones Exótico y Extranjero, con valores de 0.71 y 0.61. Por esta razón a este factor se le denominó Gourmet.

La siguiente tabla resume las denominaciones de cada factor y su descripción.

Factor	Constructo	Descripción
1	Saludable	Ponderadores altos las dimensiones Verde y Muy Saludables, bajo en productos No Saludables.
2	Vegetariano	Alto las dimensiones Verde y Muy Saludable, pero bajo en Carnes.
3	Fiestero	Destaca por tener altos valores en la dimensión Fiesta.
4	Cocinable	Alto valores para las dimensiones Carnes y Lenta Preparación.
5	Infantil	Destaca por tener un alto valor en el ponderador de Niño
6	Económico	Muy alto en el nivel de precio más barato y alto en niño.
7	Gourmet	Alto valor para las dimensiones Extranjero y Exótico. Además sobresale el nivel de precio más alto.

Figura 22: "Resumen Dimensiones"; Fuente: Elaboración Propia

10.7.2 – Segmentación

Utilizando los constructos antes definidos se procedió a realizar la segmentación de los clientes, para esto se utilizó el algoritmo de K-Medias vía el software SPSS. El número total de clientes con los que se trabajará es cercano a los 8600.

Lo primero a realizar fue determinar el número adecuado de clústeres, esto se determinó usando la regla del codo. Se realizaron pruebas para 3, 4, 5, 6, 7 y 8 clústeres para luego graficar el máximo del indicador F de cada caso, obteniendo la serie que indica la figura.

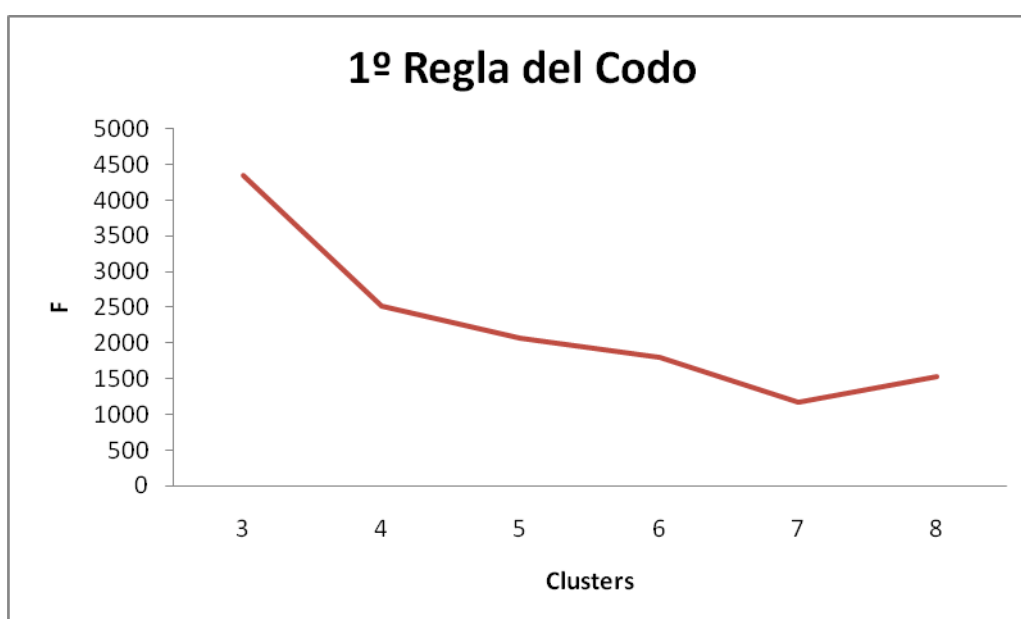


Figura 23: "Regla del Codo Inicial"; Fuente: Elaboración Propia

Se observa que la serie sufre dos cambios bruscos en su curvatura, en 4 y 7 clústeres, que vendrían siendo los números de clústeres óptimos para la segmentación, acorde a la regla del codo. Estudiando cada caso se obtuvieron los siguientes datos.

- Para el caso de 4 clústeres, ocurre que uno de estos posee sobre el 60% de los casos que representarían el cliente tradicional que tiene un comportamiento de compra plano por lo que no destaca en ningún constructo en comparación con el resto.
- Para el caso de 7 clústeres, ocurrió que un clúster poseía menos de un 1% de los clientes, con 58 casos.

Frente a este escenario se optó por excluir los 58 casos antes mencionados y volver a realizar el mismo procedimiento. El resultado de las pruebas se muestra en la siguiente figura.

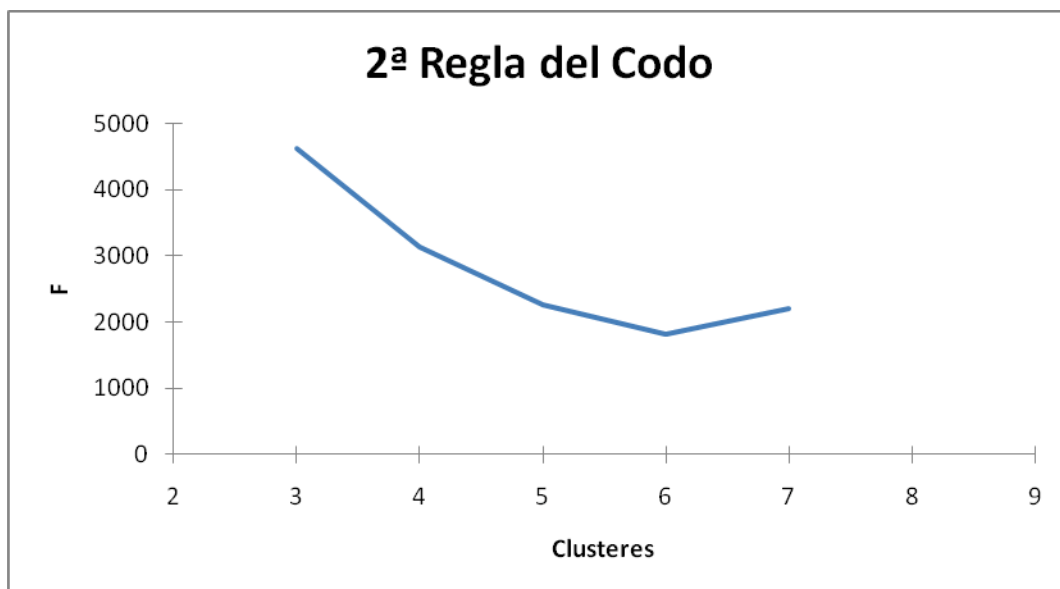


Figura 24: "Regla del Codo Final"; Fuente: Elaboración Propia

Se observa en la figura que la curva cambia bruscamente su curvatura en 6 clústeres. Bajo el criterio de la regla del codo, la figura indica que 6 clústeres es el número indicado para la segmentación.

Definido el número de clústeres, se procede a analizarlos para la realización de la segmentación. Lo primero a observar es el número de casos de cada clúster, los datos indicarían que el clúster más pequeño contendría cerca del 5% de los clientes, mientras que el más grande contendría cercano al 31%, lo cual es aceptable para la segmentación.

El siguiente paso en la segmentación viene siendo la identificación de los clústeres. Para esto se elaboró la siguiente tabla, la cual contiene los promedios de cada constructo para cada clúster y el promedio general.

Cluster	Segmento	Dimensiones						
		Saludable	Vegetariano	Cocinable	Carnes	Infantil	Económico	Gourmet
1	Tradicional	-0,12	0,29	0,02	0,42	0,09	-0,32	0,11
2	Express	-0,69	-1,05	-0,37	-0,53	-0,98	-1,2	-0,64
3	Familiar	-0,71	-0,39	0,48	-0,39	0,49	0,86	-0,51
4	Parrillero	-0,43	-0,08	0,08	2,59	0	0,04	-0,27
5	Gourmet	-0,36	-0,19	-0,26	-0,2	-0,07	0,18	2,12
6	Saludables	0,91	0,33	-0,23	-0,34	0	0,12	-0,18
Promedio		0	0	-0,04	0,01	0,01	0	-0,02

Figura 25: "Construcción Segmentos"; Fuente: Elaboración Propia

A continuación se detallará la identificación de cada clúster.

Clúster 1

Se observa de la tabla que el clúster 1 no destaca en ningún constructo en particular en comparación con el resto de los clústeres, es por esto que a este clúster se le denominó Tradicional. Este clúster representa a los clientes que compran similar al promedio y que no destacan en ninguna dimensión en particular, esto se puede ver en la Anexo 9.

Clúster 2

Se puede ver en la tabla que el clúster tiene valores menores y negativos al promedio en todos los constructos, además es como se verá más adelante es el clúster de menor ticket promedio y como se observa en del Anexo 9, este clúster tiene promedios de lifts positivos y alto para la dimensión *Instantáneo* y negativa para *Lenta Preparación*, es decir, estos clientes compran más productos tipo Instantáneos y menor productos de lenta preparación que el promedio. Por estas razones a este clúster se le denominó de Conveniencia.

