



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**METODOLOGÍA PARA EL REDISEÑO DEL MIX DE PRODUCTOS EN
TIENDAS MINORISTAS UTILIZANDO REGLAS DE ASOCIACIÓN**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

OCTAVIO CORNEJO VEGA

**PROFESOR GUÍA:
RICARDO MONTOYA MOREIRA**

**MIEMBROS DE LA COMISIÓN:
MANUEL REYES JARA
NANCY VERONICA LARA RODRIGUEZ**

**SANTIAGO DE CHILE
NOVIEMBRE 2011**

Tabla de contenido

I.	INTRODUCCIÓN	1
1.1	Historia de la empresa	3
1.2	Antecedentes	4
1.2.1	CMR	4
1.2.2	Customer Business Intelligence	5
1.2.3	Mix de productos	6
II.	DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN	8
2.1	El problema de selección de productos	8
2.2	Descripción del proyecto	8
2.3	Justificación	9
2.3.1	Necesidades de los clientes	9
2.3.2	El valor de tomar buenas decisiones	9
2.3.3	Conocimiento del cliente	10
III.	OBJETIVOS	11
3.1	Objetivo general	11
3.2	Objetivos específicos	11
IV.	METODOLOGÍA	11
4.1	Revisión bibliográfica	11
4.2	Análisis de la situación actual	12
4.2.1	Composición de la cadena	12
4.2.2	Identificación de información disponible	12
4.2.3	Consolidación de la información	12
4.2.4	Caracterización de la tienda y su mix	12
4.2.5	Canastas frecuentes	13
4.2.6	Selección del surtido	13
4.2.7	Resultados y recomendaciones	13
V.	MARCO CONCEPTUAL	13
5.1	Data Mining	13
5.2	Escalamiento multidimensional	14
5.3	Clasificación Jerárquica	14
5.4	Análisis de canasta	15
5.4.1	Indicadores de asociación	15
5.5	PROFSET	17
5.5.1	¿Qué incluye y deja de incluir PROFSET?	17
5.5.2	Especificaciones del modelo	18
5.5.3	Modelo de optimización PROFSET	24
VI.	MARCO METODOLÓGICO	26
6.1	Estandarización de las transacciones	26

6.2	Base de datos y herramientas computacionales	27
6.3	Selector de información	28
6.4	Métodos estadísticos y exploratorios	28
6.5	Análisis de Canasta	29
6.6	Implementación de PROFSET	31
6.7	Validación de la metodología	31
VII.	ALCANCES	32
VIII.	ANÁLISIS DEL MIX DE PRODUCTOS (AMP)	33
8.1	Etapas del AMP	33
8.2	Descubrir intenciones de compra	34
8.2.1	Selección de Data	35
8.2.2	Escalamiento multidimensional	39
8.2.3	Análisis Jerárquico	44
8.3	Determinar Frequent Sets y valorizarlos	46
8.3.1	Encontrar Frequent Sets	46
8.3.2	Valorización Frequent Sets	48
8.4	Selección de producto en base a sus correlaciones	49
8.4.1	PROFSET	49
IX.	APLICACIÓN DE LA METODOLOGÍA	55
9.1	Descripción de la tienda	55
9.2	Información a utilizar	56
9.3	Descubrir las intenciones de compra	59
9.3.1	Selección de Data	59
9.3.2	Análisis Multidimensional	60
9.3.3	Análisis Jerárquico	61
9.3.4	Caracterización de las intenciones de compra	62
9.4	Determinar Frequent Sets y valorizarlos	67
9.4.1	Selección de la intención de compra	68
9.4.2	Agrupación de productos	68
9.4.3	Canastas frecuentes	70
9.4.4	Valorizar las canastas frecuentes	71
9.5	Selección de productos en base a sus correlaciones	72
9.5.1	PROFSET	73
9.5.2	El verdadero valor de los productos	73
9.5.3	Productos candidatos a ser eliminados del mix	76
9.5.4	Productos candidatos a ser incluidos en el mix	76
X.	COMENTARIOS Y CONCLUSIONES	79
XI.	PROPUESTA PARA TRABAJOS FUTUROS	82
XII.	BIBLIOGRAFÍA	84

<i>XIII. ANEXOS</i>	86
Anexo A: Script para calcular las correlaciones entre productos	86
Anexo B: Diversas Categorías de Tottus Kennedy	88
Anexo B: Categorías y Sub-Categorías en Food	89
Anexo C: Matriz de distancia entre sub categorías	90
Anexo D: Resultado del análisis multidimensional	91
Anexo E: Indicadores de las diversas categorías del estudio	92
Anexo F: Modelo y datos (resumido) para AMPL.	93
Anexo G: Productos seleccionados por PROFSET	95

I. INTRODUCCIÓN

S.A.C.I. Falabella es una sociedad anónima abierta, que tiene como principal negocio la venta al detalle de: vestuario con sus tiendas Falabella, accesorios y productos para el hogar a través de Sodimac, *malls* y tiendas de especialidad, así como alimentos a través de hipermercados como Tottus. Además, se ha desarrollado el área de Servicios Financieros (emisión de tarjetas de crédito, corretaje de seguros, banco y agencia de viajes) y la manufactura de textiles¹.

CMR es la empresa financiera de Falabella encargada de la administración de la tarjeta de crédito del “holding”². Cuya tarjeta es utilizable en todas las tiendas del grupo Falabella. Además, gracias a una asociación con Visa (en septiembre del 2009) se creó la nueva CMR Visa, la cual permite que los clientes de CMR compren en todos los locales establecidos. Permitiendo a CMR poder asociar las compras realizadas con el cliente que la realiza, siempre y cuando este pague con la tarjeta del holding.

Debido a la gran envergadura del negocio y a la diversidad de estos, a modo de ejemplo, si se consideran solamente las tiendas de Falabella, Sodimac y Tottus de Chile, se genera mensualmente más de 30.000.000 de transacciones y se reportan ingresos por más de 80.000 millones de pesos mensuales, y solamente en las tiendas de Falabella más de un millón de personas compra al menos una vez con su tarjeta CMR.

La competitividad de la industria y la gran cantidad de oferentes en los distintos rubros ha generado que cada vez sea más difícil satisfacer al complejo cliente que visita las tiendas. Por lo que distintas acciones se han tomado con el fin de mantener la posición dominante que tiene el grupo Falabella, dentro de las cuales, últimamente destaca la contratación de una importante consultora de renombre mundial quienes tuvieron como tarea, entre otras, segmentar a los clientes de Falabella y además sentaron las bases para que dicho trabajo sea replicado tanto en Sodimac como en Tottus. No hay que olvidar que segmentar a los clientes por segmentar reporta cero utilidades, lo que se necesita ahora es gestionar y tomar decisiones utilizando el conocimiento que se generó tanto durante el proceso de segmentar a los clientes de tiendas Falabella, como también, el conocimiento generado finalmente. Extrayendo lo esencial de este conocimiento generado, se desprende la importancia de entender y analizar a los clientes tomando en consideración las diferencias que estos presentan.

Antiguamente los dueños de retail veían que el negocio era comprar productos y ponerlos a disposición del público en sus tiendas. Si estos productos se vendían, había que ordenar más de estos, de lo contrario, eran eliminados del mix. Blischok (1995) describe a los retailers según este paradigma como “product-orientes”, donde se apoyaban de talentosos comerciantes que con sus habilidades distinguían si un producto iba a ser un éxito de ventas. Pero hoy en día no se puede ser exitoso solo con esa forma de operar. Tal como señala Blischok los retailers deben tender a ser “customer-oriented” y para orientarse de una buena forma al cliente, deben manejar de forma excelente el conocimiento que se va obteniendo del cliente. Este conocimiento se obtiene entendiendo e comportamiento de compra de los clientes a través de sus compras, es decir, *análisis de canasta de compra*.

La complejidad de estudiar lo previamente descrito, surge por un lado debido al explosivo aumento en el número de productos que se ofrecen en cada sala y por otro lado debido a la gran

¹ Para más información sobre Falabella visitar la página web <http://www.falabella.com>

² Conglomerado de empresas que dependen todas de una misma empresa matriz.

cantidad y diversidad de los consumidores que compran constantemente en las tiendas del grupo Falabella. Clientes quienes esperan encontrar en dichas tiendas exactamente lo que buscan.

Junto con lo anterior, los negocios han ido presenciando un aumento cada vez más acelerado de la cantidad de información que se genera día a día, haciendo cada vez más difícil el obtener información útil, ya que los métodos tradicionales no pueden lidiar con la gran cantidad de información. Es aquí donde las técnicas de KDD³ entran en juego.

Dentro de las técnicas de KDD más populares está la extracción de reglas de asociación desde grandes bases de datos. Muchos investigadores han trabajado en pos de mejorar estas técnicas creando algoritmos cada vez más eficientes. El problema es que estas técnicas de asociación por si solas no permiten solucionar problemas del negocio. Ya en el año 1998 (Kleinberg, Papadimitriou, & Raghavan) se señalaba que la extracción de reglas de asociación para soportar decisiones de negocio, solo puede ser abordada teniendo en consideración una mirada microeconómica de la empresa. Lo que significa que los patrones obtenidos en la información solo son útiles si pueden ser usadas en la toma de decisiones de la empresa para aumentar las utilidades de esta.

Luego este trabajo persigue la creación de una metodología de análisis del mix de productos (llamada en adelante AMP, análisis mix de productos) que permita evaluar el desempeño de la categoría y gestionarla con una mirada microeconómica del negocio, al introducir en el estudio un modelo de optimización entera llamado PROFSET (Brijs, Swinnen, Vanhoof, & Wets, 2004) que utiliza sets de productos frecuentes para determinar la selección óptima de productos con foco en maximizar las utilidades. Y permitiendo así la detección y evaluación de oportunidades de quitar o agregar productos del mix.

El AMP que se creará, constará con tres etapas principales. En primer lugar se agruparán los productos según distintas necesidades que ellos satisfacen. Luego dentro de estos grupos de productos se determinará cual analizar y finalmente se proseguirá a realizar el modelo de PROFSET con estos productos.

Cabe recordar que debido al dinamismo de la industria, este análisis (AMP) debe ser fácilmente replicable para así poder estudiar las diversas categorías y en los distintos negocios del grupo Falabella⁴. Por lo que será necesario junto con la creación de la metodología el diseño de herramientas computacionales que agilicen y estandaricen dicho proceso.

En la primera parte del trabajo se introducirá al lector en el problema en cuestión, luego se definirán las partes que componen al AMP, con especial énfasis en los métodos exploratorios que se utilizan a lo largo del análisis. Junto con lo anterior, se explicará el modelo de PROFSET. Para finalizar el trabajo poniendo a prueba la metodología en una categoría en particular del supermercado Tottus.

³ KDD: “*Knowledge discovery in databases*”, son técnicas que permiten encontrar patrones útiles para el negocio desde las bases de datos.

⁴ La metodología y las herramientas computacionales que se realicen solo tomarán en cuenta los tres principales retailers del grupo (Tiendas por departamento Falabella, Sodimac y Tottus)

1.1 Historia de la empresa

Los orígenes de Falabella se remontan al año 1889 cuando Salvatore Falabella inauguró la primera sastrería en Chile. Casi 50 años tuvieron que pasar para que Alberto Solari se integrara a la compañía y le diera un renovador cambio. Integrando artículos de moda y nuevos productos en los distintos puntos de venta. Transformando así a Falabella en una importante tienda de vestuario.

El éxito alcanzado por la nueva estrategia, les permitió incorporar nuevos productos orientados al hogar, y de esta forma el año 1958 poder transformarse en una tienda por departamentos.

Cabe destacar como hito importante, y como respuesta a una demanda creciente por créditos, que es el año 1980 Falabella creó su propia tarjeta de crédito, CMR.

Ilustración 1: Tarjeta de crédito CMR Falabella



Fuente: www.falabella.cl

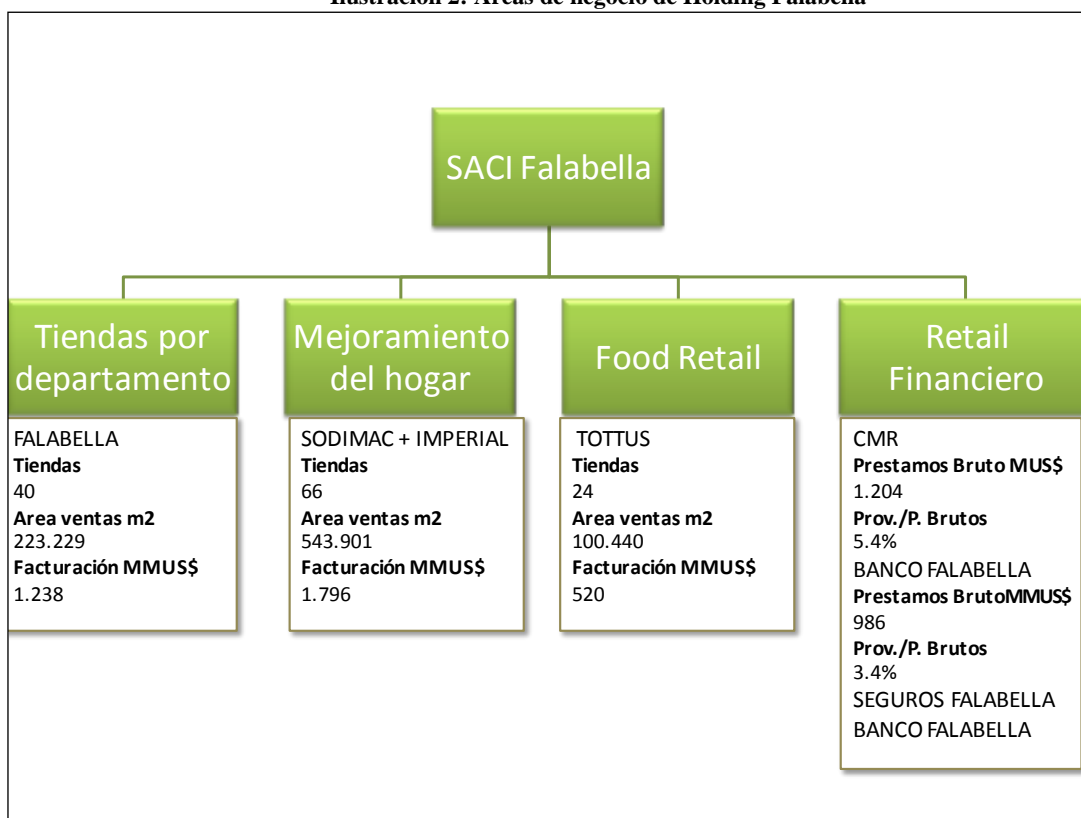
El año 1990 Falabella ingresó al negocio de los centros comerciales, con una estrategia de posicionamiento enfocada principalmente a los sectores emergentes de la economía. Inaugurando así centros como Mall Plaza Vespucio en La Florida. Años más tarde la empresa dio un gran paso al empezar su internacionalización, abriendo su primera tienda en Mendoza, Argentina. Luego al adquirir la cadena Saga, empezó sus operaciones en Lima Perú.

Continuando con su plan de aumentar sus líneas de negocio Falabella ingresó al área de mejoramiento del hogar al asociarse con la cadena norteamericana líder mundial en el rubro, *The Home Depot*. Junto con lo anterior, ese mismo año, Falabella aumentó su oferta de servicios con la creación de Viajes Falabella, Seguros Falabella y al año siguiente, se creó Banco Falabella. Pocos años más tarde, Falabella entró al negocio de los hipermercados con la creación de Tottus. Y de esta forma expandió considerablemente la cantidad de áreas de negocio en las cuales Falabella está presente.

En los últimos años Falabella ha ido aumentando su presencia en sus distintas áreas de negocio, fortaleciendo su participación en Perú, Colombia y Argentina.

Como se ha podido apreciar, existe una gran cantidad de negocios en los cuales Falabella es actor. A continuación se ilustran las principales áreas de negocio de Falabella. (Ver Ilustración 2: Áreas de negocio de Holding Falabella)

Ilustración 2: Áreas de negocio de Holding Falabella



Fuente: Elaboración propia.

1.2 Antecedentes

1.2.1 CMR

CMR es una de las empresas financieras del grupo Falabella y está encargada principalmente de la administración de la tarjeta de crédito CMR, la tarjeta CMR Falabella nació el año 1980, como respuesta a las tarjetas de la competencia, Almacenes Paris y Ripley, y se presentó como un nuevo medio de pago en los servicios de Falabella.

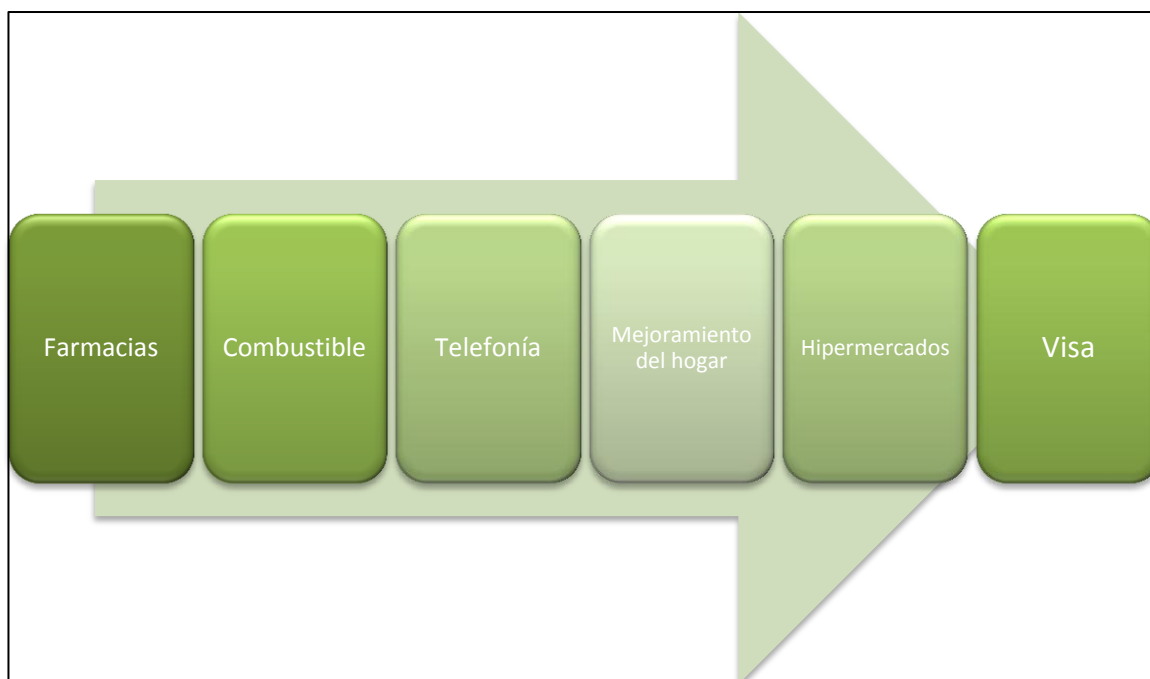
En la década del 90 Falabella se internacionalizó, con lo cual CMR comenzó sus operaciones en Argentina y luego en Perú.

Teniendo la mayor participación de mercado, CMR centró sus objetivos en aumentar el alcance de la tarjeta. Para lograrlo, comienza su plan de alianzas asociándose con Farmacias Ahumada (adquiriendo además el 20% de la propiedad) y de esta forma permitiendo que la tarjeta se pueda utilizar en este negocio, pasando la tarjeta a ser la primera tarjeta de tiendas por departamento aceptada en el mercado de las Farmacias en Chile. Años más tarde, se crea una alianza con Copec, Movistar y Entel, lo que permitió usar las tarjetas en estos negocios también.

Continuando con su plan de expansión, y gracias a las fusiones que realiza Falabella, la tarjeta CMR fue entrando a un mayor número de negocios, y por consiguiente adquiriendo mayor información, tanto de sus distintos negocios como de sus clientes.

A todo lo anterior, se le suma el acuerdo firmado este año (2009) entre Visa y CMR. El cual gracias a la creación de una nueva tarjeta CMR Visa, permite que los clientes compren en cualquier comercio que acepte tarjetas Visa. En la ilustración 3 se resume las alianzas que ha ido realizando CMR con objetivo de expandir su campo de acción, y de esta forma permitir que sus clientes utilicen la tarjeta en los distintos negocios. Esto último permite tener una visión transversal en las compras de un cliente, ya que se tiene información de las compras de clientes en alimentación, reparación del hogar, vestuario, entre otras. Dando la posibilidad de no solo entender a cabalidad a un cliente sino que además poder estimar de mejor forma el valor de este para el negocio y su posible potencial para este.

Ilustración 3: Cuadro resumen de alianzas de la tarjeta CMR.



Fuente: Elaboración propia

1.2.2 Customer Business Intelligence

Customer Business Intelligence (en adelante CBI) es el área encargada de administrar y obtener conocimiento de la información que se obtiene de los distintos negocios, la cual incluye principalmente la data transaccional (de los clientes cuando utilizan la tarjeta CMR) y la información personal de los clientes. CBI tiene tareas como, por ejemplo, evaluar el desempeño de las campañas publicitarias, segmentar a los clientes, realizar campañas de marketing directo.

Las principales áreas de negocio del “holding” Falabella (Sodimac, Falabella, CMR, Banco Falabella) cuentan con su propia área de CBI y, a pesar de que Falabella trabaja como “holding”, existen problemas de interés entre estas áreas lo que dificulta que se comparta la información que cada cual obtiene en su negocio.

Es por esto que los estudios que realizan las distintas áreas de CBI son en base a la información solo de su negocio y se ven imposibilitados de realizar estudios transversales entre

los distintos negocios del “holding”, los cuales serían de gran valor para el “holding”, debido a que incluirían una visión más integra de todos los negocios.

Las distintas áreas de CBI han ido aumentando su complejidad, tanto en número como en información a manejar, debido al aumento de información que surge naturalmente a causa del progreso de las tecnologías y al crecimiento del negocio. El aumento de complejidad del área se ha visto impulsada por el crecimiento global de la informática y la información, que ha llevado a que los altos directivos apuesten por obtener una ventaja competitiva mediante la utilización de los datos, para así obtener conocimiento. Para ejemplificar la situación, en la tabla 1 se muestra cómo ha cambiado el área de CBI en CMR en los últimos 10 años.

Tabla 1: Evolución del área de inteligencia de cliente de CMR.

	2000	2010
Número de tarjetas	1,5 MM	+2MM
Cantidad empleados	4	8
Origen de la información	Solo Falabella	Variadas líneas de negocio
Campañas	+ 2 campañas	+5 campañas simultaneas
Información mensual	12 millones líneas	60 millones líneas

Fuente: Ejecutivo CMR.

Actualmente en CBI se está comenzando a estandarizar procesos de evaluación, así como también se busca unificar la información de los distintos negocios. Con el fin de permitir que los estudios sean fácilmente replicables en los distintos negocios.

1.2.3 Mix de productos

El constante y rápido dinamismo del mercado obliga a la industria del *retail* a estar constantemente actualizando y monitoreando su mix de productos. Se pudiese llegar a pensar que esto último es una preocupación exclusiva de las tiendas Falabella causado por la rapidez actual de la moda, lo que lo obligaría a estar constantemente actualizando su mix de productos. Pero esta urgencia por mantener actualizado su mix de productos también ocurre en las tiendas de Tottus y Sodimac, donde por ejemplo en Tottus deben estar al día con las últimas tendencias en alimentación. Solamente en Tottus se vende mensualmente más de 20.000 productos distintos, dentro de los cuales 1.000 son que no estaban en el mix de productos el mes anterior. Debido a la gran cantidad de productos distintos que se manejan en estas tiendas⁵, y para cumplir de buena forma la tarea de entregar un preciso mix de productos, es que los negocios cuentan con Gerentes de Línea encargados de gestionar una porción del mix total de productos, los cuales poseen características en común. Son ellos los principales responsables de seleccionar que colocar y que eliminar de las góndolas.

⁵ En las tiendas de Falabella se venden más de 100.000 productos distintos cada mes.

Los Gerentes luchan constantemente por aumentar el *Share of wallet*⁶ de sus clientes y para satisfacer de mejor forma las necesidades de ellos. Para conseguir se han visto forzados a aumentar cada vez más la variedad de productos que ofrecen, pero existe un problema: la limitación de espacio que presentan las góndolas los obliga a seleccionar cuidadosamente los productos a ofertar.

Se cree que enfocarse en el cliente es necesario para enfrentar de buena forma estas necesidades, y que las técnicas de análisis de canasta en su conjunto pueden ayudar a esta tarea, al aportar conocimiento valioso sobre el comportamiento de compra de los diversos clientes. El que puede ser utilizado por ejemplo para enfocar de forma eficiente las diversas campañas de marketing.

Por otro lado, estudios han demostrado que existen fuertes correlaciones entre los diversos productos que se ofertan en las tiendas. A simple vista se puede observar que en promedio un cliente compra nueve productos distintos⁷ en cada visita. Es por esto que los Gerentes de Tienda buscan seleccionar para el mix a productos que: sus ventas no sean a costa de la disminución de la demanda de otros productos (“canibalismo”) y que por el contrario presenten un alto grado de complementariedad entre los distintos productos de la tienda.

Desde un punto de vista metodológico, este trabajo aprovecha técnicas de minería de datos (reglas de asociación) para generar una metodología que utiliza la información de las transacciones de los diversos clientes para descubrir sus intenciones de compra y luego valorizar los productos en base a las correlaciones que poseen con los demás.

Cabe destacar que la metodología de este trabajo, a diferencia de la mayoría de las tesis de rediseño de mix como la de Solervicens (2009), que solo toman en consideración los efectos generados por cada producto en particular, o el trabajo de “Metodología para apoyar la toma de decisiones en surtido de supermercados” (Piña,2007) que considera la entropía de los productos pertenecientes a la categoría de productos en estudio, este estudio toma en consideración la correlación existente entre los productos de todo el retail. Para así crear una metodología que apoyándose en técnicas de minería de datos y el modelo PROFSET, evalúa el desempeño de los productos tomando en consideración el aporte de este en particular como también dentro del negocio, al considerar las interrelaciones con otros productos.

⁶ Share of Wallet, es la proporción de dinero que un cliente invierte en una determinada necesidad del 100% del dinero que este cliente invierte en esta necesidad.

⁷ Dato obtenido en base a las transacciones del mes de marzo del 2011 del supermercado Tottus de Kennedy.

II. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN

En este capítulo se describe en términos generales el proyecto explicando por qué la necesidad y oportunidad para llevarlo a cabo. Junto con lo anterior se verá el por qué CMR posee una posición privilegiada dentro de la industria para realizar este estudio y como este trabajo puede ayudar a mejorar la gestión de categorías en las diversas tiendas del holding Falabella.

2.1 El problema de selección de productos

El sueño de todo gerente de categoría es determinar el mix de productos ideal. Aun no existe ni la metodología ni los modelos que permitan conseguir este objetivo. Debido a la complejidad de los retail, esto es una tarea cada vez más difícil, a modo de validar lo anterior se puede observar la gran cantidad de productos distintos que maneja el supermercado Tottus, más de 40.000 productos distintos. A pesar de ser una tarea difícil se sabe, desde la literatura del marketing (Van der Ster & Van Wissen, 1993), que el mix de productos óptimos debe cumplir dos criterios importantes.

Los productos, en términos cualitativos, que se venden en las tiendas deben ser consistentes con la imagen que este desea reflejar, lo que le permite diferenciarse de la competencia lo que involucra aspectos como diseño, *layout*, servicios y por supuesto los productos adecuados. Dentro de los productos, los retail generalmente distinguen entre productos básicos y adicionales. Los primeros son productos que no pueden ser removidos del mix, ya que son parte indispensable de este, mientras que los segundos complementan la imagen de retail y aportan como productos complementos de los básicos, por lo que deben ser seleccionados buscando maximizar la utilidad del negocio y las correlaciones que tienen estos con los productos básicos del mix.

Como segundo criterio, debido a que los negocios buscan generar utilidades, es que el mix de productos se debe determinar tomando en cuenta una mirada cuantitativa de utilidades que genera para el retail.

Tomando en consideración los dos criterios previamente descritos, surge evidente la importancia de la interdependencia entre productos, para evaluar el desempeño de los productos. Por lo que la mirada de interrelación debe ser considerada al momento de determinar el mix de productos óptimo. Y no solo la interrelación que existe entre los productos de una misma categoría sino que también la que tienen estos con los productos de otras categorías.

2.2 Descripción del proyecto

Alineándose con la estrategia de centrarse en el cliente, y buscando mejorar el mix de productos, es necesario entregarles a los clientes una oferta de productos atractiva. Por lo cual surge indispensable en primer lugar descubrir las distintas intenciones de compra que los clientes tienen al momento de ir a cada negocio de Falabella para luego gestionar el mix de productos a ofertar, teniendo como foco la utilidad para negocio.

Teniendo claro todo lo anterior, y observando que existe una necesidad y oportunidad latente de gestionar en base al cliente, es que se creará una metodología de análisis de mix de productos (Llamada en este estudio AMP) que permita en primer lugar detectar las intenciones de

compra de los clientes, para así gestionar en base a estas intenciones de mejor forma el mix de productos.

Una vez que se tenga descubiertas las necesidades del cliente, se proseguirá a analizarlas en forma general todas ellas buscando donde pueda existir alguna oportunidad de rediseño.

La metodología además de permitir la detección de diversas oportunidades de rediseño, será una excelente herramienta para la evaluación de los diversos productos del mix, ya que se estará utilizando información transaccional actualizada. Lo que permite obtener rápidamente diversos indicadores de desempeño de los productos analizados.

Esta metodología como se esbozó recientemente, utiliza la información transaccional generada al momento que los clientes realizan sus compras en los negocios y la información que CMR posee de los diversos clientes.

Gracias a herramientas computacionales que se crearán para que AMP sea fácil de utilizar y a la vez rápidamente replicable con distintos productos y distintas categorías, es que esta metodología podrá ser utilizada tanto por analistas como por gerentes de negocios, con lo que se podrá aprovechar el gran conocimiento que poseen estos últimos del negocio. Logrando así que este trabajo apunte a la creación de una herramienta metodológicamente robusta que permita evaluar y detectar diversas oportunidades de rediseño.

2.3 Justificación

2.3.1 Necesidades de los clientes

Actualmente los negocios no buscan simplemente vender productos a sus clientes, sino más bien satisfacer necesidades de estos. Diversas son las necesidades que busca satisfacer el grupo Falabella en su conjunto, ya que posee diversos tipos de negocios y formatos. Esto presenta una gran oportunidad para el área de inteligencia de negocios ya que dispone de una amplia y variada información del comportamiento de compra de sus clientes, lo que permite realizar diversos análisis y modelos.

Ahora bien, se observa que tanto Falabella como Sodimac y Tottus han ido siguiendo la tendencia mundial de algunas grandes tiendas de ampliar cada vez más la variedad de productos distintos que ofertan en sus tiendas. Es por lo mismo que ya no nos extrañamos al observar como los grandes Supermercados dejaron hace ya bastante tiempo de ofrecer solo alimentos, y empezaron a incursionar en la venta de artículos tan fuera del rubro como vestuario.

Por lo anterior se desprende que los distintos negocios del *holding* buscan cubrir diversas y cada vez más variadas necesidades. Luego para gestionar eficazmente el mix de productos en base al cliente, es necesario descubrir cuáles son las diversas necesidades que el cliente busca satisfacer en cada negocio y en base a estas gestionarlas para generar una grata experiencia de compra en el cliente y lograr obtener la mayor utilidad en cada visita del cliente.

