

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

**LEVANTAMIENTO DE PATRONES DE COMPORTAMIENTO DE
PRECIOS UNA CADENA DE SUPERMERCADO**

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

VANJA VALENTINA DEMETRIO CERDA

**SANTIAGO DE CHILE
ABRIL 2012**



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

**LEVANTAMIENTO DE PATRONES DE COMPORTAMIENTO DE
PRECIOS DE UNA CADENA DE SUPERMERCADO**

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

VANJA VALENTINA DEMETRIO CERDA

PROFESOR GUIA:
LUIS ABURTO LAFOURCADE

MIEMBROS DE LA COMISIÓN:
MARCEL GOIC FIGUEROA
EMILIO POLIT CORVALÁN

SANTIAGO DE CHILE
ABRIL 2012

LEVANTAMIENTO DE PATRONES DE COMPORTAMIENTO DE PRECIOS DE UNA CADENA DE SUPERMERCADO.

La fijación de precios en la industria del retail se desarrolla hoy en día en base a definición de márgenes objetivos, en función de la competencia, o usando modelos de optimización que son alimentados con estimaciones de demanda que requieren de extensos tiempos de tratamiento de datos. Con el presente trabajo, se busca generar herramientas alternativas que mejoren sustancialmente el desempeño de la categoría en un tiempo menor.

Las reglas buscan restringir el accionar de quienes toman decisiones en los retailers, alertándolos de las posibles consecuencias que conllevan promociones aplicadas para levantar las ventas de un producto, en sus productos sustitutos y en la competencia. Por lo anterior, el objetivo principal del proyecto es definir reglas de precios para los productos de una categoría, que permitan apoyar las decisiones tomadas a nivel operativo.

La metodología se enfoca en determinar las relaciones de orden existentes entre los precios de productos de una misma categoría, para levantar patrones de comportamiento de precios y rentabilidad en base a datos transaccionales pasados. Posteriormente, éstos se establecen como reglas de precios en tres niveles: sucursal, cadena y en comparación con la competencia; las cuales son evaluados con indicadores de técnicas de Reglas de Asociación.

Los indicadores utilizados para la selección de las reglas son tres: soporte, confianza, y peso, que respectivamente, se consideran como: la frecuencia de ocurrencia de la regla, la proporción de veces en que la regla representa una mejora en la contribución, y la significancia de los indicadores con respecto al periodo evaluado.

El resultado obtenido con el desarrollo de la metodología, es un set de 40 reglas de precios a nivel producto, a nivel de atributos y en comportamiento frente a la competencia. Estas reglas pueden aplicarse como filtros reguladores en sus distintos niveles de agregación, y constituirse como pre pricing para la fijación de precios de una compañía.

De la evaluación económica realizada a una regla aplicable a toda la cadena, se tiene que la pérdida por no cumplirla es de aproximadamente \$16.918.158 anuales.

Finalmente, del trabajo realizado es posible concluir que las relaciones entre los precios de los productos pueden ser correctamente evaluadas con los indicadores construidos, ya que éstos poseen capacidad de discriminación entre las reglas que representan una mejora para la contribución de los productos, de las que no; y porque las reglas seleccionadas permanecen en distintos periodos de datos.

Agradezco a mi familia por todo el apoyo.

A mi madre por sus palabras,
a mi padre por su incondicionalidad,
y a mi hermana por su compañía.

ÍNDICE

1	INTRODUCCIÓN.....	9
1.1	ANTECEDENTES GENERALES	9
1.2	DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN	11
1.3	OBJETIVOS	12
1.3.1	Objetivo General.....	12
1.3.2	Objetivos Específicos	12
1.4	ALCANCES	13
1.5	RESULTADOS ESPERADOS	13
2	MARCO TEÓRICO	13
2.1	CONCEPTOS.....	14
2.2	REGLAS DE ASOCIACIÓN.....	14
3	METODOLOGÍA.....	16
3.1	PREPROCESAMIENTO DE DATOS.....	16
3.2	LEVANTAMIENTO DE REGLAS	16
3.2.1	A nivel Sucursal.....	17
3.2.2	Comparación entre Locales	18
3.2.3	Comparación con la Competencia.....	18
3.2.4	Reglas Atributadas	19
3.3	SELECCIÓN DE REGLAS.....	19
3.4	EVALUACIÓN DEL MODELO	20
4	DESARROLLO METODOLÓGICO Y PRESENTACIÓN DE RESULTADOS.....	21
4.1	CATEGORÍA Y ATRIBUTOS ESCOGIDOS.....	21
4.2	DATOS UTILIZADOS Y ANÁLISIS PRELIMINAR.....	24
4.3	PRE PROCESAMIENTO DE DATOS.....	29
4.4	LEVANTAMIENTO DE REGLAS	31
4.4.1	Reglas a Nivel Sucursal.....	31
4.4.2	Reglas de Comparación entre Locales	38
4.4.3	Reglas en Comparación con la Competencia	40
4.5	SELECCIÓN DE REGLAS.....	41
4.5.1	Reglas Seleccionadas a Nivel Sucursal.....	42
4.5.2	Reglas Seleccionadas de Comparación entre Locales	49
4.5.3	Reglas Seleccionadas en Comparación con la Competencia	51

4.6	Evaluación del Modelo.....	52
4.7	Aplicación	54
5	CONCLUSIONES	55
5.1	Conclusiones del Trabajo	55
5.2	Discusión de Resultados	56
5.3	Trabajos Futuros	57
6	REFERENCIAS	59
7	ANEXOS.....	60
7.1	ANEXO A: Maestro de Productos Categoría Leches Liquidas	60
7.2	ANEXO B: Regresión Lineal de Niveles de Atributos.....	62
7.3	ANEXO C: Ranking de Sucursales	63
7.4	ANEXO D: Estudio de Outliers	64
7.5	ANEXO E: Tabla Resumen Reglas a nivel Sucursal.....	66
7.6	ANEXO F: Reglas de comparación entre locales Atributadas.....	67
7.7	ANEXO G: Reglas contradictorias encontradas en reglas de comparación entre locales atributadas	69
7.8	ANEXO H: Reglas atributadas seleccionadas a nivel sucursal	70
7.9	ANEXO I: Reglas seleccionadas en comparación con la competencia	70
7.10	ANEXO J: Set resumen de reglas a aplicar	72

ÍNDICE DE TABLAS

Tabla 1:	Resumen de Indicadores de Reglas de Asociación.....	15
Tabla 2:	Resumen de indicadores de reglas de asociación, adaptados al problema.....	17
Tabla 3:	Precio por ml promedio ponderado atributo Marca.....	25
Tabla 4:	Precio por ml promedio ponderado atributo Tipo.....	26
Tabla 5:	Precio por ml promedio ponderado atributo Sabor	26
Tabla 6:	Precio por ml promedio ponderado atributo Packaging	27
Tabla 7:	Segmentos de Sucursales	28
Tabla 8:	Formato regla a nivel sucursal	32
Tabla 9:	Formato ramificación reglas a nivel sucursal.....	35
Tabla 10:	Variación Confianzas con Ramificación	37

Tabla 11: Estadísticas de resultados obtenidos en Reglas de Comparación entre Locales Atributadas	39
Tabla 12: Descripción productos estudiados de la competencia	40
Tabla 13: Resultados a Nivel Sucursal	42
Tabla 14: Estadísticas reglas de Tipo, marca Colun	43
Tabla 15: Estadísticas reglas de Tipo, marca Soprole	44
Tabla 16: Resultados Ramificación a Nivel Sucursal	49
Tabla 17: Resultados Reglas de Comparación entre Locales	49
Tabla 18: Resultados Reglas de Comparación entre Locales Atributadas	50
Tabla 19: Reglas de Orden	50
Tabla 20: Porcentajes de reglas que pasan testeo	53
Tabla 21: Estimación de pérdida en regla ejemplo.....	53
Tabla 22: Porcentaje de pérdida estimado en regla ejemplo.....	54

ÍNDICE DE GRÁFICOS

Gráfico 1: Porcentaje Acumulado de la Venta y la cantidad de productos que la componen.....	21
Gráfico 2: Ventas productos más vendidos	22
Gráfico 3: Análisis de Sucursales.....	28
Gráfico 4: Dispersión de pares (confianza, soporte) de Reglas a nivel Sucursal.....	32
Gráfico 5: Porcentajes de Tipos de Reglas a Nivel Sucursal.....	33
Gráfico 6: Histograma de variaciones de precios.....	35
Gráfico 7: Porcentajes de Tipos de Reglas Ramificación	36
Gráfico 8: Histogramas Confianzas Reglas Originales y Ramificadas.....	37
Gráfico 9: Dispersión de pares (confianza, soporte) de Reglas de Comparación entre Locales - Productos	38
Gráfico 10: Porcentajes de Tipos de Reglas en Comparación con la Competencia	41
Gráfico 11: Estudio para replicar reglas N°1, N°2, y N°3 (LECHE ENTERA COLUN CAJA 1L más cara que LECHE COLUN SEMI DESCREMADA CAJA 1L).	45
Gráfico 12: Estudio para replicar reglas N°5, N°6, y N°7 (LECHE ENTERA COLUN CAJA 1L más cara que LECHE PARMALAT ENTERA CAJA 1LT).....	45
Gráfico 13: Estudio para replicar reglas N°14, N°15, N°16, y N°17 (LECHE SOPROLE ENTERA CAJA 1L más cara que LECHE PARMALAT ENTERA CAJA 1LT).....	46

Gráfico 14: Estudio para replicar reglas N°12 y N°13 (LECHE SOPROLE DESCREMADA CAJA 1L más cara que LECHE SOPROLE ENTERA CAJA 1L).....	46
Gráfico 15: Estudio para replicar reglas potenciales N°1, N°2, y N°3 (LECHE SOPROLE CHOCOLATE CAJA 200ML más cara que LECHE SOPROLE FRUTILLA CAJA 200ML)	48
Gráfico 16: Estudio para replicar reglas potenciales N°4, N°5, N°6, N°7, N°8, N°9, y N°10 (LECHE SOPROLE CHOCOLATE CAJA 200ML más cara que LECHE LONCO CHOCOLATE CAJA 200ML)	48
Gráfico 17: Histograma de contribuciones reglas seleccionadas y resto	52

ÍNDICE DE FIGURAS

Figura 1: Impacto en los beneficios de una compañía, al mejorar en un 1% ciertas variables. ...	9
Figura 2: Etapas de la fijación de precios.....	10
Figura 3: Estructura lógica de la metodología	20
Figura 4: Distribución de Venta por Marca	22
Figura 5: Distribución de Venta por Tipo	23
Figura 6: Distribución de Venta por Sabor	23
Figura 7: Distribución de Venta por Packaging	24