Clúster 3

Se observa en la tabla que este clúster tiene altos valores para los constructos Infantil y Económico, además del constructo Cocinable que es mucho mayor al promedio. En suma a lo anterior del Anexo 9, se observa que tiene lifts promedios alto para los productos tipo No Perecibles y bajos para productos de alto precio. En base a esto a este clúster se le denominó Familiar.

Clúster 4

Se observa en la tabla que el constructo Carnes sobresale con respecto de los demás y tiene un valor muy por sobre el resto. Por esta razón a este clúster se le denominó Parrillero.

Clúster 5

De la tabla que puede ver que este clúster destaca por el alto valor que tiene el constructo Gourmet, adicionalmente se puede ver del Anexo 9, que el promedio de los lifts de los productos de alto precio es positivo y el mayor entre los clústeres. Por esta razón a este clúster se le identificó como Gourmet.

Clúster 6

Se observa de la tabla que este clúster es el que más altos promedios tiene para los constructos Saludable y Vegetariano, además de la tabla del Anexo 9 se observa que estos clientes compran menos que el promedio productos de las dimensiones Fiesta y Carnes. A este clúster se le denominó Saludables.

Ya identificados los segmentos, a continuación se muestra un gráfico que indica cómo se distribuyen los clientes entre los segmentos.

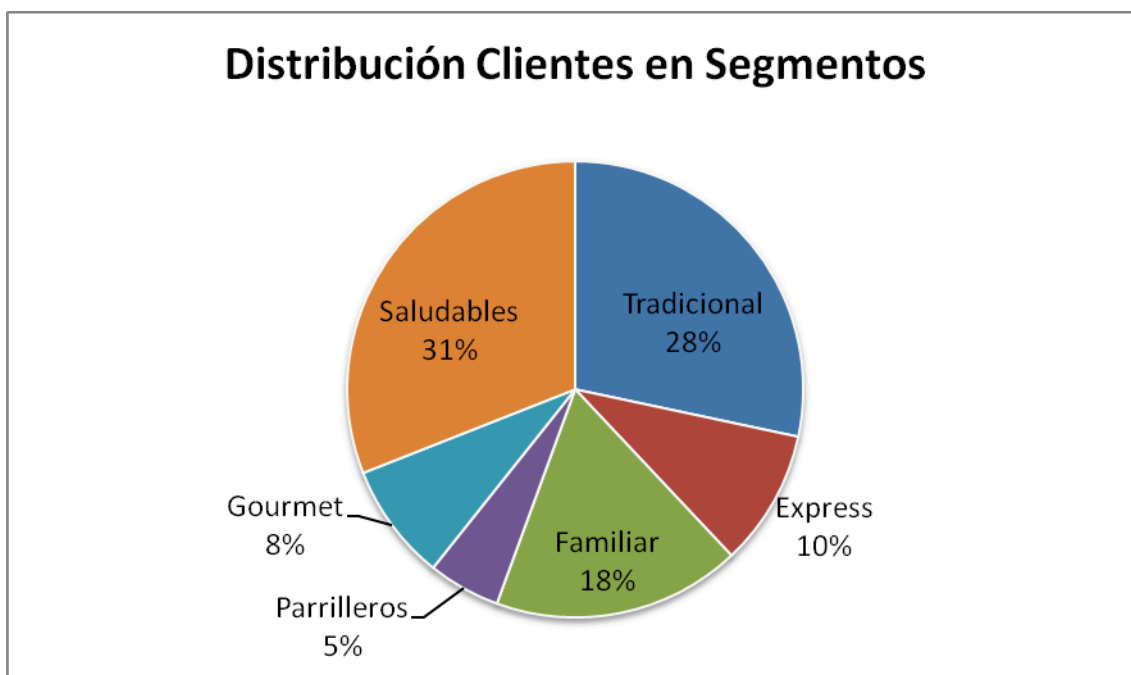


Figura 26: "Distribución Segmentos"; Fuente: Elaboración Propia

Se observa de la figura que el segmento definido como Saludables son los que tienen mayor participación, lo cual es consecuente con el mapa de roles del local ya que según este la categoría de productos Frutas y Verduras es de destino y es de las que mayor ventas genera. Se observa también que los menores segmentos son los Gourmet y Parrilleros, pero aún así representan una masa crítica abordable para la realización de marketing focalizado.

Nota: Alternativamente a realizar la segmentación con el método K-Means, se puede realizar con el método Fuzzy-C Means, el cual es un algoritmo de clasificación difuso, es decir, por cada cliente entrega una probabilidad de pertenencia a cada cluster. Resultados de la aplicación de Fuzzy C-Means en Anexo 12.

Finalmente, se presenta la siguiente tabla, que representa el porcentaje de las compras de cada segmento con respecto de algunas dimensiones. Tabla completa en Anexo 10.

	Tradicional	Express	Familiar	Parrillero	Gourmet	Saludables	Promedio General
Instantáneo	19%	42%	20%	21%	25%	16%	21%
Niño	14%	11%	28%	10%	14%	12%	15%
No Perecible	19%	15%	31%	21%	16%	16%	20%
Carnes	13%	12%	8%	28%	11%	8%	11%
Precio Bajo-Bajo	28%	22%	41%	32%	28%	37%	33%
Precio Alto-Alto	11%	6%	5%	9%	18%	8%	9%
Extranjero	1%	0%	0%	0%	3%	0%	1%
Exótico	2%	0%	1%	1%	8%	1%	2%
Muy Saludable	43%	23%	26%	35%	37%	58%	42%
Verde	22%	8%	12%	16%	19%	38%	23%

Figura 27: "Distribución Compras"; Fuente: Elaboración Propia

De la tabla se obtienen las siguientes conclusiones:

- Se observa que el segmento Tradicional posee compras similares al promedio de los clientes, por ejemplo: En la dimensión Muy saludable el segmento posee un 43% de sus compras y el promedio es de 42%.
- Se tiene que el segmento Express posee un porcentaje de sus ventas igual a 42% para la dimensión Instantáneo, mayor que el promedio, de 21%, lo cual respaldaría la identificación del segmento.
- Del mismo modo que lo anterior, se observa que el segmento Familiar posee los mayores porcentajes de sus compras en las dimensiones Niño, No Perecible y Precios Bajo-Bajo. Ídem con el segmento Parrillero con la dimensión Carnes, el segmento Gourmet con las dimensiones Precios Alto-Alto, Extranjero y Exótico y el segmento Saludable con las dimensiones Muy Saludable y Verde.

10.7.3 - Descripción Segmentos

Esta etapa tiene por objetivo caracterizar cada segmento en función de las características de los clientes. Para ello se cuenta con la figura 28, la cual entregará la información relacionada al comportamiento de compra de cada segmento, y la figura 29, que entrega las características demográficas disponibles de los segmentos.

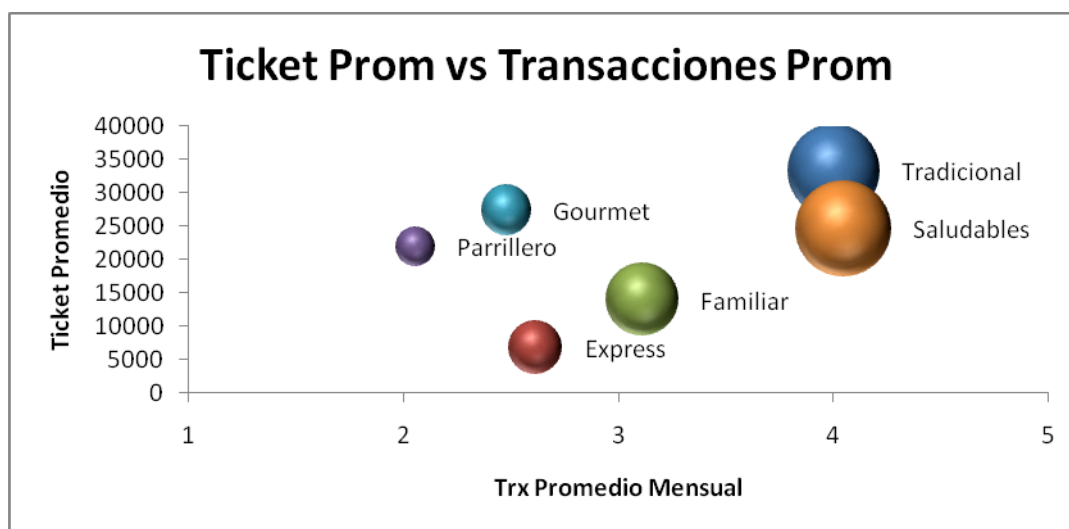


Figura 28: "Ticket Promedio vs Trxs Promedio"; Fuente: Elaboración Propia

Nota: El tamaño de las esferas en la figura X indica la participación de cada segmento.

Nº	Segmento	Clientes con Información			
		% Clientes	% Hombres	% Mujeres	Promedio Edad
1	Tradicional	29%	29%	71%	47
2	Express	8%	47%	53%	46
3	Familiar	17%	34%	66%	44
4	Parrillero	5%	38%	62%	45
5	Gourmet	8%	36%	64%	47
6	Saludables	32%	19%	81%	50
TOTAL		100%	29%	71%	46

Figura 29: "Información demográfica de los segmentos"; Fuente: Elaboración Propia

Nota: Los clientes con información corresponden al 45% de los clientes en estudio, más detalle se encuentra en Anexo 11.