2.3.2 El valor de tomar buenas decisiones

Tomar decisiones acertadas respecto a los productos que se ofrecen en cada tienda es fundamental para mantener la permanencia del negocio. Actualmente el tomar decisiones acertadas se considera más un arte que una ciencia, y son los mismos tomadores de decisiones quienes confirman que estas son tomadas prácticamente en su totalidad en base a la experiencia y el conocimiento que ha generado la “prueba y error”.

Este proyecto no busca eliminar ni cambiar en su totalidad la forma en que estas decisiones son tomadas, buscar esto último sería perseguir una utopía, ya que el dinamismo del negocio impide evaluar a cabalidad todas las acciones. Lo que se busca es apoyar dichas decisiones con datos empíricos y facilitar la detección de oportunidades en el rediseño del mix.

Detectar anticipadamente oportunidades de rediseño puede llegar a ser un gran ahorro para el negocio o inclusive un aumento en las utilidades, ya que un por un lado se tiene el costo de tener en góndola productos con baja rotación, los cual generan costo de stock y el costo de oportunidad inherente debido a la posibilidad de tener algún otro producto en su lugar. Por otro lado, detectar oportunidades para expandir el mix de productos como por ejemplo alguno de marca propia, puede traer inmediatas utilidades al negocio.

Debido a la segmentación previamente echa, es posible analizar las oportunidades en base a los distintos segmentos de clientes, lo que permitiría saber si se está cubriendo de forma adecuada la necesidad para cada segmento. Y así detectar oportunidades de expansión donde se esté dejando de lado algún segmento.

2.3.3 Conocimiento del cliente

No hay que olvidar que quienes compran los productos son los clientes, los cuales presentan distintas características. Debido a esta variabilidad es que existen clientes más y menos valiosos para la empresa, tanto debido a que son buenos clientes actualmente, como también debido a que tienen un alto potencial de ser buenos clientes en el futuro. Esto último es conocido como el Life Time Value⁸, y en palabras simples, representa el valor de un cliente considerando su potencial actual y futuro.

Debido a que existen clientes más y menos valiosos para la empresa, es fundamental considerar al cliente al momento de decidir acciones sobre el mix de productos. Una de las tareas del negocio debe ser atraer y retener clientes valiosos, y esta tarea se puede ver afectada debido a un mal manejo del mix de productos que se está ofreciendo. A pesar de que en este proyecto no se calcula un valor exacto del Life Time Value, da pie a trabajos futuros que permitirían realizar estimaciones más exactas sobre el verdadero impacto que causan las acciones en términos por ejemplo de fidelidad del cliente hacia la marca.

Por último no hay que olvidar la posición privilegiada que tiene CMR al momento de categorizar al cliente, debido principalmente a que es poseedor de mucha información de cada uno de sus dos millones de clientes activos. Dentro de la cual cuenta entre otras cosas, *relocalización* del cliente, información socio económica, comportamiento de compra en otros negocios. Todo lo anterior permite tener una buena apreciación del cliente y poder tomar mejores decisiones.

La posición de CMR además evita el roce que existe entre las distintas empresas del holding en el compartir la información. Debido a que CMR posee completo acceso al *Data Warehouse* de la empresa y posee información privilegiada adquirida de los clientes de la tarjeta CMR al momento de que adquirieron su tarjeta.

⁸ Para mayor referencia sobre el cálculo del *Life Time Value*: Shawn, R., & Stone, M. (1990). *Database Marketing*. Wiley US Edition.

III.OBJETIVOS

3.1 Objetivo general

El objetivo general se puede enunciar de la siguiente manera:

“Diseñar e implementar una metodología de análisis y diseño de herramientas para la detección de oportunidades de rediseño del mix de productos en base a las correlaciones de estos”

utilizando reglas de asociación obtenidas de la información transaccional de diversos retails”.

3.2 Objetivos específicos

- 1) Se crearán herramientas computacionales que faciliten y agilicen cada análisis.
- 2) Para lograr que la metodología y las herramientas sean aplicables a los distintos negocios de Falabella, se estandarizará la data transaccional de cada uno de estos negocios.
- 3) Se creará un ranking de los productos del mix en base al aporte en ventas de estos considerando el efecto de correlación de estos.

IV.METODOLOGÍA

A continuación se presentan los principales puntos relevantes que componen este trabajo.

4.1 Revisión bibliográfica

La primera etapa de la metodología de trabajo contempla la revisión bibliográfica sobre los distintos trabajos que han abordado el análisis de mix de productos. Centrándose particularmente en los métodos que consideran las diversas interacciones entre los productos.

4.2 Análisis de la situación actual

4.2.1 Composición de la cadena

Se describe la distribución de unidades y sub unidades de negocio dentro del holding Falabella. Se estandarizarán estas distribuciones para luego trabajar indistintamente del negocio al cual se haga referencia.

Además se caracterizará físicamente la tienda en la cual se realizará el estudio del mix de productos.

4.2.2 Identificación de información disponible

Es necesario tener certeza del tipo y calidad de la información que se dispone al momento de realizar el estudio. Si bien este estudio solo utilizará información transaccional, esta debe ser lo más actual posible y recolectada en un horizonte de tiempo lo suficientemente amplio (mínimo un mes), para mitigar así los efectos provocados por anomalías en el negocio, como estacionalidades.

4.2.3 Consolidación de la información

Para el rápido trabajo de la información será necesaria la consolidación de esta en una base de datos. La cual permita realizar los diversos filtros y cruces de información necesaria que servirá de input para los distintos pasos de la metodología.

4.2.4 Caracterización de la tienda y su mix

Debido a que el estudio se enmarca en una tienda en particular, la tienda seleccionada será caracterizada tanto a nivel de composición de productos como además del comportamiento de compra de los clientes en esta.

Al estudiar el comportamiento de compra de los clientes, y el cómo los diversos productos de la tienda son comprados por los clientes, se logrará identificar las diversas intenciones de compra de estos. Lo que permitirá fijar la frontera del análisis al considerar una de estas intenciones de compra como el conjunto de productos que se gestionará.

Por otro lado, si bien escapa del alcance de este trabajo, esta agrupación de los productos bajo intenciones de compra, permite realizar segmentaciones de clientes basadas en el comportamiento de compra de estos.

4.2.5 Canastas frecuentes

Continuando el estudio del comportamiento de compra de los clientes, se proseguirá a descubrir las diversas canastas frecuentes que estos compran. Llegando así a contar con una primera aproximación del efecto de correlación existente entre los productos.

4.2.6 Selección del surtido

La información de las canastas frecuentes será aprovechada para que mediante una optimización lineal se seleccionen el surtido más eficiente al estar considerando las correlaciones existentes entre los distintos productos.

4.2.7 Resultados y recomendaciones

Para finalizar el trabajo se presentan los resultados obtenidos por la metodología y se realizan recomendaciones que permitan mejorar a futuro el estudio del mix de productos.

V. MARCO CONCEPTUAL

En este capítulo se le da a la investigación un sistema coordinado y coherente de conceptos y proposiciones que permitirán abordar el problema. Se resume a continuación las recopilaciones que se han hecho de la bibliografía.

La metodología de análisis de este proyecto se enmarca en las técnicas de minería de datos y técnicas exploratorias. Siendo más específicos, por el lado de minería de datos se realizara estudio de correlaciones entre serie de datos y por técnicas exploratorias se pueden contar el escalamiento multidimensional y el análisis jerárquico.

5.1 Data Mining

Data mining es el proceso en el cual se extraen patrones de la información. Debido al aumento exponencial en la cantidad de información que las empresas manejan y a la necesidad de aprovechar al máximo dicha información es que esta técnica ha aumentado su popularidad entre las distintas empresa, debido a que les da una ventaja competitiva al transformar la información disponible que se obtiene por ejemplo desde el punto de venta (en adelante POS) en conocimiento para el negocio.

Tal como dice Fayyad; *“Data Mining es el proceso no trivial de identificación de patrones entendibles, válidos y potencialmente útiles para el negocio, desde grandes bases de datos”* (Fayyad, 1996). Estos patrones pueden servir por ejemplo para:

- Estudia los ciclos de vida de los clientes.
- Segmentar a los clientes.
- Entender comportamiento de compra.
- Detectar fraude.

Basándose en lo que Fayyad dice, se desprende que este trabajo se enmarca en las técnicas de *Data Mining* por el hecho de buscar en grandes bases de datos patrones entendibles y aprovechables.

El grupo Falabella es una de las entidades que más información genera y almacena en Chile, pero sin duda es la entidad que posee la mayor riqueza de datos en términos de diversidad de negocios y detalle de las transacciones. Como ya se ha señalado con anterioridad, la gran cantidad de información proviene de sus diversos negocios, y es esta diversidad y magnitud la que crea una oportunidad única para que el analista genere un sin número de información y oportunidades relevantes para el negocio. Debido a que el trabajo en cuestión persigue ser una herramienta útil para el negocio, y debido a la gran cantidad de datos que es necesario procesar cada vez que se pretenda realizar un estudio, urge la necesidad de tener especial preocupación por los tiempos de espera que el usuario enfrenta al momento de realizar algún estudio.

Por otro lado, como se señaló con anterioridad en este informe, en primer lugar se buscaran las distintas necesidades de los clientes, lo cual se realizará utilizando la información transaccional generada por los clientes y se aplicará en primer lugar un análisis multidimensional que permite encontrar patrones entre la información al disminuir en dimensiones el problema.

Para aprovechar el conocimiento del usuario y facilitarle a este la tarea de encontrar las distintas necesidades de los clientes, se utilizará un análisis jerárquico.

Finalmente para entender de mejor forma el roll que cumple cada producto, se utilizarán técnicas de Análisis de canasta para encontrar diversas correlaciones entre el producto en cuestión y algún otro.

5.2 Escalamiento multidimensional

El escalamiento multidimensional (en adelante MDS) es una técnica exploratoria que utiliza la distancia entre distintos objetos y busca encontrar la estructura subyacente de dichos objetos en un espacio multidimensional. Más particularmente en este trabajo, el MDS utilizará la distancia entre diversos productos, y para calcular la distancia entre dos productos se hará en base a la cantidad de veces que dichos productos se compran juntos. Luego productos que son constantemente llevados juntos deberán presentar una baja distancia entre ellos.

Una vez que se ha utilizado el MDS para encontrar factores entre los productos y por lo tanto se tiene cada producto representado como una coordenada en un espacio multidimensional, se proseguirá con una clasificación jerárquica para así mostrarlos gráficamente.

5.3 Clasificación Jerárquica

El análisis de conglomerados es un procedimiento multivariado que a partir de un conjunto de objeto crea una cierta cantidad de segmentos llamados conglomerados, basándose en diversos criterios de similitud. Al momento de generar las agrupaciones el procedimiento tiene dos objetivos, por un lado busca que los elementos de un grupo sean lo más homogéneo posible y

a la vez que entre los grupos, estos sean lo más heterogéneos posible, todo esto utilizando los mismos criterios de similitud.

La similitud de los objetos se mide principalmente en forma de distancia entre los objetos, donde la distancia más utilizada es la euclidiana. En este trabajo la lógica que se utilizará para asignar los objetos a los distintos grupos es del tipo jerárquica. Y gráficamente esto se representa mediante un dendrograma o conocido también como árbol jerárquico. El procedimiento para asignar un grupo a un elemento puede comenzar con todos los elementos en un grupo, he ir realizando una división a este gran grupo en cada iteración, o bien, comenzar con todos los elementos en distintos grupos e ir juntando en cada iteración un grupo. Este último método tiene la ventaja de no necesitar un número preestablecido de grupos a generar, dándole la facultad al analista de determinar a posteriori el número de grupos con los que se va a trabajar.

En este trabajo la clasificación jerárquica se utilizará en conjunto con el MDS para encontrar distintas necesidades de los clientes en el mix de productos.

5.4 Análisis de canasta

El estudio de correlaciones se ha utilizado mucho en el área del retail y se ha popularizado gracias al Análisis de Canasta. Se puede señalar, a grandes rasgos, que el Análisis de Canasta busca encontrar relaciones existentes entre los productos presentes en un ticket de compra, pudiendo así, encontrar patrones de compra de los clientes. Esta tesis utiliza el Análisis de Canasta para obtener conclusiones, pero como se verá más adelante se utiliza de una nueva forma. Por ejemplo, se realiza una nueva definición sobre lo que el Análisis de Canasta entiende por “canasta de compra”.⁹

Diversos estudios se han realizado sobre Análisis de Canasta en los retails del mundo. Y se ha obtenido diversa información sobre el comportamiento de compra de los clientes en dicho negocio y de la afinidad que existe entre los distintos productos dentro de la tienda. Los resultados de estos estudios permiten tomar decisiones de negocios con un sustento empírico. Estas decisiones, van desde la configuración del *layout* en las tiendas hasta la asignación de crédito a los distintos clientes.

Al analizar la información proveniente del POS, se busca medir la asociación existente entre los distintos productos, para cuantificar estas asociaciones se utilizan diversos indicadores¹⁰. Los más utilizados son: Soporte, Confianza, Ganancia y Convicción.

Para lograr generar estos indicadores, es fundamental tener una matriz en la cual cada fila represente un ticket de compra. Y la información de esta debe ser los productos que integran dicho ticket de compra.

5.4.1 Indicadores de asociación

A continuación se explicará la forma de calcular cada uno de los indicadores y la información que se logra obtener de ellos. Las variables X e Y representan un producto o un pack de productos.

⁹ Para mayor información sobre la definición de la canasta ver: Capítulo: 6.5 Análisis de Canasta.

¹⁰ *Mining Association Rules Between Sets of Items in Large Databases*, R. Agrawal.

1. Soporte

Este indicador mide qué tan frecuente es una relación entre los productos. Del total de tickets mide la probabilidad de que se encuentre un conjunto de productos en ella. Por lo tanto, este indicador no busca medir si existe o no vínculo entre los productos.

$$\text{Soporte}(X, Y, \dots) = P(X \cap Y \cap \dots)$$

2. Confianza

Este indicador mide el vínculo o sinergia entre los productos. Y se puede interpretar como la probabilidad de tener un producto al saber que ya se cuenta con otro.

$$\text{Confianza}(X \rightarrow Y) = \frac{\text{Soporte}(X, Y)}{\text{Soporte}(X)}$$

3. Ganancia

Este indicador mide el aumento de probabilidad del consecuente al ser comprado el antecedente. Este indicador difiere del anterior, ya que mientras la Confianza nos determina la probabilidad exacta de que el segundo producto se encuentre en el ticket de compra al momento de ya contar con el anterior, este nos señala en cuanto aumenta la probabilidad de aparecer en el ticket el artículo cuando se compra el antecedente.

$$\text{Ganancia}(A \rightarrow B) = \frac{\text{Soporte}(X, Y)}{\text{Soporte}(Y) \times \text{Soporte}(X)}$$

Este indicador es muy útil para el ámbito del retail ya que tiene una interpretación económica que permite encontrar relaciones de complementariedad entre productos dependiendo del valor de la ganancia. (Silverstein, Brin, & Montwani, 1998)

Tabla 2: Interpretación económica de la ganancia

<i>Resultado</i>	<i>Interpretación</i>
Ganancia > 1	Efecto de complementariedad entre X e Y
Ganancia = 1	Independencia entre X e Y
Ganancia < 1	Efecto de sustitución entre X e Y

Fuente: (Brijs, Swinnen, Vanhoof, & Wets, 2004)

4. Convicción

La convicción es la razón de la frecuencia esperada de que X ocurra sin Y, es decir, que la regla genere una predicción errada.

$$\text{Convicción}(X \rightarrow Y) = \frac{1 - \text{Soporte}(Y)}{1 - \text{Confiabilidad}(X \rightarrow Y)}$$

5.5 PROFSET

PROFSET (Brijs, Swinnen, Vanhoof, & Wets, 2004) es una estructura de optimización para la selección de productos, que se captura la interdependencia de los productos gracias a la utilización de reglas de asociación entre productos, al utilizar la información obtenida del análisis de canasta, particularmente los sets de ítems frecuentes que se obtienen de este análisis. (Mannila, 1997). PROFSET es considerado como una estructura de optimización y no solo como un modelo de optimización ya que permite la resolución de diversas optimizaciones, según sea el problema concreto de surtido que el retail quiera abordar.

Antes de entrar a describir la estructura en sí, se comparará la contribución echa de esta estructura con lo realizado anteriormente en términos de selección de productos y de gestión del espacio en las góndolas.

5.5.1 ¿Qué incluye y deja de incluir PROFSET?

Los modelos de selección de productos se pueden dividir en dos grandes grupos, los que se encargan de seleccionar a los diversos productos que deben ir en las góndolas y los que estudian la cantidad de espacio que hay que atribuirle a cada producto. PROFSET por su parte, solo se enmarca en el problema de selección de productos, y no en el de plano grama de góndolas.

Como segundo punto cabe destacar que la ecuación que PROFSET busca maximizar incluye tanto el efecto de las ventas individuales del producto como también el efecto de venta cruzada de este con otros productos. Como se ha señalado con anterioridad, considerar solo la utilidad generada por cada producto como indicador de su aporte al negocio, es una mirada “miope” al no estar considerando todo el efecto indirecto que genera en las ventas de los demás productos. Pero no hay que olvidar que dicho efecto puede ser tanto de carácter positivo (aumentando las ventas del otro producto) como negativo. Y este modelo solo considera el efecto de interdependencia positivo (*cross-selling*) entre productos.

Por otro lado el modelo posee flexibilidad tanto en su función objetivo como también en sus restricciones. La función objetivo, además de considerar las ventas de cada producto, permite incluir información sobre los costos de inventario y los costos unitarios de compra de estos, consiguiendo con la inclusión de costos un análisis más preciso. Por el lado de las restricciones, en primer lugar se incluye una restricción que permite restringir al modelo a que seleccione solo a algunos productos, esto puede ser útil con productos que contribuyan principalmente a la imagen que el retail desee reflejar y que a pesar de no generar un aporte significativo en las ventas, son productos primarios que no pueden ser eliminados. Además, el modelo permite incluir

restricciones que fijen cotas tanto superiores como inferiores para el número de productos a seleccionar para cada categoría.

Haciendo finalmente referencia a lo que no incluye el modelo, como punto más importante se tiene que el modelo no incluye el efecto de cambio de marca que se genera por algún quiebre de stock. Esto es claramente una limitación del modelo, pero nuevamente hay que señalar que debido a la gran cantidad de productos que se pretende estudiar, esta consideración es prácticamente no factible.

5.5.2 Especificaciones del modelo

Como se ha hablado con anterioridad, el modelo presenta una gran flexibilidad para abordar diversos problemas de marketing. Pudiendo variar en tres criterios distintos según la optimización que se busca, la regla que se utilizará para atribuir la venta a cada canasta frecuente y si se incluyen o no las diversas restricciones. A continuación se busca resumir en la Tabla 3 las diversas versiones que puede tener el modelo en base a la selección de diversos criterios.

Tabla 3: Especificaciones del modelo PROFSET

<i>Criterio</i>		
Optimización	Dentro de productos seleccionados	Dentro + Fuera
Regla de asignación	Basado en el soporte de las canastas	Basado en significancia estadística
Restricciones	Sin restricciones	Con restricciones

Fuente: (Brijs, Swinnen, Vanhoof, & Wets, 2004)

En primer lugar se selecciona el objetivo de la optimización, el cual está directamente relacionado con el problema de marketing que se busca resolver. Cuando la optimización se realiza solamente dentro de los productos que el modelo selecciona, se considera que no existen cruces con los productos fuera del set de productos seleccionados. Si por el contrario la optimización considerase los productos dentro y fuera de los seleccionados, a pesar de seleccionar solo algunos productos del mix, la función de utilidad considera los cruces generados tanto entre los productos seleccionados como también de estos con los que no quedaron seleccionados.

También existe una posibilidad de variar la heurística que se utiliza para asignar la utilidad de cada canasta frecuente. Una forma es mediante el uso del soporte de cada canasta frecuente, mientras que una segunda opción es utilizando una significancia estadística obtenida mediante un análisis log linear.

Finalmente el modelo permite la utilización o no de diversas restricciones, principalmente al fijar las cotas en la cantidad de productos que se seleccionan.

A continuación se desarrollaran en mayor profundidad cada uno de los distintos criterios.

5.5.2.1 Optimización

Basándose en el tipo de problema de marketing a resolver es que se selecciona el criterio de optimización. Si bien ambos criterios buscan encontrar un conjunto de productos optimo, estos difieren en la forma que tienen de evaluar la utilidad final que estos generan, al considerar de distinta manera la venta cruzada.

Dentro de los productos seleccionados

Al seleccionar este criterio de optimización, lo que se busca es encontrar un mix de productos que maximicen la rentabilidad sin considerar a los productos que quedan fuera de esta selección.

A modo de entender de mejor forma este criterio, se analizará un problema de marketing en particular: En la actualidad los compradores disponen cada vez de menos tiempo para realizar las compras, y como respuesta a esto es que han surgido cada vez más tiendas de conveniencia y mini mercados. Sin entrar en grandes detalles de las características y el enfoque estratégico de cada uno de estos formatos, lo que sí es importante es que poseen un número reducido de productos distintos.

A pesar de que la superficie de estas tiendas es relativamente pequeña (de 15 a 150 m²), contar con el mix de productos específicos que maximicen la rentabilidad total de la tienda es fundamental.

El objetivo es que el modelo de PROFSET utilice la información que se tiene de los *híper* mercados para ahí seleccionar un mix de productos reducido, los cuales generen la mayor rentabilidad para la tienda al considerar la venta en particular de cada producto como también la venta cruzada que cada uno genera. Pudiendo así seleccionar un grupo reducido de productos valorizando solo los cruces generados dentro del mix ya seleccionado. De esta forma este mix de productos podrá ser utilizado para optimizar la composición de productos presentes en una supuesta tienda de conveniencia.

Dentro y fuera de los productos seleccionados

La optimización previamente descrita optimiza considerando como válidas solo las correlaciones existentes entre los productos seleccionados. Esta nueva consideración a diferencia de la anterior, considera las correlaciones existentes entre los productos seleccionados y las existentes entre los productos seleccionados y los que no lo han sido, pero no entre los productos no seleccionados.

Para entender de mejor forma este modelo de optimización, supongamos el problema al que se enfrenta que el supermercado Tottus con CMR se enfrentan: Desean enviar a sus clientes cupones de descuentos en diversos productos, los cuales se hacen efectivos solo al momento de pagar la compra con la tarjeta CMR. El objetivo de esta acción es aumentar la participación

CMR, con lo cual se espera que los clientes no solo utilicen los cupones de descuento en comprar únicamente el producto en promoción, sino además compren otros productos que no estén en promoción. Con lo cual en definitiva lo que se está buscando es potenciar la venta cruzada entre productos.

Luego, idealmente se espera que los productos que se seleccionan para ser promocionados generen la mayor venta cruzada posible. Por consiguiente, el modelo en este caso selecciona los productos considerando la utilidad que deja cada producto en el retail, pero además considera las correlaciones que genera cada producto con los otros del supermercado y con los otros en promoción.

Por lo tanto bajo esta formulación del problema, la función objetivo y las restricciones adquieren otra forma. Ya que a diferencia de la formulación anterior donde la utilidad de cada canasta frecuente era considerada solo si se seleccionaban todos los productos pertenecientes a esta, en este caso la utilidad de cada canasta es considerada si existe al menos un producto seleccionado de la canasta frecuente.

A continuación se profundiza el concepto de asignar utilidad a las canastas frecuentes y se muestran dos criterios que PROFSET utiliza para asignar utilidad a las distintas canastas.

5.5.2.2 Reglas de asignación

Junto con haber definido qué criterio de optimización se utilizará, es necesario determinar cuál se usará para asignar la utilidad de los productos a cada canasta frecuente. El problema radica en determinar a cuál de todas las canastas frecuentes se le atribuye la venta de una transacción en particular. A continuación se describen dos métodos distintos para asignar la venta a cada canasta frecuente, donde el primero de ellos se basa en el soporte y tamaño de cada canasta frecuente donde todos sus productos están contenidos en la transacción. Y el segundo método utiliza un criterio estadístico para determinar a cual canasta asignar la venta.

Basado en el soporte de las canastas

Lo que se pretende con este criterio es asignar la venta de una determinada transacción a las diversas canastas frecuentes que se incluyen en ella. Ya que una transacción en particular, por ejemplo compuesta por varios productos, puede contener un número muy grande de canastas frecuentes. De hecho, para que una canasta frecuente pertenezca a una transacción, todos los productos que pertenecen a la canasta frecuente deben estar incluidos en la transacción del cliente.

Luego en lo que se basa este criterio de asignación es en intentar determinar las diversas intenciones de compra de los clientes que están presentes en la transacción que realizaron. Para de esta forma asignarle la venta de esos productos a las intenciones de compra correspondientes. En un escenario ideal, se le preguntaría al cliente al momento de abandonar la tienda cual fue su intención de compra y así asignarle dicha venta a esas canastas, pero como esto es no se puede hacer, se opta por utilizar el soporte de cada canasta frecuente perteneciente dentro del ticket de compra como medida de probabilidad de ocurrencia.

Por lo tanto si por ejemplo una canasta frecuente posee un soporte que es dos veces mayor que otra canasta se asumirá que es el doble de probable que la intención de compra del cliente sea

la primera canasta, luego se le atribuirá la venta a la primera canasta en doble de las veces que a la otra.

Antes de ejemplificar la situación previamente descrita, es necesario introducir el concepto de canasta frecuente máxima. Para esto se utilizará la definición utilizada por Tom Brijs.

Definición V.1: Canasta frecuente máxima

Sea F el conjunto de todas las canastas frecuentes incluidas en la transacción T . La canasta frecuente $X \in F$ será una canasta frecuente máxima, denotada como X_{max} , si y solo si $\forall Y \in F: |Y| \leq |X|$.

Utilizando la definición previa, la venta de cada transacción será atribuida a cada canasta frecuente según el siguiente criterio. En caso de existir solo una canasta frecuente máxima se le atribuirá toda la venta de la transacción generada por los productos que pertenecen a la canasta frecuente máxima. Pero en caso de existir más de una canasta frecuente máxima, se utilizará el soporte de cada una de estas para definir en base a una función de probabilidad a cuál de ellas atribuirle la venta de los productos. Luego la función de probabilidad que se utilizará en este caso será:

Definición V.2: Función de probabilidad para seleccionar la canasta frecuente máxima

$$\Theta_T(X_{max}) = \frac{Soporte(X_{max})}{\sum_{Y_{max} \in T} Soporte(Y_{max})}$$

Luego una vez que se ha seleccionado una canasta frecuente en base al algoritmo previamente explicado, se prosigue a atribuirle la venta correspondiente a los productos de dicha canasta frecuente que estaban en la transacción. Para después eliminar de la transacción los productos que ya fueron asignados a una canasta frecuente y volver a realizar el proceso hasta que no queden canastas frecuentes en la transacción.

Para entender mejor lo anterior se muestra a continuación el pseudo código. Donde X es una canasta frecuente y $M(X)$ es la utilidad (puede ser venta o margen) ya atribuida a la canasta frecuente X . Y $m(X)$ es la utilidad de la canasta frecuente X en la transacción T , lo cual no es más que la suma de utilidades dentro de la transacción de todos los productos pertenecientes a la canasta frecuente.

Código V.1: Pseudo código de algoritmo para asignar utilidad a las canastas frecuentes basándose en el soporte de las canastas frecuentes.

```

For cada transacción T en periodo de evaluación do{
    while (T≠∅) or (∃X ⊆ T) do{
        if ∃! Xmax ⊆ T
            then M(X):= M(X) + m(Xmax );
        else seleccionar Xmax de todas las canastas
    
```

```

        frecuentes máximas utilizando la función
        de probabilidad  $\Theta T$ ;
         $M(X) := M(X) + m(X_{max})$ 
        con  $m(X_{max})$  la utilidad de  $X_{max}$  en  $T$ ;
        A la transacción  $T$  se le sustraen los
        productos de  $X_{max}$ ;
    }
}
return all  $M(X)$  ;

```

Con este algoritmo se logra recorrer todas las transacciones realizadas en algún periodo y atribuir toda la venta (o margen) generada en dicho periodo a las diversas canastas frecuentes.

Esto generará al finalizar que se tengan todas las canastas frecuentes valorizadas para el periodo en estudio. Estas canastas valorizadas se utilizarán en el proceso de optimización, en donde PROFSET las utiliza para seleccionar los productos.

A continuación se explica otro criterio para asignar la utilidad a las canastas frecuente, que si bien el resultado al igual que este criterio es la valorización de las canastas, debido a que el algoritmo para asignar la utilidad es distinto, se puede llegar a obtener diversos resultados.

Basado en significancia estadística

Este criterio de asignación de utilidad si bien es muy similar al anterior basado en el soporte de los productos, se basa en una metodología estadística para asignar la utilidad de cada transacción a las distintas canastas frecuentes. Lo que se pretende conseguir es seleccionar a que canasta frecuente atribuirle la utilidad dentro de todas las canastas frecuentes estadísticamente significativas, a diferencia del método anterior que atribuía la utilidad en base al soporte de todas las canastas frecuentes máximas.

Lo que busca este método es determinar si la utilidad de una transacción se le atribuye a la canasta frecuente de uno, dos o tres productos. Ya que puede ser que la canasta frecuente que contiene tres productos, lo sea ya que dos de ellos son muy frecuentes. Por ejemplo, supongamos que la combinación de productos {cola, maní, jamón} es muy frecuente, luego la pregunta es que este trío de productos puede ser frecuente ya que {cola, maní} y/o {maní, jamón} y/o {cola, jamón} son muy frecuentes, o es que efectivamente el trío de productos es frecuente en sí.

Luego el análisis log linear busca determinar si es que el trío de productos es estadísticamente significativo por sí solo, o es que se pudo haber deducido de la combinación de canastas frecuentes que se forman dentro de este. En este último caso, no se considerará la canasta frecuente con tres productos al momento de asignar la utilidad.

En el trabajo de Brijs se demostró que la diferencia al utilizar estos dos criterios era prácticamente nula. Debido a esto, y a modo de simplificar el modelo es que en este trabajo se utilizará la asignación de utilidad basada en soporte. De todas formas si el lector desea profundizar en este criterio puede hacerlo en el trabajo de (DuMouchel & Pregibon, 2001).

5.5.2.3 Restricciones

Por último, la estructura de PROFSET permite incluir en el modelo diversas restricciones que permiten modelar los diversos requerimientos del problema en cuestión.

Uno de los aspectos más importantes de incluir restricciones en el modelo, es la facultad de incluir el concepto de categoría, es decir, de poder agrupar productos y tratarlos estos de una manera diferente. Pudiendo así restringir al modelo al número de productos a seleccionar por categoría, como también por ejemplo forzarlo a seleccionar determinados productos.

Sin restricciones de categorías

Al no asignarle restricciones al modelo y por consiguiente no agrupar los productos en categorías, se trabajará a nivel de SKU todos los productos y se estará desaprovechando todo el potencial de modelo.

Desde un punto de vista práctico al no incluir restricciones en el modelo, este explotará todo el efecto individual y cruzado de los productos, y debido a que todos serán seleccionados se podrá observar la utilidad final del modelo al seleccionar todos los productos.