1 INTRODUCCIÓN

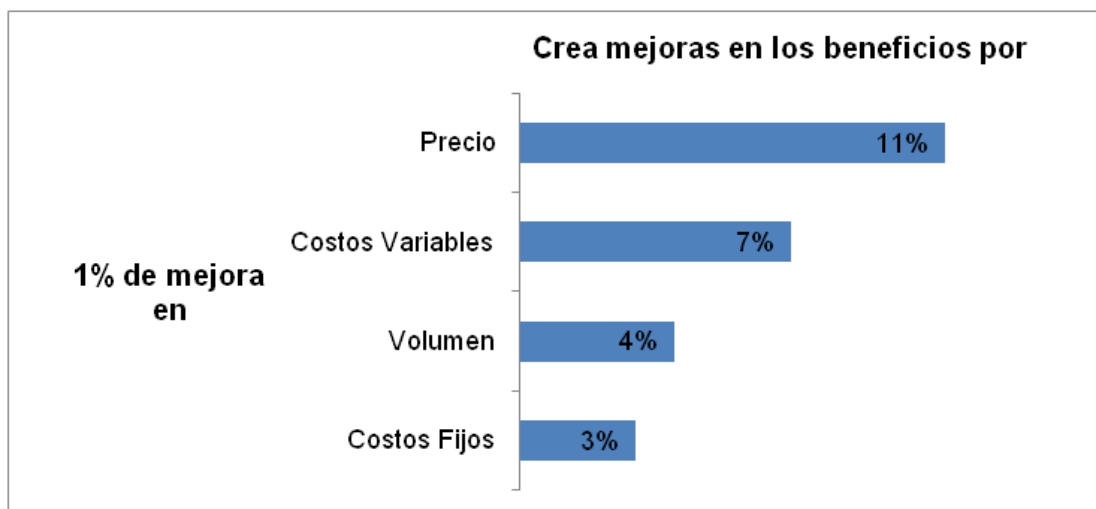
1.1 ANTECEDENTES GENERALES

La industria del retail en Chile representa hoy el foco principal del gasto de la demanda interna, y sus ventas representan cerca del 9% del PIB del país. El crecimiento experimentado por la industria en Enero 2012 en doce meses fue de un 6,3%, superando levemente la expectativa que se tiene para el presente año de un 6% anual ([1], [2],[3], y [4]).

Lograr modelos de negocios rentables en el largo plazo es uno de los principales desafíos que se presentan para la gestión de retail, donde las decisiones que se deben enfrentar para mejorar los resultados de las compañías son principalmente la variedad (gama de productos disponibles), el surtido (profundidad de productos en las categorías), el nivel de servicio que se busca entregar, la experiencia de compra, y el precio. En especial este último constituye una variable crítica para los resultados económicos.

Estudios indican que entre las variables claves para mejorar los resultados de una compañía (unidades vendidas, precio, costos fijos y costos variables) el instrumento más poderoso es la fijación de precios[5], [6], y [7], ya que una buena estrategia en este ámbito asegura traducirse en resultados con un mayor porcentaje de impacto que las demás variables. Como se muestra en la figura a continuación, una mejora de un 1% en la fijación de precios, crea una mejora de un 11% en los beneficios totales de una compañía.

Figura 1: Impacto en los beneficios de una compañía, al mejorar en un 1% ciertas variables.



*Basado en el promedio de la economía de 2.463 compañías.

Fuentes: Exhibit 1. Comparison of Profit Levers[6], and Exhibit A: Profit drivers, [7].

Además, en la industria del retail, el precio es uno de los factores más influyentes a la hora de establecerse como tienda predominante para el cliente[8], por lo que la gestión de éstos se conforma como una herramienta estratégica para capturar valor e incrementar la rentabilidad de la empresa.

Es por esto que las variaciones en los precios provocadas por promociones o por otras decisiones tácticas pueden traer consigo una serie de consecuencias que afecten la venta y/o el margen de ciertos productos, impactando directamente a la rentabilidad de la compañía. Entre las cuales se encuentran:

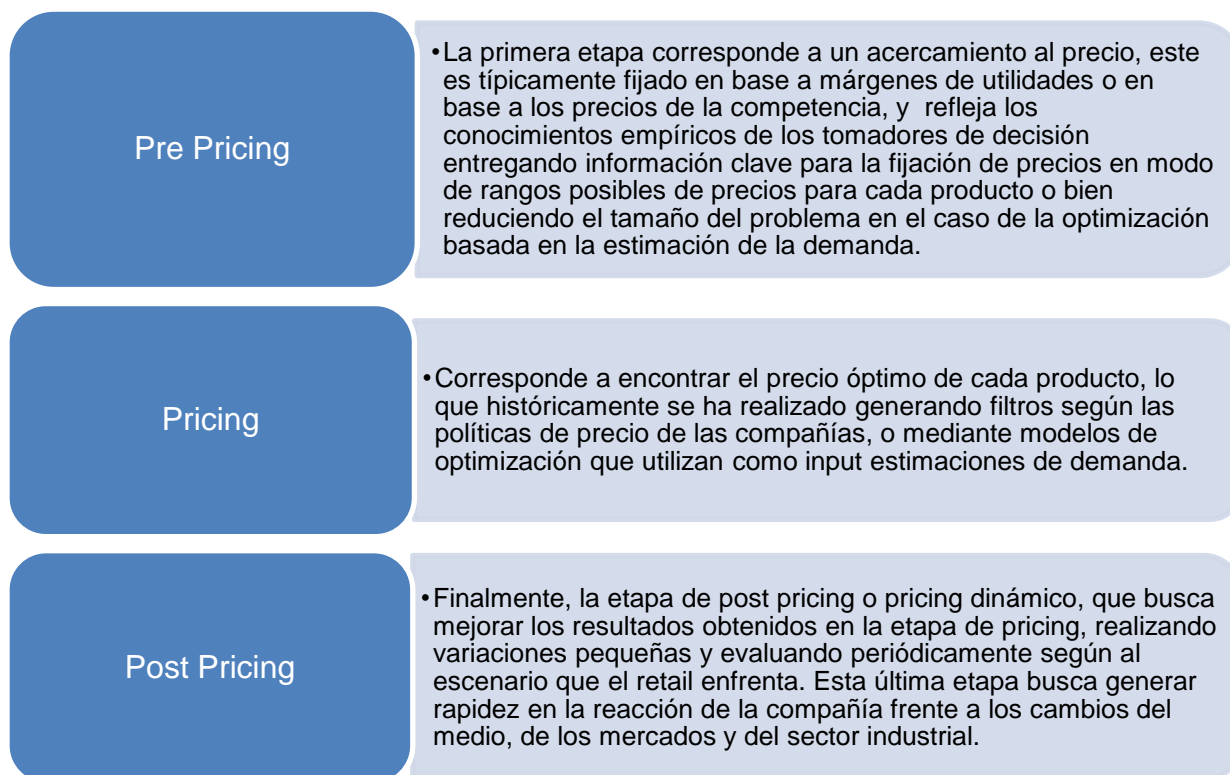
- Disminución de la competitividad frente a la competencia
- Destrucción de una marca
- Generación de quiebres de stock en formatos individuales, si los formatos familiares (o a granel) están más caros, etc.

En la industria del retail, es frecuente que los proveedores impulsen promociones de sus productos. Entre éstas, las promociones de descuentos directos y las de compensaciones conforman ofertas que deben ser oportunamente evaluadas por quienes gestionan las categorías, de modo de no dañar las ventas de los otros productos.

La visión global de lo que ocurre con las ventas de la categoría en general, si se mueve el precio de un cierto producto, es una tarea muchas veces no bien lograda por los managers debido a la alta complejidad que tiene en su práctica. Es por esto que para apoyar estas decisiones resulta necesario construir herramientas de control que adviertan de los efectos producidos con cada movimiento en los precios.

El proceso de fijación de precios en una empresa se puede separar en tres etapas diferentes[9], con distintos objetivos y grados de complejidad.

Figura 2: Etapas de la fijación de precios.



Fuente: Elaboración propia

Lo esencial en el proceso completo, es que los managers se resistan a la tentación de dejar elementos o variables influyentes fuera de las evaluaciones, por lo difícil que éstas sean de medir y/o calcular. Es importante que se haga ingeniería en el manejo de los precios, y que se renueve constantemente, de modo de adaptarse mejor a las nuevas preferencias y necesidades del cliente. Cualquier esfuerzo en mejorar las decisiones tomadas en este ámbito, tiene un pago extremadamente alto para las compañías [6].

El presente trabajo se ubica en la etapa de pre pricing, y se enfoca en determinar las relaciones existentes entre los precios de productos de una misma categoría, estableciendo reglas que restrinjan el accionar de los tomadores de decisiones de los retailers. Se busca alertarlos de los posibles perjuicios que puedan generar las activaciones para levantar las ventas de un producto, en el resto de los productos de la categoría.

La metodología que se presenta, encuentra la información necesaria de una categoría para hacer pricing con las políticas de precios de la compañía. Conformando iteraciones o ciclos, se puede verificar el cumplimiento de las reglas, y se constituye así una alternativa al pricing de optimización. Las reglas pueden tener distintas prioridades dependiendo de los productos que afecten (típicamente éstas se establecen por la presencia en boletas y/o las ventas de los productos).

1.2 DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN

Como se menciona anteriormente, la fijación de precios es una de las herramientas más influyentes cuando se busca mejorar los resultados de alguna compañía, llegando a generar con un 1% de mejora en el manejo de los precios, un 11% de mejora en la ganancia de la empresa [6] y [7].

Históricamente, la fijación de precios se ha realizado en base al conocimiento del negocio, definiendo márgenes de utilidades objetivos, apuntando a posicionarse de algún modo respecto de la competencia, o en base a modelos de optimización que requieren de estimaciones de demanda con tiempos extensos de tratamiento de datos. Esta última razón, ha motivado la inversión de recursos y esfuerzos en estudios de investigación de distintas metodologías que mejoren los resultados y los tiempos de procesos obtenidos.

El desarrollo y aplicación de los modelos que requieran de estimación de demanda en la gestión de precios ha tenido buenos resultados para los supermercados, logrando mejoras para los márgenes e ingresos de alrededor de un 30% ([10] y [11]). Sin embargo, una de sus limitaciones es el tiempo (entre 1 y 2 semanas por categoría) que estos estudios requieren, si se proyecta el tiempo de estudio por cada una de las categorías (generalmente más de veinte por local) el periodo necesario para realizar pricing excede el tiempo que la gestión de los supermercados requiere para la toma de decisiones oportuna.

El objetivo de este nuevo trabajo, es generar herramientas alternativas que mejoren sustancialmente el desempeño de la categoría en un tiempo menor. Se busca desarrollar una metodología que sea capaz de disminuir los tiempos de ejecución y tener resultados prácticos que constituyan apoyo a la gestión, además de conformar un estudio factible para cualquier compañía que debe lidiar con la rapidez que caracteriza a la industria.

El modelo que se propone en el proyecto es el levantamiento de reglas de decisión, como una metodología nueva, que busque patrones en los datos históricos, los evalúe, y construya con estos un ordenamiento de precios para los productos de la categoría en estudio.

El trabajo se enfoca en determinar las relaciones existentes entre los precios de productos de una misma categoría, de modo de restringir el accionar de los tomadores de decisiones de los retailers, alertándolos de los posibles perjuicios que puedan generar las activaciones para levantar las ventas de un producto, en el resto de los productos de la categoría.

Con el proyecto se genera también una mayor capacidad de gestión, debido a que se entrega información sobre las preferencias del consumidor. Esto, ya que se establecen relaciones entre productos, que poseen distintos atributos, según el precio que estos tengan.