Un resultado interesante es que el 70% de los *shoppers* son mujeres, lo cual ya da luces generales de cómo debieran ser las tácticas del supermercado en estudio.

A continuación el análisis de cada segmento.

1.- Tradicional

Segmento con una proporción de mujeres del 71%, con una frecuencia de compra semanal y ticket promedio cercano a los \$30mil.

Este segmento es el de mayor valor para el supermercado en términos de ingresos. La importancia de este segmento es evidente, por lo mismo su fidelización es fundamental. Dado que este segmento no destaca en ninguna dimensión en particular, se complica aún más su fidelización. Un programa de lealtad para este segmento podría actuar en función de beneficios como puntos por compras, sorteos y regalos. Aprovechando que sus *shoppers* son prioritariamente mujeres, los beneficios debieran ir en función de éstas.

2.- Express

Segmento de clientes que posee casi misma proporción entre hombre y mujeres, posee un ticket promedio de casi \$7 mil y una frecuencia de compra menor a la semanal.

Este segmento lo constituyen los clientes que buscan gastar el menor tiempo posible en el supermercado, llevan pocos productos y en particular productos de tipo Instantáneos, pero a pesar de aquello constituyen un 10% de los clientes del local, por lo que no se les debiera restar importancia. Dada las características antes mencionadas, tiene sentido pensar que para estos clientes sustituir el local tiene un costo muy bajo, por lo mismo para retener a estos clientes el local debiera facilitarles su compra. Para ello puede modificar el layout del local y habilitar cajas rápidas de manera de hacer su compra lo más expedita posible.

3.- Familiar

Segmento que se caracteriza por tener el promedio de edad más bajo, de 44 años, representa el 18% de los clientes del local con un ticket promedio cercano a los \$14 mil y de una frecuencia de compra cercana a las 3 veces por mes.

Dada la importancia de estos clientes el local debiera ofrecer una adecuada oferta en productos para niños y en consideración que este segmento compra productos precio-convenientes, un adecuado marketing focalizado debiera apuntar a ofrecerles promociones que signifiquen ahorros para estos clientes,

por ejemplo: Promociones de tipo *pack de productos* o descuentos acorde a sus preferencias.

4.- Parrilleros

Segmento de clientes que en promedio visitan 2 veces el local al mes, con un ticket promedio de sobre \$20mil. Su *shopper* es mayoritariamente mujer.

De la tabla del Anexo 9 se observa que estos clientes compran, por sobre el resto, productos de tipo Carnes, además de Anexo 2 se observa que casi el 30% de su compra es de este tipo, por lo tanto se puede decir que para lograr la fidelización de estos clientes el local debe tener una oferta de valor en carnes, que satisfaga a este segmento de clientes, de lo contrario es un 5% de los clientes que podrían cambiarse de cadena.

5.- Gourmet

Segmento que representa un 8% de los clientes, acuden entre 2 y 3 veces al mes al local y tienen un ticket promedio cercano a los \$27 mil. Su *shopper* es mayoritariamente mujer.

Este segmento es el segundo con mayor ticket promedio, esto indicaría que la disponibilidad a pagar por estos clientes es alta. El problema es que su frecuencia de compra es menor a la semanal, por lo tanto el supermercado debería realizar un marketing focalizado que tenga por objetivo aumentar su frecuencia de compra. Por otro lado, este segmento gusta de comprar productos exóticos y extranjeros, lo cuales tiene alto margen para el local, por lo mismo el local debería tener un adecuado surtido de productos que permita fidelizar este segmento.

6.- Saludables

El segmento más grande de clientes, de un ticket promedio de casi \$25mil y una frecuencia de compra semanal. Se caracteriza por ser el segmento de mayor promedio de edad, 50 años, y por tener la mayor proporción de mujeres, 81%.

Es uno de los segmentos más relevantes para el local. Se puede suponer que estos clientes gustan de productos frescos debido a su alta demanda por frutas y verduras y alta frecuencia de compra, por lo mismo es fundamental que el local tenga buena disponibilidad y exhibición de los productos preferentes por este segmento. Una campaña de marketing focalizado debiera apuntar a fidelizar a estos clientes por medio de la experiencia de compra y el posicionamiento del local en la mente de estos consumidores.

10.8 - Análisis RFM

Esta etapa consiste en el estudio de los segmentos por medio del análisis *Regency, Frecuency and Value (RFM)*.

Se calcularon los parámetros R, F y M de cada cliente y luego se calculó la razón entre R y F, para más tarde definir los clientes Leales, En Fuga y Fugados. El gráfico a continuación muestra el número de clientes para cada nivel de R/F.

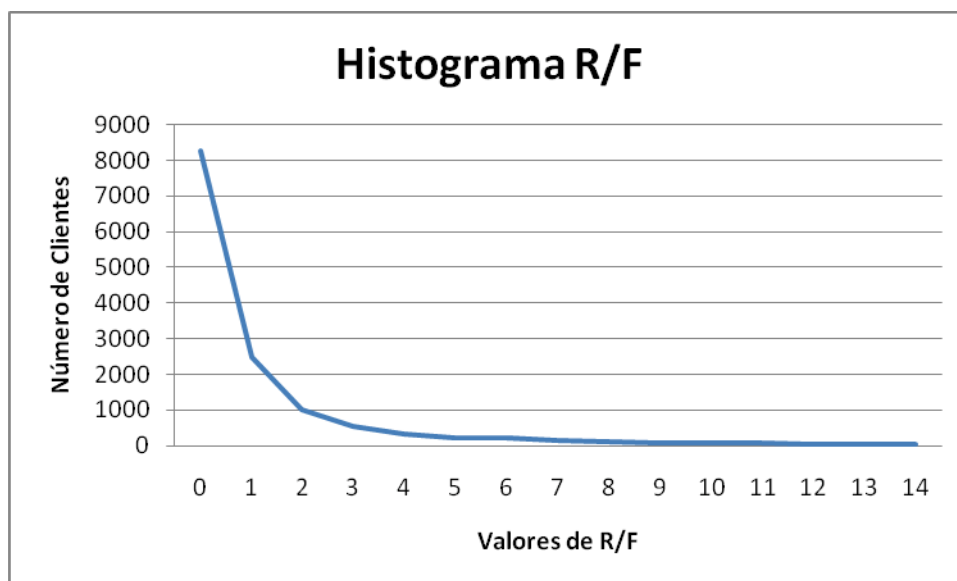


Figura 30: "Histograma R/F"; Fuente: Elaboración Propia

En consideración que los clientes del local en estudio tiene una frecuencia de compra semanal, se escogieron los siguientes criterios.

Rango R/F	Situación Cliente
[0,2)	Leal
[2,4)	En Fuga
[4,inf)	Fugado

Figura 31: "Criterios Lealtad"; Fuente: Elaboración Propia

Luego por cada segmento se tienen los siguientes resultados.

SEGMENTO	LEALES	EN FUGA	FUGADOS	%MONTO
Tradicional	80,9%	9,3%	9,8%	42,7%
Express	64,1%	12,1%	23,8%	2,4%
Familiar	69,6%	11,2%	19,2%	9,0%
Parrillero	69,2%	12,4%	18,4%	2,7%
Gourmet	76,9%	10,7%	12,3%	7,2%
Saludables	77,5%	10,2%	12,3%	36,0%
Total general	75,3%	10,5%	14,2%	100,0%

Figura 32: "Criterios Lealtad"; Fuente: Elaboración Propia

Se observa que los clientes que tienen menor lealtad y mayores fugados son los clientes Express, lo cual hace sentido ya que son clientes que buscan rapidez y comodidad en sus compras, por lo que el supermercado es fácilmente sustituible por un almacén u otra tienda de conveniencia. Estos clientes, si bien son los menos leales, sólo representan un 2,4% de las ventas de supermercado.

Es interesante ver que el segmento que mayor monto representa son los clientes Tradicionales, a pesar de ser el segmento Saludable el que tiene más clientes y tener ambos la misma frecuencia de compra, esto se explicaría por la diferencia en ticket promedio.

Otro punto de interés son los clientes Gourmet, ya que estos clientes tienen mayor porcentaje de clientes leales que el promedio, a pesar que los Familiares tienen mayor frecuencia de compra. Además se observa que los clientes Saludables son más del doble que los clientes Gourmet y las diferencias en monto son menores a un 2%.

A continuación se presenta una tabla que indica cómo se distribuye el monto en cada uno de los segmentos.

SEGMENTO	LEAL	EN FUGA	FUGADO	Total general
Tradicional	89,7%	6,5%	3,9%	42,7%
Express	82,9%	8,2%	8,9%	2,4%
Familiar	84,3%	8,9%	6,8%	9,0%
Parrillero	84,8%	8,0%	7,2%	2,7%
Gourmet	84,6%	9,7%	5,7%	7,2%
Saludables	88,5%	7,1%	4,4%	36,0%
Total general	88,1%	7,2%	4,7%	100,0%

Figura 33: "Distribución Monto en Segmentos"; Fuente: Elaboración Propia

Se observa que la distribución de los montos gastados por cada segmento es similar, en todos los clientes leales representan sobre el 80% del monto.