Luego surge indispensable la inclusión de restricciones en el modelo que permitan por un lado agrupar a los productos y a la vez personalizar el modelo para que resuelva los problemas en particular.

Con restricciones de categorías

Como ya se ha dicho, para adaptarse de buena forma al problema del retail es necesaria la inclusión de restricciones al modelo.

Las restricciones del modelo se pueden dividir en tres grandes grupos: las necesarias para que el modelo asocie las canastas frecuentes a los productos, las que fijan las cotas en la cantidad de productos a ser seleccionados y las que fuerzan al modelo a seleccionar algunos productos.

- Restricciones de asociación:

Las primeras restricciones, encargadas de asociar las canastas frecuentes con los productos, son indispensables y difieren según el criterio de optimización que se utilizará. Es decir, que según si la optimización se realizará solamente dentro de los productos seleccionados o si se hará con los productos tanto de adentro como de afuera, se utilizará una restricción distinta.

Si se desea optimizar solo considerando las correlaciones existentes dentro de los productos seleccionados, la restricción forzará al modelo a solo tomar en consideración la utilidad de una canasta frecuente si es que se seleccionan todos los productos de esta.

Por el contrario, si se busca optimizar considerando las correlaciones existentes entre los productos seleccionados y también las que se generan entre los productos seleccionados y los de afuera del mix, es que se debe forzar al modelo a considerar la utilidad de las canastas frecuentes si al menos se selecciona un producto dentro de la canasta frecuente.

- Restricciones de cantidad

Por restricciones de cantidad se refiere a forzar al modelo a seleccionar un mínimo de productos y/o un máximo de productos a ser seleccionados. Al mezclar esta restricción con la categorización de productos se logra una versatilidad y una adaptación del modelo a la realidad que permite modelar diversos escenarios.

Por ejemplo, al retail le puede interesar dejar en el mix un mínimo de tres productos distintos por categoría, o bien forzar al modelo a que no elimine por completo ninguna categoría.

- Restricciones de inclusión de productos

Este último set de restricciones fuerzan al modelo a incluir o no ciertos productos en el mix. Pudiendo así por ejemplo forzar al modelo a incluir ciertos productos que a pesar de no aportar una contribución significativa para el negocio, ayudan a mantener la imagen que este desea reflejar. También es posible que se desee forzar al modelo a no incluir algún producto en particular y poder ver así cual es el impacto que tiene esto.

El combinar las restricciones previamente descritas permite que el modelo se adapte a los diversos requerimientos del negocio y por otro lado permitir la realización de estudios de sensibilidad, al forzar al modelo a distintos escenarios.

5.5.3 Modelo de optimización PROFSET

Como se pudo apreciar en la sección 5.5.2 el modelo PROFSET puede ser adaptado según las diversas necesidades del negocio.

A continuación se describe un modelo general donde se optimiza considerando solo las correlaciones existentes dentro del conjunto de productos seleccionados y se restringe al modelo a incluir los productos básicos (restricción 2) y crear una cota superior para la cantidad de productos a elegir.

A continuación se describe el modelo:

- Utilidad:

Sea:

$I = \{i_1, i_2, \dots, i_m\}$ Productos en un negocio (Mix de productos)

$D :=$ Conjunto de transacciones donde $T \in I$

$PV(i) :=$ Precio de venta del producto i al consumidor

$PC(i) :=$ Precio de compra del producto por el retail

$f(i) :=$ Veces que el producto i es comprado en una canasta particular T

- Definición 1: $m(T)$ es la utilidad generada por las transacciones T .

$$m(T) = \sum_{i \in T} (PV(i) - PC(i)) * f(i)$$

- Definición 2: $M(X)$ es la utilidad generada por el set de ítems frecuentes X .

$$M(X) = \sum_{i \in T} m'(T) \begin{cases} m'(T) = m(T) & \text{if } X = T \\ m'(T) = 0 & \sim \end{cases}$$

- Función Objetivo:

$$MAX \left(\sum_{X \in L} M(X) * P_x - \sum_{i \in L} Cost_i * Q_i \right)$$

- Sujeto a:

$$\forall X \in L, \forall i \in X: Q_i \geq P_x \quad (1)$$

$$\forall i \text{ producto basico}: Q_i = 1 \quad (2)$$

$$\sum_{i \in L} Q_i = ItemMax \quad (3)$$

$$P_x \text{ y } Q_i \in \{0,1\} \quad (4)$$

Restricciones:

- 1) Debido a que las decisiones de mix de productos deben ser tomadas a nivel de productos y no a nivel de sets de productos frecuentes, es que se especifica que al agregar un set de productos frecuentes, todos los productos pertenecientes a este deben ser incluidos.
- 2) Es posible forzar al modelo a que seleccione ciertos productos.
- 3) La cantidad máxima de productos a seleccionar puede ser definida.
- 4) Como es lógico, las variables P_x y Q_i deben ser enteras ya que representan la inclusión o exclusión de un producto o set de productos.

Este modelo permitirá determina el set de productos óptimos que maximicen la utilidad del retail, considerando no solo el aporte a las utilidades generado por cada producto, sino también el valor generado por estos productos debido a la interacción entre ellos.

Además, el análisis de sensibilidad de esta programación entera, permitirá encontrar el efecto de eliminar cada producto en la utilidad total. Por otro lado, los indicadores de análisis de canasta previamente encontrados, permitirán darle explicación y mejorar las conclusiones del estudio.

VI. MARCO METODOLÓGICO

En este capítulo se sientan las bases en lo que respecta a la metodología que se seguirá para llevar a cabo los distintos objetivos, haciendo énfasis en las diversas herramientas que se utilizarán y crearán para conseguirlo.

Para generar una metodología que permita la rápida detección de oportunidades de rediseño de mix y evaluación de distintas categorías, en primer lugar hay que comprender y estandarizar la información con la que se dispone. Una vez estandarizada la información se puede proseguir indistintamente del negocio que se esté evaluando, permitiendo así replicar de forma fácil el modelo en los distintos negocios del holding.

Una vez estandarizada la información transaccional, se prosigue con el Análisis del Mix de Productos (AMP) el cual consta principalmente de tres etapas. En primer lugar se busca encontrar las intenciones de compra de los clientes, esto mediante un estudio de las distintas canastas de compra de los clientes. La obtención de las intenciones de compra permitirá agrupar de forma eficiente los distintos productos de la tienda, ya que la agrupación realizará basándose plenamente en las canastas que realizan los clientes.

Ahora que se cuenta con los productos agrupados según las diversas intenciones de compra de los clientes, se prosigue a estudiar su desempeño y a caracterizarlas. El estudio de las intenciones de compra permite entender la composición de la tienda y el rol de los productos dentro de esta, además al realizar un estudio entre intenciones de compra es posible descubrir cómo es que se relacionan unas con otras. Una vez hecho el estudio de las intenciones de compra, se prosigue a seleccionar una de estas para continuar con su gestión.

Habiendo hecho lo anterior, es ahora necesario encontrar las diversas canastas frecuentes de los clientes que incluyan al menos un producto que pertenezca a la intención de compra seleccionada. Para obtenerlas se utilizarán algoritmos de minería de datos, que utilizando la información transaccional logren obtener esta gran cantidad de canastas. Teniendo las canastas frecuente de los clientes, se prosigue a valorizar estas canastas y de esta forma capturar el valor de las correlaciones entre los productos.

El paso siguiente es utilizar el valor de las canastas frecuentes para llevar a cabo la optimización de PROFSET, la cual según la configuración que se utilice permitirá tomar diversas decisiones del negocio.

6.1 Estandarización de las transacciones

CMR posee actualmente alrededor de 4 millones de clientes con tarjeta, los cuales realizan sus compras en las diversas tiendas del grupo Falabella, generando más de cincuenta millones de transacciones¹¹ mensualmente. Junto con lo anterior, la nueva integración de la tarjeta CMR Visa ha permitido que los clientes realicen transacciones en una mayor gama de comercios, generando una información muy variada.

Debido a que la información proviene de diversas fuentes esta presenta diversos formatos y estructuras. Por lo que si se quiere crear una metodología que sea fácilmente replicable es

¹¹ En este contexto una transacción se define como cada producto comprado por un cliente en una ocasión de compra. Por lo que una boleta puede tener una o más transacciones asociadas.

necesario estandarizar esta información que proviene de diversas fuentes. Para realizar esto será necesaria la creación de diversos scripts computacionales que realicen la estandarización.

La información referente a las transacciones realizadas por los clientes se encuentra en el Centro de Datos de la empresa y para poder extraerla es necesaria la utilización de consultas SQL. Para mantener la estandarización se creará una base de datos para la información proveniente de los tres negocios más la información proveniente de la tarjeta CMR Visa. Esta base de datos solo almacenará información relevante para los estudios que este trabajo pretende realizar. Por lo que es necesario determinar cuál será la información que será necesaria, para así crear herramientas computacionales que automáticamente extraigan día a día dicha información del Centro de Datos de la empresa y la almacenen en esta base de datos.

Junto con la necesidad de automatizar el proceso de transformar los datos desde el Centro de Datos a la base de datos que se tendrá para realizar este estudio, será necesario crear diversas herramientas que logren manejar de manera eficiente la gran cantidad de información. A continuación se explica cómo se abordará el manejo de los datos y como se crearán herramientas computacionales.

6.2 Base de datos y herramientas computacionales

Como se señaló con anterioridad, será necesario almacenar de forma estandarizada la información transaccional de los distintos negocios. Para lograr esto, se utilizará un motor de base de datos MySQL debido a su eficiencia y a que es gratuito. Para almacenar la información será necesario crear una estructura de datos que permita realizar los cruces necesarios entre la información de las transacciones y de los clientes.

Una vez que se encuentre diseñado el modelo de datos, será necesario poblar la base de datos automática y continuamente, permitiendo así que la información esté siempre disponible a disposición de los distintos análisis que se deseen hacer. Está y otras tareas serán programadas utilizando lenguaje PHP y Java. Los scripts en PHP serán montados sobre un servidor Apache el cual también se utilizará para la interacción con los usuarios.

Junto con herramientas que permitan manejar la información de la base de datos, será necesaria la utilización de algún software estadístico que permita realizar Análisis multidimensional. En este trabajo se utilizará el software estadístico SAS¹².

Además del software comercial señalado anteriormente, será necesaria la utilización de uno que resuelva el problema de programación entera PROFSET. Para esta tarea se utilizará AMPL¹³ el cual permite la resolución del problema planteado mediante el método de *branch-and-bound*. Debido a que el software AMPL requiere que el modelo a optimizar presente una sintaxis en particular y esta diferirá según cada estudio, es que además de crear scripts que manejen los datos, se crearán encargados de generar la sintaxis que se utilizará.

¹² SAS es un conjunto de paquetes estadísticos enfocados en la utilización de la información para resolver problemas de negocio. En este trabajo se utilizará el paquete de modelos estadísticos principalmente para ejecutar un Escalamiento Multidimensional. Para mayor información visitar www.sas.com.

¹³ AMPL (Modeling Language for Mathematical Programming) es un software para realizar modelamiento matemático y optimizaciones tanto lineal como no lineal. Recopila una serie de algoritmos capaces de modelar problemas de gran escala. Para mayor información visitar www.ampl.com.

Como último punto y para facilitar la interacción de los usuarios con los distintos scripts necesarios para llevar a cabo la metodología propuesta en este trabajo, se utilizará el servidor Apache y se crearán páginas web programadas principalmente en lenguaje HTML y JavaScript.

Finalmente se contará con una base de datos relacional e indexada sobre un motor de base de datos MySQL que contiene información tanto transaccional como de los clientes. Además de diversos scripts para el manejo de estos datos y para la creación de la sintaxis para AMPL.

6.3 Selector de información

La metodología de análisis busca ser flexible en cuanto a la información que utilizará para realizar el estudio. Esta flexibilidad podrá ser aprovechada por el usuario, lo que le dará la facultad de seleccionar precisamente cuales son los datos de input del estudio. Para esto se creará una interfaz donde el usuario tendrá la opción de filtrar la información transaccional que se desee utilizar para el estudio.

Esto permitirá seleccionar las transacciones de distintos negocios y categorías de productos. Más aun, gracias a la estandarización se podrá seleccionar las transacciones de un segmento en particular de clientes.

Por ejemplo, se podrá seleccionar una categoría de productos completa dentro de una tienda solo una parte de esta. Además la selección rápida de la información es fundamental debido al dinamismo del negocio, ya que se espera que los resultados del estudio de una categoría en particular debiesen cambiar de un periodo a otro. Por lo mismo tanto la estandarización como el selector de información permitirá replicar el estudio para distintos periodos de tiempo.

6.4 Métodos estadísticos y exploratorios

Una vez que se cuenta con la información transaccional con la que se trabajará, el siguiente paso es obtener las diversas intenciones de compra de los clientes. Para esto se utilizaran diversos métodos estadísticos y exploratorios que permitirán agrupar los productos.

A pesar de que el proceso para obtener las diversas intenciones de compra se abordará en profundidad en el capítulo 0, resulta importante destacar que para esto se calculará la distancia entre productos utilizando la Confianza entre ellos¹⁴. Una vez que se tiene la matriz distancia entre productos, se prosigue a utilizar el paquete estadístico SAS para realizar un escalamiento multidimensional (MDS) que permita encontrar factores que expliquen la cercanía entre los productos. Para que el MDS sea un procedimiento estadísticamente valido será fundamental cuidar el stress de la matriz generada. Para lograr esto es fundamental trabajar con una gran cantidad de transacciones y así contar con distancias entre productos representativas, además de esto será necesario realizar un *trade-off* entre dimensiones del MDS y eficacia del procedimiento. Ya que si bien a mayor número de dimensiones menor es el stress de la matriz final, también al aumentar el número de dimensiones se produce una suerte de sobre ajuste a los datos y menor es el valor del procedimiento¹⁵.

¹⁴ Para mayor información sobre la Confianza revisar la sección 5.4.1 Indicadores de asociación.

¹⁵ Si el número de dimensiones es igual al número de productos, el stress de la matriz será cero pero el procedimiento no habrá generado información extra.

Teniendo el escalamiento multidimensional realizado y con un stress de la matriz aceptable (menor al 10%) se realizará un análisis jerárquico para agrupar los productos en base a las dimensiones obtenidas del escalamiento multidimensional. Los resultados obtenidos del análisis jerárquico se utilizarán para la creación de un dendrograma dinámico¹⁶, que permite seleccionar la distancia de agrupación que se utilizará para definir el corte que generará los grupos de productos que representarán las diversas intenciones de compra.

Esta agrupación de productos servirá para aumentar la significancia entre las relaciones de los productos, ya que si bien la metodología considera las correlaciones existentes entre todos los productos, esta considera de distinta forma los productos que se encuentran dentro de una misma intención de compra de los que se encuentran fuera de ella.

Teniendo finalmente agrupado los productos en base a las intenciones de compra de los clientes se proseguirá a realizar el análisis de canasta utilizando estos grupos de productos. Los paquetes estadísticos no implementan los algoritmos de análisis de canasta, por lo cual será necesaria la implementación de algoritmos que sean capaces de obtener resultados en la gran cantidad de datos que se tiene.

El análisis de canasta por su parte será explicado en más detalle a continuación debido a su complejidad y que no existen paquetes estadísticos que implementen estos algoritmos.

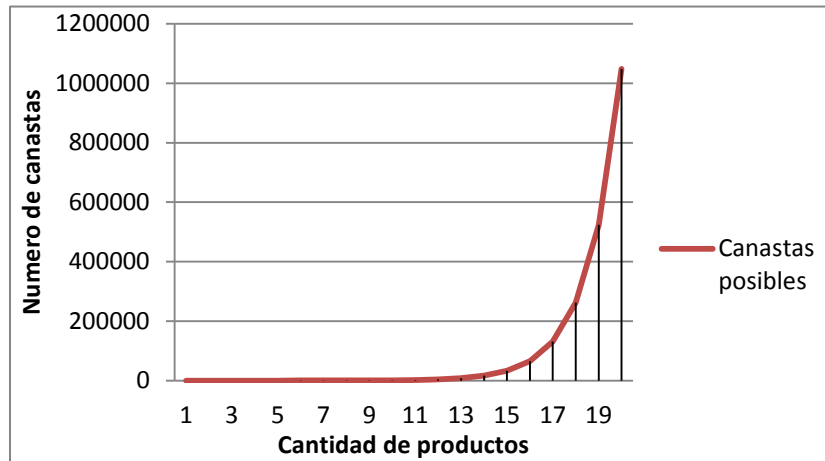
6.5 Análisis de Canasta

El análisis de canasta, en particular las técnicas de minería de dato que buscan correlaciones entre todos los productos y grupo de productos, permiten asociarlos en distintas canastas de compras llamadas canastas frecuentes o *frequent sets*. Debido a la gran cantidad de productos que se manejan en una tienda, encontrar todas las canastas frecuentes que se pueden construir podría llegar a ser una ardua e inclusive una imposible tarea (debido a capacidades de cómputo). Uno de los problemas del análisis de canasta es que la cantidad de canastas que se pueden generar con n productos es prácticamente igual a dos elevado a n , luego la complejidad (en términos de procesamiento de datos) crece exponencialmente, como se puede apreciar en la

Ilustración 4. Que con solo 20 productos se pueden formar más de un millón de canastas distintas, lo cual no es complicación gracias a los computadores actuales. Pero si se tienen 40 productos, se ese número aumenta a 1 billón de canastas distintas, lo cual si es un problema en términos de complejidad computacional. Y obliga a utilizar algoritmos eficientes para encontrar canastas frecuentes.

¹⁶ La Ilustración 7 muestra el dendrograma dinámico que se utilizará en la metodología.

Ilustración 4: Cantidad de canastas que se pueden formar a partir de un número de productos



Fuente: Elaboración propia.

Debido que obtener billones de canastas frecuente resulta sumamente complejo y más aún poco útil en términos de negocio, se trabajará solo con canastas que presenten un alto nivel de Soporte, es decir, que en términos de probabilidades, sea probable que alguna canasta de compra contenga todos los productos de la canasta frecuente.

Existen un gran número de algoritmos que encuentran frequent sets de entre las transacciones. Estos algoritmos difieren principalmente en la forma de abordar el problema y en como encuentran las canastas frecuentes.

El algoritmo más utilizado para generar los indicadores de asociación y encontrar canastas frecuentes es *Apriori* (Srikant, 1994), el cual como input recibe una solución factible. Por lo tanto es necesario haber definido de antemano las canastas tipos sobre las cuales se obtendrá la información, perdiendo así la posibilidad de que el algoritmo encuentre relaciones fuertes inesperadas. Para resolver el problema de ingresar de antemano el set de ítems sobre los cuales se desea obtener información, el algoritmo *Eclat* (Zaki, 2000) realiza una búsqueda exhaustiva inicial y determina posibles candidatos de sets de productos iniciales. Ahora bien, si el problema es que la cantidad de datos a trabajar es demasiado grande y esto genere un alto tiempo de cómputo, el algoritmo *FP-Growth* (Jiawei Han, 2004) compacta la información y esto disminuye tiempos de computo.

Todos los algoritmos previamente enunciados requieren ser calibrados para la información en particular a utilizar. Además de esta calibración, como se dijo con anterioridad es necesario realizar un *trade-off* entre el tiempo de cómputo, la complejidad del algoritmo y la eficiencia y efectividad de este.

Debido a la naturaleza del retail actualmente, se trabajará constantemente con una gran cantidad de transacciones y a la vez estas compuestas por una gran diversidad de productos. Por lo mismo se trabajará con el algoritmo Apriori pero en caso de tener un gran número de transacciones y productos se aprovecharán las ventajas del algoritmo FP-Growth.

6.6 Implementación de PROFSET

Como se habló con anterioridad, el software que se utilizará para implementar el modelo PROFSET será en su primera para un código creado en Java que valore las canastas frecuentes encontradas en el paso anterior. Para luego realizar la optimización entera en el software AMPL. Debido a la complejidad y gran cantidad de productos que se analizarán, es que también se creará un script en java que genere la sintaxis y los datos que requiere AMPL para poder ejecutar la optimización.

Si bien el problema de optimización es relativamente simple, puede resultar útil recalcar la dificultad de implementar correctamente los “sets de sets” en el lenguaje de AMPL. Un set no es más que una indexación de alguna *identidad*¹⁷, como puede llegar a ser en este caso un set de todos los productos. Luego en este trabajo las diversas *identidades* que se tiene son: productos, canastas frecuentes y categorías. Los productos tienen asociado un costo de almacenaje y corresponden a un set de todos los productos. Las categorías son sets que indexan a los diversos productos que a la vez son sets. Y finalmente las canastas frecuentes, las cuales tienen asociadas utilidades y cada canasta frecuente puede tener uno o más productos, es decir, que un producto se encuentra en el set de los productos pero a la vez en los sets de set de las canastas frecuentes.

6.7 Validación de la metodología

Una vez que se cuente con la metodología creada y las herramientas necesarias para ejecutarla de manera rápida y eficiente, esta será puesta a prueba en la categoría de Lácteos del supermercado Tottus. Con lo cual se espera determinar la eficiencia de la metodología tanto en tiempo de ejecución como en resultados obtenidos.

Para permitir mejores resultados y conclusiones del testeo del AMP, se analizará particularmente el supermercado Tottus ubicado en Avenida Kennedy 5601. Se eligió esta tienda debido a que es la que mayor utilidad genera para la cadena y además por presentar el mayor número de productos distintos. Lo cual permitirá al encontrar productos valiosos permitir la inserción de estos en las diversas tiendas de Tottus.

Para encontrar las intenciones de compra de los clientes en esta tienda, se utilizará la información transaccional de tres meses del año 2010 de este supermercado. Una vez encontrada las diversas intenciones de compra, se estudiarán estas en busca de posibilidades de rediseño del mix, y se evaluará el impacto de estas posibles acciones.

¹⁷ Una identidad es un objeto tanto físico como por ejemplo un producto o incluso un objeto abstracto como una canasta frecuente.

VII. ALCANCES

Los alcances del proyecto delimitan la frontera de la cual se ceñirá el presente trabajo, al definir el enfoque del estudio y su aporte particular al rediseño del mix de productos junto con la forma en que será puesto a prueba.

Este trabajo permitirá el estudio del mix de productos de los tres principales negocios del grupo Falabella. Permitiendo además el seleccionar específicamente que tiendas se considerarán para el estudio y que productos. Logrando así la creación de una metodología que se adapte fácilmente a las distintas necesidades de cada negocio y tienda y que a su vez sea replicable fácilmente.

El enfoque principal de la metodología se centra en la evaluación del desempeño de cada producto desde un enfoque microeconómico. Es decir, considera el aporte de cada producto a las utilidades del negocio. Pero llevando el análisis más allá de la mera consideración de las ventas de cada producto, sino también al incluir en la evaluación del desempeño de cada producto el efecto de venta cruzada que genera cada uno.

Para conseguir lo anterior en primer lugar se descubren las intenciones de compra que cada negocio satisface para luego encontrar dentro del mix de productos las diversas canastas frecuentes de productos que los clientes compran. Para determinar las intenciones de compra se utiliza información transaccional para determinar las *distancias* entre los productos para luego apoyado de un dendrograma y el juicio de expertos, determinar las diversas intenciones de compra. Y es en el estudio de estas canastas frecuentes donde se captura el efecto de la venta cruzada de cada producto.

A modo de probar la metodología, es que esta será puesta a prueba en la tienda de Kennedy del supermercado Tottus. Utilizando la información transaccional generada por esta tienda durante los tres primeros meses del año 2010. Debido a limitantes en capacidad de cómputo que se abordarán más adelante en este trabajo, es que solo se considerarán las categorías de alimentos del supermercado. Luego el estudio considerará solo las correlaciones existentes entre productos pertenecientes a la categoría de alimentos del supermercado¹⁸.

A diferencia de otros trabajos realizados en el Departamento de Ingeniería de la Universidad de Chile sobre rediseño de mix de productos, el presente no incluye clúster de tiendas, por lo que la totalidad de la información obtenida en la tienda de Kennedy no podrá ser extrapolada a otras. Aun así como se verá más adelante se utilizarán resultados de otros trabajos para aprovechar el conocimiento obtenido en la tienda en estudio para realizar inferencias sobre otras tiendas.

Como se señaló con anterioridad, la primera parte de la metodología logra descubrir las diversas intenciones de compra. Información que es utilizada para gestionar el mix de productos ya no desde una mirada clásica de categoría de productos sino que desde la perspectiva de intenciones de compra.

Junto con lo anterior es que al momento de realizar el estudio PROFSET este considera de distinta forma a los productos dentro de la intención de compra en estudio a los que se encuentran fuera. Ya que los que se encuentran en el estudio son considerados sin agregación (a nivel de SKU), mientras que los demás productos son agregados a un nivel según la jerarquización que

¹⁸ Para mayor información sobre la categoría de alimentos del supermercado ver la sección 9.2 de este trabajo.

posee el negocio. Esto como se verá, ayuda a generar un mejor estudio de los productos en análisis, aumentando la capacidad de determinar el efecto de venta cruzada que posee cada producto.

Si bien durante el presente trabajo será necesaria la creación de diversos scripts para poder trabajar con los datos, no se aborda ni el diseño ni la creación de una plataforma que incluya todos los scripts que se crearán. Por otro lado, en lo que respecta al motor de base de datos, si bien se señala cual es el utilizado, no se hace mayor profundización en el modelo de datos que se utilizará para automatizar los diversos procesos.

VIII. ANÁLISIS DEL MIX DE PRODUCTOS (AMP)

En este capítulo se definen las tres principales etapas de la metodología que se utilizará para analizar el mix de productos, haciendo especial desglose en cada una de ellas. En conjunto las diversas etapas permiten entre otras cosas analizar los diversos productos dentro de un mix considerando el efecto que presentan las correlaciones de cada producto con los otros de la tienda.

8.1 Etapas del AMP

AMP es una metodología que permite abordar diversos problemas de marketing, al tomar en consideración las correlaciones que se generan entre los productos, capturando así el efecto *cross selling* de estos. Para conseguir esto se hace uso de diversas técnicas y herramientas computacionales que facilitan la valorización de estas correlaciones.

El proceso completo comienza utilizando la información transaccional de los clientes para descubrir las intenciones de compra de estos, luego encuentra las canastas frecuentes que se realizan en la tienda y finalmente valorizar los productos en base a sus correlaciones. Si bien el proceso completo tiene un fin en particular, sus partes generan diversa información muy valiosa para el negocio.

AMP en su completitud se compone de tres etapas fundamentales, las cuales a su vez están divididas en sub etapas para facilitar la comprensión del análisis y para poder hacer referencia a los *inputs* y *outputs* de cada una.

A continuación se describe cada una de las siete sub etapas del AMP, haciendo referencia a la información que utiliza cada una, la información que genera y la utilidad de la etapa tanto para el proceso en su totalidad como para el negocio. En la Ilustración 5 se muestran las siete etapas que se verán a continuación.

Ilustración 5: Las 7 sub etapas del AMP.



Fuente: Elaboración propia.

8.2 Descubrir intenciones de compra

Como se ha hablado con anterioridad en este trabajo, las intenciones de compra incluyen por un lado la complementariedad entre productos y a su vez los ciclos de compra de estos¹⁹. Lo cual permite agrupar de manera inteligente los productos ya que permite agruparlos según cómo es que el cliente compra en la tienda.

Si bien el descubrir las intenciones de compra es útil para el negocio como información en sí, esta información también será útil para el proceso completo ya que permite aumentar la significancia entre los productos de las canastas frecuentes que se obtienen más adelante.

Descubrir las intenciones de compra permite al retail gestionar el mix de productos agrupando estos según cómo es que los clientes compran los productos y no por las jerarquías de estos. Lo que permite entre otras cosas acotar el problema y entender cuál es el horizonte del problema. Es por esto que en este trabajo se busca caracterizar las distintas intenciones de compra y estudiar a los productos que la componen junto con la interacción de estas.

Otra utilidad que se le da a las intenciones de compra descubiertas mediante este método es mezclar esta información con una segmentación previamente hecha de los clientes. Para así poder estudiar cómo es la penetración en cada segmento de clientes de las distintas intenciones de compra, lo que permitiría observar si efectivamente el mix de productos satisface las necesidades de los segmentos de clientes. Este estudio sacaría a la luz diversas oportunidades de rediseño de mix, por ejemplo el incluir productos de marca propia en las intenciones de compra que no estén llegando ha determinado segmento.

Ahora centrándose en la utilidad de descubrir las intenciones de compra para el proceso completo, este sirve fundamentalmente como se dijo con anterioridad para aumentar la significancia entre las relaciones de las canastas frecuentes. Esto para mejorar la medición de la correlación entre productos y así mejorar lo hecho por Tom Brijs en su trabajo “Building an association rules framework to improve product assortment decisions”.

Brijs se desilusionó en su trabajo al detectar que el efecto de correlación entre productos fue menor al que él estimaba. Creemos que su metodología subestimaba este efecto al no considerar todas las canastas frecuentes relevantes para un producto, ya que como se vio, la limitante en términos de capacidad de cómputo del modelo PROFSET, es la minería de las canastas frecuentes. Lo que obligó a Brijs a no trabajar con todas las canastas frecuentes posibles.

¹⁹ Para más información sobre las intenciones de compra de los clientes, ver el capítulo: **¡Error! No se encuentra el origen de la referencia..**

Lo que se propone para solucionar este problema es trabajar con algunos productos a nivel de SKU (lo más desagregado posible), y los otros trabajarlos a nivel agregado. La idea y justificación de este procedimiento queda clarificado al más fácilmente con un ejemplo. Si queremos gestionar la categoría de salsas de tomate por ejemplo, será necesario trabajar estas a nivel de SKU, luego si queremos valorar las correlaciones de la Salsa de tomate Malloa formato de 200grs (SKU: 2634123) debemos trabajar y valorizar todas las canastas donde la Salsa de tomate Malloa esté presente. Será fácil notar que probablemente si también trabajamos con los tallarines a nivel de SKU, se encontrarán todas las canastas frecuentes donde la Salsa de tomate Malloa cruce con los tallarines, debido a que estas canastas presentan una frecuencia alta por lo que el algoritmo *A priori* las considerará. Pero, ¿Si trabajamos con las masas de pizzas a nivel de SKU, estas canastas serán consideradas?, el punto es que probablemente la relación existente entre la salsa de tomate Malloa y cada una de las masas de pizza tenga un soporte muy bajo (y no será extraída del algoritmo *A priori*), pero al considerar de esta manera las correlaciones estemos subestimando una correlación que efectivamente existe, que es la salsa de tomate Malloa de 200grs con las masas de pizza en términos general, y más aún se esté dejando de considerar una correlación que es intuitivamente lógica. Por lo tanto, el agrupar las distintas masas de pizza ayudaría a aumentar la significancia de las relaciones de las salsas de tomate que se desean gestionar.

Ahora el problema es (continuando con el ejemplo), si queremos gestionar las salsas de tomate, ¿qué productos agrupar y cuáles no? Lo ideal sería no agrupar ningún producto y trabajar con todos a nivel de SKU, pero como ya se vio esto es una limitante por temas computacionales. Más aun, para trabajar con todos los productos a nivel de SKU y que sus relaciones sean significativas y no mera casualidad sería necesario trabajar con muchas transacciones, volviendo a caer en el problema de necesitar más capacidad de procesamiento.