Por otro lado, si esta información es compartida con los proveedores permite que éstos puedan generar mejores ofertas al retail, y entiendan mejor los cambios de productos que realizan sus consumidores finales cada vez que los precios no cumplen el orden ideal.

Finalmente, en cuanto al beneficio económico, este proyecto afecta directamente las contribuciones de los productos analizados mejorando en consecuencia la contribución de la categoría completa en forma sustancial.

1.3 OBJETIVOS

1.3.1 Objetivo General

Levantar patrones de comportamiento de precios y rentabilidad en base a datos transaccionales pasados, que permitan apoyar decisiones de pricing en base a reglas de precios entre los productos de una categoría para una cadena de supermercado.

1.3.2 Objetivos Específicos

- Definir metodología genérica para el levantamiento de reglas de decisión.
- Aplicar la metodología en una categoría de una cadena de supermercado, identificando reglas de los precios a nivel producto y el comportamiento frente a la competencia.
- Evaluar el modelo en base a la capacidad de discriminación de reglas efectivamente mejores y la validación de las reglas con periodos diferentes de datos.
- Estimar impacto económico del no cumplimiento de las reglas.
- Explicitar cómo generar un modelo de fijación de precios con la información obtenida a partir de las reglas.

1.4 ALCANCES

La presente memoria busca desarrollar una metodología genérica para el levantamiento de reglas de decisión para la fijación de los precios, para lo que se utilizarán datos de una categoría de una cadena de supermercados mayoristas.

- Se trabajará sobre todas las sucursales de la cadena a nivel de SKU
- En el desarrollo del trabajo, no se construirán modelos para estimar o pronosticar demanda, por lo que no se calcularán elasticidades precio, ni elasticidades cruzadas.
- En el proyecto no se levantarán variables influyentes en la determinación de la competencia por sucursal, estas se tomarán como dato en base a los archivos recibidos
- El proyecto no se hará cargo de los errores cometidos en el pasado que se detecten con los indicadores calculados a las reglas.
- No se contempla la implementación del modelo de fijación de precios, si no solo su explicitación.

1.5 RESULTADOS ESPERADOS

En el trabajo se espera encontrar patrones de comportamiento levantados en base a datos históricos. Obteniendo reglas de decisión para la fijación de precios, a nivel producto dentro de cada sucursal, reglas de comparación entre locales y reglas de comparación con la competencia de una categoría en particular de la cadena.

Se espera medir el impacto económico del quebrantamiento de las reglas en los resultados de la cadena, en especial la contribución alcanzada, pudiendo identificar la importancia que este conocimiento tiene para quienes gestionan de las categorías. Se espera también, verificar que el modelo funcione seleccionando, en base a los indicadores calculados por este mismo, las reglas que representen un mayor aporte en los beneficios de la empresa.

Finalmente, discutir sobre la forma de aplicar las reglas a un modelo de fijación de precios, de modo de apoyar la toma de decisiones de los retailers y mejorar los tiempos de las prácticas actuales.

2 MARCO TEÓRICO

Dentro del marco referencial del trabajo, se presentan las bases conceptuales de la metodología que se utilizará. En primer lugar se definirán los conceptos básicos que circunscriben los procesos de pricing en la industria del retail de modo de contextualizar al lector, y posteriormente, se hablará sobre las técnicas de *Association Rules* y *Frequent Itemsets* utilizadas para la definición y cálculo de los indicadores, que finalmente, se utilizarán para el proceso de selección de las reglas.

2.1 CONCEPTOS

- **Pricing:** El pricing es un proceso en el que se buscan identificar los niveles óptimos de precios de manera de capturar la disposición a pagar del cliente, implementando la estrategia definida por la compañía para alcanzar el posicionamiento buscado. La fijación de precios pueden determinar el éxito futuro de un negocio, por lo que se considera uno de los factores más importantes en la gestión de un retail.
- **Categorías:** Corresponden a conjuntos de productos similares. Cada categoría cumple un rol dentro de la tienda que está definido en base a la rotación y la presencia en ticket que ésta tenga. La categoría que sea escogida definirá el espacio de productos sustitutos sobre el cual se buscarán patrones de comportamiento.

En el caso de un supermercado, los roles típicos que se pueden encontrar para una categoría, son los siguientes: destino (productos que impulsan la visita a tienda), rutina (productos de necesidad diaria de los clientes que complementan la canasta de destino), estacional y conveniencia (productos que corresponden a compras no planeadas y en los que se permite extraer un mayor margen).

Esta variable condicionará el nivel de información disponible en el proyecto, ya que los productos transados en cada periodo dependen de la rotación que tenga la categoría.

Además del rol, las categorías poseen una estrategia y distintas tácticas que pueden ser utilizadas. Como estrategias se pueden encontrar la generación de tráfico o margen, el reforzamiento de imagen, y/o la defensa de territorio. Y como tácticas se encuentran la exhibición, surtido, promoción, disponibilidad, servicio, y el precio.

- **Atributos:** Los atributos son las características que describen un producto de forma única, tienen distintos niveles, los cuales corresponden básicamente a las variedades disponibles de cada atributo. Por ejemplo, se tienen como atributos las marcas, los formatos, los display, etc. y sus niveles correspondientes serían todas las marcas disponibles en la categoría, todos los tamaños de gramaje disponibles en la categoría, y todas las presentaciones tales como vidrio, doypack, lata, etc. respectivamente. Los atributos permiten comparar los productos cualitativamente, por lo tanto serán determinantes para la definición de las reglas y muy influyentes para la interpretación de estas mismas.
- **Category Management:** Corresponde al manejo de los datos de mercado para afectar el desempeño de los productos, y de este modo, mejorar los resultados de la categoría. Con el proyecto se busca apoyar este objetivo, entregando una mirada global de los efectos en los precios de la categoría al mover uno de ellos.

2.2 REGLAS DE ASOCIACIÓN

Las reglas de Asociación han sido utilizadas típicamente para el análisis de canastas de compra, donde lo que se busca son las implicancias entre ítems (o productos). De este modo, una regla de asociación es una implicancia del tipo “si cierta canasta contiene todos los ítems del set I, entonces es probable que contenga también al ítem j” y se representan $I \Rightarrow j$, donde I es el antecedente de la regla, y j el consecuente. Las canastas son evaluadas con las transacciones del retail, y cada canasta o transacción está compuesta por un cierto set de ítems o productos ([12], [14] y [15]).

Entre los set de ítems que se componen las canastas existen los llamados “frecuentes” o *Frequent Itemsets*, que corresponden a aquellos que superan un estándar en su frecuencia de ocurrencia, es decir, que su soporte, indicador definido como la cantidad de canastas en las que dicho set es un subconjunto del set transado sobre el total de canastas analizadas, se ubique por sobre un cierto umbral[16].

Encontrando estas canastas de compra frecuentes, el retailer puede gestionar mejor tanto inventarios y promociones, como a sus propios clientes, mejorando su servicio y desarrollando ventajas competitivas frente a la competencia.

Las reglas de asociación son evaluadas típicamente con dos indicadores el soporte y la confianza. El soporte como se menciona antes, representa la proporción de canastas en las que el set antecedente de la regla aparece del total de canastas analizadas, y la confianza, que representa la probabilidad de que aparezcan ambos en una canasta dado que aparece el antecedente. Matemáticamente corresponde al soporte de I unido al consecuente, sobre el soporte de I[14].

Sin embargo, cuando se tiene un ítem consecuente con alta presencia en boletas (con un soporte notablemente alto), el resultado del indicador confianza puede verse afectado debido a que la cantidad de transacciones en el caso en que se vendan ambos ítems y en el que se venda solo el consecuente será similar, resultando una confianza alta no debido a una relación entre los ítems, si no, a que el soporte del otro ítem en cuestión se vuelve marginal frente al del producto muy vendido.

Existe entonces un tercer indicador que mejora este efecto, es llamado ganancia, y corresponde al puntaje de aumento de probabilidad de que sea transado el ítem j, a que sea transado j dado que se ha transado I. Con lo anterior, se mide la relación existente entre el antecedente y el consecuente sin influir el soporte del consecuente.

En la siguiente tabla se presenta un resumen de los indicadores antes explicitados, de modo de visualizar el cálculo que cada uno significa.

Tabla 1: Resumen de Indicadores de Reglas de Asociación

INDICADOR	ECUACIÓN
Soporte	$\text{Soporte } (I \Rightarrow j) = \frac{\text{Transacciones que contienen } (I \cup \{j\})}{\text{Total Transacciones}}$
Confianza	$\text{Confianza } (I \Rightarrow j) = \frac{\text{Transacciones que contienen } (I \cup \{j\})}{\text{Transacciones que contienen } (I)}$
Ganancia	$\text{Ganancia } (I \Rightarrow j) = \frac{\text{Confianza } (I \Rightarrow j)}{\text{Soporte } (j)}$

Fuente: Elaboración propia

Finalmente para encontrar los *Frequent Itemsets*, las Reglas de Asociación son evaluadas en sus indicadores usando algoritmos de búsqueda, debido a las grandes cantidades de datos y combinaciones de productos y subgrupos de productos que pueden ser analizados en un

retailer. Algunos de los algoritmos son el Algoritmo A-Priori, el Algoritmo de Park, Chen, and Yu, el Algoritmo de Toivonen, entre otros[17].

3 METODOLOGÍA

La metodología para la búsqueda de las reglas entre los precios semanales de los productos de una categoría se basa en la técnica de reglas de asociación, y utiliza la herramienta de consultas SQL.

Las etapas principales del proceso son cuatro, el acercamiento a los datos, la extracción de reglas candidatas de relaciones entre los precios, su selección en reglas frecuentes, y la evaluación del modelo, que considera el contraste en las contribuciones de las reglas seleccionadas versus las contribuciones del resto de las reglas y la medición de la validez de las reglas seleccionadas, en un testeo con datos de la misma categoría en un periodo diferente

3.1 PREPROCESAMIENTO DE DATOS

En esta primera etapa se seleccionarán la categoría y los atributos sobre los cuales se trabajará y los periodos de datos transaccionales a utilizar para la extracción como para el testeo de las reglas. Además, se calculará el precio semanal de cada sucursal como el cociente entre el monto y las unidades de producto vendidas en cada transacción, optándose por precio representativo de la semana el precio moda.

La medida anterior se prefiere porque representa un precio real, es decir, un valor al cual efectivamente fue transado el producto en sala esa semana, y porque no considera precios con descuentos por razones ajenas al estudio.

Dado el precio representativo de cada producto en cada semana, se estandarizará por el contenido neto del producto de forma de obtener un valor comparable que sea transversal a los atributos y sus respectivos niveles. A este valor se le llamará en adelante $precio_{mi}$.

Finalmente en esta etapa, se agregan a los datos indicadores de distancia entre productos de modo de identificar los productos que sean comparables en función de la cantidad de atributos que varían entre ellos.

3.2 LEVANTAMIENTO DE REGLAS

La extracción de reglas candidatas se desarrollará como una variante de las aplicaciones de las técnicas *Association Rules* y *Frequent Itemsets*, y se llevará a cabo considerando tres objetos de comparación de precios de los productos de una categoría, la primera considera la comparación a nivel de producto dentro de una misma sucursal, la segunda se enfoca en la comparación entre locales, y la tercera, busca identificar la posición que se tiene frente a la competencia.