Un resultado importante es que los clientes leales, que son el 75%, representan el 88% del monto gastado en el supermercado, lo cual apunta a que el supermercado debe poner foco principalmente en esos clientes.

Por último se muestra el siguiente gráfico de ganancias para monto.

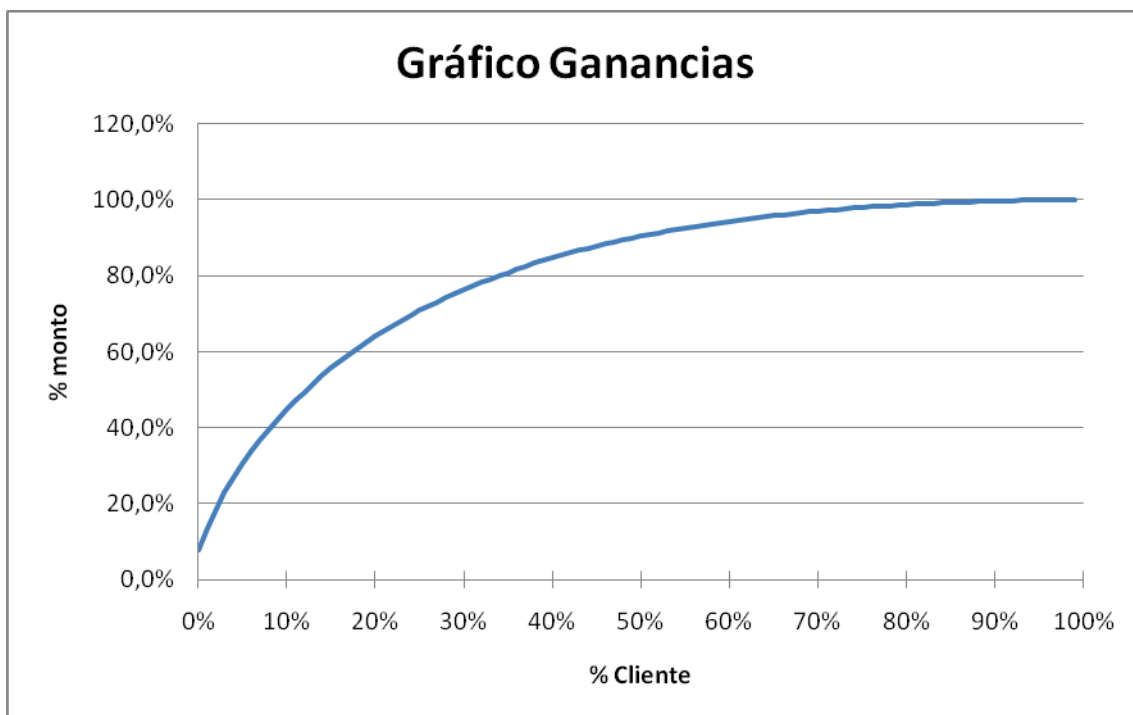


Figura 34: "Gráfico de Ganancias Monto"; Fuente: Elaboración Propia

La principal conclusión del gráfico es que un 30% de los clientes hacen casi el 80% de la venta del supermercado, entre los clientes en estudio. Esta información es tremendamente relevante, ya que conociendo los estilos de vida de tales clientes, se puede efectuar marketing focalizado que permita crear relaciones de confianza y de largo plazo con el cliente.

11.- Conclusiones

A continuación se presentan las principales conclusiones obtenidas del trabajo realizado.

1.- Se logró segmentar a los clientes e identificarlos según un estilo de vida, obteniendo un resultado similar a los resultados esperados planteados.

2.- Se determinaron los atributos con los cuales se caracterizarían los productos del maestro, los cuales correctos para identificar estilos de vida en los clientes.

3.- Se elaboraron dos métodos para marcar los productos del maestro. Los resultados indican que la predicción vía los clientes es más eficiente que la predicción vía las reglas de asociación. Resultó que el mejor método predice productos con buena efectividad, pero logra identificar una reducida cantidad. La marcación manual de los productos del maestro sigue siendo la mejor opción.

4.- Se definió exitosamente una metodología para caracterizar los clientes, en base a los atributos de los productos presentes en sus compras. Se observó que tal metodología permitió describir cada cliente en función de los atributos, para luego hacer posible la formación de clusters de clientes, identificables por estilos de vida. Tal metodología se basaría en describir cada cliente, en base a la presencia promedio de cada atributo en sus compras.

5.- Se logró determinar 6 segmentos representables por estilos de vidas, los cuales son: Tradicional (28%), Express (10%), Familiar (18%), Parrillero (5%), Gourmet (8%), Saludables (31%). Estos segmentos pueden ser alcanzados vía marketing focalizado.

6.- Se caracterizaron los segmentos antes mencionados. Para ello se utilizaron los parámetros ticket promedio, frecuencia de compra y las características demográficas. Por otro lado, utilizando el método RFM, se analizó la lealtad de cada segmento y el valor, en términos de los ingresos que representan para el supermercado. Se observó que el segmento Tradicional es el que presenta el mayor valor en cuanto a ingresos para el supermercado y es el que tiene mayor porcentaje de clientes Leales, aproximadamente 80%.

7.- Resulta interesante ver que un 30% de los clientes generan casi el 80% de los ingresos para el supermercado, de los clientes en estudio. Por otro lado el hecho que el 70% de los shoppers sean mujeres, lo cual es importante tener en consideración para las acciones de marketing del supermercado.

12.- Trabajos Futuros

Lo primero es realizar el procedimiento antes estudiado pero con datos de un mayor periodo de tiempo, con el fin de evitar estacionalidades y hacer más confiable la segmentación. En este trabajo se utilizaron los meses del primer semestre del 2010, un trabajo futuro podría incluir un año completo. Además se podría realizar la misma segmentación pero otro local, para observar las diferencias entre los segmentos que puedan existir.

Uno de las etapas más críticas del trabajo realizado, fueron los modelos para la marcación de los productos. Con respecto a este tema se proponen los siguientes trabajos futuros:

- Calcular el indicador Ganancia para un mayor periodo de tiempo.
- Tener una escala mayor por cada atributo, de esta forma podrían marcarse con la puntuación máxima los productos de las primeras iteraciones y luego ir bajando.
- Alternativamente a los modelos propuestos, se podría hacer un modelo *Producto-Boleta-Producto*, es decir, con los primeros productos buscar sus semejantes en las boletas conjuntas y luego seleccionar con algún criterio.
- Para el modelo PCP propuesto, se podría implementar un score por producto en función de las características de los clientes que lo compran. El actual modelo no considera el hecho que hay clientes más saludables que otros, por lo mismo se podría realizar un indicado similar a un presencia ponderada.

Por otra parte, para marcar los clientes en base a sus compras, se podría hacer vía las boletas. El modelo propuesto agrega por cliente la información de las boletas, por el contrario se podrían caracterizar las boletas y base a estas caracterizar a los clientes.

Con respecto de la segmentación, se podría validar con datos del segundo semestre y observar cómo cambian los segmentos. Por otro lado, se podrían formar clusters utilizando otros métodos, como el Fuzzy C-Means (Anexo 12).

Por último, con respecto de los segmentos encontrados, se podrían caracterizar utilizando más información como GSE, número de hijos, información personal, opiniones, etc.

13.- Bibliografía

- [1] **Abell Peter, Scott Kevin (2003)**, “Loyalty Programs Can Succeed without Excess Cost”, AMR Research.
- [2] **Cigliano Jim, Georgiadis Margo, Pleasance Darren, Whalley Susan (2000)**, “*The Power of Loyalty*”.
- [3] **Hayward Martin (2009)**, “Any colour you like as long it’s any colour you like”.
- [4] **Humby Clive (2004)**, “*Scoring Points: How Tesco is winning customer loyalty*”.
- [5] **Levy & Weitz (2006)**, “*Retailing Management*”.
- [6] **Mena Sebastian (2008)**, “*Clasificación del shopper en una cadena de supermercados*”. Santiago, Chile.
- [7] **Oracle (2006)**, “*Gaining Competitive Advantage through Effective Retail Loyalty Programs*”.
- [8] **Pavey Giles (2008)**, “*Customer insight in action at Tesco*”, DUG Conference.
- [9] **Pizarro Claudio (2009)**, “*Curso: Gestión de Retail*”, Universidad de Chile, Santiago, Chile.
- [10] **Sigdman Juan José (2009)**, “*Caracterización de las transacciones para una cadena de supermercados*”, Santiago, Chile.