Por lo tanto, si queremos gestionar una categoría en particular caemos nuevamente en la pregunta de: ¿qué productos agrupar y cuáles no? Para solucionar esto se plantea utilizar la agrupación de las intenciones de compra. Es decir, si queremos gestionar una categoría en particular trabajaremos con todos los productos pertenecientes a su misma intención de compra a nivel de SKU, mientras que los que se encuentran fuera la intención de compra serán agrupados lo mínimo posible utilizando la tabla jerárquica de la tienda en cuestión.

A continuación se describirán las tres distintas etapas necesarias para lograr descubrir las intenciones de compra. En primer lugar será necesario seleccionar la data con la que se va a trabajar, luego realizar un escalamiento multidimensional de los datos, para finalmente realizar un análisis jerárquico.

8.2.1 Selección de Data

En primer lugar y tal como se ha señalado en reiteradas veces a lo largo de este trabajo, para lograr llevar a cabo toda la metodología se necesita la información de las transacciones de diversos clientes. Pero es muy probable que el estudio a realizar se desee hacer solo con una porción de las transacciones, y así poder estudiar un sub segmento del problema completo. Por lo tanto, será necesario filtrar y seleccionar cuidadosamente la información a utilizar.

La información transaccional de los diversos negocios será estandarizada y almacenada en una base de datos, lo que facilitará y agilizará la obtención de estos. Esta estandarización genera que de cada transacción²⁰ se tenga:

- a) RUN del cliente: El RUN del cliente se obtiene solo en caso de que este haya realizado la transacción con un medio de pago identificable, que generalmente es la tarjeta del holding CMR. Esta información permite poder asociar la transacción a un cliente, con lo cual entre otras cosas permite agrupar las diversas compras que un cliente realiza en diversas ocasiones de compra.
- b) Fecha de la transacción: La fecha en cuando fue realizada la transacción. Esto permite estudiar diversos periodos de tiempo.
- c) Negocio: El negocio donde fue realizada la transacción. Los tres negocios principales son: Falabella, Sodimac y Tottus.
- d) Tienda: La tienda donde fue realizada la transacción, permitirá estudiar tiendas en particular dentro de un negocio.
- e) Número de boleta: Toda transacción está asociada a un código único que permite agrupar los diversos productos que un cliente compra en determinada ocasión de compra.
- f) SKU del producto: Cada producto posee un identificador único el cual será almacenado para saber exactamente cuál es el producto al que la transacción está asociada.
- g) Categorización jerárquica del producto: La gran cantidad de productos que se ofertan en los distintos negocios están agrupados jerárquicamente. A pesar que los distintos negocios utilizan distintas jerarquías e incluso poseen distinta cantidad de niveles, se estandariza la jerarquización tomando como convención que el nivel más mayor agregación de todo negocio es el cero, y luego los niveles superiores presentan menor agregación. El contar con esta información permite trabajar a distinto nivel y así gestionar por ejemplo categorías de productos.
- h) Monto cancelado: Es el precio final que el cliente debe pagar por el producto tras haber hecho efectivo (en caso de existir) los descuentos del productos.
- i) Descuento: Representa el monto descontado al cliente del precio de lista.
- j) Medio de pago: Es un código identificativo del medio de pago que utiliza el cliente al momento de cancelar la compra. Este puede ser: Pago en efectivo, cheque, tarjeta de crédito, débito, tarjeta CMR o alguna otra tarjeta.

Esta base de datos se puebla día a día a medida que los clientes realizan transacciones en los diversos negocios. Lo que permite contar en todo momento con la información al día de las ventas en los distintos negocios.

Para poder trabajar con toda esta información de forma ordenada es indispensable la creación de filtros que permitan seleccionar la información específica con la que se quiere realizar el estudio. Para filtrar se hará uso de la información extra de cada transacción, es decir, la fecha de transacción, el medio de pago, etc.

Se busca que los filtros sean simples y a la vez respondan a las distintas necesidades de estudio, por lo que se han definido seis filtros distintos para seleccionar la información. Para

²⁰ En este contexto una transacción se define como cada producto comprado por un cliente en una ocasión de compra. Por lo que una boleta puede tener una o más transacciones asociadas.

cualquier estudio no es necesario utilizar todos los filtros, por lo que estos solo serán utilizados en caso de ser necesarios. A continuación se enumeran y se explican los distintos filtros que permitirán seleccionar la información que se utilizará para llevar a cabo toda la metodología.

- 1) Negocio: En primer lugar se selecciona el negocio que se desea estudiar. En general y debido a que se gestionará el mix de productos, surge lógico estudiar un negocio en particular. Pero esto último no siempre es cierto, ya que debido a la naturaleza del negocio de CMR, sería interesante realizar un estudio entre categorías de distintos negocios para estudiar el comportamiento de compra de los clientes con CMR. Esto serviría por ejemplo para diseñar una campaña de marketing que estimule la penetración de la tarjeta CMR, activando a los clientes en los distintos negocios.
- 2) Tiendas del estudio: Una vez que se seleccionó el negocio o negocios donde se realizará el estudio, se prosigue a seleccionar las tiendas de las cuales se extraerán las transacciones. Cabe destacar, tal como se vio en los alcances, que en este trabajo no se realizará una segmentación de las tiendas, que pudiese ayudar a entender de mejor manera las diferencias entre estas, y además permitiría trabajar al mismo tiempo con más de una tienda. Esto último al asumir que las tiendas pertenecientes a un mismo segmento poseen clientes que se comportan de manera similar, por lo que sería interesante aprovechar el conocimiento y la información de una tienda en particular para gestionar a todas las del segmento. Por lo que en general lo que se realizará en este estudio es seleccionar solo la información transaccional de una tienda para luego proseguir con el resto de la metodología.
- 3) Fecha de inicio y término: Luego la información es filtrada por fecha de inicio y término del estudio, logrando extraer solo las transacciones que se efectuaron entre ambas fechas. Poder filtrar por fecha de las transacciones es fundamental debido al dinamismo del retail, ya que un mismo estudio (con los demás filtros iguales) a distinta fecha arroja distinto resultado debido principalmente al dinamismo del retail.
Por otro lado, el poder filtrar por fecha en la cual se realizó la transacción, permite estudiar por ejemplo el efecto de determinadas promociones al estudiar la composición de las canastas durante la promoción y contrastar este resultado con la composición de estas tanto antes y después de la promoción.
Además de poder seleccionar la fecha de inicio y término del estudio, en ocasiones puede resultar necesario el excluir o incluir manualmente algunas fechas. Esto resulta particularmente útil al momento de mitigar los efectos anómalos producto de fechas particulares como por ejemplo el día de la madre. Esto último es por el lado de la exclusión de fechas del análisis, pero también puede resultar útil el incluir determinadas fechas al análisis. Supongamos por ejemplo que se desea estudiar la composición de las canastas para las quincenas y días últimos de cada mes. En este caso resulta útil el poder incluir solo determinadas fechas en el análisis, en vez de seleccionar el periodo de estudio y excluir todas las fechas que no interesan.

- 4) Medio de pago: El medio de pago es de suma importancia para el negocio de CMR, por lo que poder filtrar bajo este criterio es indispensable. Luego este filtro permite seleccionar las transacciones según el medio de pago que utilizó el cliente, pudiendo así discriminar entre transacciones pagadas con la tarjeta CMR, cheque, efectivo o alguna otra tarjeta.
Este filtro además de permitir contrastar el comportamiento de los clientes según el medio de pago permitiría, al seleccionar solo las transacciones realizadas con la tarjeta CMR, asociar la transacción a un cliente en particular. Ya que como se ha señalado con anterioridad, se posee el RUN del cliente que realizó la transacción y además información de este obtenida al momento de que el cliente obtuvo la tarjeta.
- 5) Nivel del estudio: Ahora que se cuenta con los diversos filtros que permiten seleccionar, el negocio, las tiendas, las fechas y el medio de pago de las transacciones a estudiar, no hay que olvidar que lo que buscamos es gestionar el mix de productos. Luego es necesario determinar filtros que permitan estudiar los productos pertenecientes a las distintas tiendas.
El primer acercamiento a esto será determinar el nivel al cual se realizará el estudio. El concepto de nivel está directamente asociado con la clasificación jerárquica que se utiliza en los distintos negocios. Recordando que el nivel más desagregado que se trabaja es el nivel de SKU, donde los productos pueden ser diferenciados incluso por el formato de los productos. Muchas veces se querrá obtener resultados a niveles más agregado de productos, por lo que resultaría útil aprovechar la jerarquía que cada negocio tiene para agregar los productos y así obtener resultados a un nivel superior.
A la selección de a qué nivel se tendrán los resultados es a lo que se le llama nivel de estudio, y como se pudo apreciar, se ciñe a la jerarquía utilizada por cada negocio.
- 6) Filtros jerárquicos o de SKU: Finalmente surge indispensable estudiar solo algunos productos dentro del amplio mix que presentan las tiendas. Para que la selección de los productos a estudiar no sea una tarea titánica, de seleccionar los productos a utilizar dentro de los miles de productos, se utilizarán filtros nuevamente aprovechando la jerarquía de productos que posee cada negocio.
Para que esto sea simple y a la vez versátil, los filtros que se utilizarán aquí serán en los niveles jerárquicos que sean iguales o más agregados que el nivel de estudio previamente seleccionado.
Por lo tanto, si por ejemplo se decidió trabajar a nivel de sub-categoría, los filtros que se utilizarán serán a nivel de sub-categoría y/o a nivel de categoría, pero no a un nivel más desagregado que sub categoría, por ejemplo a nivel de SKU.

Todos los filtros previamente descritos, le dan amplia facultad al estudio para seleccionar adecuadamente la información con la que desea trabajar. Pudiendo por ejemplo filtrar la información para realizar estudios tales como: “El desempeño de todos los productos de la subfamilia *Lácteos, cremas y mantequilla* pero no incluir los de la sub categoría *cremas*, del

supermercado Tottus ubicado en Colina durante lo que va de este mes”, o bien simplemente “El desempeño a nivel de Familia de todas las tiendas de Falabella los últimos 3 meses”.

Luego esta información filtrada será la que saldrá de output de esta etapa y será utilizada por las etapas posteriores. Para lo cual será necesaria la creación de scripts que permitan realizar los diversos filtros y así seleccionar efectivamente las transacciones para el estudio.

8.2.2 Escalamiento multidimensional

La etapa en donde se realiza escalamiento multidimensional es la primera de las dos etapas que permiten encontrar las diversas intenciones de compra de los productos que se estudiarán. En esta etapa se busca encontrar las correlaciones existentes entre todos los productos al estudiar las transacciones, utilizando en primer lugar el indicador de Confianza²¹ entre productos. Para luego realizar un escalamiento multidimensional para encontrar los diversos factores que expliquen estas correlaciones.

8.2.2.1 Determinar alcance de la canasta de compra

Antes de comenzar a estudiar las canastas de compra de los clientes, es necesario determinar el alcance de una canasta de compra. Por alcance nos referimos a la cantidad de ocasiones de compra que se considerarán para construir una canasta de compra.

La opción más simple y probablemente la más utilizada en supermercados es considerar que cada ticket de compra es una canasta de compra particular. Pero para formatos de negocio donde la frecuencia de compra es mucho menor como es el caso de tiendas por departamento (Falabella) o mejoramiento del hogar (Sodimac), surge relevante estudiar no solo una ocasión de compra sino más bien una seguidilla de ocasiones de compra durante un periodo. Y de esta forma considerar como canasta de compra de cada cliente las compras que el realiza en diversas ocasiones de compra durante un periodo, capturando su comportamiento en el tiempo.

Para construir las canastas de compra considerando solo una ocasión de compra, se hará uso del campo Numero de boleta, para lo cual una canasta de compra será un ticket de compra. Ahora si lo que se busca es agrupar varias ocasiones de compra, será necesario contar con transacciones identificables. Para tener transacciones identificables será necesario que el cliente al momento de cancelar lo haya hecho con la tarjeta CMR. Para luego agrupar todas las transacciones de un periodo utilizando el campo RUN, y así considerar como canasta todas las transacciones realizadas por un cliente en un periodo de tiempo.

Cabe destacar que si se decide considerar la agrupación de ocasiones de compra de un cliente, existirá un error en la canasta de compra final si es que el cliente realiza sus transacciones con distintos medios de pago, ya que solamente se considerarán las hechas con la tarjeta CMR.

8.2.2.2 Correlación entre productos

Una vez que se cuenta con las transacciones agrupadas según canasta de compra se prosigue a calcular las distintas correlaciones entre productos. Se dice que dos productos presentarán mayor correlación si aparecen constantemente juntos en una canasta de compra.

²¹ El indicador de confianza y otros de análisis de canasta se pueden encontrar en la sección 5.4.1 Indicadores de asociación.

Luego nuestro objetivo es construir una matriz de $N \times N$ (donde N es el número de productos a estudiar) que contenga la correlación existente entre todos los productos, y luego calcular un indicador de asociación (que en nuestro caso será la Confianza) para todos los pares de productos.

Para construir esta matriz es necesario recorrer todas las canastas de compra enumerando la cantidad de veces que dos productos aparecen en una determinada canasta de compra. Supongamos un ejemplo donde solo se tiene seis productos (A, B, C, D, E, F) y siete canastas de compra. Para iniciar el proceso será necesario tener un archivo como el que aparece en la Tabla 4, donde cada línea representa una canasta de compra y en cada celda un 1 en caso de que el producto está presente en la canasta de compra y un 0 si no.

Tabla 4: Ejemplo de siete canastas de compra para seis productos

	A	B	C	D	E	F
1	0	1	1	0	0	
0	0	0	0	1	1	
0	0	0	0	1	1	
0	1	1	1	1	0	
1	0	1	1	1	0	
1	1	1	1	1	0	
1	0	1	1	0	0	

Ahora se busca construir la matriz con la frecuencia en que dos productos se compran en conjunto. Pero será necesario estandarizar la frecuencia en que aparecen juntos dos productos por lo que se verá a continuación.

Debido a que las relaciones entre productos (supongamos A, B, C y D) son bilaterales (AB, AC, AD, BC, BD y CD) se puede demostrar que el número de relaciones bilaterales entre N productos es igual a $N*(N-1)/2$. Como consecuencia a lo anterior es posible que productos con igual frecuencia en canastas de compra tengan mayor frecuencia con los demás productos si es que estas provienen de canastas de compra con un mayor número de productos. Ya que tendrán un mayor número de correlaciones.

El razonamiento que está detrás de este problema es que las correlaciones existentes en una canasta de compra son menos intensas mientras mayor es el número de productos en la canasta de compra. Luego será necesario ponderar las correlaciones existentes en cada canasta de compra por un factor que considere la cantidad de productos existentes en una canasta de compra. Para solucionar este problema Böcker and Merkle proponen ponderar todas las relaciones entre dos productos por un factor de $1/(m-1)$ donde m es el número de productos presentes en la canasta de compra. En el Anexo A: Script para calcular las correlaciones entre productos se tiene el código que se utilizó para aplicar el factor previamente descrito y en la tabla 5 se tiene la matriz generada tomando como input la información de la Tabla 4. Cabe destacar que la diagonal de esta matriz representa el número de canastas que contienen a dicho producto.

Tabla 5: Correlación entre productos en base a la tabla 4

	A	B	C	D	E	F

A	4	0.25	1.58	1.58	0.58	0
B	0.25	2	0.58	0.58	0.58	0
C	1.58	0.58	5	1.92	1.92	0
D	1.58	0.58	1.92	5	0.92	0
E	0.58	0.58	1.92	0.92	5	2
F	0	0	0	0	2	2

La decisión de incluir este factor, en el cálculo de la frecuencia con que dos productos están presentes en una misma canasta, se debió a que con él el stress de la matriz que se obtenía al realizar el escalamiento multidimensional era menor que al no utilizarlo²². A continuación se prosigue a calcular el indicador entre relaciones de productos que representará la distancia existente entre ellos.

8.2.2.3 Matriz de Confianza entre productos

Una vez que se cuenta con el cálculo de la correlación ajustada por el factor de Böcker and Merkle. Se prosigue a calcular un indicador que represente la distancia existente entre cada par de productos, para que este sea el input del escalamiento multidimensional.

En este trabajo se utilizará el indicador de Confianza para estimar la distancia entre dos productos. La base detrás de esta elección es que gracias a que la Confianza, entre A y B por ejemplo, representa la probabilidad de que el producto B se encuentre en una determinada canasta sabiendo que esta tiene al producto A, la distancia entre dos productos será menor mientras mayor sea esta probabilidad. Luego si la probabilidad de que en la canasta se encuentre B sabiendo que se tiene A es muy alta, estos productos están muy cercanos.

Más aun, debido a que la distancia de A a B debe ser la misma que la distancia entre B a A, es que se utilizará como factor de distancia la máxima Confianza entre los dos productos. Y a modo de tener la distancia y no la cercanía es que se utilizará como factor final la fórmula:

Sea:

N_{ij} : La cantidad de canastas en la que se encuentra el producto i y j .

N_i : La cantidad de canastas en la que se encuentra el producto i .

$$D_{ij} = 1 - \frac{N_{ij}}{\text{Min}(N_i, N_j)}$$

²² El stress de la matriz final disminuía aproximadamente en 6 puntos porcentuales.

Por construcción esto genera una matriz simétrica entre los distintos productos, si utilizamos a modo de ejemplo la Tabla 4, y construimos con esta la matriz resultante, obtenemos la Tabla 6. En la cual cada fila y columna representa un producto, y la intersección es la distancia de esos productos.

Tabla 6: Matriz de distancia cruzada a partir de la tabla 5.

	A	B	C	D	E	F
A	0	0.88	0.6	0.6	0.85	1
B	0.88	0	0.71	0.71	0.71	1
C	0.6	0.71	0	0.62	0.62	1
D	0.6	0.71	0.62	0	0.82	1
E	0.85	0.71	0.62	0.82	0	0
F	1	1	1	1	0	0

Con esta matriz es posible utilizar el escalamiento multidimensional para disminuir la dimensión de esta y así poder encontrar factores que expliquen estas distancias.

8.2.2.4 Escalamiento multidimensional

Una vez que se cuenta con la matriz de distancia generada en base a las correlaciones de los productos, se prosigue a realizar un escalamiento multidimensional. Para realizar este procedimiento se utilizará el paquete estadístico SAS, el cual supera a SPSS en términos de mayor capacidad para configurar el método.

En primer lugar hay que tener en consideración que con la matriz inicial de distancias tenemos la representación de todos los productos incluidos en el estudio, distribuidos en un espacio de N dimensiones, donde N es el número de productos. Con el escalamiento multidimensional se pretende disminuir la dimensión del problema. Pero surge inmediatamente la pregunta respecto a cuál número de dimensiones a disminuir la matriz inicial, pregunta que debe ser respondida antes de realizar el procedimiento debido a que es uno de los inputs del problema.

El problema de disminuir demasiado la dimensión de la matriz es que al realizar esto, el stress de la matriz final aumenta considerablemente. Por otro lado, la literatura no especifica un método para determinar el número óptimo de dimensiones que se deban utilizar. Lo que si se discute es respecto al stress de la matriz final.

Se espera que el stress de la matriz sea el menor posible y al mismo tiempo obtener una matriz final con un bajo número de dimensiones. Pero tal como se vio en el marco teórico estos dos objetivos son opuestos ya que a mayor número de dimensiones, menor es el stress y por consiguiente, al buscar disminuir la cantidad de dimensiones se obtiene un aumento en el stress de la matriz final. Luego es imperante encontrar un equilibrio a esta problemática.

Para determinar el número indicado de dimensiones es necesario en primer lugar observar los datos. Y pronosticar cuál será el número de intenciones de compra que se obtendrán tras

realizar el análisis multidimensional y luego el jerárquico. Luego el número óptimo de dimensiones se encuentra cercano a este, generalmente un poco mayor, pero todo depende de la naturaleza de los datos y de la cantidad de datos que se utilizaron para construir la matriz de correlación.

En lo que respecta al stress de la matriz final, se espera que este no sea superior a un 10%. Es necesario tener cuidado de no aumentar mucho el número de dimensiones finales con el fin de obtener un stress de la matriz final menor al 10%. Si este último fuese el caso, será necesaria la revisión de la información que se utilizó para calcular las distintas distancias de los productos, viendo la posibilidad de aumentar el número de canastas que se utilizó para construir la matriz de correlaciones de productos, y con esto aumentar la significancia de las distancias.

La sintaxis utilizada en SAS posee varios parámetros siendo los más relevantes el criterio de convergencia, el número de iteraciones del algoritmo, las dimensiones y la forma de que el algoritmo de mínimos cuadrados estima los parámetros. A continuación se muestra una sintaxis tipo que se utilizó en el caso práctico que se verá más adelante y se describen las partes principales de esta.

Código VIII.1: Sintaxis tipo para realizar MDS en software SAS.

```
proc mds data=Tottus CONVERGE=0.01 ITER=100 SHAPE=TRIANGLE PFINAL
DIMENSION=10 ALTERNATE=ROW PCONFIG PFINAL;
      id id;
run;
```

- **Converge:** Determina el criterio de detención del algoritmo en base al gradiente de convergencia.
- **Iter:** Determina el máximo número de iteraciones que realiza el algoritmo.
- **Shape:** Determina si se almacena toda la matriz para trabajar con toda ella (*Square*), o solo una porción triangular de la matriz (*Triangle*).
- **Dimension:** Determina el número de dimensiones que tendrá la matriz final.
- **Alternate:** Determina la forma en que el algoritmo de mínimos cuadrados estima los parámetros. El parámetro por defecto es *None*, donde todos los parámetros son ajustados de forma simultánea. *Matrix* ajusta todos los parámetros del primer producto y luego todos los para el segundo producto. *Row*, se comporta similar a *Matrix* solo que divide el problema en n sub partes. En caso de no definir n este se calcula en base a la memoria disponible.

La los parámetros a utilizar para realizar el procedimiento van directamente ligados a la capacidad de memoria RAM (Random Access Memory) disponible para realizar el procedimiento. Hay que tener en cuenta que el proceso de escalamiento multidimensional junto con la obtención de las canastas frecuentes (que se verá más adelante), son los dos pasos de la

metodología más intensivos en términos de memoria requerida. Por lo cual hay que realizar en ambos un *trade-off* entre precisión y velocidad de cálculo.

Por lo que el definir los parámetros para este procedimiento debe hacerse cuidando en primer lugar si es que es posible realizar en términos de memoria, y en segundo lugar el tiempo que se demorará en terminar el procedimiento.

En términos de parámetros, a menor criterio de convergencia mayor es la cantidad de iteraciones a necesitar y por ende mayor será el tiempo de cómputo. Lo mismo sucede lógicamente al aumentar la cantidad de iteraciones que se necesitan. En términos de memoria, según la metodología AMP, siempre el parámetro *Shape* será *Triangle* con lo que se disminuye la cantidad de memoria en comparación a definir el criterio como *Square*. Ahora en base a las dimensiones, a mayor número de dimensiones mayor será cantidad de memoria requerida para llevar a cabo el procedimiento, pero menor será el tiempo de computo.

Una vez habiendo obtenido la matriz final, se prosigue a agrupar los productos utilizando un análisis jerárquico.

8.2.3 Análisis Jerárquico

La matriz que se obtiene de los procesos anteriores, donde los productos están ubicados en una posición en un espacio de N dimensiones, donde N fue elegida arbitrariamente en el paso anterior, servirá como input en este proceso para poder agrupar a los productos.

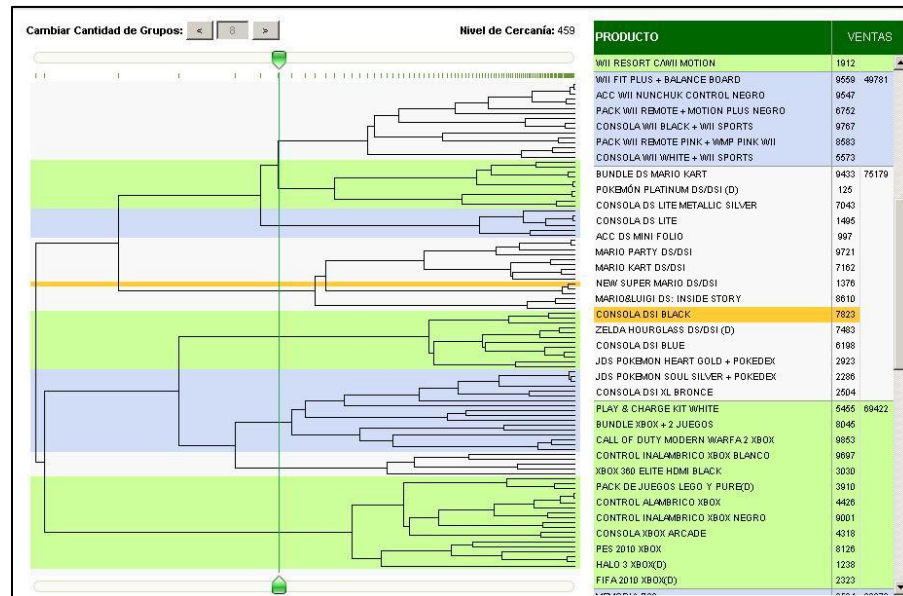
Para realizar este proceso es posible utilizar el software estadístico SPSS, el cual realiza el análisis jerárquico e incluso permite entregar en dendrograma para facilitar la interpretación de los datos. Pero debido a que este es el último paso, donde finalmente se obtendrán las intenciones de compra al agrupar los productos, es que se ha creado una interfaz web que permite al usuario cargar el dendrograma final y determinar en base a la distancia de agrupación cuantos segmentos formar.

Debido a que dicha interfaz requiere de un formato especial como input, es que el output del análisis multidimensional es entregado como input a un script²³ que realiza el proceso de determinar las distancias de agrupación de los distintos grupos de productos.

El output de este análisis es recibido por la interfaz web que muestra un dendrograma interactivo donde el analista puede variar la cantidad de grupos que desea formar. Un ejemplo del dendrograma se puede ver en la Ilustración 6: Dendrograma de la categoría video juegos en Falabella, visualizado en la interfaz web.

²³ Este script fue realizado por Daniel Wilson para CMR en lenguaje Java. El autor de este trabajo lo tradujo al lenguaje PHP y adaptó para realizar las tareas necesarias.

Ilustración 6: Dendrograma de la categoría video juegos en Falabella, visualizado en la interfaz web.



Fuente: Elaboración propia.

Una vez que se tiene el dendrograma se prosigue con determinar el número de intenciones de compra que se generarán. Esto se realiza determinando una distancia de aglomeración como corte para formar los segmentos.

Una vez que se cuenta con las diversas intenciones de compra se prosigue a seleccionar una intención de compra en particular para comenzar a trabajar con ella.

8.2.3.1 Selección de la intención de compra

A este punto de la metodología, se cuenta con los productos agrupados según las distintas intenciones de compra. Estas intenciones de compra representan productos que se llevan constantemente juntos en una misma canasta de compra.

Gracias a esto es posible estudiar las diversas intenciones de compra del mix de producto completo bajo diversos indicadores de rentabilidad y descriptivos de la propia intención de compra. Por otro lado es interesante estudiar la penetración que tienen las diversas intenciones de compra en la cartera de clientes, saber cuántos y luego cuáles son los clientes que compran ciertas intenciones de compra (*Reach*). Y además un indicador de que porcentaje de las canastas presenta cada intención de compra.

Además de estudiar las intenciones de compra por separado, surge importante, en particular para los pasos posteriores del AMP, estudiar la correlación que existe entre las intenciones de compra.

Finalmente junto con estudiar las diversas intenciones de compra como un todo, se busca caracterizar los productos que corresponden a cada intención de compra y observar en estos indicadores claves para el negocio como por ejemplo: Venta, número de unidades vendidas, ticket promedio, reach. Además de tener una visión en particular por cada producto, surge relevante tener una mirada de cuál es la participación de cada producto dentro de cada intención de compra.

Es necesario seleccionar una intención de compra para proseguir a trabajar con esta en los pasos siguientes, donde se encontrarán las canastas frecuentes y luego se valorizarán.

8.3 Determinar Frequent Sets y valorizarlos

Habiendo finalmente encontrado las distintas intenciones de compra y teniendo seleccionada una intención de compra para trabajar es posible iniciar la segunda parte del AMP. En esta etapa se aprovechará el conocimiento de las intenciones de compra generado en la primera parte del AMP para aumentar la significancia de las relaciones de los productos. Estas relaciones se capturarán gracias a la obtención de las canastas frecuentes que se minarán en esta parte y se valorizarán.

En esta segunda parte se utilizarán algoritmos de minería de datos para encontrar los *frequent sets* (FS) que incluyen a productos pertenecientes a la intención de compra seleccionada. Habiendo realizado lo anterior, se utilizará el algoritmo explicado en el marco metodológico para valorizar los diversos FS y finalmente se proseguirá a implementar el modelo PROFSET en la tercera parte del AMP.

8.3.1 Encontrar Frequent Sets

Los FS como ya se ha señalado, son grupos de productos (con cardinalidad uno o más) que se encuentran presentes en un número de canastas de compra superior a una cota fijada, esta cota se conoce por el nombre de soporte mínimo. Debido a que después de este paso se prosigue a valorizar (monetariamente) los FS, se busca encontrar aquí un gran número de FS, ya que los FS con bajo soporte tendrán tan bajo valor que no influirán significativamente en el procedimiento PROFSET.

Por lo anterior, se espera que el algoritmo encuentre el mayor número de frequent sets. Para que se encuentren FS significativo, se deberá trabajar con un gran número de transacciones, además a medida que mayor es la cantidad de productos con lo que se está trabajando, mayor deberá ser el número de transacciones que se necesitan.

- **Algoritmo Apriori**

Si bien no se pretende ahondar en profundidad en este trabajo en el algoritmo Apriori²⁴, si se verá el pseudocódigo de este. Ya que debido a que no existen paquetes estadísticos que implementen estos algoritmos, será necesaria la creación de scripts que lo implementen para poder hacer uso de estos.

El pseudo código del algoritmo Apriori se ve en el código 8.2 a continuación. Para entender el algoritmo, sea C_k un candidato para ser canasta frecuente con k productos y sea F_k el set de canastas frecuentes con k productos.

²⁴ Para mayor información respecto al algoritmo: (Tan, Steinbach, & Kumar, 2005)

En una primera etapa el algoritmo recorre todas las transacciones determinando el soporte de cada producto. Gracias a esta etapa ya se tiene F_1 , que corresponde a todas las canastas frecuentes que cumplen el soporte mínimo y poseen un producto. Luego, el algoritmo iterativamente genera los F_k utilizando los F_{k-1} ya generados.

Código VIII.2: Pseudo código del algoritmo Apriori

```

k=1
Fk = Todos los productos cuyo soporte es mayor que minsup
do{
    k = k+1;
    Ck = apriori_gen(Fk-1); //Genera todos los posibles
candidatos
    for cada transacción t en la base do{
        Ct = subset(Ck,t); //Identifica los candidatos en t
        for cada candidato c en Ct do{
            ++S(c); //Incrementa el conteo del soporte
        }
    }
    Incluir a Fk, Ct en caso de ser mayor a minsup;
}until( Fk=∅);
Resultado = Unión todos Fk;

```

- **Algoritmo FP-Growth**

A diferencia del algoritmo anterior que recorre la base de datos K veces, donde K es el número de productos en la canasta frecuente con más productos encontrada, FP-Growth recorre la base de datos solo dos veces y almacena la información en un árbol.