3.2.1 A nivel Sucursal

Las asociaciones serán estudiadas en base al cálculo de indicadores de soporte y confianza para la relación general del tipo:

$$1) \text{ precioml}(i) > \text{precioml}(j)$$

Para este tipo de regla se define soporte como la importancia relativa del cumplimiento de la regla, sobre la cantidad total de periodos en que ambos productos son transados. A diferencia del soporte estándar, el indicador en este trabajo busca identificar la frecuencia de ocurrencia de la regla.

La confianza, en este caso, buscará señalar el vínculo existente entre el cumplimiento de la regla y sus resultados positivos, entendido como las veces en que dado que se cumple la regla, la contribución del par de productos es mayor o igual al promedio histórico.

La probabilidad condicional que constituye la confianza no será entonces sobre la aparición conjunta de dos ítems o productos, si no, sobre cada regla y sus casos favorables.

Del mismo modo que la ganancia regula el efecto de un ítem muy vendido en el resultado de la confianza en un problema de canasta, en este trabajo se buscará regular el efecto de un producto muy poco transado en el resultado del soporte y la confianza de una regla. Es decir, la ganancia en el contexto del proyecto distinguirá que tan importante es la cantidad de semanas en que se transan ambos productos con respecto al total de semanas estudiadas, y se calculará como una proporción directa de ambos factores.

Tabla 2: Resumen de indicadores de reglas de asociación, adaptados al problema.

INDICADOR	ECUACIÓN
Soporte	$\text{Soporte} = \frac{\text{Semanas en que se cumple la regla}}{\text{Total de semanas en que ambos productos son transados}}$
Confianza	$\text{Confianza} = \frac{\text{Semanas en que se cumple la regla} \wedge \text{la contribución es} \geq \text{al prom histórico}}{\text{Total de semanas en que se cumple la regla}}$
Ganancia	$\text{Ganancia} = \frac{\text{Total semanas en que ambos productos son transados}}{\text{Total semanas estudiadas (54)}}$

Fuente: Elaboración propia

Como se explica anteriormente, la ganancia será el indicador que corrija los resultados de aquellas reglas que pudiesen tener un alto soporte y una alta confianza, debido a que alguno de los productos que la conforman es de muy baja y/o esporádica venta, y resulta de ello que el total de semanas en que es transado es bajo. Este indicador será llamado en adelante como 'peso'.

Para la obtención de las reglas se consideran sólo los pares de productos comparables, es decir, aquellos que tienen sólo un atributo diferente. De este modo, son fijados los otros atributos y se puede estudiar la variación de cada uno por separado.

Sobre la base de los productos que pueden ser comparados se estudiarán las semanas en que se cumpla la sentencia o relación entre los precios que constituya cada regla, con el objeto de calcular el peso, soporte y confianza de ésta.

Luego, éstas últimas serán reevaluadas con segmentos más pequeños de rangos de diferencia de precios (ya no $>$, si no, $>5\%$ por ejemplo) para indagar sobre la posibilidad de encontrar reglas más desagregadas.

3.2.2 Comparación entre Locales

Las reglas de relación de los precios de los productos entre las distintas sucursales de la cadena se abordarán con foco en un mismo producto, de modo de comparar los niveles de precios de éste en las distintas salas y poder determinar si existe diferencia significativa y/o sistemática.

La relación buscada se conformará en base a sentencias como la que sigue:

$$\text{precioml}_{(i, \text{sucux})} > \text{precioml}_{(i, \text{sucuy})}$$

Donde i representa el producto analizado, que variará según los productos que ambas sucursales transen, y sucux e sucuy corresponden a las distintas sucursales con $s = \{1, \dots, 29\}$.

Los índices de soporte, confianza y peso, se calcularán del mismo modo, pero como se menciona anteriormente, con la diferencia en que el producto analizado es el mismo y la diferencia en el precio lo darán las sucursales que conformen la regla.

3.2.3 Comparación con la Competencia

El estudio de la relación con los precios de la competencia se trabajará con la información de los 9 productos más vendidos de la categoría.

Los precios de la competencia serán resumidos en el precio moda nacional de los precios levantados, debido a que la falta de rigurosidad en el procedimiento que se realiza para la toma de precios de la competencia hace que exista una gran cantidad de datos faltantes.

Una vez definidas los precios de la competencia, éstos se someterán a estudio, conformando reglas del tipo:

$$\text{precioml}_{(i, \text{sucux})} > \text{precioml}_{(i, \text{sucuy})}$$

Donde i representa el producto analizado, sucux corresponde a las sucursales de la cadena con $x = \{1, \dots, 29\}$, y sucuy corresponde al precio moda nacional de la competencia.

Los índices de soporte y confianza serán en este caso, semejantes a los de la parte anterior.

3.2.4 Reglas Atributadas

Como una derivada de las reglas de las etapas de sucursal y cadena, se pueden encontrar las reglas atributadas que corresponden a la agregación según los distintos niveles de los atributos de reglas encontradas en cada caso.

Esta derivación se realiza para generar reglas fácilmente gestionables por los tomadores de decisión (por ejemplo “Soprole es más caro que Loncoleche” es más manejable a la hora de llevar el conocimiento a accionables que “leche entera caja 1L Soprole más cara que leche entera caja 1L Loncoleche”).

El hecho de que sean combinaciones de las reglas entre productos encontradas, quiere decir que la regla entre Soprole y Loncoleche se constituya en base a las reglas entre las leches enteras caja 1L, entre las leches chocolatadas caja 200ml, leches descremadas caja 1L, etc. existentes entre las marcas Soprole y Loncoleche.

Los indicadores para estos casos, se conforman como combinaciones de los resultados anteriores:

- $\text{Soporte} = \frac{\sum_i \text{frecuencia de ocurrencia}_i}{\sum_i \text{total veces que ambos productos son transados}_i}$
- $\text{Confianza} = \frac{\sum_i \text{frecuencia de ocurrencia}_i \cdot \text{semanas favorables}_i}{\sum_i \text{frecuencia de ocurrencia}_i}$

Donde i representa el par de productos que componen una regla en la que el atributo en cuestión sea el comparado.

Es importante en este tipo de reglas considerar la cantidad de productos que son comparables entre dos niveles de un atributo, debido a que entre dos marcas no todos los productos tiene un símil, por ejemplo, se tiene en Soprole leche entera en bolsa 500ml producto que en Loncoleche no existe, se tiene en Loncoleche envasados del tipo square que en calo no se encuentran, etc.

3.3 SELECCIÓN DE REGLAS

La selección de reglas será en base a los índices de soporte y confianza calculados para cada una de ellas. Se buscarán las confianzas con mayor valor que tengan un soporte mínimo previamente definido como criterio de selección. Además se precisará como condición higiénica un nivel de peso mínimo que tengan que cumplir las reglas seleccionadas de modo de asegurar la significancia de los valores de soporte y confianza.

De esta forma, se obtienen las reglas más favorables a la contribución que tienen una frecuencia de ocurrencia base que asegura que éstas constituyen un patrón de comportamiento de los datos históricos.

Los criterios para cada uno de los indicadores serán elegidos en base a las reglas resultantes para cada corte. Estas serán evaluadas por opinión de experto y el corte se definirá dependiendo de la credibilidad que los conjuntos tengan.

3.4 EVALUACIÓN DEL MODELO

El modelo se evaluará midiendo la capacidad de discriminar que tenga, comprobando si las reglas que selecciona son efectivamente mejores en rentabilidad que el resto de las reglas. Más aún, como segunda prueba se medirá que las reglas detectadas tengan consistencia, es decir, que independiente del periodo escogido las reglas sigan ocurriendo.

Para evaluar la capacidad que tiene el modelo de discriminar entre reglas ‘buenas’ y ‘malas’, se confrontarán las contribuciones de las reglas seleccionadas con las del resto de las reglas en el periodo de testeo, es decir, con datos diferentes a los que se usaron para construirlo. Para escoger sólo uno de los dos sentidos en que las reglas pueden aparecer, se utilizará un número aleatorio que defina entre las reglas no seleccionadas, cuál de las dos será evaluada.

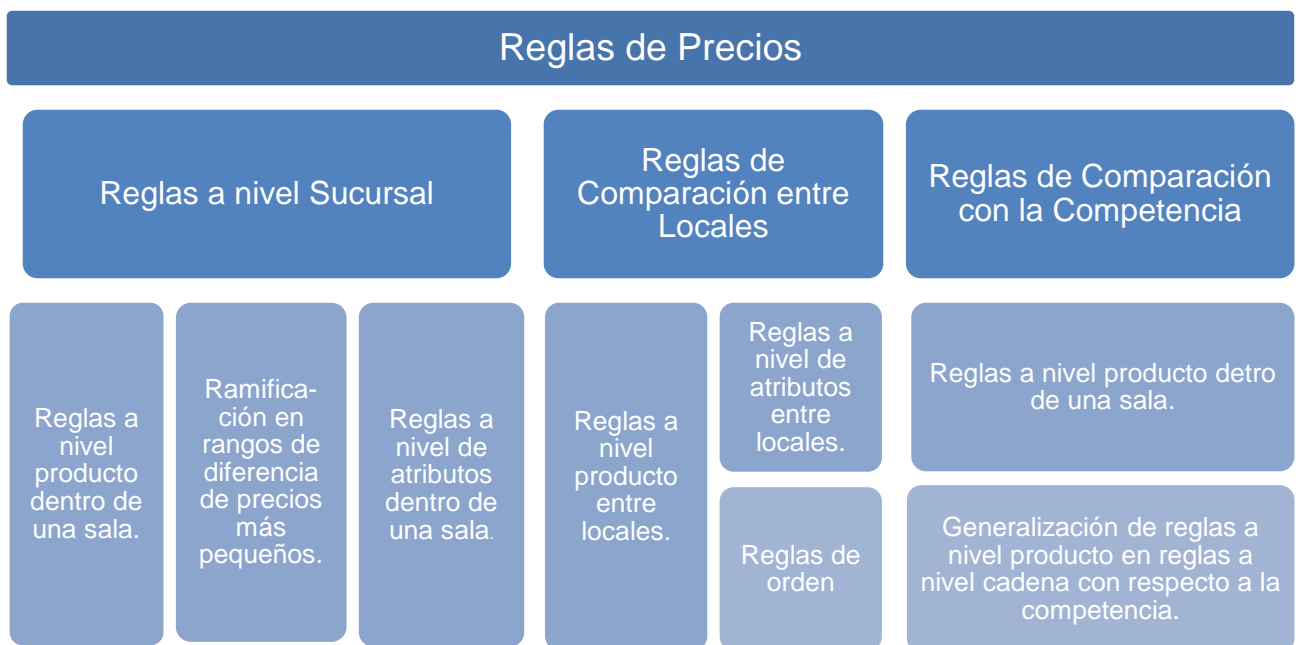
De esta manera se busca evaluar si el modelo de levantamiento de reglas genera efectivamente relaciones “buenas” entre los precios de los productos de la categoría, pudiendo establecer el comportamiento típico en base a datos histórico entre estos mismos.

Posteriormente, para evaluar la consistencia se testearán de los resultados obtenidos utilizando datos transaccionales de la misma categoría pero de diferente periodo, a los cuales se les aplicarán las reglas encontradas y se verificará si las reglas se mantienen en el tiempo. Se busca con esto, determinar si los patrones perduran y si continúan significando una mejora en la contribución.