14.- ANEXOS

ANEXO 1: Definiciones

1.- Retail: Se entiende como retail el sector económico que engloba a las empresas especializadas en la comercialización masiva de productos o servicios uniformes, a grandes cantidades de clientes. Se tienen cuatro principales categorías que son: Los supermercados como por ejemplo Lider, Jumbo y Unimarc. Las tiendas por departamento como por ejemplo Paris, Ripley y Falabella. Las tiendas de mejoramiento del hogar como por ejemplo Easy y Homecenter. Y por último las tiendas de especialidad las cuales comercializan una categoría de productos en particular, como por ejemplo las Salco Brand, Cruz Verde y FASA entre las farmacias y MTS en ferreterías. Cabe notar que por lo amplia de la definición también podrían ser considerando como retail los bancos y otros.

2.- Categoría: Corresponde a un grupo distinto y administrable de productos o servicios que los consumidores perciben como interrelacionados y que satisfacen una necesidad del consumidor. Por ejemplo: lácteos, carnes, ropa, medicamentos, electrodomésticos, etc.

3.- Shopper o Cliente Comprador: Corresponde a la persona natural que hace la compra en la sala de venta.

4.- SKU: En inglés Stock Keeping Unit, corresponde a un código único asignado por el comercializador, mediante el cual, el producto queda identificable por marca, tamaño, peso, volumen, sabor, etc.

5.- Surtido: Corresponde al número de SKU por categoría.

6.- Rol de la categoría: Se entenderá como tal, al uso e importancia que el shopper le otorga a la categoría. Existen diversas clasificaciones de roles, para el caso de este estudio se considerará la siguiente:

i.- Destino: La categoría destino es la categoría por la cual el cliente se acerca a la tienda, es decir, por la cual el cliente prefiere la tienda por sobre la competencia para satisfacer su necesidad.

ii.- Rutina: Categoría por la cual el cliente no percibe mayor valor agregado que la competencia.

iii- Ocasional/ Estacional: Categorías de la tienda que el cliente no compra con frecuencia.

7.- Ticket: Representa el registro de la compra realizada por el shopper en el punto de venta. Esta información incluye entre otras cosas la ID del shopper y el detalle de su compra.

8.- Gestión de Retail: Corresponde a las distintas acciones a realizar con el fin de mejorar las relaciones con los clientes y rentabilizar los recursos disponibles. Se tienen cuatro principales áreas de estudio como indica el siguiente cuadro.

9.- ANOVA (analysis of variance): es una recopilación de modelos estadísticos, los cuales particionan la varianza observada en las distintas variables explicativas, es decir, ANOVA provee distintos test que indican si la media es igual o no. En particular se realizan los test Student y Fisher.

10.- Regla del Codo: Es una técnica utilizada en los análisis de conglomerados que permite escoger, en forma visual, el número de clústeres a utilizar. Esta consiste en formular el gráfico Entropía (F) vs Clusters, para luego visualmente escogen el número de clústeres donde la curva presente un cambio brusco en su curvatura. Usualmente se grafica el promedio o el máximo de la entropía por cluster.

11.- Errores de Clasificación: Se tienen dos errores fundamentales en la clasificación de productos para este trabajo. El error de tipo I corresponde a dejar de clasificar productos siendo estos ciertos. El error de tipo II corresponde a clasificar productos siendo estos no ciertos. Por ejemplo si se quiere clasificar productos de tipo saludables, el error I viene siendo dejar de incluir productos saludables y el error II viene siendo incluir productos no saludables en la clasificación.

12.- Árboles de clasificación¹¹: Tipo de técnica de clasificación catalogada como supervisada, la forma por nodos y arcos con las siguientes restricciones:

1. Hay un solo nodo raíz, es decir un nodo que no tiene padre.
2. Cada nodo distinto de la raíz tiene un solo padre

Cada ramificación del árbol se define como nivel. Las primeras ramificaciones se llevan a cabo con la variable que mejor discrimina. La metodología que usan los árboles para ramificar es clasificada como Greedy. Es por eso que se da la posibilidad de que un atributo aparezca en diferentes niveles de las ramificaciones.

La variable supervisada está compuesta por valores nominales que indican la pertinencia de cada registro a cada clase. Se destacan los siguientes tipos de árbol.

¹¹ Fuente: Vargas, María Fernanda. (2009), Rediseño de un modelo de incentivos para vendedores de una tienda por departamentos. Santiago, Chile.

- Chi Squared Automatic Interaction Detection (CHAID): fue diseñado por J.A. Hartigan. Utiliza la prueba chi-cuadrado para determinar si se debe continuar con la ramificación y en caso afirmativo identifica qué variable ramificar.
- C4.5: basada en ID3 (Interactive Dichotomiser 3) de J. Ross Quinlan. Este tipo de árbol selecciona las variables para la ramificación de acuerdo a indicadores de entropía o desorden, es decir, utiliza la variable que proporcione mayor información en cada ramificación.
- CART: Utiliza el criterio basado en GINI index para el caso de la clasificación. Cuando se pretende construir un árbol de regresión, los criterios se basan en la mínima suma de las desviaciones cuadráticas. Para realizar la post-poda realiza una estimación de error, bien mediante un conjunto de datos diferente del que se ha utilizado para construir el árbol, o bien aplicando una metodología de validación cruzada.

Se deben sensibilizar una serie de parámetros, tales como qué criterio usar para la ramificación, criterio de poda, es decir, cuántos terminales generar, y también cuántos casos debe haber como máximo en los nodos terminales, Mientras más nodos, menor error de entrenamiento se tendrá, pero suele incurrir en un mayor error de test, lo que provoca un sobre ajuste del modelo calibrado.

13.- Reglas de Asociación: Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos. Se tienen 3 variables relevantes:

i.- Soporte: Corresponde a presencia del producto en el local, se calcula como el N° de transacciones de un producto sobre el N° de transacciones totales.

ii.- Confianza: Corresponde a la probabilidad de compra condicionada de un producto, se calcula a partir de: Soporte de dos productos sobre el Soporte de uno de los productos anteriores.

iii.- Ganancia: Indicador que señala en cuanto aumenta la probabilidad de compra de un producto dado que se lleva otro, se obtiene a partir de Confianza dividida por su Soporte. Para su utilización, fue determinada de la siguiente manera.

$$G_{AB} = \frac{TrxConj \times TrxCad}{TrxA \times TrxB}$$

Donde:

- G_{AB} : Ganancia entre A y B
- TrxConj: Transacciones conjuntas entre A y B.

- TrxCad: Transacciones totales de la cadena.
- TrxA: Transacciones totales de A.
- TrxB: Transacciones totales de B.

ANEXO 2: Extracto Cuestionario VALS

Selección de las preguntas más interesantes VALS:

- Sexo
- Edad
- Nivel de estudios
- ¿Sigue las últimas tendencias?
- ¿Le gusta estar a cargo de un grupo?
- ¿Prefiere hacer algo que comprarlo?
- ¿Se considera un intelectual?
- ¿Le gusta tener emoción en su vida?
- ¿Le gusta hacer trabajos manuales?

ANEXO 3: Limpieza Maestro

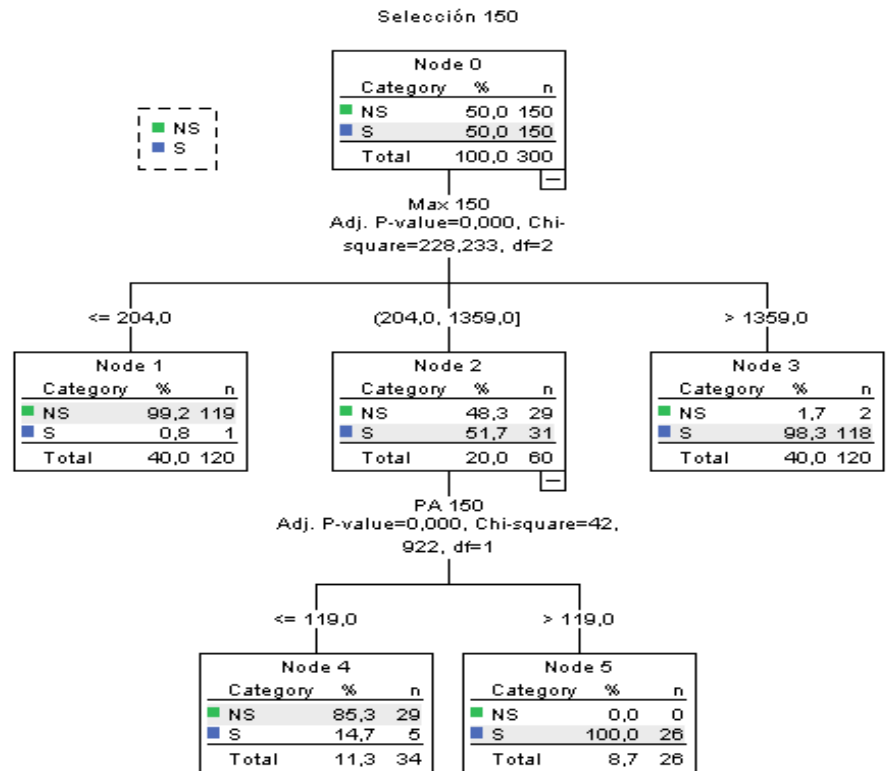
Categorías de Productos Descartadas		
PASTA DENTAL	TINTURAS	TEXTIL COCINA
LIMPIADORES BANO	SHAMPOO	COLONIA
TOALLA HIGIENICA	ARTICULOS ESCOLARES	ACCESORIOS COCINA
AUTOMOVIL	INSECTICIDAS Y ANTIPLAGA	ARTICULOS BEBE
LAVALOZAS	FOSFOROS	CREMAS AFEITAR
DETERGENTES	JABON	VELAS E INCIENSOS
ENJUAGUE BUCAL Y FIJADORES	FIJADORES	ADORNOS CASA
ALIMENTO ANIMALES	TOALLAS FACIALES Y ALGODON	CUADERNOS
BOLSAS BASURA	LIMPIADORES HOGAR	ALCOHOLES Y AGUA OXIGENADA
PILAS	CEPILLOS Y SEDAS	AFTER SHAVE
CREMAS	ACCESORIO PELO	FIESTAS PATRIAS
PANO	TAMPONES	ARTICULOS ELECTRICOS
DESODORANTES AMBIENTALES	BALSAMO-ACONDICIONADORES-CREMAS	SERVICIOS GENERALES
CLOSOS	PANUELOS DESECHABLES	ACCESORIOS DEPILATORIOS
DESODORANTES	AMPOLLETAS	CRISTALERIA Y VAJILLA
ESPONJAS	ACCESORIOS PRIMEROS AUXILIOS	ACCESORIOS ANIMALES
CUMPLEANOS-FIESTAS	ACCESORIOS JARDIN	LIBROS
ACCESORIO ASEO	PANAL ADULTO	FERRETERIA
PAPEL HIGIENICOS	CERAS	ACCESORIOS TOCADOR
TOALLA DE PAPEL	PROTECTORES Y BRONCEADORES	ROPA
ROLLOS FILM Y BOLSAS HERM	BETUNES	JUGUETES
PRENSA Y REVISTAS	MAQUINAS Y REPUESTOS	ACCESORIOS CALZADO
BELLEZA FACIAL Y CUERPO	CARBON ESPINO	NAVIDAD