A grandes rasgos el algoritmo trabaja de la siguiente manera. En la primera lectura de la base de datos el número de ocurrencias de cada producto es determinado y se descartan los productos con bajo soporte. En la segunda lectura, las posibles canastas frecuentes son almacenadas en lo que se llama una estructura de árbol FP, de esta forma la información es podada y almacenada en la memoria. Buscando así almacenar toda la información útil en un árbol para encontrar patrones en la información de forma eficiente. Para mayores referencias sobre el algoritmo ver (Jiawei Han, 2004).

A continuación se muestra el pseudo código de este algoritmo que en este trabajo fue implementado en lenguaje Java.

Código VIII.3: Pseudo código del algoritmo FP-Growth.

```

Función FPGrowth(Tree, α)
    if Tree contiene algún camino P then
        for each β = comb. de nodos en P do

```

```

        Pattern =  $\beta \cup \alpha$ ;
        Sup = min(suporte de nodos en  $\beta$ );
    else
        for each ai en cabeza de Tree do
            patrongenerado =  $\beta \cup \alpha$ ;
            sup = ai.soporte;
            construir  $\beta$ 's patrón base condicional;
            FPTree = construir  $\beta$ 's conditional FP-tree;
            if FPTree !=  $\emptyset$  then
                FPGrowth(FPTree,  $\beta$ );

```

Habiendo utilizado alguno de los algoritmos previamente señalados, se tiene gracias a estos una gran cantidad de grupos de productos que se encuentran reiteradas veces en alguna canasta de compra. A la cantidad de veces que una canasta frecuente aparece en alguna canasta de las transacciones, se le llama soporte de la canasta frecuente. Y recíprocamente, para que un grupo de productos sea considerado como frecuente, este debe aparecer en un número de canastas mayor a un soporte mínimo pre fijado.

Ahora que se cuenta con esta gran cantidad de canastas frecuentes es necesario atribuirles a ellas algún valor económico para así poder proceder aplicar PROFSET. Para lo cual a continuación se describe el algoritmo utilizado para valorizar las canastas frecuentes.

8.3.2 Valorización Frequent Sets

Como se ha señalado anteriormente este trabajo logra capturar el efecto de correlación entre los distintos productos, aprovechando las técnicas de análisis de canasta para descubrir canastas frecuentes. Gracias a lo anterior solo es posible saber si un grupo de productos (con cardinalidad 1 o más) se encuentra presente en un número de canastas suficientemente alto, pero no se conoce el valor de cada una de estas canastas.

Para poder valorizar estas canastas frecuentes se utilizan las transacciones realizadas en determinado periodo, y se le atribuye la venta de cada transacción a algunas canastas frecuentes. El problema es que una transacción en particular puede contener más de una canasta frecuente, por lo cual el problema de atribuir la venta de una transacción a determinada canasta frecuente deja de ser trivial. Más aún si para determinada transacción las canastas frecuentes dentro de esta poseen una intersección no vacía, significa que algún producto de una transacción se encuentra en más de una canasta frecuente, por lo cual es necesario determinar algún criterio que discrimine a que canasta frecuente asignarle un valor.

Los criterios que se pueden utilizar como se vio en 5.5.2.2 Reglas de asignación son en base al soporte de las canastas o utilizando un modelo logarítmico. Del trabajo realizado por Brijs, se concluye que realmente no se obtienen resultados muy distintos al utilizar una u otra metodología para asignar la utilidad a las canastas frecuentes. Por lo que por motivos de simplicidad se utilizará el criterio de asignación en base al soporte. Cuyo pseudo código se puede

ver en el Código V.1: Pseudo código de algoritmo para asignar utilidad a las canastas frecuentes basándose en el soporte de las canastas frecuentes..

8.4 Selección de producto en base a sus correlaciones

Habiendo hecho todo lo anterior, en este momento se cuenta con una lista de canastas frecuentes valorizadas en base a la información transaccional de un periodo. Por lo que la venta generada por un producto en particular durante el periodo estudiado, se encuentra distribuida entre una gran cantidad de canastas frecuentes que incluyen a este producto. Logrando de esta forma valorizar el efecto *cross-selling* de cada producto.

Para aprovechar esta información se utilizará el modelo PROFSET el cual seleccionará los diversos productos maximizando la utilidad generada por las canastas frecuentes que se logran formar con los productos seleccionados. Para lo cual será necesario ajustar el modelo PROFSET según las necesidades del estudio y luego interpretar los resultados obtenidos.

8.4.1 PROFSET

El modelo PROFSET es configurable bajo distintos criterios: en las restricciones que utiliza, en el horizonte para considerar la utilidad de las canastas y en cómo se valoran las canastas. Esto permite resolver diversos problemas de marketing según el ajuste que se haga a estos criterios y según la información que se utilice para el modelo.

A continuación se verán diversas aplicaciones del modelo, explicando en cada una de ellas los criterios utilizados y cuál es el problema de marketing que buscan resolver.

La primera aplicación busca encontrar los productos más y menos rentables dentro de una intención de compra, considerando sus correlaciones con todos los productos de una tienda, para así aprovechar esta información y poder obtener productos candidatos a ser incluidos en otras tiendas.

La segunda aplicación del modelo PROFSET busca aprovechar el conocimiento obtenido de una tienda con gran surtido para generar el mix óptimo de productos que debiese tener una tienda de conveniencia. Teniendo en cuenta que el mix de productos necesario para una tienda de conveniencia si bien debe cubrir un diverso número de categorías, estas debiesen tener un menor mix. Por consiguiente, el modelo seleccionará los productos que maximicen las ventas al considerar el aporte individual de cada producto y a al mismo tiempo sus correlaciones con los demás productos.

Como tercera aplicación del modelo se verá un caso particular donde el modelo es utilizado para seleccionar los productos de una cuponera de descuentos, con el fin de generar un cupón de descuento que aplique a los productos seleccionados por el modelo. Con lo cual se logrará que los productos seleccionados una alta correlación con otros productos de la tienda, lo que se espera que el ticket de los clientes que utilizan los cupones no solo tenga al producto en descuento, sino que también otros productos de la tienda.

Para cada una de estas aplicaciones y basándose en las especificaciones del modelo PROFSET de la Tabla 3: Especificaciones del modelo PROFSET), se verá en ellas la optimización, la regla de asignación y las restricciones que utilizan para lograr su objetivo.

a) Productos más y menos rentables dentro de una intención de compra:

Esta aplicación del modelo PROFSET busca encontrar y a la vez ordenar los productos más y menos rentables. Donde por rentabilidad se toma en consideración el aporte a las ventas de cada producto en particular como también a las correlaciones que este genera con los demás productos de la tienda. Todo esto considerando a todos los productos pertenecientes una intención de compra calculada anteriormente, lo que permitirá gestionar los productos en base a su aporte en las ventas no solo considerando una categoría como se hace comúnmente, sino que abarcando diversas categorías según la composición de la intención de compra.

Para lograr esto y considerando los criterios de Tabla 3: Especificaciones del modelo PROFSET) que se presentan a continuación se tienen:

Optimización:

Si bien el modelo busca solo gestionar los productos que se encuentran dentro de una intención de compra, se aprovechará al máximo el potencial de él al considerar no solo las correlaciones existentes entre los productos dentro de una intención de compra, sino también las existentes entre estos y los productos que se encuentran fuera de la intención de compra.

Por lo que se desea que la función de optimización considere la utilidad de todos los productos que se encuentren fuera de la intención de compra, los cuales siempre serán seleccionados, y solo la utilidad de los productos seleccionados por el modelo dentro de la intención de compra.

Debido a esto, la optimización debe considerar como utilidad solo a los productos seleccionados por el modelo. Y será tarea de las restricciones del modelo determinar y restringir cuales productos se encuentran fuera de la intención de compra (por lo que deben estar siempre seleccionados) de los que hay que seleccionar y se encuentran dentro de la intención de compra.

Regla de asignación:

La regla de asignación que se utilizará para valorizar las canastas frecuentes será basándose en el soporte de estas canastas. Pero resulta útil destacar en este punto que antes de valorizar las canastas frecuentes y aplicar ahí la regla de asignación, se considerarán como input solo las canastas frecuentes que contengan algún producto dentro de la intención de compra.

Por lo que al momento de asignar la utilidad de una transacción a determinada canasta frecuente, estas siempre contendrán al menos un producto perteneciente a la intención de compra que se desea gestionar. Y por ende existirán transacciones que no aporten utilidad a alguna canasta frecuente, generando que tras el proceso de asignación de utilidad, exista un descuadre al sumar la utilidad de todas las transacciones de un periodo con la utilidad de todas las canastas frecuentes del periodo.

Restricciones:

Con lo que respecta a las restricciones, en primer lugar es necesario contar con la restricción que liga la inclusión de productos con la de canastas frecuentes. Logrando de esta forma que la función objetivo considere la utilidad de una canasta frecuente si y solo si todos los productos de esta son seleccionados por el modelo.

Como segunda restricción es necesario incluir en estas la diferenciación entre los productos que pertenecen o no a la intención de compra que se desea gestionar. Ya que se restringirá a que el modelo siempre seleccione los productos que se encuentran fuera de la intención de compra. Para poder distinguir a los productos que corresponden a la intención de

compra de los que no corresponden, será necesario crear “*Sets of Sets*” en la sintaxis del AMPL, lo cual se explica en la metodología de este trabajo.

Luego puede ser interesante incluir restricciones que limiten el número máximo y mínimo de productos a incluir dentro de la intención de compra.

Y finalmente una restricción que restrinja la naturaleza de las variables decisión a ser binarias.

Finalmente el modelo de optimización queda como sigue:

- Función Objetivo: (Si se incluye o no el costo de bodegaje de los productos)

$$MAX \left(\sum_{X \in L} M(X) * P_x - \sum_{i \in L} Cost_i * Q_i \right) \vee MAX \left(\sum_{X \in L} M(X) * P_x - \sum_{i \in L} Q_i \right)$$

- Sujeto a:

$$\forall X \in L, \forall i \in X: Q_i \geq P_x \quad (1)$$

$$\forall i \text{ producto fuera de la intención de compra: } Q_i = 1 \quad (2)$$

$$\sum_{i \in L} Q_i \leq ItemMax \quad (3)$$

$$\sum_{i \in L} Q_i \geq ItemMin \quad (4)$$

$$\forall P_x, \forall Q_i \in \{0,1\} \quad (5)$$

Luego para obtener el ranking de los productos será necesario iterar la resolución de la optimización variando el parámetro *ItemMax* de la restricción (3). Con lo que el ranking de los productos se obtendrá al observar cuales productos son seleccionados en cada una de las iteraciones de la resolución de la optimización. Si bien este proceso puede resultar engorroso, será particularmente esta utilización del modelo la que se llevará a la práctica en el capítulo **¡Error! No se encuentra el origen de la referencia. ¡Error! No se encuentra el origen de la referencia..**

b) Selección de productos para tienda de conveniencia:

Con esta aplicación del modelo se busca utilizar la información obtenida de alguna tienda con un amplio surtido y aprovecharla en la creación del mix de producto para una tienda de conveniencia. Por consiguiente solo se considerará la correlación que existe entre los productos seleccionados y no la que estos tienen con los productos no seleccionados, ya que como es evidente en la nueva tienda de conveniencia no se incluirán en el mix los productos no seleccionados, por lo que no existirán las correlaciones existentes con estos productos.

El conocimiento generado en los pasos anteriores del AMP permite categorizar los productos en las distintas intenciones de compra, con lo cual se podrá determinar tanto un máximo y un mínimo de productos para cada intención de compra. Permitiendo que el modelo

considere, al momento de optimizar, las restricciones del negocio que fijen cuales intensiones de compra incluir y cuantos productos incluir en cada una de estas.

Los criterios utilizados para resolver este problema son:

Optimización:

En este caso, la optimización solo se realiza dentro de los productos seleccionados por el modelo. Por lo tanto solo aportarán utilidad las canastas frecuentes de las cuales todos sus productos hayan sido seleccionados. Y por consiguiente, las canastas frecuentes que incluyen a productos que no fueron seleccionados por el modelo, para ser parte del nuevo mix de productos, no aportarán utilidad a la función objetivo.

Regla de asignación:

En este caso nuevamente se utilizará como regla de asignación el soporte de las diversas canastas frecuentes. Pero a diferencia del caso anterior, aquí se considerarán todos los productos que se utilizaron para generar las canastas frecuentes en el paso anterior.

Restricciones:

La primera restricción nuevamente es la encargada de ligar a los productos con sus canastas frecuentes correspondientes. También puede ser necesario restringir al modelo a que este incluya algunos productos en particular, estos serán llamados productos básicos.

En este caso, a diferencia del anterior donde los productos eran clasificados entre pertenecientes o no a la intención de compra que se desea gestionar, aquí los productos serán clasificados en base a la intención de compra a la cual pertenecen. Teniendo de esta forma n grupos distintos de productos, donde cada grupo será una intención de compra.

Finalmente puede ser necesario incluir una restricción para el número total de productos que se desean incluir en la tienda.

Por lo que el modelo que se utilizará en este caso es:

- Función Objetivo: (Si se incluye o no el costo de bodegaje de los productos)

$$MAX \left(\sum_{X \in L} M(X) * P_x - \sum_{i \in L} Cost_i * Q_i \right) \vee MAX \left(\sum_{X \in L} M(X) * P_x - \sum_{i \in L} Q_i \right)$$

- Sujeto a:

$$\forall X \in L, \forall i \in X: Q_i \geq P_x \tag{1}$$

$$\forall i \text{ producto basico}: Q_i = 1 \tag{2}$$

$$\forall c \in \text{IntenciónCompra} \sum_{i \in C} Q_i \leq \text{ItemMax}_c \tag{3}$$

$$\forall c \in \text{IntenciónCompra} \sum_{i \in C} Q_i \geq \text{ItemMin}_c \tag{4}$$

$$\sum_{i \in L} Q_i \leq ItemMaxTotal \quad (5)$$

$$\forall P_x, \forall Q_i \in \{0,1\} \quad (6)$$

El punto fuerte de esta optimización es que logramos tener a los productos categorizados según a la intención de compra que pertenecen, y es posible gestionar cada intención de compra de forma independiente.

c) Cuponera de descuentos:

La cuponera de descuentos es una acción de marketing directo, que consiste en un conjunto de descuentos que se les envían a los clientes en forma de cupones canjeables al momento de realizar sus compras. Uno de los objetivos de esta acción es aumentar el ticket de compra de los clientes.

Se pretende aprovechar la metodología de AMP para seleccionar los productos que serán promocionados en la cuponera. Esto con el fin de que los productos promocionados sean productos que presenten un alto grado de efecto correlación con los demás productos, lo que aumentaría la probabilidad de que un cliente compra más que solamente el producto promocionado.

Para esto se seleccionarán productos pero se medirán las correlaciones que existen no solo entre ellos, sino que además se incluirán las correlaciones con los demás productos no seleccionados por el modelo.

Puede resultar útil categorizar los distintos productos para generar una cuponera variada en términos de tipos de productos ofertados. Lo que además permitiría generar restricciones según estas categorías, como por ejemplo restringir al modelo para que seleccione solo un producto de cada categoría.

Los criterios utilizados para esta aplicación son:

Optimización:

En esta aplicación de la metodología la optimización es distinta a las dos previamente vistas, principalmente porque aquí se considerarán tanto las interrelaciones que se generan entre los productos seleccionados como la generada entre los productos seleccionados con los de afuera. Respecto a esto último, es muy importante distinguir que no se considerarán todas las correlaciones entre todos los productos, sino que solo las que incluyan a al menos un producto dentro de los seleccionados.

La función de optimización por su parte, difiere un poco de las dos anteriores, debido a que no incluye el costo de tener el producto. La razón para excluir el costo en la optimización es que independiente de cuales productos serán seleccionados, se seguirá incurriendo en el costo de todos los productos.

Ahora bien, si existiese un costo distinto por realizar un determinado descuento en los productos, sería buena la inclusión de este costo en la función de optimización. Esta diferencia en el costo de descuento pudiese estar gatillada por los distintos proveedores de los productos, y el poder de negociación entre la tienda y estos.

Reglas de asignación:

Como se ha dicho antes, surge eficiente considerar el soporte de las canastas frecuentes para asignar la utilidad de las transacciones a estas. En este caso, al igual que en la segunda aplicación se utilizarán todas las canastas frecuentes posibles.

Restricciones:

En primer lugar, la restricción que liga productos con las canastas frecuentes es distinta en este caso. Esto debido a que se incluirá la utilidad de cada canasta frecuente siempre y cuando se incluya algún producto de esta y no como era antes, solo al incluir todos los productos pertenecientes a la canasta frecuente.

Luego, como se pretende evitar que los cupones de descuento sean de productos muy similares y menos que sean muchos de una misma categoría, se categorizarán los productos para así poder restringir el número de productos a seleccionar por cada una de estas categorías.

El modelo de optimización es el siguiente:

- Función Objetivo: (Si se incluye o no el costo de bodegaje de los productos)

$$MAX \left(\sum_{X \in L} M(X) * P_x \right)$$

- Sujeto a:

$$\forall X \in L: \left[\frac{1}{|X|} \sum_{i \in X} Q'_i \right] = P_x \tag{1}$$

$$\forall c \in IntenciónCompra \sum_{i \in c} Q_i \leq ItemMax_c \tag{2}$$

$$\forall P_x, \forall Q'_i \in \{0,1\} \tag{3}$$

Con esta optimización se logra seleccionar productos que tengan altas ventas y alto efecto correlación con los demás productos. Lo cual no solo es útil para generar cuponerías de descuentos como se ve aquí, además es posible utilizar esta información para diseñar de mejor forma el *layout* de los productos, colocando en posiciones más visibles los productos que sean seleccionados por el modelo, aumentando de esta forma el posible *cross-selling* de los productos.

IX. APLICACIÓN DE LA METODOLOGÍA

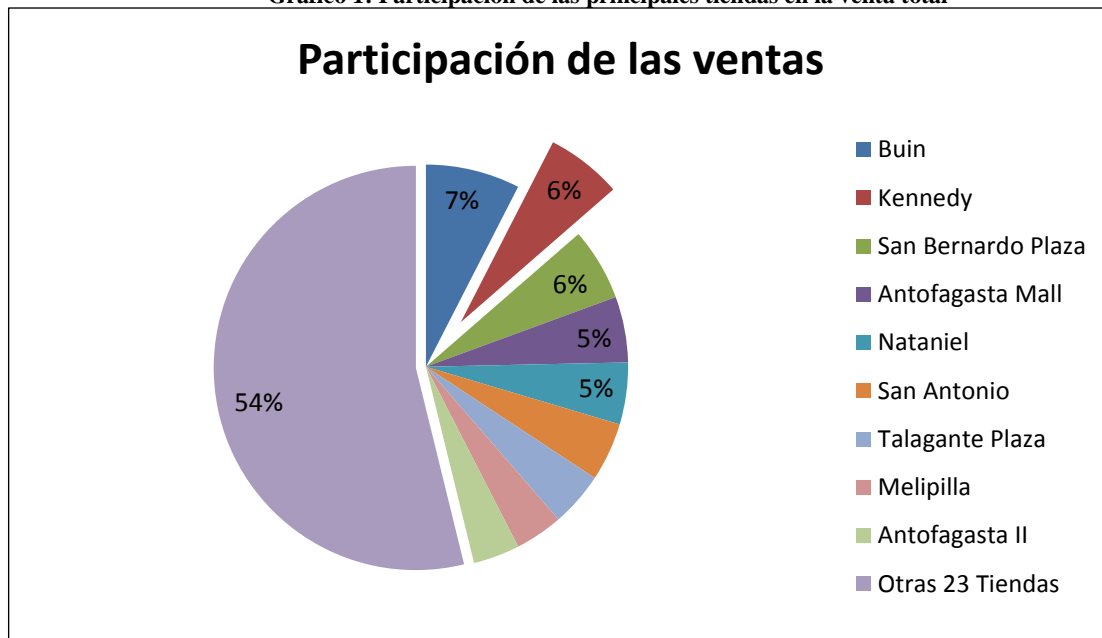
En este capítulo se valida la metodología del AMP para una tienda de supermercado en particular. En la cual se estudiará el mix de un conjunto de productos para finalmente tomar decisiones de agregación o sustracción de productos de ese mix.

9.1 Descripción de la tienda

Se dispone de la información transaccional de los diversos negocios del holding, lo cual hace posible llevar a cabo la metodología en cualquiera de estos. Por motivos principalmente de que la estacionalidad de los productos en las tiendas es más baja en Tottus que en las tiendas de Falabella, y a que la frecuencia de compra es mayor en Tottus que en Sodimac, es que se eligió el supermercado Tottus para realizar el estudio.

Tottus cuenta con más de 30 tiendas a lo largo de todo Chile, y dentro de las más importantes se destaca las tiendas de Buin, Kennedy y San Bernardo. Es posible ver en el Gráfico 1 la participación de las ventas total que tiene cada tienda.

Gráfico 1: Participación de las principales tiendas en la venta total



Fuente: Elaboración propia, en base a las ventas de Tottus en el mes de Marzo del 2011

La tienda en la cual se probará la AMP es la de Avenida Kennedy, ubicada en Avenida Kennedy #5601, comuna de las Condes en Santiago de Chile. Esta tienda presenta una ubicación privilegiada por ubicarse a menos de 100 metros de una tienda de Falabella y de un Sodimac. Estas tiendas por consiguiente se consideran como actores principales en la creación de una imagen corporativa e individual de cada negocio. Por lo que es en la tienda de Kennedy (en adelante la tienda) donde los gerentes de esta tienen principal preocupación por la imagen que esta tienda refleja, y el mix de productos tiene un papel muy importante en la construcción de esta imagen.

Una tienda promedio maneja alrededor de diecisiete mil productos distintos, mientras que la tienda de Kennedy presenta más veinticinco mil productos distintos²⁵, esto significa que existen en promedio ocho mil productos que se encuentran en la tienda de Kennedy que podrían ser incluidos en el mix de las demás tiendas de Tottus. Es principalmente debido a la gran cantidad de productos y a su posición estratégica que se seleccionó la tienda de Kennedy para probar la metodología.

9.2 Información a utilizar

Habiendo ya seleccionado la tienda en la cual se realizará el estudio, se proseguirá contextualizando la tienda y finalmente la información a utilizar.

Como se señaló anteriormente, la tienda presenta más de 25.000 productos distintos, estos agrupados en 35 categorías, las cuales se pueden apreciar en el Anexo B.

Para obtener mejores resultados y para poder trabajar con la gran cantidad de productos y transacciones que se tiene es que se decidió solo trabajar con productos comestibles. Las categorías consideradas como comestibles (Categoría Food) se encuentran en la Tabla 7.

Tabla 7: Categorías jerárquicas pertenecientes al grupo Food.

FOOD	
LIQUIDOS	CONGELADOS
CARNICERIA	PESCADERIA
LACTEOS	FIAMBRERIA
ABARROTOS	PANADERIA Y PASTERIA
PLATOS PREPARADOS	FRUTAS Y VERDURAS

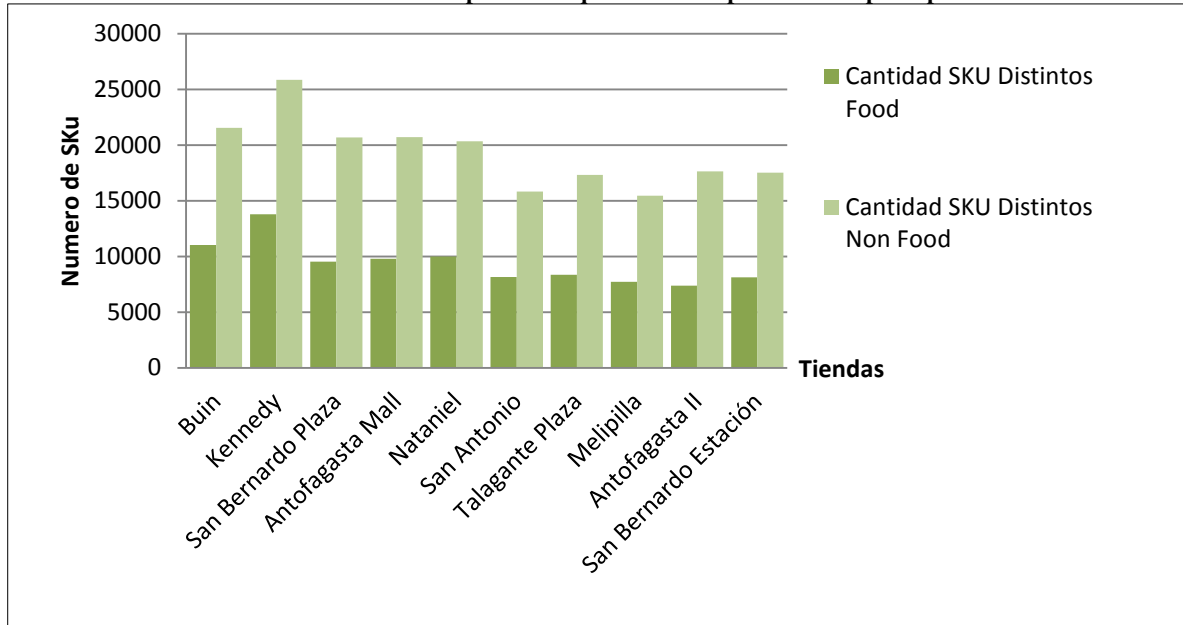
Fuente: Elaboración propia.

Es importante para el análisis, comprender cuál es la importancia de las categorías en estudio para el retail, ya que a simple vista se observa que existen 10 categorías en *Food* y 34 categorías que fueron eliminadas del estudio, y esto llevaría a pensar que se está considerando una porción muy pequeña de las ventas de la tienda. Lo cual es falso ya que todos los productos considerados en el estudio representan el 75% de la venta de la tienda y el 94% de todos los tickets contienen algún producto de *Food*.

Por otro lado, es importante apreciar que a pesar de estar considerando solo 10 categorías, en Kennedy esto es tomar en estudio 12.000 productos distintos. La diferencia entre la cantidad de todos los productos que maneja una tienda, y los considerados en la categoría de Food, se puede apreciar en el Gráfico 2.

²⁵ Información obtenida de las transacciones realizadas en todas las tiendas de Tottus el mes de marzo del año 2011.

Gráfico 2: Cantidad de productos que venden en promedio las principales tiendas.



Fuente: Elaboración propia

Ya se ha señalado que se trabajara con la tienda Kennedy y particularmente con productos pertenecientes a las 10 categorías ya señaladas. Ahora bien, las transacciones que se tomaran en cuenta para todo el proceso que viene son las que se realizaron en dicha tienda el mes de enero, febrero y marzo del año 2010. Además, como ya se señaló solo se considerarán los productos de alimentación de la tienda en cada ticket.

A pesar de estos filtros, se trabajará con más de cinco millones de transacciones y con casi veinte mil clientes distintos identificables²⁶. Lo anterior se puede observar más detalladamente en la Tabla 8, en la que se observan los números de días del estudio, la cantidad de clientes inidentificables distintos que compraron en este periodo, el número de tickets, el número de transacciones y finalmente la cantidad de productos distintos que se compraron al menos una vez en este tiempo.

Tabla 8: Características de la información transaccional con que se trabajará

Número de Días	Clientes Identificables	Número de Tickets	Número Transacciones	Cantidad de productos distintos
87	19.887	429.310	5.206.351	13.784

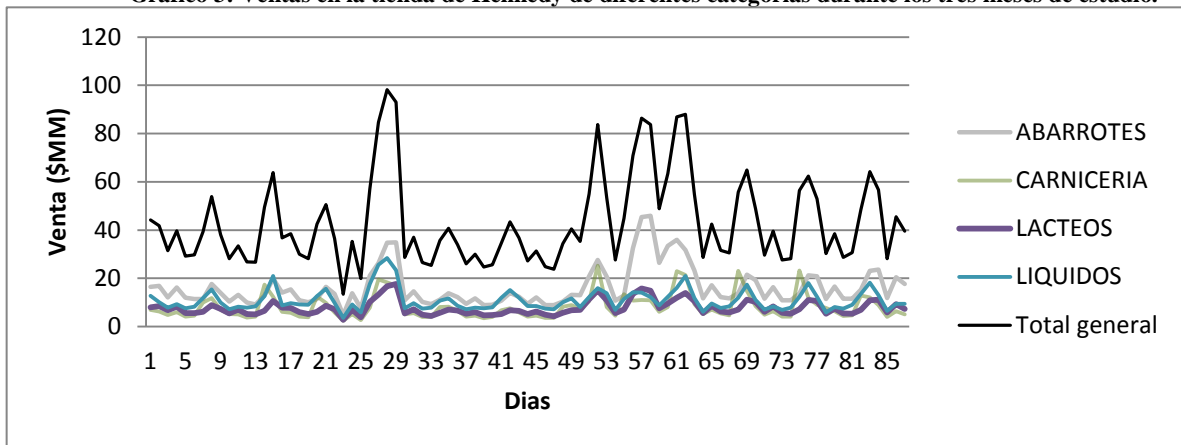
Fuente: Elaboración propia.

Teniendo clara la composición general de la información que se utilizará es que a continuación se detallará dicha información, con el fin de dar una mejor interpretación a los resultados futuros. En primer lugar resulta importante observar el comportamiento de las ventas durante los meses que se considerará el estudio, sin olvidar que solo se está considerando los

²⁶ Para que la transacción de algún cliente sea identificable, este debió haber pagado con la tarjeta CMR o algún otro medio identificable como cheque u otra tarjeta de crédito.

productos pertenecientes a la categoría *Food*. A continuación se muestra el comportamiento de las ventas en los tres meses para cuatro de las categorías más importantes y para el total.

Gráfico 3: Ventas en la tienda de Kennedy de diferentes categorías durante los tres meses de estudio.



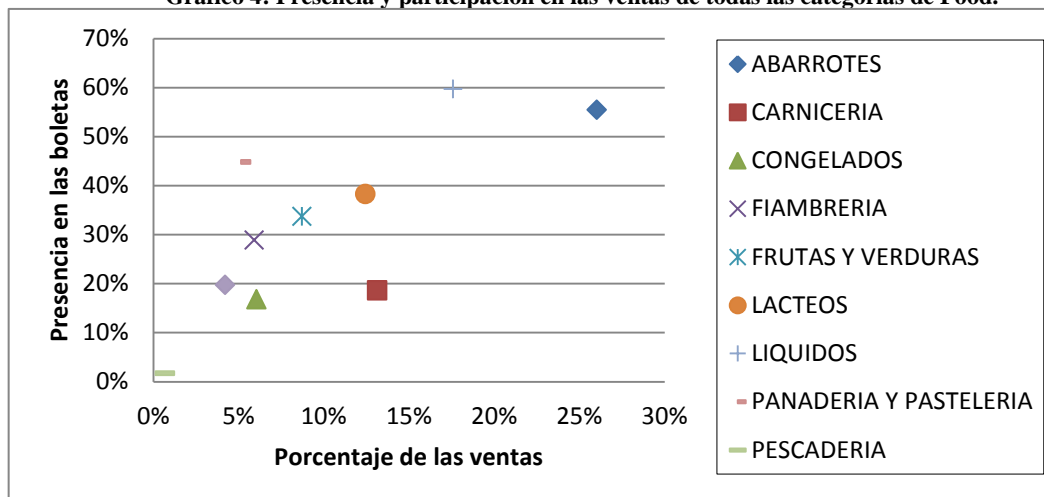
Fuente: Elaboración propia.