Para asegurar la validez de las confianzas, éstas se calcularán con respecto al promedio histórico del periodo de calibración, ya que este por ser calculado con más semanas representa un valor más estable y representativo.

A modo de resumen, se muestra en la Figura 3 la estructura lógica de la metodología.

Figura 3: Estructura lógica de la metodología



4 DESARROLLO METODOLÓGICO Y PRESENTACIÓN DE RESULTADOS

4.1 CATEGORÍA Y ATRIBUTOS ESCOGIDOS

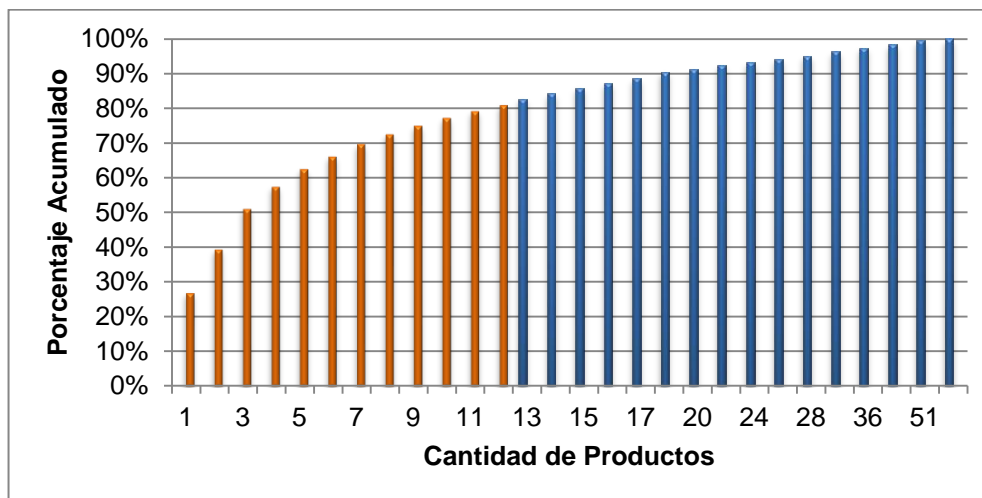
La categoría y subcategoría escogidas para analizar fue lácteos y leches líquidas, debido a que corresponde a una categoría con rol de destino y por ende tiene altos niveles de venta y alta presencia en boletas, permitiendo tener datos suficientes para el estudio.

Además, leches líquidas cumple con otras dos condiciones que son necesarias para facilitar el trabajo:

- Que las ventas se concentren en un porcentaje bajo del total de sus productos, es decir, que pocos productos (12 de 65) representan el 80% de las ventas y conforman la base de los volúmenes vendidos cada mes. Con esto es más fácil identificar las razones de cualquier cambio brusco o anomalía.
- Y ser una subcategoría fácilmente atribuible cumpliendo con las características básicas de los atributos y sus niveles, es decir, ser reconocible (fácilmente identificable por el consumidor), objetivo (no debe ser ambiguo en su definición), y aplicables a todos los productos de la misma categoría.

La concentración se evidencia en el Gráfico 1 a continuación, donde se muestra el porcentaje acumulado de las ventas en un año en función de la cantidad de productos que componen la venta. Como se puede apreciar, solo 12 productos de la categoría representan el 80% de las ventas de esta

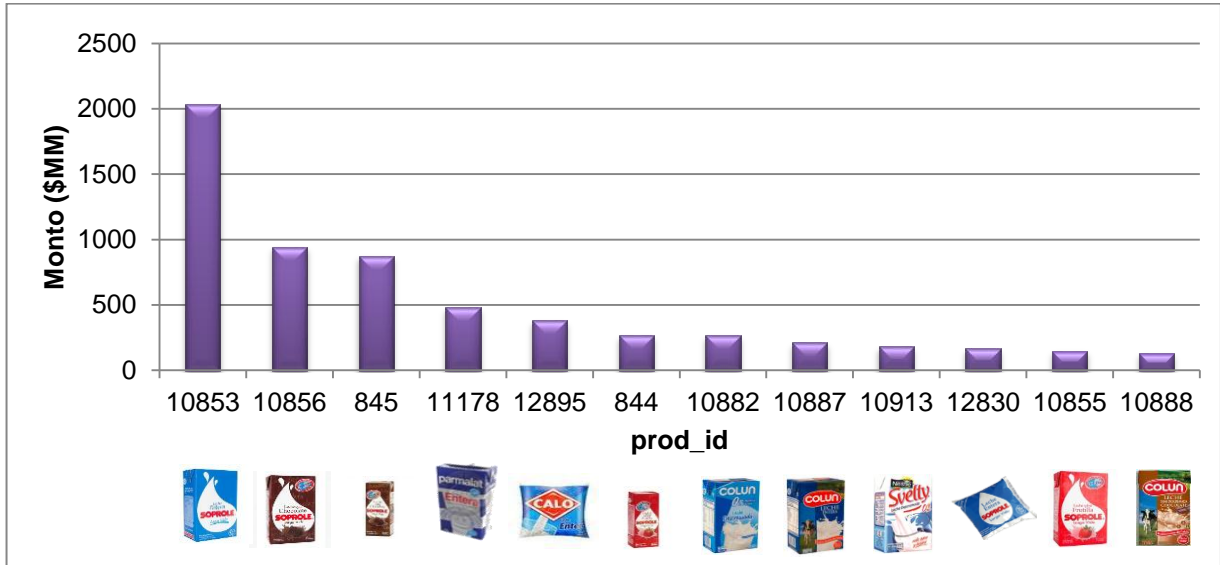
Gráfico 1: Porcentaje Acumulado de la Venta y la cantidad de productos que la componen



Fuente: Elaboración propia

Los niveles de venta en \$MM de los 12 productos más vendidos se muestran a continuación en el Gráfico 2, indicados con figuras de cada producto.

Gráfico 2: Ventas productos más vendidos

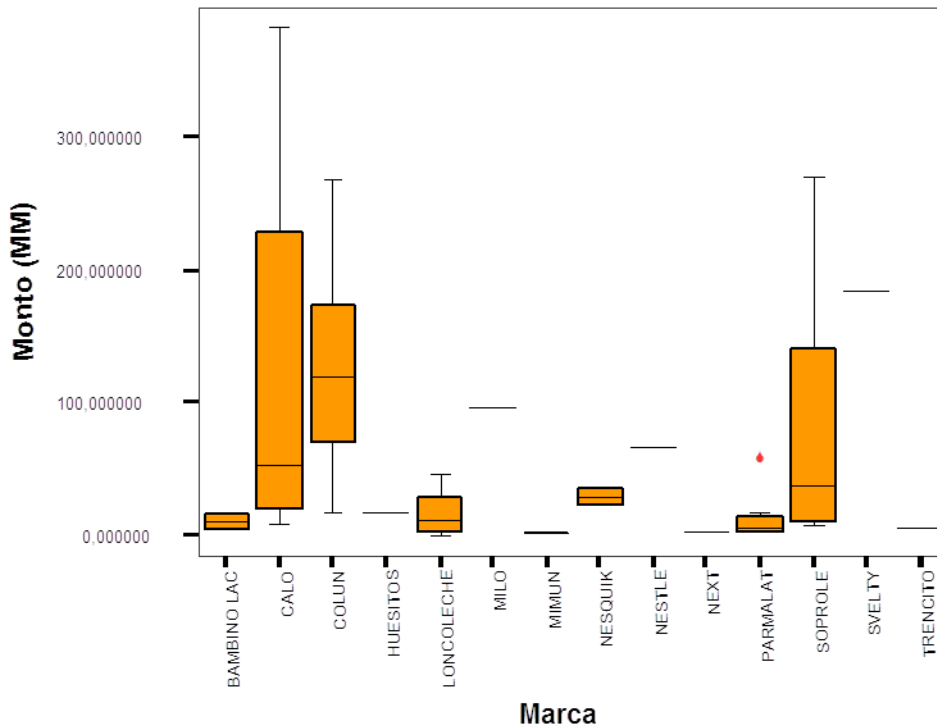


Fuente: Elaboración propia

A continuación se muestran los atributos escogidos para la caracterización de los productos y las figuras que exponen gráficos de distribuciones de venta por atributo se muestran los niveles preferidos de cada uno, medidos durante el periodo de Julio 2010 a Junio 2011a nivel cadena.

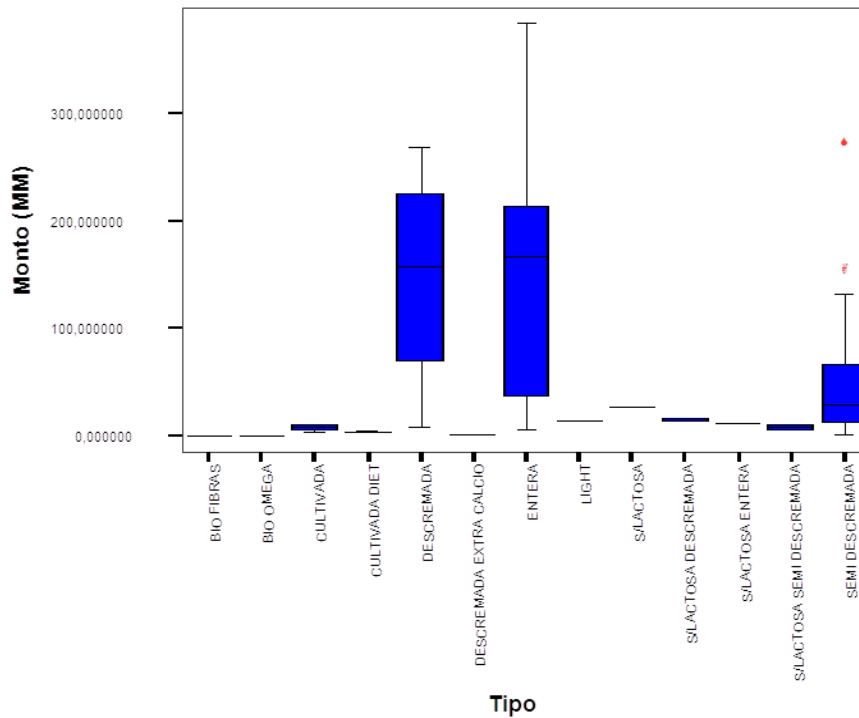
- Marca, con 14 niveles busca especificar más allá del proveedor, el perfil o target distinto que se tenga.