Estas categorías de productos fueron descartadas para el desarrollo del presente trabajo.

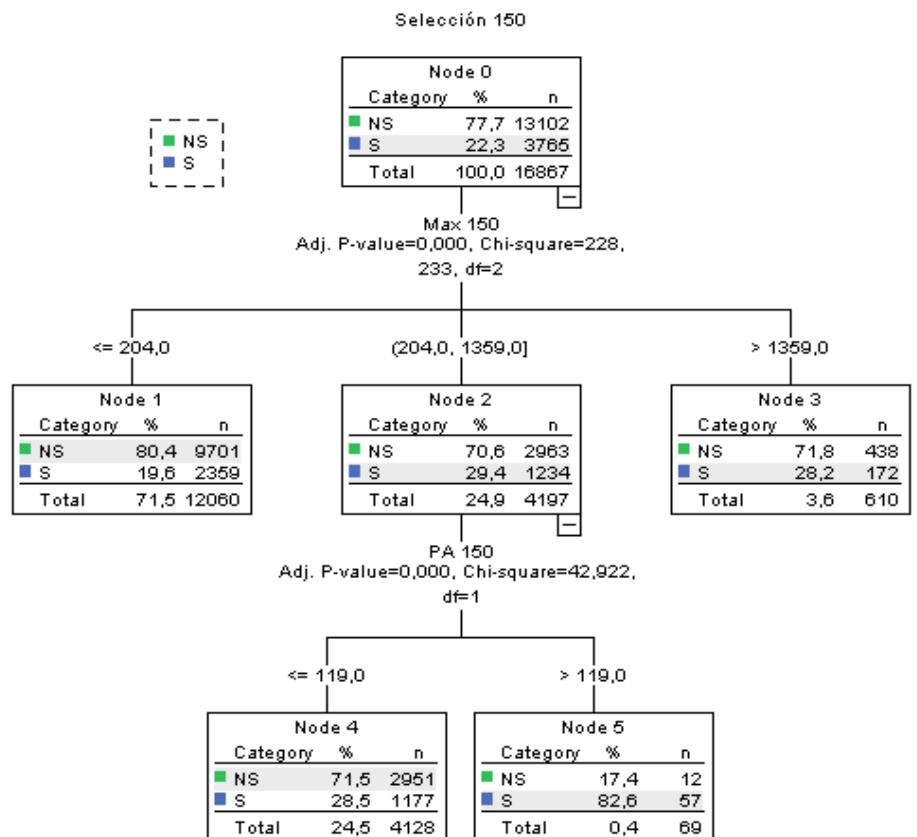
ANEXO 4: Diagrama CHAID

Prueba con 150 productos iniciales.

ARBOL DE TEST



ARBOL DE TRAINING



ANEXO 5: Matrices de Confusión del modelo PPC

70 Prods Iniciales

Classification

Sample	Observed	Predicted			PRECISIÓN	SOPORTE
		No saludables	Saludables	Percent Correct		
Training	No saludables	69	2	97,2%	PRECISIÓN	97,0%
	Saludables	4	65	94,2%		
	Overall Percentage	52,1%	47,9%	95,7%		
Test	No saludables	12790	431	96,7%	PRECISIÓN	40,8%
	Saludables	3554	297	7,7%		
	Overall Percentage	95,7%	4,3%	76,7%		

150 Prods Iniciales

Classification

Sample	Observed	Predicted			PRECISIÓN	SOPORTE
		NS	S	Percent Correct		
Training	NS	148	2	98,7%	PRECISIÓN	98,6%
	S	6	144	96,0%		
	Overall Percentage	51,3%	48,7%	97,3%		
Test	NS	12652	450	96,6%	PRECISIÓN	33,7%
	S	3536	229	6,1%		
	Overall Percentage	96,0%	4,0%	76,4%		

300 Prods Iniciales

Classification

Sample	Observed	Predicted			PRECISIÓN	SOPORTE
		NS	S	Percent Correct		
Training	NS	280	20	93,3%	PRECISIÓN	93,5%
	S	14	286	95,3%		
	Overall Percentage	49,0%	51,0%	94,3%		
Test	NS	11590	1403	89,2%	PRECISIÓN	37,2%

S	2789	830	22,9%	SOPORTE	22,9%
Overall Percentage	86,6%	13,4%	74,8%		

500 Prods Iniciales

Classification

Sample	Observed	Predicted			PRECISIÓN	
		NS	S	Percent Correct		
Training	NS	466	34	93,2%	PRECISIÓN	93,3%
	S	26	474	94,8%	SOPORTE	94,8%
	Overall Percentage	49,2%	50,8%	94,0%		
Test	NS	11806	987	92,3%	PRECISIÓN	33,2%
	S	2928	491	14,4%	SOPORTE	14,4%
	Overall Percentage	90,9%	9,1%	75,9%		

1000 Prods Iniciales

Classification

Sample	Observed	Predicted			PRECISIÓN	
		NS	S	Percent Correct		
Training	NS	883	117	88,3%	PRECISIÓN	89,1%
	S	40	960	96,0%	SOPORTE	96,0%
	Overall Percentage	46,2%	53,9%	92,2%		
Test	NS	5710	6583	46,4%	PRECISIÓN	27,7%
	S	398	2521	86,4%	SOPORTE	86,4%
	Overall Percentage	40,2%	59,8%	54,1%		

ANEXO 6: Ejemplo Cuadro Análisis PCP

		Número de Productos					
		Lift-Presencias					
		0	2	4	6	8	10
Presencia-Clientes	0%	6743	5495	3547	1969	1110	643
	1%	3911	3535	2403	1314	684	332
	2%	2555	2382	1692	911	440	190
	3%	1567	1480	1116	571	242	95
	4%	1191	1131	872	449	183	67
	5%	846	811	651	325	115	40
	6%	696	666	541	270	88	28
	7%	530	509	413	204	65	21
	8%	463	446	364	183	55	18
	9%	355	345	285	137	42	12
	10%	317	308	255	123	37	12
	11%	268	261	215	101	32	8
	12%	240	233	196	90	28	6
	13%	200	193	163	76	22	5
	14%	184	178	150	66	19	3
	15%	169	163	138	58	17	3
	16%	161	156	133	55	17	3
	17%	144	140	119	47	12	2
	18%	131	128	112	41	10	2
	19%	119	116	101	38	10	2
	20%	109	106	92	35	10	2

Tabla 1: Productos encontrados con prueba A

		Efectividad					
		Lift-Presencias					
		0	2	4	6	8	10
Presencia-Clientes	0%	26%	29%	35%	36%	34%	32%
	1%	32%	34%	41%	42%	41%	41%
	2%	37%	39%	46%	47%	47%	46%
	3%	43%	45%	52%	53%	55%	57%
	4%	46%	48%	54%	57%	58%	67%
	5%	50%	51%	56%	59%	58%	60%
	6%	52%	54%	59%	65%	67%	75%
	7%	57%	58%	64%	70%	71%	76%
	8%	59%	61%	66%	71%	73%	78%
	9%	63%	64%	69%	73%	76%	83%
	10%	63%	64%	69%	72%	73%	83%
	11%	64%	65%	71%	73%	75%	88%
	12%	65%	66%	71%	71%	71%	83%
	13%	71%	73%	78%	78%	77%	80%
	14%	72%	74%	79%	79%	79%	100%
	15%	73%	75%	80%	81%	82%	100%
	16%	76%	78%	81%	82%	82%	100%
	17%	76%	78%	82%	81%	83%	100%
	18%	76%	78%	81%	80%	80%	100%
	19%	78%	80%	84%	84%	80%	100%
	20%	77%	79%	84%	86%	80%	100%

Tabla 2: Efectividad de Productos encontrados con prueba A

Las tablas contienen el número de clientes y la efectividad al aplicar dos filtros simultáneamente, por ejemplo: En la Tabla 1, al aplicar los filtros presencia clientes mayor o igual a 1% y Lift Presencias mayor o igual a 2 se obtienen 3.535 productos. De la tabla 2 se puede ver que tales filtros conllevan a una efectividad de 34%.