Se puede apreciar en el Gráfico 3 que los meses de enero y febrero las ventas no decrecen considerablemente, pero si se puede observar un cierto tipo de suavizado en las ventas, ya que no se observan los picos característicos de los fines de semana que si se aprecian en el mes de marzo. Otra observación importante es el efecto del fin de mes, el cual genera un alza en la demanda principalmente en abarrotes. A pesar de estos efectos, se espera que la composición de las canastas no se vea afectada, y por consiguiente no afectaría las conclusiones finales del estudio.

Además de las ventas y principalmente debido a que en este trabajo se considerará el efecto de *cross-selling* de los productos, es que resulta importante observar la participación aproximada²⁷ en las ventas totales de cada mes y la presencia en tickets de cada una de las categorías que se estudiarán. Esto se puede ver en el Gráfico 4.

²⁷ Los valores son aproximado debido a fueron calculados considerando datos de tres meses. Por lo que pueden variar de mes a mes.

Gráfico 4: Presencia y participación en las ventas de todas las categorías de Food.



Fuente: Elaboración propia.

Una de las observaciones importantes del gráfico anterior es el cómo “panadería y pastelería” posee una alta presencia en las boletas, más de un 40%, pero a su vez una baja participación en las ventas, cerca del 5%. Este puede ser un caso en donde al analizar su venta de forma particular (sin analizar el efecto de *cross-selling*) se puede estar subestimando su importancia en las ventas. Este efecto suele pasar con los productos generadores de tráfico²⁸ de cada supermercado. Y sin precisamente estos de los cuales la tienda espera que generen un alto *cross-selling* con los demás.

9.3 Descubrir las intenciones de compra

Habiendo ya definido las características de la información que se va a utilizar para continuar con el estudio, sigue descubrir las intenciones de compra. Esto permitirá crear una nueva agrupación de productos que generará un mejor estudio de las correlaciones de estos.

Para conseguir esto, en primer lugar se agruparán los productos según una agrupación utilizada por el supermercado. Y se proseguirá a encontrar las canastas típicas que los clientes realizan en el supermercado considerando esta agrupación como si fueran productos. Luego como ya se habló en la metodología se utilizará un indicador de cercanía entre productos. Posteriormente se realizará un análisis multidimensional para encontrar factores que expliquen estas cercanías y finalmente los productos serán agrupados mediante un análisis jerárquico. Para así encontrar las distintas intenciones de compra de los clientes.

9.3.1 Selección de Data

Ya se ha hablado de la información que se utilizará, pero es relevante abordar en este punto (y durante el transcurso de la metodología) la agrupación que se irá haciendo de los

²⁸ Los productos (o categoría) generadores de tráfico son los que incentivan al cliente a ingresar a la tienda, generalmente ubicados en el extremo más lejano a la entrada de la tienda. Son identificables por su alta presencia en las boletas, y esta alta presencia no está correlacionada con su alto aporte en las utilidades.

productos. Ya que no siempre se trabajará a nivel de SKU ni a nivel de categoría, más adelante, en particular al momento de llevar a cabo el PROFSET, se utilizará una agrupación especial de productos.

Para obtener las intenciones de compra, se trabajará a nivel de sub-categoría. Por lo que se tendrán 32 sub categorías pertenecientes a la categoría *Food*. Estas sub categorías son observables en el Anexo B. Luego el estudio será llevado a cabo determinando las cercanías existentes entre estas sub categorías. Estas sub categorías así como cualquier agrupación llevada a cabo más adelante serán considerados como si fuese un producto y se llamarán macro productos para diferenciarlos de los productos que corresponden a un SKU en particular. Es decir, para encontrar las intenciones de compra, se observarán las transacciones de los clientes como si estos comprasen sub categorías y por consiguiente no se discriminará por qué es efectivamente lo que el cliente está comprando, tal como si la tienda solo vendiese 32 productos distintos.

Se decidió trabajar a nivel de sub-categoría debido a que al trabajar a un nivel más desagregado de productos, es decir a nivel de línea, se tenían 260 líneas de productos distintas, y esto generaba dos problemas. Primero, el tiempo de cómputo principalmente para calcular el análisis multidimensional crecía exponencialmente con la cantidad de productos en el análisis. El segundo problema era generado principalmente por líneas con baja frecuencia de compra, y a pesar de estar considerando tres meses de transacciones, en los cuales se compraron más de cinco millones de productos, existían líneas de productos que se compraban en menos de 10 ocasiones, lo que generaba mucho ruido al momento de buscar encontrar la cercanía de estas líneas con las demás.

Es por todo esto que se decide trabajar a nivel de sub-categoría en todo lo que respecta a encontrar las distintas intenciones de compra.

9.3.2 Análisis Multidimensional

Como se acaba de señalar se trabajará a nivel de sub-categoría dentro de *Food*. Ahora utilizando la información transaccional se determina la matriz de cercanía entre los macro productos.

Para determinar la cercanía entre los macro productos se generó una matriz de 32x32 donde A_{ij} representa el número de tickets donde el cliente compró al mismo tiempo algún producto de i y alguno de j . Teniendo esa matriz se utilizó el factor que se explica en 8.2.2.3 como indicador de distancia. Con lo cual se generó la matriz que se puede encontrar en el Anexo C: Matriz de distancia entre sub categorías.

Luego se utilizó el software SAS para realizar el análisis multidimensional y así poder encontrar patrones que expliquen la cercanía de los productos.

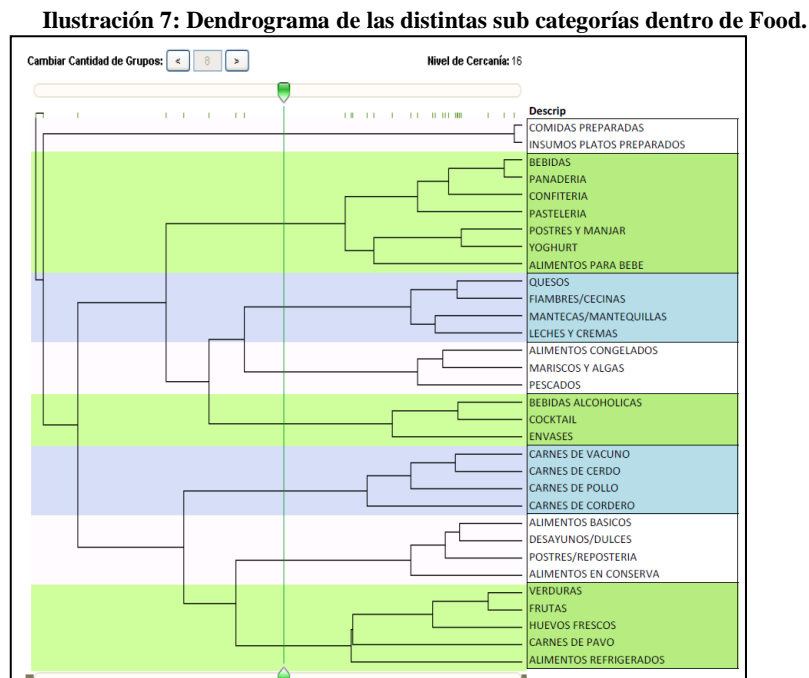
Se ejecutó el análisis en un inicio buscando obtener cinco dimensiones pero el stress de la matriz resultante era demasiado alto, un 25%. A lo cual se decidió aumentar el número de dimensiones hasta obtener un stress menor al 10% (tal como se discutió en la sección 8.2.2.4), lo cual fue conseguido al llevar a cabo el Análisis Multidimensional con 10 dimensiones, obteniendo un stress de un 8%. Las coordenadas de cada macro producto en el espacio de 10 dimensiones pueden verse en el Anexo D: Resultado del análisis multidimensional.

En este momento, se tiene a los 32 macro productos posicionados en un espacio de 10 dimensiones, con lo cual se proseguirá a agruparlos utilizando un análisis jerárquico.

9.3.3 Análisis Jerárquico

Para realizar el análisis jerárquico se utilizó un script propio programado en el lenguaje de programación Java, el cual considera la distancia euclidiana entre los productos para determinar las distintas agrupaciones. Se decidió utilizar este script debido a que se disponía de un software diseñado en la empresa para graficar el dendrograma que se obtiene, junto con diversos indicadores (venta, participación, margen...) de los productos.

El dendrograma final obtenido se puede observar la Ilustración 7. En el cual finalmente se decidió obtener 8 grupos distintos, por variadas razones.



Fuente: Elaboración Propia.

Si bien tal como señala (Shalizi, 2009) existe poca literatura y estudios referente al número indicado de clústeres a generar tras el análisis jerárquico, si se desprende de los diversos estudios que la diferencia en la distancia de aglomeración entre los clústeres es un factor importante a la hora de decidir el número de estos a formar. Por esta razón y tal como se observa en la figura, existe una gran distancia para pasar de siete a ocho grupos, luego ocho grupos es un fuerte candidato a ser elegido. Lo mismo pasa al tener tres grupos distintos, pero se descarta el formar solo tres grupos ya que sería una agrupación muy agregada.

Finalmente este corte fue validado por analistas del área y se concluyó que rescata de buena forma las diferentes intenciones de compra de los clientes del supermercado. Por lo que ahora, ya teniendo definida las diversas intenciones de compra de los clientes, se proseguirá a caracterizarlas, colocándole un nombre identificador a la intención de compra y observando diversos indicadores de estas.

9.3.4 Caracterización de las intenciones de compra

Habiendo ya definido que se trabajará con ocho grupos distintos de intenciones de compra, se prosigue a analizar estos distintos grupos. Tomando en consideración en primer lugar la composición de estos, las correlaciones que existen entre ellos y luego calculando diversos indicadores que permitan generar una mirada sobre el desempeño de estas.

En primer lugar se estudiará a cada uno de los 32 macro productos por separado, pero manteniendo la relación existente de estos con la intención de compra previamente encontrada. Luego se estudiarán las intenciones de compra que se generaron como si estas fueran un macro producto. Es decir, a continuación se enumerarán las 8 intenciones de compras y se estudiará las categorías que quedaron dentro de cada una de esta.

9.3.4.1 Estudio de categorías dentro de cada intención de compra

1. Comida preparada:

Intención Compra	SUB CATEGORÍA	Distintos SKU en Categoría	N° Transaccion es (Miles)	Reach	Presencia	%Vent a Total
Comida Preparada	COMIDAS PREPARADAS	261	126	17%	14%	2.7%
	INSUMOS PLATOS PREPARADOS	2	7	1%	1%	0.0%

Se observa en el dendrograma que esta intención de compra está completamente separada de las demás, lo que significa que esta categoría prácticamente no “gatilla” la compra de ningún otra categoría dentro de la tienda. Por otro lado se observa que prácticamente de las dos categorías que se encuentran dentro de esta intención de compra, *insumos platos preparados* solo posee dos SKUs distintos, con un *reach* de un 1%. Por lo que en esta intención de compra, solamente tiene importancia la categoría *comidas preparadas*, la cual tiene un *reach* bajo de un 17% y una presencia igualmente baja. A pesar de ser muy pocos SKUs distintos, el porcentaje de la venta al que *comidas preparadas* aporta, no es despreciable.

2. Conveniencia

Intención Compra	SUB CATEGORÍA	Distintos SKU en Categoría	N° Transacciones (Miles)	Reach	Presencia	%Venta Total
Conveniencia	BEBIDAS	1.218	761	73%	52%	9.3%
	PANADERIA	193	282	65%	41%	3.5%
	CONFITERIA	1.358	461	64%	31%	5.0%
	PASTELERIA	388	64	28%	9%	1.7%
	POSTRES Y MANJAR	139	76	21%	6%	0.7%
	YOGHURT	349	255	43%	16%	2.4%
	ALIMENTOS PARA BEBE	76	19	5%	1%	0.5%

En esta intención de compra se encuentran categorías en los cuales los clientes utilizan la tienda como tienda de conveniencia²⁹. En esta intención de compra, se tiene a categorías con un gran número de SKUs distintos, como por ejemplo bebidas, donde ese alto surtido se explica principalmente por los diversos formatos que existen. El *reach* bebidas panadería y confitería es muy alto, al igual que el aporte de estas categorías a las ventas, lo cual permite entender a estas categorías como las principales dentro de esta intención de compra.

3. Agregado once

Intención Compra	SUB CATEGORÍA	Distintos SKU en Categoría	N° Transacciones (Miles)	Reach	Presencia	%Venta Total
Agregado Once	QUESOS	337	183	55%	22%	4.9%
	FIAMBRES/CECINAS	579	206	56%	25%	5.0%
	MANTECAS/MANTEQUILLAS	97	47	29%	7%	0.8%
	LECHES Y CREMAS	284	204	47%	17%	3.8%

Esta intención de compra posee categorías que si bien se esperaría que su cercanía con la categoría *panadería* fuese alta, y por consiguiente estuviesen juntas al momento de realizar el análisis jerárquico, se encuentran bastante alejadas. Esta lejanía se explica por la diferencia que existe en los tiempos de recompra de las categorías que están en *Agregado once* con *panadería* (es común que un cliente compre diariamente pan, pero no así leches y mantequillas por ejemplo).

4. Pescado y Congelados

Intención Compra	SUB CATEGORÍA	Distintos SKU en Categoría	N° Transacciones (Miles)	Reach	Presencia	%Venta Total
Pescados y Congelados	ALIMENTOS CONGELADOS	629	185	47%	16%	5.7%
	MARISCOS Y ALGAS	33	4	2%	1%	0.2%
	PESCADOS	57	7	5%	1%	0.5%

Esta intención de compra es un poco heterogénea, debido a que *alimentos congelados* se comporta muy distinto que las otras dos, presentando la primera un alto *reach* y una gran cantidad de productos distintos. Junto con lo anterior, el aporte a las ventas de las dos últimas es

²⁹ Se llaman tiendas de conveniencia a los establecimientos con menos de 500 m², tienen un amplio surtido de productos, centrado en bebidas y alimentación.

ínfimo, esto principalmente debido al bajo consumo de pescados y mariscos que existe en Chile³⁰.

5. Alcohol y Coctel

Intención Compra	SUB CATEGORÍA	Distintos SKU en Categoría	N° Transacciones (Miles)	Reach	Presencia	%Venta Total
Alcohol y coctel	BEBIDAS ALCOHOLICAS	1.777	173	39%	17%	8.5%
	COCTEL	556	123	37%	13%	2.2%
	ENVASES	14	52	10%	3%	0.0%

Las sub categorías de esta intención de compra se agruparon debido a la alta afinidad que existe entre bebidas alcohólicas y coctel. La sub categoría coctel incluye por ejemplo: maní, almendras, papa fritas en bolsa, entre otras. Mientras que envases hace referencia a todas las devoluciones de envases que se realizan en la tienda.

Es posible observar como bebidas alcohólicas y coctel tienen un *reach* una presencia similar, lo que es otro punto a favor para demostrar que son sub categorías muy afines. Observando esto se puede llegar a suponer que los clientes en un alto porcentaje al momento de comprar bebidas alcohólicas, llevan algún producto de coctel.

6. Carnes

Intención Compra	SUB CATEGORÍA	Distintos SKU en Categoría	N° Transacciones (Miles)	Reach	Presencia	%Venta Total
Carnes	CARNES DE VACUNO	232	122	42%	13%	8.5%
	CARNES DE CERDO	41	17	13%	3%	1.0%
	CARNES DE POLLO	77	73	36%	10%	3.3%
	CARNES DE CORDERO	9	0	0%	0%	0.1%

En esta intención de compra se agruparon prácticamente todas las carnes salvo las carnes de pavo, ya que las últimas tienen mayor cercanía con verduras. Se observa que existe una gran cantidad de SKU distintos, principalmente en las tres primeras carnes, esto debido a la gran variedad de cortes, si la carne es o no importada y si el cliente la solicita en el mesón.

Como observación se puede apreciar que la carne de cordero se encuentra en la tienda para ayudar a generar la imagen que se quiere de Tottus, ya que a pesar de que prácticamente no se vendan los productos dentro de esta categoría, aun la tienda sigue teniéndola.

³⁰ Según el Instituto de Nutrición y Tecnología de alimentos (INTA), Chile es uno de los países con más bajo consumo de pescado en el mundo: no más de 7 kilogramos por habitante al año en los últimos 30 años, frente a los 75 kg que cada chileno consume al año en otras carnes (bovinos, cerdos, aves).

7. Abarrotes

Intención Compra	SUB CATEGORÍA	Distintos SKU en Categoría	N° Transacciones (Miles)	Reach	Presencia	%Venta Total
Abarrotes	ALIMENTOS BASICOS	2.171	526	62%	24%	8.3%
	DESAYUNOS/DULCES	1.390	293	59%	22%	7.0%
	POSTRES/REPOSTERIA	291	39	16%	3%	0.3%
	ALIMENTOS EN CONSERVA	669	165	41%	12%	3.2%

En esta intención de compra se encuentran sub categorías principalmente de abarrotes, donde se incluyen: arroz, aceite, tallarines. Si bien su *reach* es alto, su presencia en tickets de compra es bajo, esto principalmente debido a que prácticamente todos los clientes alguna vez (durante los tres meses de datos) compraron esta sub categoría, pero son compras muy esporádicas, cada siete, quince o treinta días.

Su aporte en las ventas salvo *postres/repostería* no es despreciable y existe un amplio surtido en esta intención de compra.

8. Frutas, Verduras y huevos.

Intención Compra	SUB CATEGORÍA	Distintos SKU en Categoría	N° Transacciones (Miles)	Reach	Presencia	%Venta Total
Frutas, Verduras y huevos	VERDURAS	270	391	53%	24%	4.5%
	FRUTAS	166	277	54%	26%	4.3%
	HUEVOS FRESCOS	22	42	29%	8%	1.0%
	CARNES DE PAVO	32	13	9%	2%	0.5%
	ALIMENTOS REFRIGERADOS	67	14	10%	2%	0.4%

Finalmente la última intención de compra incluye frutas y verduras, las cuales se comportan de manera muy similar. Junto con estas dos se encuentra la sub categoría *huevos frescos* la cual a pesar de estar en una categoría jerárquica distinta a estas, posee mucha cercanía con ellas. Esto podría explicar el hecho de que en muchos supermercados los huevos estén cerca de las frutas y las verduras.

Otro aspecto importante es que la carne de pavo quedo en esta categoría, esto se puede atribuir al hecho que es considerado como una carne sana.

Habiendo caracterizado la composición de cada intención de compra, se prosigue a analizar cada una de estas de forma agrupada.

9.3.4.2 Estudio de las intenciones de compra

A continuación, en la Tabla 9 se pueden apreciar diversos indicadores de cada una de las intenciones de compra. Para calcularlos se consideró que si el cliente estaba comprando algún producto dentro de la intención de compra, esa compra en su totalidad aportaba a dicha intención de compra. Es decir, si un cliente compra una fruta (intención de compra número 8) y pan (intención de compra número 2), se considerará esa boleta en ambas intenciones de compra.

Tabla 9: Indicadores de las Intenciones de compra

Nº	Intención de compra	Monto Ticket Prom	Nº SKU Distintos Prom	IC distintas Prom	Categorías Distintas Prom	Reach	Presencia	Ventas
1	Listo para comer	9.909	10	3	4	17%	14%	3%
2	Destino	13.681	14	3	5	93%	83%	23%
3	Agregado Once	22.809	22	4	7	76%	41%	15%
4	Pescados y congelados	33.532	31	5	9	49%	17%	6%
5	Alcohol y Coctel	24.350	22	4	7	56%	27%	11%
6	Carnes Rojas	35.143	31	5	9	53%	18%	13%
7	Abarrotos	24.790	24	4	8	76%	37%	19%
8	Frutas, Verduras y Huevos	23.364	23	4	8	68%	36%	11%

Fuente: Elaboración propia.

Debido al supuesto anterior es que no se pueden apreciar grandes diferencias entre las distintas intenciones de compra. Pero esta baja varianza en los datos se debe principalmente a que los valores tienden a aproximarse al promedio, luego una pequeña diferencia en los valores promedio de la Tabla 9 significa una mayor diferencia en la realidad.

Es posible observar como las intenciones de compra difieren en el promedio del monto del ticket en que se están llevando. La intención de compra 1 presenta un bajo ticket promedio, atribuido principalmente por que el cliente simplemente compra algún producto de consumo que esté listo para comer y algún bebestible y algunas veces pocas cosas más. Lo que hace que el ticket promedio de esta intención de compra sea muy bajo. Por otro lado, esta intención de compra presenta un bajo *reach* y una baja presencia. Por lo que se puede concluir que son pocos los clientes que van a la tienda con esta intención de compra, y cuando lo hacen tienen muy poco cross con los demás productos de la tienda.

Algo similar sucede con la intención de compra de destino, en el hecho de que el monto del ticket promedio sigue siendo bajo y la cantidad de productos distintos que el cliente lleva al momento de comprar algún producto de esta intención de compra, también lo es. Por lo que se podría concluir nuevamente que el efecto cross de esta intención de compra con las demás es muy bajo. Pero esta intención de compra tiene un alto *reach*, es decir, prácticamente todos los clientes llevan algún producto de esta intención de compra, una altísima presencia en los tickets y a su vez la participación de esta intención de compra a las ventas totales es muy alta. Por consiguiente esta intención de compra está presente en una gran cantidad de tickets pero generalmente en tickets con baja cantidad de productos.

La intención de compra *Agregado once* está presente en boletas relativamente grandes y es la tercera intención de compra que más aporta a las ventas.

Finalmente para terminar con la caracterización de las distintas intenciones de compra, en la Tabla 10 se puede tener un primer esbozo del efecto cross que existe entre las distintas intenciones de compra. La tabla en cuestión presenta la probabilidad de que un cliente sabiendo que: está comprando algún producto en las intenciones de compra de la primera columna, lleve al menos un producto de las intenciones de compra ubicadas en la primera fila.

Tabla 10: Efecto cross selling entre las distintas intenciones de compra

INTENCION DE COMPRA	Listo para comer	Conveniencia	Agregado Once	Pescados y congelados	Alcohol y Cocktail	Carnes Rojas	Abarrotes	Frutas, Verduras y Huevos
Listo para comer	-	76%	24%	11%	16%	10%	26%	29%
Conveniencia	13%	-	45%	18%	27%	19%	38%	37%
Agregado Once	8%	91%	-	29%	36%	32%	57%	55%
Pescados y congelados	9%	89%	71%	-	47%	45%	69%	71%
Alcohol y Cocktail	8%	81%	55%	29%	-	30%	51%	48%
Carnes Rojas	8%	88%	73%	42%	45%	-	70%	71%
Abarrotes	10%	86%	63%	31%	37%	34%	-	59%
Frutas, Verduras y Huevos	11%	85%	63%	33%	36%	36%	60%	-

Fuente: Elaboración propia

Luego se pueden obtener interesantes conclusiones. En primer lugar no existe ninguna intención de compra que genere que el cliente lleve al mismo tiempo un producto listo para comer. Como ejemplo contrario al anterior, el hecho de que un cliente compre algún producto de la intención de compra *Pescados y congelados* o de *Carnes Rojas* implica que con una alta probabilidad el cliente comprará productos de las demás intenciones de compra, llevando así una canasta de compra con un gran número de productos, por consiguiente alto cross-selling.

Además de estas, otras conclusiones se pueden obtener viendo esta tabla de correlaciones entre las distintas intenciones de compra.

9.4 Determinar Frequent Sets y valorizarlos

Ahora que se trabajó con la información transaccional y se lograron encontrar las diversas intenciones de compra para todo lo que es *Food* dentro de la tienda de Kennedy de Tottus, además de haber caracterizado y visto a grandes rasgos las correlaciones existentes entre estas, se prosigue en esta segunda etapa a encontrar oportunidades de rediseño. En primer lugar se debe seleccionar una intención de compra, posterior a esto se requerirá de una nueva agrupación de productos. A continuación utilizará nuevamente la información transaccional para encontrar las canastas frecuentes y así poder iniciar la aplicación del PROFSET.

Finalmente se continuará con la tercera etapa de la metodología en donde se obtiene el verdadero valor de los productos considerando su efecto de cross-selling y se obtienen posibilidades de rediseño del mix de productos.

9.4.1 Selección de la intención de compra

La intención de compra que se elegirá para trabajar es la de *Agregado Once*, esto debido a que la empresa decidió que esta pudiese resultar ser una intención de compra interesante a investigar. Lo que significa que el objetivo será encontrar oportunidades de rediseño para los productos que incluye esta intención de compra, estos son, todos los productos que están en las sub categorías de *quesos, fiambres/cecinas, mantecas/mantequillas y leches y cremas*.

9.4.2 Agrupación de productos

Con esta implementación de PROFSET se busca encontrar oportunidades de rediseño de productos dentro de la intención de compra elegida. Pero considerando, además de las interrelaciones que existen entre los productos de una misma intención de compra, las que generan estos con los productos de fuera.

Debido al procedimiento anterior para identificar las intenciones de compra, se posee dentro de cada intención productos muy afines a ser comprados en conjunto y por ende tiene gran valor aprovechar la capacidad de estudiar estos productos a nivel de SKU. Pero, debido a que por construcción las correlaciones de los productos entre diferentes intenciones de compra son bajas, se explicará a continuación el por qué surge provechoso trabajar con los productos fuera de la intención de compra agrupados y con los que se desea realizar acciones de rediseño no.

Esta decisión de agrupar los productos fuera de la intención de compra fue realizada para mejorar el problema de la baja significancia del efecto cross-selling entre productos que encontró Brijs (Brijs, Swinnen, Vanhoof, & Wets, 2004). Para entender el problema se analizará un caso práctico, en donde se observa el valor de trabajar con distintos niveles de agrupación.

Surge evidente suponer que existen correlaciones entre el queso rallado Dos Álamos 40 gramos (SKU: 2292259) por ejemplo y los diversos fideos. Pero si por ejemplo se analiza de forma particular todas las correlaciones entre el queso rallado ya mencionado y cada uno de los diversos fideos, y se analiza la confianza³¹ del queso rallado Dos Álamos con cada uno de los fideos, se puede ver en la Tabla 11, que muchas de estas correlaciones presentan un muy baja confianza, como por ejemplo el fideo “87 al huevo 400 GRS Chef”. Este tallarín por ser un tallarín Premium cuya rotación es muy baja presenta una baja confianza con el queso, al igual que sucede con prácticamente todos los fideos de la categoría Especiales.

Tabla 11: Confianza entre el queso rallado (SKU:2292259) y alguno de los fideos.

SKU	Categoría	Descripción Producto	Confianza de queso rallado con fideos
2293484	ESPECIALES	FIDEO TALLARIN 87 AL HUEVO 400 GRS CHEF.	0,08%
2292205	ESPECIALES	FIDEOS 65 GRS. SABOR POLLO	0,22%
2298595	INTEGRAL	FIDEO ESPIRAL 3 PUNTAS NUTRITIVO 250 GRS.	1,13%
2294111	INTEGRAL	FIDEO SPAGHETTI # 5 SLOW 250 GRS LUCCHETTI.	0,59%
2292633	VITAMINIZADOS	FIDEO SPAGHETTI #5 400 GRS. #101409 C	6,38%

³¹ La confianza en este caso se puede interpretar como la probabilidad de que el cliente compre un determinado fideo sabiendo que ya está comprando el queso rallado Dos Álamos.

2292209	VITAMINIZADOS	FIDEO ESPIRAL 400 GRS. P #113122	3,13%
2292262	ESPECIALES	FIDEO TRATTORIA FUSILLI 400 GR	1,43%
2292297	VITAMINIZADOS	FIDEO SPAGHETTINI 400 GRS. P #113403	0,72%
2292319	VITAMINIZADOS	FIDEO SPAGHETTI #5 400 GRS. DV	2,79%
2292907	RELLENOS	FIDEO RAVIOLI VIVO 300 GRS. #103607	0,40%
2292331	VITAMINIZADOS	FIDEO SPAGHETTI #5 400 GRS. #106402 AC	2,14%
2292378	ESPECIALES	FIDEO TRATTORIA FARFALLA 400 GR	1,81%
2292515	VITAMINIZADOS	FIDEO ESPIRAL 400 GRS. DV	2,23%
2292554	VITAMINIZADOS	FIDEO ESPIRAL 400 GRS. CHEF	1,21%
2292566	PASTINAS/CORTOS	FIDEO PASTINA MARIPOSA #81 250 GRS. L	1,38%

Fuente: En base a las transacciones realizadas durante Enero, Febrero y Marzo del 2010 en el supermercado Tottus de Kennedy.

Debido a que la confianza posee una alta correlación con el soporte se estarían dejando de lado prácticamente todas las canastas frecuentes que incluyen al queso rallado en cuestión con los fideos especiales, por lo que se estaría sub estimando el efecto que tiene el queso rallado con los fideos, y más grave aún, dejando completamente fuera del estudio a productos como el fideo ya mencionado que sí presenta correlaciones con los productos dentro de la intención de compra pero debido a su bajo soporte no estaría en el estudio.

Ahora si se agrupan los productos que se encuentran fuera de la intención de compra a estudiar, es decir, que los productos que no están en la intención de compra que se estudiará se trabajarán a nivel de categoría³². Al realizar esto se mitiga en parte este problema y se aumenta las correlaciones de los productos dentro de la intención con los de fuera. Aumentando así el soporte entre el queso rallado anterior con los fideos tal como se observa en la Tabla 12.

Tabla 12: Confianza entre el queso rallado (SKU:2292259) y las distintas categorías de fideos.

Categorías Fideos	Confianza de queso rallado con categoría fideos
ESPECIALES	31%
VITAMINIZADOS	69%
PASTINAS/CORTOS	8%
RELLENOS	6%
INTEGRAL	5%

Fuente: En base a las transacciones realizadas durante Enero, Febrero y Marzo del 2010 en el supermercado Tottus de Kennedy.

Se observa ahora como tras haber agrupado los fideos según su categoría, la confianza entre el queso rallado y cada una de las categorías de fideos es de sobre el 5%. En definitiva se trabajará con los productos dentro de la intención de compra a nivel de SKU y con los de fuera de esta agrupados por la categoría de cada uno, mejorando así la confianza y por consiguiente el soporte entre los productos.

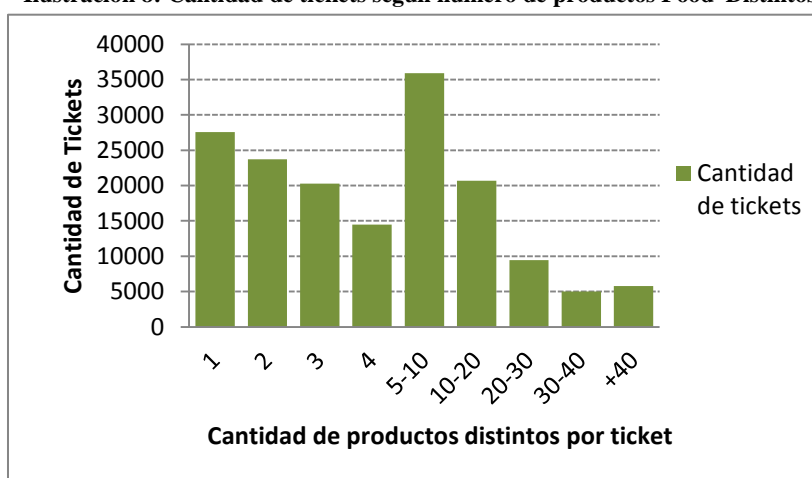
³² Categoría, es la primera agrupación de la jerarquía en Tottus y tiene una cardinalidad aproximada de 900 para los productos en Food.