Figura 4: Distribución de Venta por Marca



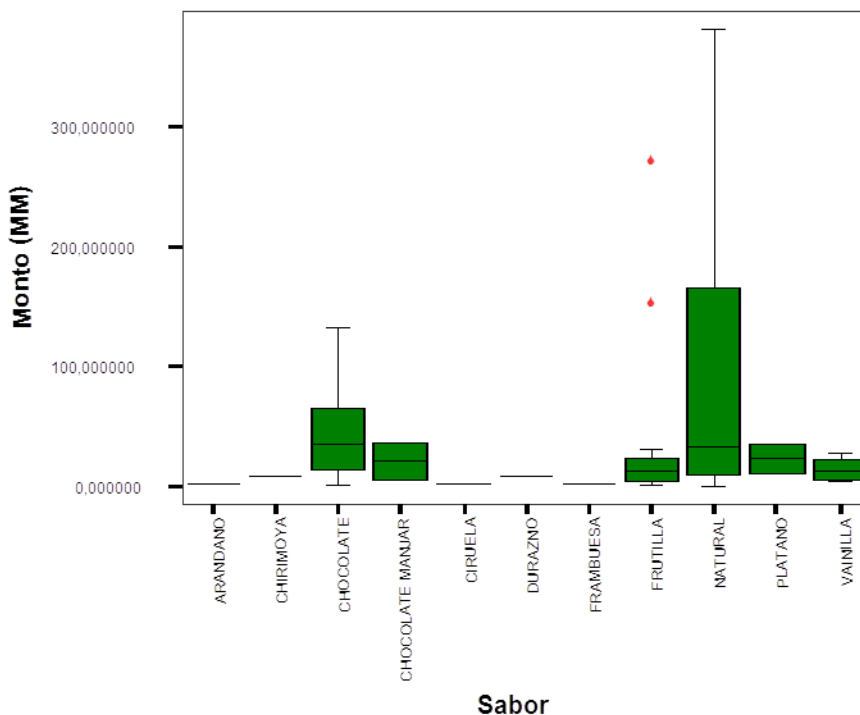
- Tipo, con 10 niveles busca identificar preferencias por productos de mayor especificidad.

Figura 5: Distribución de Venta por Tipo



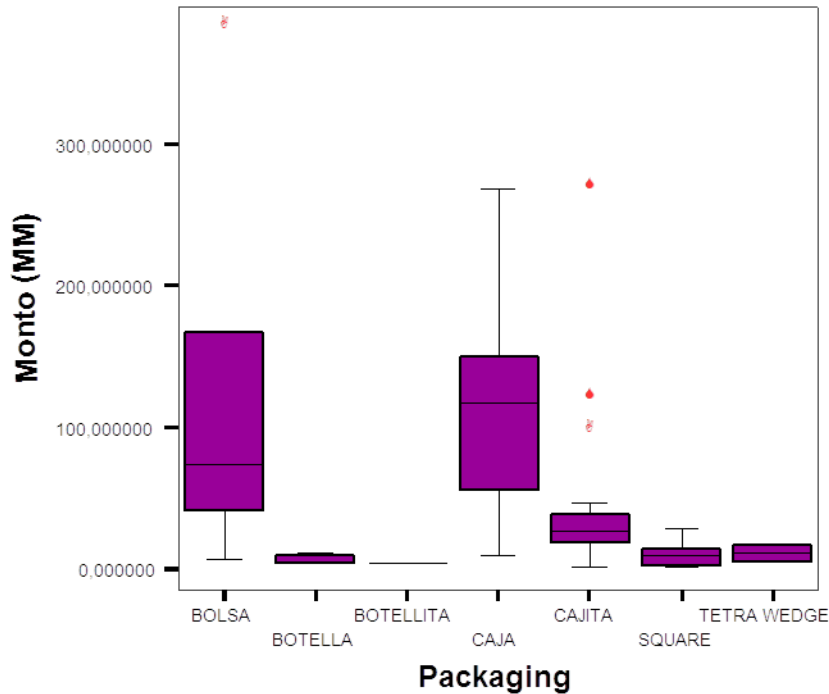
- Sabor, con 11 niveles diferencia los valores que se dan a las leches naturales y a las saborizadas.

Figura 6: Distribución de Venta por Sabor



- Packaging, con 7 niveles explicita la diferencia entre productos individuales y familiares.

Figura 7: Distribución de Venta por Packaging



4.2 DATOS UTILIZADOS Y ANÁLISIS PRELIMINAR

Los datos utilizados son los datos transaccionales obtenidos de la base de datos de una cadena de supermercado mayorista con 41 sucursales. De éstas, fueron consideradas sólo 29 que corresponden a salas de tipo “no institucionales”, es decir, que venden a público general y no en grandes volúmenes como lo hacen las de esta definición.

El periodo en estudio para el levantamiento de reglas corresponde a un año considerado desde el mes de Julio de 2010 a Junio de 2011, evitando así el primer semestre de 2010 que puede haberse afectado por el terremoto ocurrido en Chile, el día 27 de Febrero de 2010. Los días en los que se realizaron transacciones suman 359, con los cuales conforman 54 semanas.

El periodo considerado para el testeo de los datos se extiende desde Julio de 2011 a Septiembre de 2011, que corresponde al tercer trimestre de dicho año. En total se utilizan para esta etapa, 13 semanas de transacciones.

Los productos a considerar en el estudio fueron seleccionados en base a dos criterios:

1. Los productos debían ser transados al menos 100 días
2. Los productos debían haber sido transados en el último trimestre de estudio (Abril-Junio 2011)

En consideración a los criterios de selección recién mencionados, el número de productos totales a analizar resultó ser de 65 (Ver ANEXO A: Maestro de Productos Categoría Leches Liquidadas), de los cuales por limpieza de outliers (eliminación de semanas de ventas) quedan 59.

En cuanto a los atributos, como se menciona en capítulos anteriores, se evaluarán la marca, el tipo de leche, el sabor de la leche, y el *packaging* del producto.

Con el fin de tener una visión previa de los valores por ml de cada uno de los atributos se realiza un estudio de los precio por ml promedio de cada nivel de atributo, ponderado por las unidades vendidas de cada producto que posee cada nivel de atributo.

Previamente, los precios promedio de cada producto fueron calculados ponderando los precios moda semanales por las unidades vendidas en esa semana de manera de tener un valor representativo y comparable del producto.

El análisis por atributo considera su precio por ml promedio ponderado por unidades vendidas, el porcentaje de las ventas totales del atributo que representa cada nivel, y el porcentaje de productos de los 59 en estudio que representa.

Las tablas que siguen presentan los resultados de manera gráfica y ordenada de forma descendiente en su precio por ml promedio ponderado.

Tabla 3: Precio por ml promedio ponderado atributo Marca

Marca	Precio por ml prom pond	% de la Venta TOTAL	% de los productos TOTALES N=59
NEXT	1,1443	0,04%	1,69%
TRENCITO	1,0345	0,08%	1,69%
MILO	0,9595	1,30%	1,69%
HUESITOS	0,9121	0,24%	1,69%
NESQUIK	0,9051	0,79%	3,39%
MIMUN	0,8425	0,04%	1,69%
BAMBINO LAC	0,7468	0,28%	3,39%
LONCOLECHE	0,7054	4,12%	20,34%
CALO	0,5491	6,67%	6,78%
NESTLE	0,5164	0,88%	1,69%
SOPROLE	0,5140	63,27%	23,73%
COLUN	0,4783	11,86%	11,86%
SVELTY	0,4699	2,46%	1,69%
PARMALAT	0,4227	7,97%	18,64%

Fuente: Elaboración propia

Como se observa, las marcas de mayor valor son las especialistas como Next una marca liviana enfocada en la mujer con un estilo de vida light, le siguen todas las de leches de sabores para niños, y finalmente se encuentran las marcas con leches naturales, de las que se destaca Loncoleche por su diferenciación en *packaging*.

Tabla 4: Precio por ml promedio ponderado atributo Tipo

Tipo	Precio por ml prom pond	% de la Venta TOTAL	% de los productos TOTALES N=59
CULTIVADA	0,7678	0,42%	6,78%
S/LACTOSA SEMI DESCREMADA	0,6563	0,23%	3,39%
S/LACTOSA DESCREMADA	0,6544	0,41%	3,39%
S/LACTOSA ENTERA	0,6425	0,16%	1,69%
S/LACTOSA	0,6418	0,37%	1,69%
SEMI DESCREMADA	0,6192	45,97%	54,24%
CULTIVADA DIET	0,5961	0,24%	8,47%
LIGHT	0,5192	0,19%	1,69%
ENTERA	0,4550	44,03%	11,86%
DESCREMADA	0,4402	7,94%	6,78%

Fuente: Elaboración propia

En los tipos de leches se sigue el mismo comportamiento y los tipos especiales como la leche cultivada y las leches sin lactosa tienen valores más altos.

Las leches normales, se ordenan siendo la semi descremada la más cara, la entera de precio medio, y la descremada como la más barata. Lo anterior puede responder a las preferencias de los consumidores, específicamente a la demanda de cada tipo.

Tabla 5: Precio por ml promedio ponderado atributo Sabor

Sabor	Precio por ml prom pond	% de la Venta TOTAL	% de los productos TOTALES N=59
CHOCOLATE MANJAR	0,7883	0,57%	3,39%
VAINILLA	0,7600	0,78%	6,78%
DURAZNO	0,7528	0,11%	1,69%
CHIRIMOYA	0,7484	0,12%	1,69%
PLATANO	0,7265	0,64%	3,39%
FRUTILLA	0,6918	7,60%	18,64%
CHOCOLATE	0,6221	34,02%	30,51%
ARANDANO	0,5993	0,04%	1,69%
CIRUELA	0,5985	0,04%	1,69%
FRAMBUESA	0,5963	0,05%	1,69%
NATURAL	0,4546	56,02%	28,81%

Fuente: Elaboración propia

Los sabores se pueden ordenar según aquellos que son más típicos como los de mayor valor, el típico favorito (chocolate) con un valor un poco menor, los extraños (pertenecientes a leches cultivadas), y finalmente como era de esperarse, el de menor valor es el natural.

Tabla 6: Precio por ml promedio ponderado atributo Packaging

Packaging	Precio por ml prom pond	% de la Venta TOTAL	% de los productos TOTALES N=59
BOTELLA 190ml	1,1443	0,04%	1,69%
CAJA 200ml	0,8065	23,15%	33,90%
TETRA WEDGE 190ml	0,7468	0,28%	3,39%
BOTELLA 1L	0,6905	0,62%	13,56%
SQUARE 1L	0,6384	1,47%	11,86%
BOLSA 500ml y 1L	0,5478	8,98%	8,47%
CAJA 1L	0,4569	65,46%	27,12%

Fuente: Elaboración propia

(*)Tetra Wedge, es un envase de cartón Tetra Pack con forma triangular utilizado para envasar alimentos líquidos.

En los envasados, se encuentran los individuales con un valor superior, y a continuación los familiares ordenados por la calidad del envase. La bolsa, como se identifica en la tabla supera a la caja ya que dentro del nivel hay productos de tipo individual y familiar que ensucian el resultado de su posición.