ANEXO 7: Comentarios Pruebas Modelo PCP

Prueba	Comentarios
A	Descartado debido a que los clientes seleccionados dificultan la identificación de los productos saludables.
B	Descartado debido a que, utilizando los criterios, se seleccionan más de 4mil clientes.
C	Descartado debido a que se encontraron muy pocos productos en la primera iteración.
D	Descartado debido a que se encontraron muy pocos productos en la primera iteración.
E	Descartado debido a que se disminuyó mucho el nivel de efectividad.
F	Resultados relativamente buenos, no se siguió adelante porque se estaban seleccionando demasiados clientes para la tercera iteración.
G	Descartado debido a que, utilizando los criterios, se seleccionan más de 4mil clientes.
H	Entrega buenos resultados, pero se empieza a disminuir mucho en efectividad.
I	Entrega buenos resultados, se detuvieron las iteraciones debido a que en la tercera iteración se encontraron pocos productos y con baja efectividad.
J	Descartado debido a que se encuentran pocos productos y la efectividad sigue disminuyendo.
K	El mejor resultados, no se siguió debido a los bajos niveles de efectividad en las últimas iteraciones.
L	Buen resultado, no se siguió debido a los bajos niveles de efectividad en las últimas iteraciones.

ANEXO 8: Varianza Explicada por Factores

ACP1

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,018	19,132	19,132	4,018	19,132	19,132	3,294	15,686	15,686
2	3,300	15,712	34,844	3,300	15,712	34,844	2,680	12,761	28,447
3	2,100	10,001	44,845	2,100	10,001	44,845	2,176	10,360	38,808
4	1,838	8,753	53,598	1,838	8,753	53,598	1,952	9,295	48,102
5	1,386	6,602	60,200	1,386	6,602	60,200	1,692	8,058	56,161
6	1,166	5,554	65,754	1,166	5,554	65,754	1,635	7,784	63,945
7	1,139	5,426	71,180	1,139	5,426	71,180	1,519	7,235	71,180
8	,981	4,671	75,851						
9	,915	4,357	80,208						
10	,848	4,039	84,247						
11	,790	3,761	88,008						
12	,715	3,405	91,413						
13	,612	2,912	94,325						
14	,458	2,181	96,506						
15	,319	1,520	98,027						
16	,225	1,072	99,099						
17	,189	,901	100,000						
18	,000	,000	100,000						
19	,000	,000	100,000						
20	,000	,000	100,000						
21	,000	,000	100,000						

ACP2

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Saludable	714,115	5	,585	8563	1221,370	,000
Vegetariano	335,340	5	,803	8563	417,413	,000
Cocinable	128,416	5	,723	8563	177,594	,000
Carnes	834,233	5	,505	8563	1652,822	,000
Infantil	236,087	5	,827	8563	285,532	,000
Barato	524,115	5	,673	8563	779,042	,000
Gourmet	813,716	5	,449	8563	1812,949	,000

ANEXO 9: Promedio Dimensiones por Clusters

Cluster	Tradicional	Express	Familiar	Parrillero	Gourmet	Saludables	Total general
Extranjero	0,10	-0,90	-0,66	-0,54	3,23	-0,38	-0,05
Exótico	0,05	-0,78	-0,61	-0,22	3,22	-0,29	0,00
Instantáneo	-0,10	1,02	-0,03	-0,02	0,20	-0,24	0,00
Indiferente	0,00	-0,22	0,03	-0,28	-0,03	0,10	0,00
Preparación	0,22	-0,55	-0,11	1,71	-0,20	-0,18	0,00
Lenta Preparación	-0,09	-0,26	0,84	-0,34	-0,09	-0,22	0,00
Niño	-0,21	-0,78	-0,51	-0,41	-0,47	0,93	0,01
Chileno	0,04	0,24	0,28	0,09	0,22	-0,38	-0,01
No Saludable	-0,10	0,42	0,26	0,13	-0,08	-0,17	0,00
Indiferente	0,04	-0,45	-0,38	-0,15	-0,11	0,39	0,01
Saludable	-0,04	0,04	-0,19	-0,05	0,05	0,15	0,01
Muy Saludable	0,20	0,15	0,10	0,15	0,07	-0,33	0,00
Muy Percible	-0,02	-0,25	0,55	0,05	-0,21	-0,22	-0,02
Indiferente Percible	0,16	0,01	0,01	-0,11	0,33	-0,42	-0,06
No Percible	0,51	-0,90	-0,60	2,80	0,00	-0,28	0,01
Fiesta	-0,06	-0,64	-0,48	-0,30	-0,18	0,65	0,01
Congelado	0,19	0,07	-0,28	1,50	-0,03	-0,26	0,00
Verde	0,23	-0,35	-0,44	-0,01	0,95	-0,11	0,00
Carnes	0,05	0,73	-0,25	-0,01	-0,05	-0,12	0,00
Precio Alto-Alto	0,05	-0,33	0,12	0,05	-0,08	0,01	0,00
Precio Alto-Bajo	-0,15	-0,31	0,25	-0,03	-0,15	0,14	0,00
Precio Bajo-Alto							
Precio Bajo-Bajo							

Promedio de los Lifts de las dimensiones para cada segmento.

ANEXO 10: Distribución Gastos de Clientes en Atributos

Cluster	Tradicional	Express	Familiar	Parrillero	Gourmet	Saludables	Promedio General
Extranjero	1%	0%	0%	0%	3%	0%	1%
Exótico	2%	0%	1%	1%	8%	1%	2%
Instantáneo	19%	42%	20%	21%	25%	16%	21%
Indiferente							
Preparación	67%	52%	70%	49%	66%	75%	67%
Lenta Preparación	14%	5%	10%	31%	9%	9%	11%
Niño	14%	11%	28%	10%	14%	12%	15%
Chileno	1%	0%	1%	1%	1%	3%	2%
No Saludable							
Indiferente	32%	38%	40%	34%	38%	19%	31%
Saludable	25%	39%	34%	31%	25%	23%	27%
Muy Saludable	43%	23%	26%	35%	37%	58%	42%
Muy Percible							
Indiferente	61%	67%	52%	61%	67%	74%	65%
Percible	19%	18%	18%	18%	17%	11%	16%
No Percible	19%	15%	31%	21%	16%	16%	20%
Fiesta	9%	7%	7%	7%	10%	4%	7%
Congelado	3%	0%	1%	7%	2%	1%	2%
Verde	22%	8%	12%	16%	19%	38%	23%
Carnes	13%	12%	8%	28%	11%	8%	11%
Precio Alto-Alto	11%	6%	5%	9%	18%	8%	9%
Precio Alto-Bajo	32%	53%	23%	30%	29%	27%	31%
Precio Bajo-Alto	29%	19%	31%	29%	25%	28%	28%
Precio Bajo-Bajo	28%	22%	41%	32%	28%	37%	33%
Número de Clientes	2428	827	1508	439	713	2654	8569
% Clientes	28,30%	9,70%	17,60%	5,10%	8,30%	31,00%	100,00%

Distribución de Compras por Dimensión Fuente: Elaboración Propia

ANEXO 11: Información Demográfica de los Segmentos

Nº Segmento	Todos los Clientes		Clientes con Información				% Clientes con Info	
	Numero Clientes	% Clientes	Número Clientes	% Clientes	% Hombres	% Mujeres		Promedio Edad
1 Tradicional	2428	28%	1127	29%	29%	71%	47	46%
2 Express	827	10%	323	8%	47%	53%	46	39%
3 Familiar	1508	18%	654	17%	34%	66%	44	43%
4 Parrillero	439	5%	211	5%	38%	62%	45	48%
5 Gourmet	713	8%	311	8%	36%	64%	47	44%
6 Saludables	2654	31%	1240	32%	19%	81%	50	47%

TOTAL	8569	3866	29%	71%	46	45%
-------	------	------	-----	-----	----	-----

ANEXO 12: FUZZY C-MEANS

Como alternativa a al método K-Means, es posible realizar la segmentación utilizando métodos difusos, como el método Fuzzy C-Means, el cual permite conocer la pertenencia de cada cliente a cada segmento pero en probabilidad.