9.4.3 Canastas frecuentes

Una vez que se cuenta con los productos definidos tanto los pertenecientes a la intención de compra como los que no, se prosigue a estudiar las transacciones para así encontrar canastas frecuentes.

Se trabajó con 162.000 tickets, correspondientes a transacciones realizadas durante los meses de enero, febrero y marzo. Se observa en primer lugar que incluso sacando todos los productos que no pertenecen a la categoría Food, solo el 15% de los tickets presentan solo un producto y que el 50% de éstas están compuestas por más de 5 productos distintos. La Ilustración 8 presenta la cantidad de productos distintos que se compran por cada ticket en la muestra utilizada.

Ilustración 8: Cantidad de tickets según número de productos Food Distintos



Fuente: Elaboración propia. En base a las transacciones del Tottus de Kennedy.

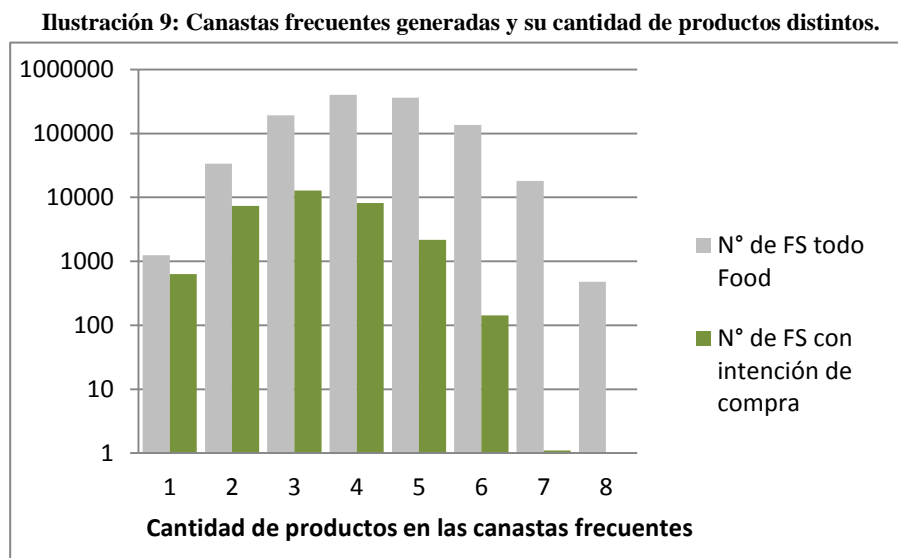
Con estas transacciones se prosigue a utilizar el algoritmo Apriori para encontrar las canastas frecuentes que existen en estas transacciones. Para esto es necesario determinar un soporte mínimo para el cual el algoritmo clasificará una canasta como frecuente.

En este paso se buscaría encontrar la mayor cantidad de canastas frecuentes (definir un soporte mínimo bajo del algoritmo), para así poder capturar todos los efectos de correlaciones entre los productos. Si bien lo anterior presenta una limitación en la capacidad de cómputo para encontrar todas las canastas frecuentes, esto no es el único inconveniente. Si el soporte que se define es demasiado bajo, el algoritmo considerará como canastas frecuentes a aquellas que ocurrieron por mera coincidencia, ensuciando así los resultados. Gracias a que en el siguiente paso se proseguirá a valorizar estas canastas frecuentes en base a su soporte, las canastas con bajo soporte que fueron elegidas por mera casualidad, no serán valorizadas por el algoritmo, previniendo de esta forma el error que pudiese causar el seleccionar un soporte demasiado bajo y con esto incluir canastas no frecuentes.

Al correr el algoritmo con un soporte mínimo de 0.0014, es decir, que una canasta de productos será considerada frecuente si esta aparece en al menos el 0.14% de las boletas. Con esto se encontraron más de un millón de canastas frecuentes. Pero dentro de estas muchas no

incluían a ningún producto de los pertenecientes a la intención de compra que se está estudiando. A lo cual, se decidió eliminar estas canastas frecuentes que no posean algún producto dentro de la intención de compra.

La Ilustración 9 muestra la cantidad de canastas frecuentes encontradas considerando todas y luego solo las que presentaban algún producto dentro de la intención de compra.



Fuente: Elaboración propia.

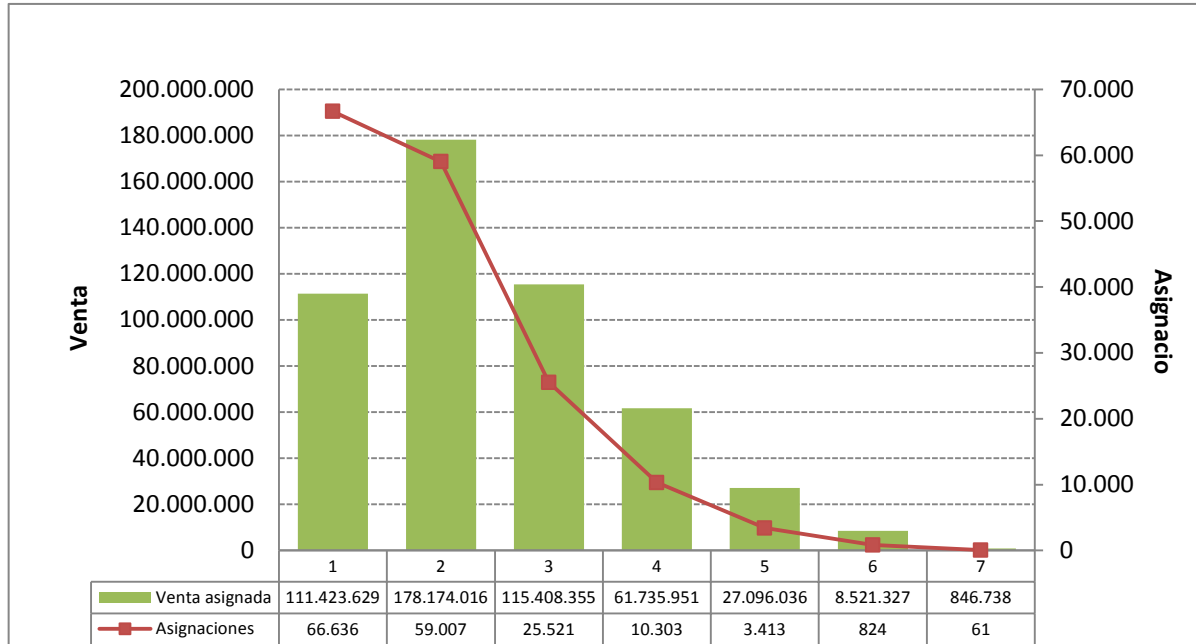
Al haber eliminado las canastas frecuentes que no poseen ningún producto dentro de la intención de compra se bajó de 1 millón de canastas, a solo 31.000 canastas frecuentes distintas. Surge interesante observar como la mayoría de las canastas frecuentes con dos, tres o cuatro productos.

9.4.4 Valorizar las canastas frecuentes

Una vez que se cuenta con las 31.000 canastas frecuentes se prosigue a valorizar estas canastas utilizando nuevamente la información transaccional de los tres meses señalados con anterioridad.

El algoritmo que se utilizará para atribuirle la utilidad de cada canasta frecuente será en base a al soporte de cada canasta, explicado en 5.5.2.2 Reglas de asignación. Para esta implementación se programó el algoritmo en lenguaje Java. El algoritmo junto con asignar la venta de cada transacción a las diversas canastas frecuentes calcula diversos indicadores que muestran el proceso de asignar utilidad. Por lo que una vez llevado a cabo el algoritmo de asignación de utilidad a las canastas, lo que se obtiene son las 31.000 canastas frecuentes valorizadas y los indicadores del proceso que se resumen en la Ilustración 10.

Ilustración 10: Venta asignada y número de asignaciones por tamaño de la canasta frecuente.



Fuente: Elaboración propia.

En la figura anterior vemos como el algoritmo asigna la venta de las transacciones a las diversas canastas frecuentes. Estas canastas frecuentes están agrupadas por cantidad de productos distintas que la componen, y es posible observar como el mayor número de asignaciones corresponden a canastas frecuentes con un solo producto (representado por el gráfico de línea). Y si bien las canastas con dos productos no presentan la mayor cantidad de asignaciones, si acumulan la mayor cantidad de venta.

Como se había dicho con anterioridad en 9.4.3 el riesgo de incluir canastas frecuentes poco significativas se ve mitigado gracias a la valorización de las canastas, ya que las canastas muy poco frecuentes, generalmente las con gran número de productos, representan una muy baja participación de las ventas. Con lo cual, en el paso siguiente, al utilizar PROFSET, prácticamente no influirán significativamente en los resultados.

A continuación se procederá a utilizar el modelo de optimización para seleccionar los productos en base al valor de las diversas canastas frecuentes.

9.5 Selección de productos en base a sus correlaciones

En esta etapa final de la metodología se utiliza la optimización entera PROFSET para valorizar a los diversos productos de la intensidad de compra “Agregados once”, tomando en consideración todas las correlaciones que tiene cada producto dentro de la intensidad de compra con todos los productos de la categoría “Food” del supermercado.

9.5.1 PROFSET

En esta implementación del modelo se utilizará la primera configuración descrita en 8.4.1 (letra a). Para llevar a cabo la optimización se utilizará el módulo simplex de AMPL debido a su gran potencia y escalabilidad. La sintaxis utilizada es generada automáticamente por el mismo código en Java que realiza la valorización de las canastas. En el anexo F se puede apreciar tanto el modelo implementado en AMPL como un resumen de los datos utilizados.

Cabe destacar en este punto que el modelo utilizado no consideró los costos atribuibles al almacenaje de los productos debido a que no se pudo contar con esta información. Por lo que se corrió el modelo con un costo fijo para todos los productos. Lo mismo sucedió para el costo de adquisición de los productos, el cual no fue considerado debido a que solo se contaba con la información actual de los costos de adquisición y existía una alta fluctuación en los precios de los productos durante los tres meses de evaluación.

A pesar de no haber incluido los costos, gracias a conversaciones realizadas con especialistas del negocio se estimó que los resultados no debiesen discrepar demasiado al incluir o no estos costos. En primer lugar los costos de bodegaje son muy similares entre los productos, lo que implicaría que el no incluirlos no debiese afectar demasiado. En segundo lugar el costo de adquisición del producto hace variar el margen que aporta este, y se observó que los productos de mayor precio generalmente presentaban un mayor margen porcentual, por lo cual es posible que se esté sub estimando el aporte al negocio de estos productos al considerar solamente el precio de venta al público como utilidad.

Una vez echas estas consideraciones se definió el número de productos máximo que deberían incluirse en la unidad de necesidad y se optimizó utilizando esta restricción. Los productos seleccionados por el modelo se pueden observar en el Anexo G.

El valor final de la función objetivo del problema no tiene mucho valor en términos prácticos, ya que como se ha visto con anterioridad, entre otras razones se tiene que el set completo de canastas frecuentes no captura la venta total del periodo. Y por consiguiente sería erróneo considerar que el valor de la función objetivo fuese una estimación de la utilidad que generaría los productos seleccionados por la metodología.

9.5.2 El verdadero valor de los productos

Gracias a la metodología seguida hasta ahora, es posible contar con un conjunto de productos valorizados no solo por su aporte individual a los ingresos del retail, sino que además se considera el efecto de *cross selling* de ellos.

A pesar de esto, si bien se sabe que los productos seleccionados son los más rentables (y los no los menos), sería útil contar con un ranking de los productos en términos de más a menos rentables. El problema es que la optimización solo entrega el conjunto final de productos seleccionados y no así el orden de estos.

Como se señaló con anterioridad la utilidad que aporta cada producto a la función de optimización podría ser una buena aproximación considerar que el aporte de cada producto es igual a su contribución marginal echo a la función de optimización, y aprovechar este valor para renquear a los productos más y menos rentables. Pero caemos en el problema de que el aporte

marginal de cada producto a la función de optimización está influenciado por los demás productos que se incluirán en el mix óptimo.

La solución a esto es llevar a cabo varias veces la optimización variando el número de productos máximos a incluir en cada optimización. Y capturando en cada iteración cuales son los productos seleccionados. Permitiendo así generar un ranking de los productos.

Teniendo el ranking de productos echo mediante está metodología, se contrastó este resultado con un ranking de los mismos productos pero construido considerando solo las ventas de estos y no su efecto *cross selling*. Pudiendo de esta forma contrastar el efecto de venta cruzada que captura la metodología.

Tabla 13: Ranking en Ventas Vs PROFSET

Nº	Descripción del producto	Ranking	
		Ventas	PROFSET
1	QUESO RALLADO PARMESANO 40 GRS.	19	1
2	LECHE NATURAL COLUN S/T 1 LT	2	2
3	QUESO RALLADO PARMESANO 40G	37	3
4	QUESO PHILADELFIA 226 GRS.	5	4
5	LECHE DESCREMADA COLUN S/T 1 LT	6	5
6	QUESO RANCO LAMINADO X 5 KG	3	6
7	QUESO LAMINADO GRANEL	4	7
8	QUESILLO QUILLAYES 300 GRS.	18	8
9	QUESO RALLADO 40 GRS. REGGI	61	9
10	LECHE DESCREMADA SURLAT 1 LT	11	10
11	QUESO REGGIANITO 80 GRS.	33	11
12	JAMON COLONIA WINTER	8	12
13	JAMON COCIDO PAVO SOPRAVAL	14	13
14	JAMON PIERNA PREFERIDA	7	14
15	QUESO RALLADO PARMESANO 80 G	29	15
16	LECHE NATURAL COLUN S/T 1 LT X 12	1	16
17	PECHUGA PAVO COCIDA SOPRAVAL	9	17
18	LECHE SEMIDESCREMADA SURLAT 1 LT	17	18
19	QUESO CHACRA QUILLAYES 500 GRS.	15	19
20	MANTEQUILLA SOPROLE C/S PAN 250 GRS	27	20
21	MANTEQUILLA SOPROLE UNTABLE POTE 250 GRS	31	21
22	PECHUGA PAVO ASADA SOPRAVAL	12	22
23	JAMON AHUMADO PAVO	22	23
24	QUESO GAUDA LAMINADO PITRUFQUEN 2 KG	16	24
25	LECHE SEMIDESCREMADA COLUN S/T 1 LT	26	25
26	LONGANICILLA DE CAMPO LA CRIANZA 280 GR	24	26
27	LECHE TOTTUS ENTERA 1LT	36	27
28	PICHANGA SAN FRANCISCO	21	28
29	LECHE TOTTUS DESCREMADA 1LT	28	29
30	MANTEQUILLA COLUN PAN 125 GRS.	88	30
31	LECHE NATURAL SOPROLE S/T 1 LT	40	31
32	QUESO MANTECOSO 500 GRS	23	32
33	LECHE DESCREMADA SOPROLE BOTELLA 1 LT	50	33
34	PECHUGA PAVO AHUMADA SOPRAVAL	35	34
35	JAMON DE PAVO COCIDO WINTER	46	35
36	PATE TERNERA PREFERIDA 160 GRS.	153	36

Fuente: Elaboración propia.

Surge interesante observar la diferencia que existe entre un simple ranking por ventas en contraste a uno utilizando la metodología actual. Basta con mirar el primer producto para ver como este si bien es un producto cuyas ventas lo sitúan en el lugar número 19 en el ranking de ventas, el nuevo ranking PROFSET lo considera como el producto más rentable. Esto debido principalmente a que el queso rallado genera mucha venta cruzada, y además está presente en muchas canastas frecuentes.

Para analizar de mejor medida cual es el verdadero valor del producto que se está considerando en la tabla 12 se describe a los seis primeros productos del ranking de la metodología actual junto con el producto que más vende y que está en la posición número 16 por el ranking PROFSET.

Tabla 14: Indicadores de desempeño de los mejores productos según PROFSET.

Marca	Descripción	R. PROFSET	R. Venta	Venta (MM\$)	Ticket Prom Total (M\$)	Numero de canastas frecuentes
DOS ALAMOS	QUESO RALLADO PARMESANO 40 GRS.	1	19	1.9	48	7017
COLUN	LECHE NATURAL COLUN S/T 1 LT	2	2	3.5	30	1612
COLUN	QUESO RALLADO PARMESANO 40G	3	37	1.4	49	3218
KRAFT	QUESO PHILADELFIA 226 GRS.	4	5	2.9	52	818
COLUN	LECHE DESCREMADA COLUN S/T 1 LT	5	6	2.6	31	894
COLUN	QUESO RANCO LAMINADO X 5 KG	6	3	3.3	20	320
COLUN	LECHE NATURAL COLUN S/T 1 LT X 12	16	1	3.9	58	59

Fuente: Elaboración propia.

Si se analiza el primer producto (del ranking PROFSET) es posible observar que este es el producto número 19 si solo se consideran las ventas, por lo que el ranking PROFSET lo sobreestima por otros efectos además de las ventas. El primer producto se observa que se compra principalmente en tickets de alto monto (\$48.000 pesos en promedio), el segundo, es al mismo tiempo el segundo producto con mayores ventas y el segundo producto en el ranking de PROFSET. Luego se podría llegar a pensar que la metodología actual esta simplemente valorando más a productos que se venden en tickets altos, lo que es directamente proporcional con la cantidad de productos pertenecientes en el ticket de compra, por lo que significaría un mayor número de canastas frecuentes (el producto número uno del ranking está presente en 7017 canastas frecuente).

Pero basta con analizar el producto con mayores ventas, para encontrar un contra ejemplo a lo señalado anteriormente. La caja de 12 leches Colun que es el producto con mayores ventas, es comprado en tickets de \$58.000 en promedio pero solo está presente en 59 canastas frecuentes. Luego hace suponer que este producto a pesar de ser comprado en canastas con gran cantidad de productos, no genera efecto de cross selling con los demás productos lo que se ve reflejado en el ranking PROFSET, que a diferencia del de ventas, aquí está número 16 en el ranking.

En conclusión la metodología si bien considera el efecto de ventas de un producto, también toma en consideración el efecto de correlación que existe entre los productos y logró así valorizar los productos incluyendo estos dos efectos al mismo tiempo.

9.5.3 Productos candidatos a ser eliminados del mix

Una de las aplicaciones de esta metodología es el determinar productos candidatos a ser eliminados del mix. Gracias a que todo el procedimiento se encuentra automatizado y es rápidamente replicable, es posible estar constantemente monitoreando los productos y encontrar así productos que han bajado su desempeño o que simplemente no están aportando en ventas lo que se esperaba de ellos, tanto en ventas directas como indirectas (efecto cross selling).

Si bien la metodología actual determina el orden en que los productos aportan a las ventas totales, sería una decisión miope el asegurar que es recomendable eliminar un producto del mix con solo mirar el ranking generado por PROFSET. A pesar de esto, con la ayuda de un experto sería posible determinar qué productos debiesen eliminarse del mix apoyado del ranking generado por PROFSET ya que este captura un efecto que es difícilmente medible pero muy importante a la hora de tomar decisiones de *assortment*.

9.5.4 Productos candidatos a ser incluidos en el mix

El problema de agregar productos al mix siempre ha sido un problema complejo y muy poco abordado, ya que es imposible evaluar el desempeño de un producto en un mix al cual no pertenece. Por lo que la única solución a esto es la de estimar cual podría llegar a ser el desempeño de algún nuevo producto en el mix al que se desea incluir.

Para abordar este problema, se utilizará el Tottus de Kennedy como laboratorio, aprovechando que es la tienda con el mayor número de productos distintos, tanto considerando todos los productos como también al considerar los productos estudiados en la intención de compra “Agregados once”. Por lo que se aprovechará la información obtenida en el Tottus de Kennedy respecto a las correlaciones entre los productos para determinar que productos son potenciales para ser incluidos en algún otro supermercado con “similares” características que actualmente no posea tal producto.

Para determinar que tiendas son similares entre sí se utilizarán las conclusiones obtenidas tanto por Passalacqua (Passalacqua, 2008) y por Solervicens (Solervicens Rojas, 2009). Que tras haber realizado un proceso para encontrar *clusters* de tiendas mediante k-medias, encontraron que los factores más importantes para considerar a dos tiendas similares son el monto de ventas de la tienda (para un determinado periodo de tiempo) en conjunto con el tamaño de la tienda. Es decir, dos tiendas son consideradas similares si estas poseen ventas similares y una superficie de tienda similar.

La tienda elegida en esta ocasión es el supermercado Tottus de Antofagasta ubicado en la calle Balmaceda #2352, que al igual que la tienda de Kennedy se encuentran ambas a un costado de un centro comercial, y además las ventas de la tienda en Antofagasta son solo un 9% menor que las de Kennedy.

Si observamos en ambas tiendas solamente la categoría “Agregados once”, vemos que existen 171 productos que fueron seleccionados por PROFSET en los pasos anteriores y que no están presentes en la tienda de Antofagasta. Por otro lado y para demostrar que efectivamente

existe una oportunidad de aprovechar la tienda de Kennedy para encontrar oportunidades de inclusión de productos en otras tiendas, vemos que solo existen 39 productos que se encuentran en la tienda de Antofagasta y no en Kennedy³³.

Por lo tanto se cuenta con una tienda con un amplio surtido y con características similares a otra tienda con un menor surtido. Por lo que si aprovechamos la información obtenida por PROFSET encontraremos productos que no se encuentran en la tienda de Antofagasta y que tanto su venta individual como cruzada son muy altas.

En la Tabla 15 es posible ver algunos productos que son candidatos a ser incluidos en la categoría “Agregados once” de la tienda de Antofagasta, ordenados según el ranking PROFSET que obtuvieron estos productos en la tienda de Kennedy, estos son productos con un muy buen desempeño en la tienda de Kennedy y que no se encuentran en la tienda de Antofagasta. Pero, si bien es posible encontrar productos candidatos a ser incluidos en la nueva tienda de Antofagasta, no es posible asegurar que el desempeño del producto será el mismo que el de Kennedy, ya que existen variados factores que incluyen en el desempeño de este.

Tabla 15: Productos candidatos a ser incluidos en la tienda de Antofagasta.

Ranking en Kennedy	Marca	Descripción
7	LA ROTUNDA	QUESO LAMINADO GRANEL
14	LA PREFERIDA	JAMON PIERNA PREFERIDA
26	LA CRIANZA	LONGANICILLA DE CAMPO LA CRIANZA 280 GR
43	HUENTELAUQUEN	QUESO MANTECOSO HUENTELAUQUEN PIEZA
54	SOPRAVAL	PECHUGA AHUMADA PAVO 150 GR
58	QUILLAYES	QUESILLO QUILLAYES 0% 300 GRS.
67	CRIADORES	QUESO MANTECOSO LOS CRIADORES PIEZA
72	SURLAT	LECHE DESCREMADA SURLAT 1 LT X 12
81	CRIADORES	QUESO MANTECOSO LOS CRIADORES TROZO
85	SOPRAVAL	PECHUGA COCIDA PAVO 150 GR
86	LA PREFERIDA	JAMON PIERNA PREFERIDA 200 GR
91	MULPULMO	QUESO CHANCO LAMINADO MULPULMO 1 KG
102	SOPROLE	LECHE DESCREMADA SOPROLE BOTELLA 1 LT X6
104	SOPRAVAL	PECHUGA ACARAMELADA PAVO 150 GR

Fuente: Elaboración propia.

No hay que olvidar que el ranking PROFSET considera a la vez la venta individual del producto junto con la venta cruzada de este. Y debido a que el mix de productos en ambas tiendas no es el mismo, es posible que tanto la venta individual del producto como su venta cruzada sean distintas. Para estimar la venta individual del producto en el nuevo mix sería necesario el estudio

³³ Esta afirmación no es del todo correcta, ya que cabe la posibilidad de que alguno de los 39 productos si esté presente en Kennedy pero debido a que no presento ventas durante el periodo que se observaron las transacciones, da la impresión de que no existe. De todas formas 39 pasa a ser una cota superior a esta afirmación.

del efecto de sustitución y complementariedad entre los productos ya existentes en el mix, en conjunto con otros atributos. Pero es posible estimar en cierta medida el efecto de venta cruzada del producto candidato a ser incluido en el nuevo mix.

Debido a que como se señaló con anterioridad, la tienda de Kennedy posee un amplio mix, y que todos los productos del mix están incluidos en el modelo que ya se utilizó para generar la optimización PROFSET, es posible eliminar del modelo los productos que se encuentran en Kennedy y no en Antofagasta y correr nuevamente el modelo. Al realizar esto sería posible estimar en cuanto disminuye el valor del producto al haber eliminado del modelo todos los productos que no se encuentran en la nueva tienda, y poder tener una estimación más acertada de cuál sería el valor del producto.

X. COMENTARIOS Y CONCLUSIONES

En el presente trabajo se ha logrado crear una metodología para analizar el mix de productos considerando el efecto de venta cruzada que presentan los diversos productos. Y poder con este análisis realizar diversas tareas útiles para el retail como por ejemplo: Determinar los productos más y menos rentables dentro de un conjunto de productos, determinar el mix de productos óptimos para la creación de una tienda de conveniencia, determinar productos candidatos a ser exhibidos en las cabeceras de las góndolas he incluso productos candidatos a ser promocionados en descuento.

Todas las tareas anteriormente señaladas utilizan información transaccional para descubrir en ellas las diversas asociaciones de productos que realizan los compradores y poder así valorizar los productos no solo por su aporte individual a la utilidad del negocio sino además por su aporte al considerar la venta cruzada de estos con otros productos.

Esta metodología llamada AMP está compuesta principalmente por tres etapas: descubrir las intenciones de compra, determinar y valorizar las canastas frecuentes y finalmente una etapa de selección de productos en base a sus correlaciones. Si bien las tres etapas son necesarias para llevar a cabo el proceso completo ya descrito en este trabajo, fundamentalmente las dos primeras etapas son necesarias para poder realizar la última. Pero además de ser necesarias, ambas entregan información relevante que es posible utilizar independientemente del resto de la metodología, es por lo mismo que las tres etapas serán abordadas independientemente en este capítulo.

Descubrir intenciones de compra

En esta primera etapa de AMP se analizan las transacciones de un determinado periodo para encontrar las diversas intenciones de compra que se presentan en la tienda. Para conseguir esto se asume que debido a la gran cantidad de tickets que se analizan, es más probable que productos de una misma intención de compra aparezcan juntos. Por eso se utiliza un indicador de análisis de canasta para determinar la cercanía entre productos, el cual determina en cuando aumenta la probabilidad de encontrar un segundo producto en el ticket en el que ya se encuentra un determinado producto. Si esa probabilidad es alta, entonces los productos son muy cercanos.

Una vez que se cuenta con la matriz de distancia entre todos los productos a analizar, se prosigue a realizar un análisis multidimensional para encontrar en esta matriz factores que expliquen la cercanía de los productos. Finalmente el resultado del análisis multidimensional es agrupado mediante una técnica de agregación jerárquica que consigue agrupar a los productos en clúster y así determinar las diversas intenciones de compra.

Si bien se probaron diversos indicadores para determinar la distancia entre productos, se utilizó el juicio de expertos para determinar que distancia era la que entregaba un resultado que más se aproximaba a la realidad y también el que presentara menor ruido. Obteniendo finalmente que el indicador que mejor resultados entrega es el que la distancia entre los productos era igual a la mínima confianza entre ellos, el cual es utilizado por primera vez por Böckel en 1978.

Una vez contando con la distancia entre los productos, la literatura presenta diversas técnicas que se utilizan para agrupar estos productos. En esta metodología se utiliza un análisis multidimensional seguido de un agrupamiento jerárquico. Si bien esta metodología entrega buenos resultados a juicio de expertos, es muy intensiva en procesamiento computacional debido

al elevado requerimiento de memoria que implica el análisis multidimensional. Esta dificultad crece exponencialmente a medida que aumenta la dimensión de la matriz de distancia entre los productos. Debido a esta complejidad es que se decidió utilizar el software SAS ya que otros software similares no eran capaces de llevar a cabo la tarea.

Por lo mismo, en caso de tener una matriz de productos de más de 150x150 distancias, se recomienda utilizar la técnica de K-medias para agrupar los productos y así determinar las intenciones de compra.

El entender las diversas intenciones de compra que son cubiertas por el mix de productos presentes en la tienda es útil para desarrollar estrategias de *layout* por ejemplo, juntando productos que corresponden a una determinada intención de compra. Sin olvidar que si bien dos productos pueden pertenecer a la misma intención de compra hay que analizar el ciclo de compra de estos. Ya que si bien es claro que el pan y la mantequilla corresponden a la misma intención de compra por ser productos muy afines, el ciclo de compra de estos dos productos es muy distinto, luego no es directo el asumir que por que dos productos se encuentran en una misma intención de compra estos siempre son comprados juntos.

Esta etapa se utilizó en esta metodología para mejorar la captura del efecto de venta cruzada entre los productos. Ya que se trabajó de distinta forma a los productos que se encuentran dentro y fuera de la intención de compra a estudiar (en este trabajo se estudió la intención de compra “Agregados Once” del supermercado Tottus).

Otra utilidad que tiene el descubrir las diversas intenciones de compras de un negocio es la de segmentar a los clientes según las intenciones de compra que satisfacen en un determinado negocio y así entender de mejor forma el comportamiento de los clientes. Si bien este trabajo no aborda la segmentación de clientes en base a necesidades, Falabella ha utilizado este conocimiento para entender mejor a sus clientes.

Determinar Frequent Sets y valorizarlos

Al igual que la etapa anterior, la información obtenida en esta etapa se puede utilizar independiente de haber llevado a cabo toda la metodología. Es aquí donde aprovechando la información de las intenciones de compra obtenida en la etapa anterior, se descubren las diversas canastas frecuentes y se les determina un valor.

Para llevar a cabo el proceso de encontrar las canastas frecuentes se utilizó el algoritmo Apriori el cual demostró llevar a cabo el proceso de forma eficiente, encontrando un gran número de canastas frecuentes en un pequeño tiempo. Cabe recordar que el parámetro utilizado para considerar determinada canasta como frecuente, el soporte de cada canasta, es el único parámetro que sirve como input para el algoritmo. Y es este parámetro el que determina la cantidad de memoria que se requiere para almacenar las distintas canastas frecuentes, a menor soporte mínimo, más canastas frecuentes se encuentran y mayor es la cantidad de memoria requerida para llevar a cabo el proceso.

Tal como señala Brijs en su trabajo, el error de considerar a demasiadas canastas como frecuentes errando en catalogar a algunas que no lo son, es mitigado por la siguiente parte de esta etapa donde se valoran estas canastas frecuentes. Ya que las canastas frecuentes con un bajo soporte prácticamente no son valorizadas con el algoritmo que se utiliza tras haberlas encontrado.

Una vez teniendo valorizado las diversas canastas frecuentes, es posible directamente observar cuales son las canastas tipos que se forman y contrastar al mismo tiempo el valor de estas.

Debido a que este proceso es rápidamente replicable, esta parte de la metodología sirve incluso para evaluar acciones de marketing como por ejemplo el rediseño del *layout*. Si el objetivo del rediseño fue aumentar el cross-selling entre determinados productos, como por ejemplo colocando una pequeña góndola de vinos blanco junto a la sección de pescadería, esta parte de la metodología responderá inmediatamente a la pregunta de si fue o no efectiva la acción. Al contrastar el valor de las canastas compuestas entre el vino y los pescados tanto antes como después del rediseño.