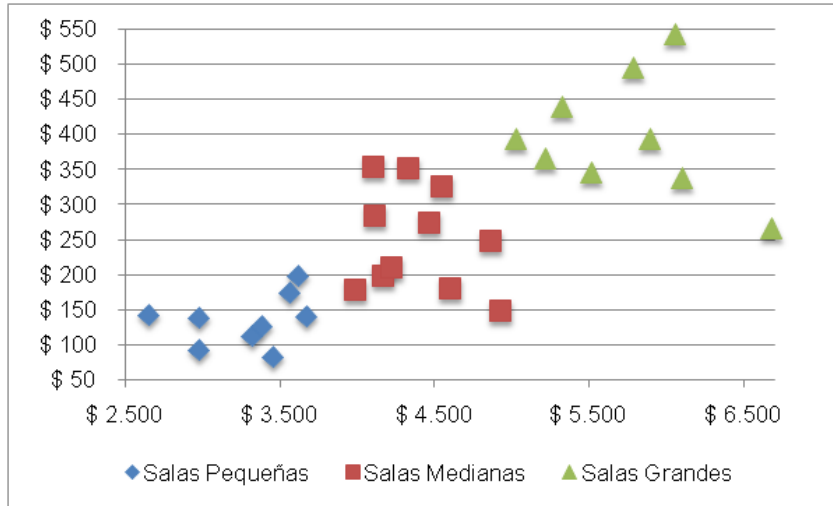
Finalmente, para separar el efecto de los distintos atributos que tiene un producto se realiza una regresión lineal con todos los niveles de todos los atributos, y los resultados respaldan las conclusiones antes presentadas (Resultados de la regresión pueden verse en ANEXO B: Regresión Lineal de Niveles de Atributos.).

Por otro lado, las salas de la cadena con mayores volúmenes de venta en orden decreciente son las sucursales José Joaquín Pérez, Concha y Toro, Vespucio, Recoleta y Maipú. Y las salas que presentan menores volúmenes de ventas son, también en orden decreciente, Puerto Montt Temuco y Talca Mercado. El ranking de las 29 sucursales es calculado en el periodo de estudio, es decir, entre Julio de 2010 y Junio de 2011 y se exhibe en el ANEXO C: Ranking de Sucursales.

En cuanto al ticket promedio de compra de los clientes de cada sucursal se tiene que las salas con mayores índices son Las Condes, Departamental y José Joaquín ordenados en forma decreciente, y las salas con menores índices son Los Ángeles, Temuco y Talca. En el Gráfico 3 se segmentan las salas en base a su ticket promedio y se presentan en las clasificaciones de salas pequeñas, medianas y grandes.

La segmentación se presenta con el objetivo de contextualizar al lector sobre los tamaños de las sucursales y la importancia que cada una tiene en la cadena, de modo de hacer más intuitivos los análisis de los resultados en las reglas a nivel sucursal.

Gráfico 3: Análisis de Sucursales



Fuente: Elaboración propia

Como se observa en el gráfico, los montos de ventas y los ticket promedio de las sucursales están relacionados, y los segmentos resultantes son conformados por 9, 11, y 9 locales respectivamente. El ticket promedio de las salas pequeñas varía entre \$2.651 y \$3.668, los de las salas medianas entre \$3.992 y \$4.929, y las salas grandes entre \$5.028 y \$6.683. En la siguiente Tabla 7 se muestran las sucursales que pertenecen a cada segmento.

Tabla 7: Segmentos de Sucursales

Salas Pequeñas	Salas Medianas	Salas Grandes
TALCA	LO PRADO	MAIPU
TEMUCO	COQUIMBO	EL BOSQUE
LOS ANGELES	SAN PABLO	VESPUICIO
PUERTO MONTT	PENALOEN	INDEPENDENCIA
LINARES	LA PINTANA	CONCHA Y TORO
TALCA MERCADO	QUILICURA	RECOLETA
CONCEPCION	PAJARITOS	J.J. PEREZ
VALPARAISO	RANCAGUA	DEPARTAMENTAL
CHILLAN	BELLOTO	LAS CONDES
	PUENTE ALTO	
	RENCA	

Fuente: Elaboración propia

Como se puede apreciar en la tabla, las salas grandes se ubican todas en la región Metropolitana a diferencia de las salas pequeñas que corresponden a salas ubicadas en otras regiones.

4.3 PREPROCESAMIENTO DE DATOS

La etapa de pre procesamiento de datos, comienza encontrando las posibles combinaciones de productos que en cada sucursal se venden en el periodo de estudio y en especial, se transan en el último trimestre del periodo de estudio. Las combinaciones fueron extraídas desde los datos transaccionales y corresponden al cruce de los productos seleccionados por local, con las semanas contenidas en el periodo de estudio. De esta forma, la dimensión de la matriz de datos queda:

$$[\text{semanas} \times \text{productos} \times \text{productos} \times \text{sucursal}]$$

Donde la cantidad de semanas es 54, igual para las 29 sucursales, y el número de productos varía según la sucursal, desde una base de 65 productos.

Sobre esta base de productos se obtienen los precios moda semanal de cada uno de ellos, calculados en base a la construcción de dos tablas. La primera, contiene los precios de cada producto en cada semana y la cantidad de veces que este aparece, y la segunda selecciona las combinaciones semanas con productos y elige el precio con máxima frecuencia de ocurrencia de la tabla anterior.

La moda se prefiere sobre el promedio porque es un valor al cual fue transado efectivamente el producto en sala esa semana, y porque no considera precios con descuentos por razones ajenas al estudio.

Es importante mencionar que los precios estudiados fueron calculados en base a los datos transaccionales, debido a que no se pueden obtener los precios reales en sala de cada día en cada sucursal. Así, los precios corresponden a la división del monto total por las unidades vendidas de cada producto en cada boleta.

Luego, se construye una tabla con las combinaciones de semanas con productos, que en adelante serán llamados $prod_i$ y $prod_j$ según su orden de aparición, los precios moda de los productos correspondientes, y los ml que contiene cada producto.

Con lo que además son calculados y agregados en la tabla, dos indicadores comparativos entre ellos:

- $precioml$, que corresponde a la estandarización del precio representativo de cada producto en cada semana, por el contenido neto del mismo. Se busca con esto obtener un valor comparable que sea transversal a los atributos y sus respectivos niveles.

$$precioml_i = \frac{\text{precio}_i}{ml_i}$$

- $delta_{precioml}$, que corresponde al porcentaje de diferencia que tiene el precio i sobre el j, matemáticamente:

$$delta_{precioml} = \frac{(\text{precioml}_i - \text{precioml}_j)}{\text{precioml}_j}$$

Paralelamente al pre procesamiento de los datos transaccionales, los productos de la categoría elegida son atributados con matrices de tipo binarias, donde cada atributo representa una tabla en la que las columnas se encuentran los niveles de dicho atributo y en las filas los distintos productos de la categoría. Con estas matrices se genera el cruce con las combinaciones encontradas.

Finalmente en esta etapa, se agregan a los datos indicadores de distancia entre productos, de modo de identificar cuántos atributos varían entre ellos. Este valor es calculado como la suma de las diferencias absolutas de todos los niveles de cada tabla de atributos; dado que las tablas son binarias (1 si el productos tiene el nivel en el atributo, y 0 si no) este valor sumará dos en cada atributo en el que la tupla de productos posea niveles distintos.

El valor de la distancia total se muestra en la siguiente ecuación:

$$\text{distancia}_{(i,j)} = \sum_{\text{atributos}} \sum_{\text{niveles}} \text{abs}(\text{nivel}_i - \text{nivel}_j)$$

Para terminar la etapa de acercamiento a los datos se debe identificar aquellos datos que correspondan a *outliers*, es decir, datos que sean significativamente diferentes al resto de los datos y que puedan ensuciar los resultados.

El estudio se centrará en las contribuciones semanales por sucursal que tiene cada producto, debido a que la consideración de valores *outliers* podría afectar directamente el cálculo de los promedios históricos y consecuentemente al de las confianzas de las reglas que se generen a partir de estos productos. Los precios por otro lado, como se menciona anteriormente, son representados por las modas de los precios a los que son transados los productos cada semana con lo que se descartan los valores extremos, pues estos no son valores frecuentes o que se repitan.

El criterio típicamente utilizado (Programa de análisis de datos SPSS) para detectar outliers es el de 1,5 desviaciones estándar sobre y bajo la media, sin embargo, para el presente trabajo este criterio podría estar capturando errores en los precios o diferencias entre sucursales que las reglas quieren incorporar en el estudio.

Además de un criterio en base al porcentaje de variación que tiene el punto extremo sobre la media histórica de las contribuciones, se define un segundo criterio que hace referencia a la variación porcentual que tiene el punto extremo con el punto inmediatamente anterior en la serie de datos de cada producto.

Este último criterio, se agrega debido a que existen productos con varios casos de contribuciones negativas en sus datos históricos¹ en los que la contribución promedio resulta baja. Cuando esto ocurre, limpiar un dato puede generar una gran diferencia sin necesariamente ser un valor lejano a la serie (un outlier).

¹Los productos con contribuciones bajas y, o negativas pueden responder a objetivos estratégicos que construyan valor al *shopper* sean distintos al de maximizar utilidad, en especial debido al rol de la categoría escogida en el estudio. Por ejemplo pueden responder a un objetivo de surtido que busque aumentar la profundidad de la categoría, o de tener precios bajos por ser categoría de destino y por ende un factor clave en la elección de la tienda como de preferencia predominante.

Para lo anterior se realizó un estudio de las contribuciones de cada producto en el periodo de estudio y de la variación de la media que se genera en cada producto cuando se quitan los puntos extremos (máximo y mínimo). Los resultados se muestran en el Anexo3 y se reconocen 6 productos con puntos extremos, que fueron posteriormente estudiados en su variación con respecto a sus puntos anteriores.

Con los resultados obtenidos y en base a los objetivos de la limpieza de datos, se define punto extremo aquel que varíe el promedio histórico de la contribución semanal de un producto en más de un 4%, y su variación con el punto anterior sea mayor al 70%.

Clasifican así 3 puntos pertenecientes a 2 productos. Estos fueron removidos de los datos de precios, de manera de desaparecer la semana de dicho producto en la sucursal correspondiente del estudio.

4.4 LEVANTAMIENTO DE REGLAS

Para la presentación de resultados se ordenan las etapas por nivel de agregación de la información que entregan. Las más específicas son las reglas entre productos a nivel sucursal, le siguen, su ramificación en segmentos de diferencia de precios más pequeños y su combinación en reglas por atributos, luego las reglas de comparación entre locales que comparan un mismo productos entre las salas, su respectiva combinación por atributos, y finalmente las reglas en comparación con la competencia.

4.4.1 Reglas a Nivel Sucursal

El levantamiento de reglas a nivel sucursal se realiza tomando todas las combinaciones de productos que cumplan la relación $\text{preciom}_i > \text{preciom}_j$, restringidas a que la distancia total entre los productos comparados sea igual a dos. De esta forma se buscan todos los pares de productos comparables (que tengan solo un atributo distinto) y que cumplan alguna de las dos relaciones una cierta cantidad de veces.

Como resultado de lo anterior, se obtiene un set de reglas en cada sucursal que establecen los productos involucrados, el número de semanas que se cumple la relación, y el número de semanas que ambos productos son transados. Con el cociente de estos últimos, es calculado el soporte de cada regla, y con el número de semanas en que ambos productos son transados sobre 54 (total de semanas estudiadas) es calculado el peso.

Para el cálculo de las confianzas, es necesario agregar a la tabla de los precios semanales las contribuciones de cada producto en dicha semana y la suma del par, de manera de calcular y comparar con el promedio histórico.

Una vez agregadas las contribuciones, se obtiene la cantidad de semanas que cumplen la regla y que la contribución del par de productos es mayor o igual al promedio histórico de ésta. A éstas se les llama en adelante semanas favorables.

La información de las reglas a nivel de sucursal es calculada como se muestra en la tabla a continuación.