Para la utilización de este método, lo primero es definir el Índice de Difusividad, el cual indica el grado determinístico del modelo, siendo 1 el valor más determinístico, es decir, el modelo ajustado con un índice igual o cercano a 1 entregará probabilidades 1 o 0 de pertenencia a los segmentos para cada cliente, por el contrario valores mayores a 2 entregarán probabilidades entre 0 y 1, teniendo el caso extremo de entregar la misma probabilidad para cada segmento.

A continuación, se muestra la siguiente figura, que grafica el número de clientes según su máxima probabilidad de pertenencia a un segmento.

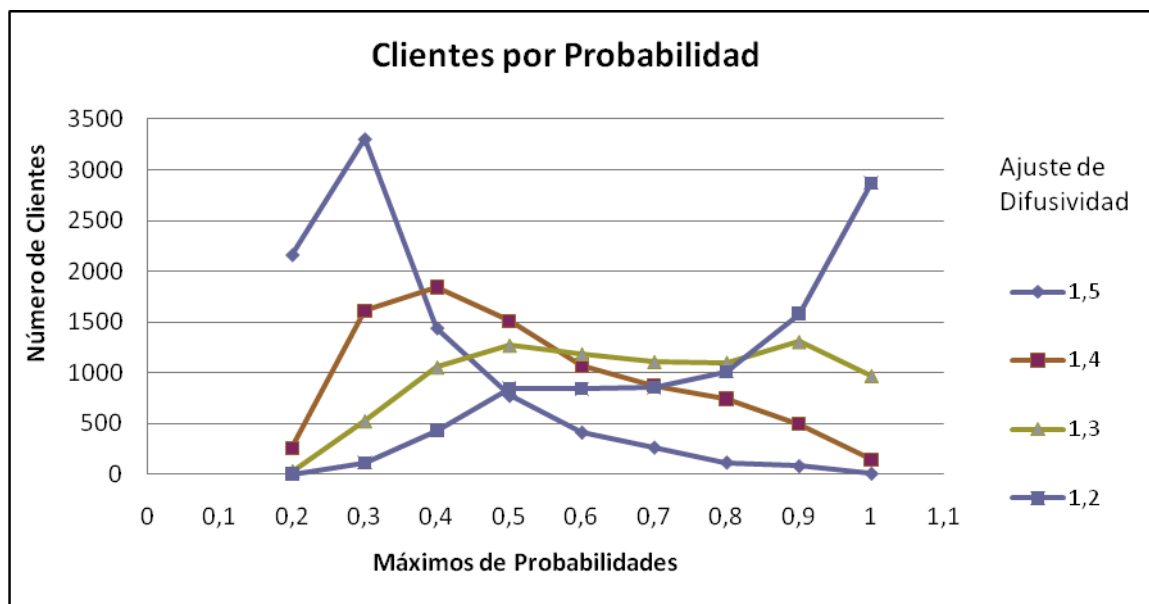


Figura 35: "Clientes por Probabilidad"; Fuente: Elaboración Propia

Del la figura 35 es posible observar que a mayor Índice de Difusividad, menores son los clientes con probabilidad de pertenencia a un segmento igual a 1 y por el contrario muchos son los clientes con probabilidades 0,2 y 0,3.

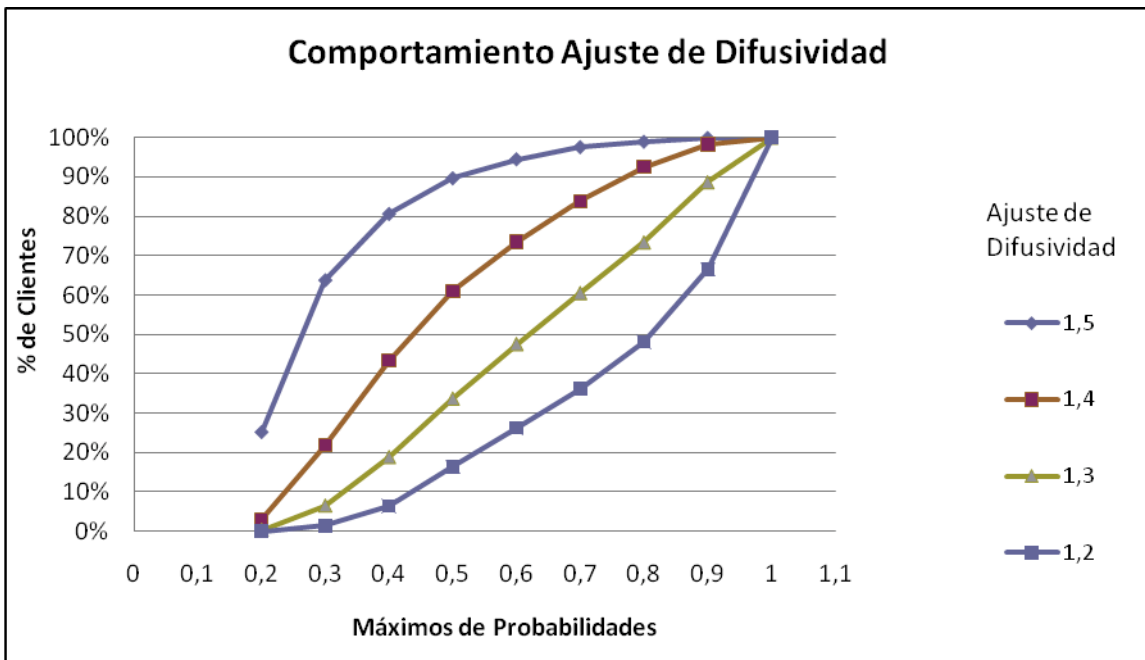


Figura 36: "Comportamiento Ajuste de Difusividad"; Fuente: Elaboración Propia

Se observa en la figura 36 que los modelos que tienen mayores índices de difusividad alcanzan casi el 100% de los clientes más rápido, esto debido a que a mayor índice de difusividad menores son los clientes con probabilidades de pertenencia cercanas a 1.

El segundo paso es identificar los clusters, en este caso con respecto a los segmentos encontrados vía K-Means, para ello se tomó el modelo con un índice de difusividad igual a 1.2, la siguiente tabla resumen presenta los resultados.

La tabla a continuación, muestra el promedio de cada dimensión para cada segmento.

	Segmentos						Total general
	FAMILIAR	TRADICIONAL	SALUDABLE	PARRILLERO	GOURMET	EXPRESS	
Numero Clientes	1091	1473	2422	674	2297	612	8569
Porcentaje de Clientes	13%	17%	28%	8%	27%	7%	100%
Saludable	-0,80	-0,05	0,99	-0,58	-0,22	-0,83	0,00
Vegetariano	-0,14	-0,58	0,38	-1,52	0,23	0,91	0,00
Cocinable	-0,19	0,68	-0,27	-0,79	-0,31	1,29	-0,04
Carnes	-0,10	0,07	-0,13	-0,25	0,28	-0,15	0,01
Infantil	-0,32	1,07	-0,18	-0,95	0,21	-0,88	0,01
Barato	1,39	0,08	0,18	-1,10	-0,32	-0,95	0,00
Gourmet	-0,18	-0,27	-0,11	-0,44	0,50	-0,21	-0,02
Extranjero	-0,40	-0,07	-0,27	-0,74	0,59	-0,10	-0,05
Exótico	0,19	-0,49	-0,16	-0,55	0,65	-0,41	0,00
Instantáneo	0,10	-0,09	-0,24	1,30	-0,01	-0,31	0,00
Lenta Prep.	-0,16	0,41	-0,06	-0,37	0,08	-0,26	0,00
Niño	0,49	0,50	-0,36	-0,35	0,00	-0,24	0,00
No Saludable	0,25	-0,03	-0,41	0,11	0,07	0,69	-0,01
Saludable	-0,26	-0,19	0,43	-0,44	0,02	-0,32	0,01
Vencimiento Rápido	-0,07	-0,15	0,17	0,17	0,02	-0,36	0,01
No Perecible	-0,02	0,69	-0,28	-0,48	-0,20	0,51	-0,02
Verde	-0,31	-0,31	0,73	-0,63	-0,08	-0,52	0,01
Fiesta	-0,07	-0,30	-0,41	-0,58	-0,09	2,64	-0,06
Animal	-0,07	-0,04	-0,14	0,41	0,14	-0,13	0,00
Precio AltoAlto	-0,31	-0,26	-0,06	-0,25	0,42	0,05	0,00
Precio AltoBajo	-0,28	-0,17	-0,11	0,70	0,04	0,44	0,00
Precio BajoBajo	0,50	-0,07	0,17	-0,29	-0,21	-0,28	0,00

Figura 37: "Segmentos Fuzzy C-Means"; Fuente: Elaboración Propia

Realizado lo anterior, ahora es posible responder preguntas tales como:

- ¿Cuántos clientes son Gourmet con un 70% de probabilidad?
- ¿Qué clientes son Gourmet y Familiares a la vez?
- ¿Cuáles son los clientes más puros de un segmento?
- ¿A qué probabilidad se está dispuesto a llegar para focalizar una promoción?
- ¿Cuál es el número máximo de clientes saludables que utilizan el local?