Selección de productos en base a sus correlaciones

A diferencia de las dos etapas anteriores, para llevar a cabo esta si es necesario el haber realizado las etapas anteriores. Ya que esta etapa aprovecha la información obtenida en la primera etapa para aumentar la significancia entre las relaciones y la etapa dos para contar con el valor de cada canasta frecuente para poder optimizar en base a esto.

Debido a la complejidad de mezclar tanta información variada es que se requiere indudablemente el apoyo de diversos scripts computacionales para llevar a cabo esta tarea. Lo cual presenta una ventaja debido a que los scripts son creados especialmente para realizar la tarea requerida, estos son fácilmente modificables en caso de ser necesario a futuro. Pero también la creación de estos genera una dificultad para transmitir y replicar la metodología en otro lugar.

Habiendo resuelto las complejidades previamente descritas, la optimización en sí converge rápidamente al óptimo global. Lo que permite replicar la optimización para diversos parámetros y poder realizar un análisis de sensibilidad de estos.

El objetivo general del trabajo fue conseguido, recordando que la metodología planteada sirve para detectar oportunidades de rediseño en el mix de productos pero no como una metodología para determinar el mix de productos óptimo. Esto debido a la complejidad de determinar un mix de productos óptimo. Si bien trabajos anteriores se habían enfocado en analizar de forma particular una categoría determinada dentro de una tienda, y estudiar de esta sus productos bajo diversas perspectivas. Ninguno había dado el paso de estudiar el problema del mix de producto con una mirada global y transversal entre diversas categorías.

Por lo anterior, y a causa de que este trabajo solo estudia las interrelaciones entre todos los productos de la tienda, es que no se puede utilizar como un método para determinar el mix de productos óptimos si no como el estudio de una arista de este complejo problema.

XI. PROPUESTA PARA TRABAJOS FUTUROS

El problema de determinar el mix de productos óptimo como se ha visto a lo largo de este trabajo al parecer es un problema más que complejo. Si bien obtener el mix óptimo es prácticamente una utopía, debido al dinamismo y complejidad del negocio, se busca acercarse cada vez más a este.

Los avances computacionales y la gran cantidad de información que se recolecta de distintas fuentes, abre un abanico de oportunidades para que el investigador desarrolle diversos trabajos. Cada vez más se busca apoyar las decisiones de negocio con datos, desde determinar donde abrir una nueva tienda hasta el decidir la eliminación de algún producto del mix.

Si bien los encargados de cada categoría se ven asustados ante la eliminación de algún producto dentro del mix por miedo a causar descontento en los consumidores, diversos estudios han demostrado que es posible eliminar gran cantidad de los productos de un mix sin afectar las ventas de dicha categoría. Siempre y cuando los clientes sigan encontrando las marcas que buscan y que los metros lineales asignados a la categoría sigan constantes. Si bien este trabajo aborda la problemática de eliminar productos dentro del mix, se cree que se pueden realizar mejoras a la heurística utilizada en este trabajo que consiste en eliminar un porcentaje de los productos con un peor ranking según PROFSET.

Para realizar lo anterior de mejor manera, es posible determinar en cuanto disminuye el valor de la función objetivo de la optimización al ir eliminando cada producto. Es decir, calcular el valor de la función de utilidad con todos los productos y luego el valor de esta con todos los productos menos el primero, luego todos los productos menos el segundo, y así sucesivamente. Luego este “costo” de eliminar cada producto podría ser incluido en el modelo de optimización de PROFSET para que de esta forma la optimización tomase en consideración el costo de eliminar algún producto del mix.

En segundo lugar se cree que es posible mejorar tanto el criterio para determinar si una canasta es frecuente o no como el criterio para asignar la utilidad a cada canasta, ya que los utilizados en este trabajo parecen ser muy arbitrarios. Con lo que respecta al primer criterio, en este trabajo se buscó utilizar el mínimo soporte posible para el algoritmo que encuentra las canastas frecuentes, argumentando que las canastas poco frecuentes no serían valorizadas en el proceso de valorización de canastas debido a su bajo soporte y por consiguiente baja probabilidad de asignación de utilidad. Luego este criterio es posible mejorarlo, y así evitar considerar como frecuentes a canastas que son estadísticamente poco significativas. Para esto (Castelo, Feelders, & Siebes, 2001), se han utilizado técnicas basadas en Markov blankets, para encontrar canastas frecuentes estadísticamente significativas y al mismo tiempo valorizarlas.

En tercer lugar es posible mejorar aún más el modelo utilizado actualmente haciendo que este no solo considere los efectos positivos de venta cruzada entre los productos, sino que además el efecto negativo entre los productos. Es decir, incluir en el costo de tener un producto el efecto de canibalización (entre otros efectos negativos) que produce este sobre los demás. Un esbozo a este problema podría ser utilizar el indicador de asociación de Ganancia (5.4.1) para entender el efecto de complementariedad y sustitución entre los productos.

Durante el trabajo se señaló que con esta metodología era posible capturar los efectos de promociones y cambios de *layout*, al contrastar la composición de canastas en los periodos durante, antes y después de la promoción. Si bien se espera que existan cambios en la composición de las canastas de los clientes gatilladas por estas acciones, no fue probado durante el presente trabajo. Luego surge necesario el testeo de dicha hipótesis como un trabajo a futuro y

en caso de poder capturar el cambio, cuantificarlo. Para así tener una nueva forma de evaluar este tipo de acciones.

Finalmente, se ve necesario el desarrollo de algún trabajo que incluya la mirada interna de una categoría de productos como una mirada inter categoría. En lo que respecta a gestión del mix de productos dentro de una categoría, gran cantidad de estudios se han llevado a cabo al respecto, abordando complejos indicadores como la entropía de los productos dentro de una categoría. Sería interesante la integración de estos indicadores de desempeño de una categoría en conjunto con la mirada presente en este trabajo del efecto de correlación existente entre los productos de distintas categorías.

XII. BIBLIOGRAFÍA

- Agrawal, R. (1993). Mining Association Rules Between Sets of Items in Large Database. *SIGMOD Conference*, (pp. 207-216).
- Blischok, T. (1995). *Every transaction tells a story. Chain Store Age Executive with Shopping Center Age*.
- Böcket, F. (1978). Determination of the Bond purchase of products. *Marketing Conference*.
- Brijs, T., Swinnen, G., Vanhoof, K., & Wets, G. (2004). Building an association rules framework to improve product assortment decisions. *Data mining and Knowledge Discovery*, 8, 7-23.
- Castelo, R., Feelders, A., & Siebes, A. (2001). MAMBO: Discovering association rules bases on conditional independencies. En F. Hoffmann, *Advances in Intelligent Data Analysis* (págs. 289-298). Springer.
- Cortés Valenzuela, J. P. (2006). *Metodología para la evaluación de la gestión de categorías en tiendas de conveniencia*. Tesis Universidad de Chile.
- Dréze, J. C. (2006). Your Loyalty Program is Betraying You. *Harvard Business review*.
- DuMouchel, W., & Pregibon, D. (2001). Empirical bayes screening for multi-item associations. *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (págs. 67-76). San Francisco CA (USA).
- Fayyad, U. M. (1996). *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- Hossain, R. (2006). Market Basket Analysis: An Approach to Association Rule Mining in Multiple Store Enviroment.
- Jiawei Han, J. P. (2004). Mining frequent patterns without candidate generation. En *Data Mining and Knowledge Discovery* (págs. 53-87).
- Kleinberg, J., Papadimitriou, C., & Raghavan, P. (1998). A microeconomic view of data mining. *Data Mining and Knowledge Discovery Journal*, 311-324.
- Mannila, H. (1997). Methods and problems in data mining. *In proceedings of the international conference on database theory*, 44-55.
- Passalacqua, A. (2008). *Metodología de apoyo a la toma de decisiones en surtido, espacio y ubicación de productos en una cadena de supermercado*. Santiago: Tesis Universidad de Chile.
- Shalizi, C. (2009). *Clustering II - Distances between partitions*. Pittsburgh.
- Shawn, R., & Stone, M. (1990). *Database Marketing*. Wiley US Edition.
- Silverstein, C., Brin, S., & Montwani, R. (1998). Beyond market baskets: Generalizing association rules to dependence rules. *Data Mining and Knowledge Discovery journal*, 39-68.
- Solervicens Rojas, M. J. (2009). *Determinación de surtido para la gestion de categoría en una cadena de tiendas de especialidad*. Tesis Universidad de Chile.
- Srikant, R. A. (1994). Fast algorithms for mining association rules in large databases. En M. J. Jorge B. Bocca, *Proceedings of the 20th International Conference on Very Large Data Bases* (págs. 487-499). Santiago, Chile.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to data mining*.
- Van der Ster, W., & Van Wissen, P. (1993). *Marketing & Detailhandel*. Wolters-Noordhoff.
- Yasemin Boztug, L. H. (s.f.). *A Market Basket Analysis Based on the Multivariate MNL Model*. Berlin.

Zaki, M. J. (2000). Scalable algorithms for association mining. En *Transactions on Knowledge and Data Engineering* (págs. 372-390).

XIII. ANEXOS

Anexo A: Script para calcular las correlaciones entre productos

Lenguaje PHP:

```
$archivos = Array("transac_enero.txt","transac_febrero.txt");
$salida_file = "salida_a_Excel.txt";
$out = Array();
$t = Array();
    for($j = 0;$j < count($ids_productos);++$j){
        $out[$ids_productos[$j]] = Array();
        for($k = 0;$k < count($ids_productos); ++$k){
            $out[$ids_productos[$j]][$ids_productos[$k]] = 0;
        }
    }
$logfile=fopen("log.txt", "w");
foreach($archivos as $archivo){
    $mi_archivo=file("".$archivo);
    fwrite($logfile,"Corriendo->".$archivo."\n");
    foreach($mi_archivo as $linea){
        $l =explode(";", trim($linea));
        //Se cuenta el numero de productos en la boleta
        $np = 0;
        for($i = 0;$i < count($l);++$i){
            $l[$i] = trim($l[$i]);
            if($l[$i] != 0) ++$np;
        }
        for($j = 0;$j < count($l);++$j){
            if($l[$j] != 0){
                for($k = ($j+1);$k < count($l);++$k){
                    if($l[$k] != 0){
                        if($np != 1){
                            $out[$ids_productos[$j]][$ids_productos[$k]] += (1/($np - 1));
                            $out[$ids_productos[$k]][$ids_productos[$j]] += (1/($np - 1));
                        }
                    }
                }
            }
        }
    }
}
```

```
$nuevo_archivo=fopen($salida_file, "w");
fwrite($nuevo_archivo,"IDs;");
foreach($out as $id1 => $v1){
    fwrite($nuevo_archivo,$id1.";");
}
fwrite($nuevo_archivo, "\n" );
foreach($out as $id1 => $v1){
    foreach($v1 as $id2 => $cuenta){
        fwrite($nuevo_archivo, $out[$id1][$id2].";"
);
    }
    fwrite($nuevo_archivo, "\n" );
}
fwrite($nuevo_archivo, "\n" );
fclose($nuevo_archivo);

}
```

Anexo B: Diversas Categorías de Tottus Kennedy

FOOD	NON FOOD	
LIQUIDOS	SERVICIOS Y DONACIONES	DEPORTES
CARNICERIA	LAVADO Y MANTENCION	PROMOCIONES
LACTEOS	PERFUMERIA	LIBRERIA
ABARROTES	HOMBRES	ARTICULOS DE TEMPORADA
CONGELADOS	JUGUETERIA	TECNOLOGIA
PESCADERIA	FERRETERIA	JUVENIL DAMAS
FIAMBRERIA	VESTUARIO NIÑOS	LINEA BLANCA
PANADERIA Y PASTELERIA	MENAJE	AUDIO/VIDEO
FRUTAS Y VERDURAS	VESTUARIO DEPORTIVO	JUVENIL HOMBRES
PLATOS PREPARADOS	ROPA INTERIOR	MALETERIA
	CALZADO	JARDINERIA
	ELECTRODOMESTICOS	S/JERARQUIA
	ENTRETENIMIENTO	MATERIALES Y DESECHOS
	AUTOMOVIL	ACCESORIOS VESTUARIO
	DAMAS	DECORACION
	TEXTIL HOGAR	CALZADO
	ALIMENTOS Y ARTICULOS MASCOTAS	DECORACION/AMBIENTACION CASA

Anexo B: Categorías y Sub-Categorías en Food

Categoría	Sub-Categoría
ABARROTES	ALIMENTOS BASICOS
	ALIMENTOS EN CONSERVA
	DESAYUNOS/DULCES
	CONFITERIA
	POSTRES/REPOSTERIA
	COCTEL
	ALIMENTOS PARA BEBE
CARNICERIA	CARNES DE VACUNO
	CARNES DE POLLO
	CARNES DE CERDO
	CARNES DE PAVO
	CARNES DE CORDERO
CONGELADOS	ALIMENTOS CONGELADOS
	ALIMENTOS REFRIGERADOS
FIAMBRERIA	FIAMBRES/CECINAS
	HUEVOS FRESCOS
FRUTAS Y VERDURAS	VERDURAS
	FRUTAS
LACTEOS	QUESOS
	MANTECAS/MANTEQUILLAS
	LECHES Y CREMAS
	POSTRES Y MANJAR
	YOGHURT
LIQUIDOS	BEBIDAS
	BEBIDAS ALCOHOLICAS
	ENVASES
PANADERIA Y PASTELERIA	PANADERIA
	PASTELERIA
PESCADERIA	PESCADOS
	MARISCOS Y ALGAS
PLATOS PREPARADOS	COMIDAS PREPARADAS
	INSUMOS PLATOS PREPARADOS

Anexo C: Matriz de distancia entre sub categorías

SUB CATEGORIAS	BEBIDAS	CARNES DE VACUNO	QUESOS	ALIMENTOS BASICOS	ALIMENTOS CONGELADOS	PESCADOS	FIAMBRES/CECINAS	MANTECAS/MANTEQUILLAS	PANADERIA	ALIMENTOS EN CONSERVA	BEBIDAS ALCOHOLICAS	LECHES Y CREMAS	VERDURAS	DESAYUNOS/DULCES	CONFITERIA	POSTRES Y MANJAR	FRUTAS	CARNES DE POLLO	YOGHURT	PASTELERIA	POSTRES/REPOSTERIA	COCKTAIL	COMIDAS PREPARADAS	CARNES DE CERDO	ENVASES	ALIMENTOS REFRIGERADOS	ALIMENTOS PARA BEBE	HUEVOS FRESCOS	CARNES DE PAVO	MARISCOS Y ALGAS	CARNES DE CORDERO	INSUMOS PLATOS PREPARADOS	
BEBIDAS	0	0.97	0.97	0.97	0.96	0.97	0.96	0.98	0.96	0.97	0.96	0.97	0.97	0.97	0.94	0.97	0.97	0.97	0.97	0.96	0.98	0.95	0.90	0.97	0.90	0.97	0.97	0.98	0.98	0.97	0.97	0.89	
CARNES DE VACUNO	0.97	0	0.98	0.98	0.99	0.99	0.98	0.99	0.97	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.98	0.98	0.99	0.99	0.99	0.99	1.00	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.98	1.00	
QUESOS	0.97	0.98	0	0.98	0.98	0.98	0.97	0.98	0.96	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.98	0.98	0.98	1.00	
ALIMENTOS BASICOS	0.97	0.98	0.98	0	0.98	0.97	0.98	0.98	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	
ALIMENTOS CONGELADOS	0.96	0.99	0.98	0.98	0	0.98	0.98	0.99	0.97	0.99	0.98	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.98	1.00	
PESCADOS	0.97	0.99	0.98	0.97	0.98	0	0.98	0.99	0.97	0.99	0.98	0.99	0.97	0.98	0.98	1.00	0.97	0.99	0.99	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	0.99	1.00	0.98	1.00	1.00	
FIAMBRES/CECINAS	0.96	0.98	0.97	0.98	0.98	0.98	0	0.98	0.94	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.99	0.98	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	1.00	
MANTECAS/MANTEQUILLAS	0.98	0.99	0.98	0.98	0.99	0.99	0.98	0	0.97	0.99	0.99	0.98	0.98	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99	1.00	0.99	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	
PANADERIA	0.96	0.97	0.96	0.97	0.97	0.97	0.94	0.97	0	0.98	0.98	0.97	0.97	0.97	0.98	0.98	0.97	0.97	0.97	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	
ALIMENTOS EN CONSERVA	0.97	0.99	0.98	0.97	0.99	0.99	0.98	0.99	0.98	0	0.99	0.99	0.98	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	
BEBIDAS ALCOHOLICAS	0.96	0.99	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.99	0	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.97	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.98	0.98	1.00	
LECHES Y CREMAS	0.97	0.99	0.98	0.98	0.99	0.99	0.98	0.98	0.97	0.99	0.99	0	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.99	1.00	
VERDURAS	0.97	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.98	0.98	0	0.98	0.98	0.96	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.99	0.97	0.98	0.97	1.00	
DESAYUNOS/DULCES	0.97	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.99	0.98	0.98	0	0.97	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.98	0.97	0.98	0.98	0.99	0.98	1.00	
CONFITERIA	0.94	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0	0.98	0.98	0.99	0.98	0.98	0.98	0.97	0.98	0.99	0.98	0.98	0.98	0.97	0.99	0.99	0.99	0.99	0.99	
POSTRES Y MANJAR	0.97	0.99	0.99	0.98	0.99	1.00	0.99	0.99	0.98	0.99	0.99	0.98	0.98	0.98	0	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	0.99	
FRUTAS	0.97	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.98	0.98	0.96	0.98	0.98	0.98	0	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.99	
CARNES DE POLLO	0.97	0.98	0.98	0.98	0.98	0.99	0.98	0.99	0.97	0.99	0.99	0.99	0.98	0.98	0.99	0.98	0	0.99	1.00	0.99	0.99	1.00	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	
YOGHURT	0.97	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.97	0.99	0.99	0.98	0.98	0.98	0.98	0.99	0	0.99	0.99	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	1.00	
PASTELERIA	0.96	0.99	0.98	0.99	0.99	1.00	0.98	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	1.00	0.99	0	0.99	0.99	0.99	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	
POSTRES/REPOSTERIA	0.98	0.99	0.98	0.98	0.99	1.00	0.99	0.99	0.98	0.99	0.99	0.98	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
COCKTAIL	0.95	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.98	0.99	0.97	0.99	0.98	0.98	0.97	0.99	0.98	0.99	0.99	0.99	0.99	0	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00
COMIDAS PREPARADAS	0.90	1.00	0.99	0.99	0.99	0.99	0.99	1.00	0.97	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.98	1.00	0.99	0.99	1.00	1.00	0	1.00	0.99	1.00	1.00	1.00	1.00	0.99	1.00	0.83	
CARNES DE CERDO	0.97	0.98	0.99	0.98	0.99	1.00	0.98	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.99	1.00	0.98	0.98	0.99	1.00	1.00	0.99	1.00	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
ENVASES	0.90	0.99	0.99	0.98	0.99	1.00	0.98	1.00	0.97	0.99	0.98	0.99	0.98	0.99	0.98	1.00	0.98	0.99	0.99	0.99	1.00	0.99	0.99	1.00	0	1.00	1.00	0.99	1.00	1.00	1.00	1.00	
REFRIGERADOS	0.97	0.99	0.98	0.98	0.98	1.00	0.98	0.99	0.98	0.99	0.99	0.98	0.98	0.98	0.98	0.99	0.98	0.99	0.99	0.99	1.00	0.99	1.00	1.00	1.00	0	1.00	0.99	1.00	1.00	1.00	1.00	
ALIMENTOS PARA BEBE	0.97	0.99	0.99	0.98	0.99	1.00	0.99	0.99	0.98	0.99	0.99	0.99	0.97	0.97	0.99	0.98	0.99	0.98	0.99	1.00	0.99	1.00	0.99	1.00	1.00	1.00	0	0.99	1.00	1.00	1.00	1.00	
HUEVOS FRESCOS	0.98	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.97	0.99	0.99	0.98	0.97	0.98	0.99	0.99	0.98	0.99	0.99	1.00	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0	0.99	0.99	0.99	1.00	
CARNES DE PAVO	0.98	0.99	0.98	0.98	0.99	1.00	0.98	0.99	0.98	0.99	0.99	0.98	0.98	0.99	1.00	0.98	0.99	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0	1.00	1.00	1.00	
MARISCOS Y ALGAS	0.97	0.99	0.98	0.98	0.98	0.98	0.98	1.00	0.98	0.99	0.98	0.99	0.97	0.99	0.99	1.00	0.97	0.99	0.99	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	0.99	1.00	0	1.00	1.00	
CARNES DE CORDERO	0.97	0.98	0.98	0.98	0.99	1.00	0.98	1.00	0.98	0.99	0.98	0.99	0.98	0.98	0.99	1.00	0.98	0.99	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0	1	
PREPARADOS	0.89	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1	0	

Anexo D: Resultado del análisis multidimensional

Nombres	Dim1	Dim2	Dim3	Dim4	Dim5	Dim6	Dim7	Dim8	Dim9	Dim10
BEBIDAS	-0.724	-0.481	0.135	-0.204	-0.018	-0.211	-0.083	0.033	-0.124	-0.273
CARNES DE VACUNO	1.482	-0.228	-1.086	-1.82	-1.209	0.029	-0.853	-0.427	0.747	0.212
QUESOS	0.978	0.295	1.39	0.535	1.234	0.365	-1.12	-0.308	0.011	1.126
ALIMENTOS BASICOS	1.28	0.115	0.04	0.512	-1.32	-1.111	-0.018	-0.099	0.192	-0.261
ALIMENTOS CONGELADOS	0.272	-0.168	0.487	1.002	0.167	0.544	-0.306	2.439	0.869	-0.221
PESCADOS	1.097	-0.728	0.309	1.424	0.309	0.807	1.518	0.097	1.797	1.128
FIAMBRES/CECINAS	1.099	-0.787	1.176	-1.325	0.498	0.544	0.547	0.29	-0.458	0.832
MANTECAS/ MANTEQUILLAS	0.219	0.99	-0.169	-0.151	-0.039	-0.61	-0.234	0.038	-1.051	3.098
PANADERIA	-0.351	-0.266	-0.251	-0.156	0.537	0.101	0.451	0.029	-0.067	0.86
ALIMENTOS EN CONSERVA	0.419	0.313	-0.532	2.023	0.288	-1.29	-0.537	-1.713	0.666	0.197
BEBIDAS ALCOHOLICAS	0.11	-2.181	1.669	0.163	-0.581	0.088	0.257	-1.158	-0.278	-0.654
LECHES Y CREMAS	0.605	2.088	0.564	-0.624	0.695	0.251	0.319	0.284	-1.654	0.365
VERDURAS	1.578	-0.177	-1.456	0.501	0.23	0.269	-0.398	0.601	0.194	0.299
DESAYUNOS/DULCES	-0.415	1.768	0.145	0.069	-1.134	-0.955	0.754	-1.358	-0.11	0.105
CONFITERIA	-1.529	1.026	1.187	0.307	-0.296	0.118	-0.182	0.197	0.653	-0.957
POSTRES Y MANJAR	-1.947	0.918	-0.349	-0.538	0.279	2.618	-0.323	0.479	-0.297	-0.194
FRUTAS	0.35	-0.179	-1.204	1.036	0.539	1.401	-0.025	-0.029	0.319	-0.533
CARNES DE POLLO	1.537	-0.356	-0.864	-1	-0.605	0.253	1.608	0.396	1.291	-1.35
YOGHURT	-1.022	1.97	-0.72	-0.018	1.282	1.169	0.732	-0.421	0.446	-0.524
PASTELERIA	-1.549	0.345	1.349	-1.284	1.006	-0.841	-0.406	0.828	1.997	0.536
POSTRES/REPOSTERIA	-0.484	1.387	0.078	1.006	-1.161	-2.491	0.692	1.222	-0.201	-0.3
COCTEL	-0.318	-0.416	2.397	0.33	0.583	-0.428	-0.57	-0.443	-0.564	-1.981
COMIDAS PREPARADAS	-2.565	-1.955	-1.584	0.125	-0.188	-0.214	-0.216	-0.178	-0.28	0.332
CARNES DE CERDO	0.87	-0.836	-0.202	-2.465	-1.112	-0.578	-0.695	1.591	-0.739	-0.33
ENVASES	-0.76	-1.278	1.036	-0.36	1.087	-0.435	2.285	-0.161	-1.766	0.043
ALIMENTOS REFRIGERADOS	0.379	0.576	-0.269	0.251	0.733	-0.242	-3.23	-0.029	-0.673	-0.858
ALIMENTOS PARA BEBE	-1.059	1.739	0.629	-0.167	-2.707	1.15	0.248	-0.376	0.498	0.149
HUEVOS FRESCOS	0.85	0.334	-1.803	1.026	-0.275	0.643	0.732	0.096	-2.203	-1.08
CARNES DE PAVO	0.658	0.194	-1.423	-0.548	2.398	-1.757	0.502	0.366	0.618	-0.767
MARISCOS Y ALGAS	1.092	-1.753	0.91	1.772	-1.037	0.9	-0.631	0.755	-0.398	0.52
CARNES DE CORDERO	1.082	-0.373	-0.098	-1.44	0.208	0.508	-0.33	-2.832	0.494	-0.094
INSUMOS PLATOS PREPARADOS	-3.234	-1.897	-1.492	0.021	-0.392	-0.595	-0.484	-0.207	0.069	0.575

Anexo E: Indicadores de las diversas categorías del estudio

				Numero de clientes = 19887		Tickets Emitidos = 429		Venta Total (MM\$) = \$ 5352					
Categoría		Distintos SKU	N°Transac (Miles)	N° Transac IC (Miles)	Distintos clientes por corr	Reac h	Distintos clientes por IC	Tickets (Miles)	Pres	Tickets (Miles)	Venta (MM\$)	%venta	Mnt_venta UN (MM\$)
Comida Preparada	COMIDAS PREPARADAS	261	126	132	3330	17%	3348	59	14%	60	\$ 147	2.7%	\$ 148
	INSUMOS PLATOS PREPARADOS	2	7		170	1%		6	1%		\$ 1	0.0%	
Conveniencia	BEBIDAS	1218	761		14563	73%		225	52%		\$ 499	9.3%	
	PANADERIA	193	282		12973	65%		176	41%		\$ 189	3.5%	
	CONFITERIA	1358	461		12633	64%		134	31%		\$ 267	5.0%	
	PASTELERIA	388	64	1917	5618	28%	18501	40	9%	355	\$ 91	1.7%	\$ 1239
	POSTRES Y MANJAR	139	76		4105	21%		26	6%		\$ 37	0.7%	
	YOGHURT	349	255		8560	43%		69	16%		\$ 131	2.4%	
ALIMENTOS PARA BEBE	76	19		945	5%		6	1%		\$ 24	0.5%		
Agregado Orice	QUESOS	337	183		10907	55%		94	22%		\$ 262	4.9%	
	FIAMBRES/CECINAS	579	206	640	11215	56%	15022	108	25%	176	\$ 268	5.0%	\$ 776
	MANTECAS/MANTEQUILLAS	97	47		5830	29%		30	7%		\$ 43	0.8%	
LECHES Y CREMAS	284	204		9394	47%		71	17%		\$ 203	3.8%		
Pescados y Congelados	ALIMENTOS CONGELADOS	629	185		9378	47%		68	16%		\$ 307	5.7%	
	MARISCOS Y ALGAS	33	4	196	483	2%	9679	3	1%	72	\$ 8	0.2%	\$ 345
	PESCADOS	57	7		1000	5%		5	1%		\$ 29	0.5%	
Alcohol y cocktail	BEBIDAS ALCOHOLICAS	1777	173		7758	39%		74	17%		\$ 457	8.5%	
	COCKTAIL	556	123	347	7353	37%	11102	55	13%	117	\$ 117	2.2%	\$ 575
	ENVASES	14	52		2020	10%		15	3%		\$ 1	0.0%	
Carnes	CARNES DE VACUNO	232	122		8312	42%		54	13%		\$ 454	8.5%	
	CARNES DE CERDO	41	17	213	2547	13%	10479	12	3%	78	\$ 53	1.0%	\$ 685
	CARNES DE POLLO	77	73		7242	36%		42	10%		\$ 176	3.3%	
	CARNES DE CORDERO	9	0		90	0%		0	0%		\$ 3	0.1%	
Abarrotes	ALIMENTOS BASICOS	2171	526		12340	62%		104	24%		\$ 444	8.3%	
	DESAYUNOS/DULCES	1390	293	1024	11674	59%	15040	94	22%	159	\$ 374	7.0%	\$ 1006
	POSTRES/REPOSTERIA	291	39		3254	16%		14	3%		\$ 18	0.3%	
	ALIMENTOS EN CONSERVA	669	165		8133	41%		51	12%		\$ 171	3.2%	
Frutas, Verduras y huevos	VERDURAS	270	391		10597	53%		103	24%		\$ 243	4.5%	
	FRUTAS	166	277		10809	54%		110	26%		\$ 231	4.3%	
	HUEVOS FRESCOS	22	42	738	5839	29%	13480	36	8%	155	\$ 54	1.0%	\$ 578
	CARNES DE PAVO	32	13		1865	9%		8	2%		\$ 28	0.5%	
	ALIMENTOS REFRIGERADOS	67	14		2071	10%		10	2%		\$ 22	0.4%	

Anexo F: Modelo y datos (resumido) para AMPL.

```
# SETS
param NPRODUCTOS >=0, integer;
set PRODUCTOS := 1..NPRODUCTOS;
param NFRECSETS >=0, integer;
set FRECSETS := 1..NFRECSETS;
param NCAT >=0, integer;
set CAT := 1..NCAT;
set PINCAT {CAT} within PRODUCTOS;
set PINFS {FRECSETS} within PRODUCTOS;

# VARIABLES
var p {x in FRECSETS}, binary; #Colocar o no el frecuentset
var q {i in PRODUCTOS}, binary; #Colocar o no el PRODUCTO

# PARAMETERS
param M {x in FRECSETS} >= 0; #Margen del frec itemset
param ITEM_MAX >= 0; #Maximo de productos a colocar
param C {i in PRODUCTOS} >= 0; #Costo productos
param MIN_CAT {mc in CAT}; #Minimo de productos a elegir por
categoría
param R {r in PRODUCTOS}, binary;

# Función Objetivo
MAXIMIZE utilidad: (sum {x in FRECSETS} M[x]*p[x]) - sum{i in
PRODUCTOS} C[i]*q[i];

# RESTRICCIONES
subject to Maximo_productos:sum{i in PRODUCTOS}q[i] <= ITEM_MAX;
subject to RELACION_ITEMSET_PRODUCTO {x in FRECSETS,i in
PINFS[x]}: q[i] >= p[x];
subject to MINIMO_EN_CATEGORIAS {ca in CAT}:
sum{i in PINCAT[ca]} q[i] >= MIN_CAT[ca];
```

```

param NFRECSETS:=XXX;
param NPRODUCTOS:=YYY;
param NCAT:=Z;

/*Productos en categoría
set PINCAT[1] :=1    2    3...;
set PINCAT[2] :=855  856  857  858...;
...

/*Productos en Frequent Sets
set PINFS[1] :=12    14    15    31    33    51    867;
set PINFS[2] :=14    15    31    51    69    1740;
set PINFS[3] :=14    15    33    51    96    1564;
...

/*Utilidad asignada a cada Frequent Set
param M:=
1    846738
2    698324
3    353809
...;

/*Costo del producto
param C:=
1    1
2    1
3    1
...;

```