Tabla 8: Formato regla a nivel sucursal

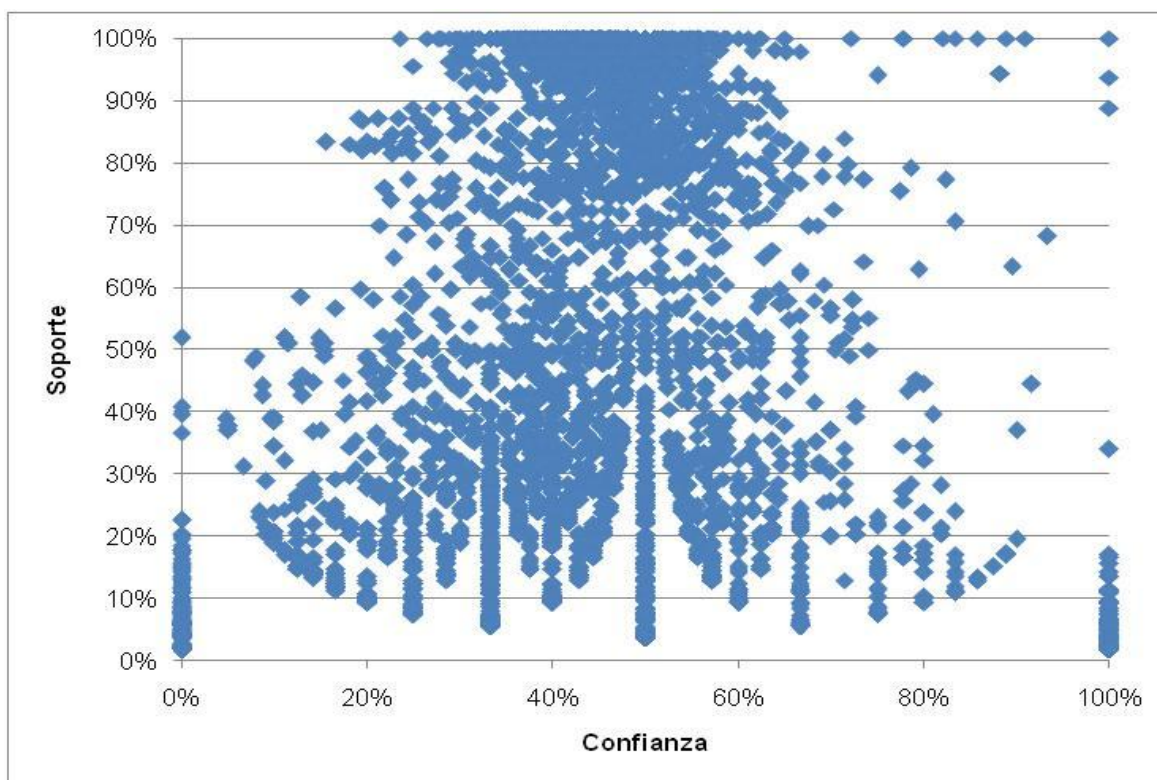
Prod i	Prod j	Frec. ocurrencia	Total tran	Soporte	Peso	Prom hist	Sem fav	Confianza
LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE SOPROLE FRUTILLA CAJA 200ML	11	54	20,37%	100%	131.692,4	9	81,82%

Fuente: Elaboración propia

Donde soporte es la 'Frec. ocurrencia' sobre 'Total tran', el peso es 'Total tran' sobre '54', y finalmente la confianza corresponde a las 'Sem fav.' sobre la 'Frec. ocurrencia'.

En el siguiente gráfico se puede ver la dispersión de los soportes versus las confianzas de las reglas obtenidas en todas las sucursales. Además de observar la constitución de distintos tipos de resultados:

Gráfico 4: Dispersión de pares (confianza, soporte) de Reglas a nivel Sucursal



Fuente: Elaboración propia

Las reglas que se observan en el Gráfico 4, resultan de comparar los precios de todos los productos en cada sucursal. Las comparaciones se conforman en cada par de productos de cada sucursal, en sus dos sentidos: $precio_{l_i} > precio_{l_j}$ y $precio_{l_j} > precio_{l_i}$, no resultando necesariamente complementarios ya que no se están considerando los casos que puedan existir en que los precios sean iguales.

De los resultados obtenidos se pueden observar tres segmentos o clases de reglas con distintos objetivos y funcionalidades:

- Reglas Evaluables, son aquellas reglas que al menos ocurren la mitad de las veces (soporte $\geq 50\%$) y se les puede comparar el desempeño con y sin regla ya que existen semanas en las que a regla no se cumple. Son las ubicadas en la mitad superior sin considerar el límite superior.
- Reglas No Evaluables, son aquellas con un soporte de 100% en las que no hay casos “sin regla” con qué contrastar su desempeño.
- Reglas Potenciales, son aquellas con un soporte insuficiente y confianza alta. Son las reglas ubicadas en el cuadrante inferior derecho del gráfico.

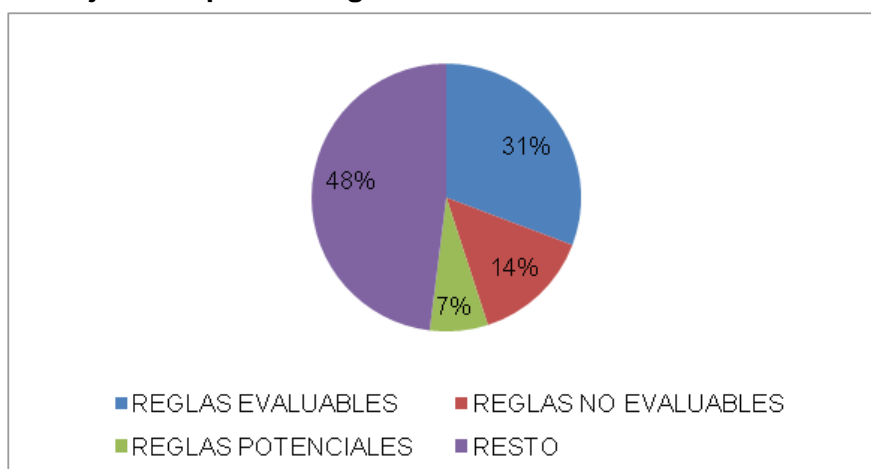
Dentro de la clasificación anterior, el foco del estudio son las Reglas Evaluables, ya que corresponden a reglas cuantificables en su aporte o beneficio. Sin embargo, las reglas No Evaluables tienen como objeto concientizar al tomador de decisión de dicho comportamiento institucionalizado, de modo de ser éste, quien resuelva sobre su debida aplicación y su potencial en cuanto al beneficio que tiene su explotación.

Las Reglas Potenciales por otro lado, dan el espacio al reconocimiento y estudio de patrones de comportamiento no habituales en la cadena, que puedan tener buen desempeño en cuanto a la contribución.

El total de reglas obtenidas a nivel sucursal es de 3.908, de las cuales 1.202 resultan reglas Evaluables, 557 corresponden a reglas No Evaluables, y 273 a reglas Potenciales por estudiar (Véase detalle por sucursal en ANEXO E: Tabla Resumen Reglas a nivel Sucursal).

En el gráfico a continuación se muestran los porcentajes que representan cada tipo de regla en los resultados obtenidos, y se identifica que existe un resto de reglas que no pertenecen a los segmentos definidos que serán descartables para el estudio. Para éstas, se recomienda re calcularlas con más data transaccional, o una vez implementadas las reglas que resulten seleccionables del conjunto de evaluables.

Gráfico 5: Porcentajes de Tipos de Reglas a Nivel Sucursal



Fuente: Elaboración propia

Dada la distancia entre los productos, que obliga a que las tuplas de productos difieran sólo en un atributo, las reglas se pueden agrupar según el atributo en cuestión. El total de reglas encontradas a nivel sucursal por atributo se muestra en la tabla a continuación:

ATRIBUTO	TOTAL REGLAS	% TOTAL REGLAS
Marca	1.902	49%
Sabor	1.330	34%
Tipo	503	13%
Packaging	173	4%
TOTAL	3.908	100%

Fuente: Elaboración propia

En el subcapítulo de Reglas a nivel Sucursal Atributadas (4.4.1.2) se ahonda en esta conclusión y se presenta el resultado obtenido a partir de la data en estudio.

4.4.1.1 Ramificación Reglas a nivel Sucursal

La Ramificación de las reglas, tiene por objeto rankear las variaciones porcentuales de los precios semanales de las reglas encontradas en cada sucursal. Por ejemplo, si una regla establece que el producto *i* es más caro que el producto *j* un 90% de las veces con una confianza del 60%, el siguiente análisis determinará cuánto más caro es, calculando indicadores respectivo para las distintas diferencias de precios que conforman las sentencias iniciales.

La discusión que se busca abordar en esta etapa se enfoca en el fraccionamiento en intervalos de atributos cuantitativos como es el precio/ml, para abordar el problema de pérdida de información que se produce debido a la disminución del valor de la confianza cuando se consideran intervalos grandes como en este caso (mayor que, menor que).

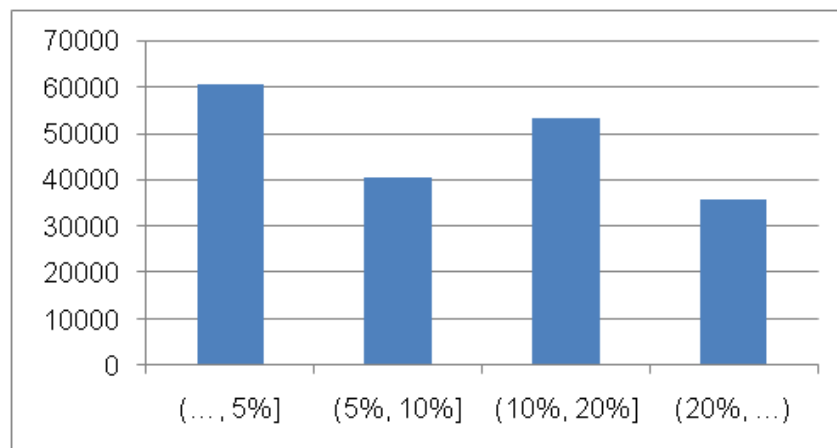
Como se explica en [19] para el problema de las canastas de productos, la confianza medida en intervalos grandes disminuye con respecto a la de intervalos pequeños, debido a que la probabilidad condicional que representa, corresponderá a un porcentaje menor que en el caso en que se consideren intervalos pequeños como antecedentes. Aplicado al proyecto, si se tienen intervalos más pequeños debido a sentencias más restrictivas (serán menos las semanas que cumplan), como por ejemplo “entre 5% y 10% mayor”, las semanas favorables que se tengan en cada caso representarán un porcentaje mayor del que pueda representar en intervalos grandes de sentencias tales como “mayor que”.

Como es de esperar, la mejora en los resultados de las confianzas en intervalos pequeños, viene acompañada de soportes más bajos, ya que al imponer reglas más restrictivas se tendrán menos casos que cumplan. En esta sección, se busca identificar si los resultados obtenidos con esta variación son mejores en sus indicadores.

A los rangos de variación porcentual de precios por ml, se les llamará Ramas. Las ramas escogidas para evaluar las reglas a nivel sucursal se eligieron en base al estudio de las frecuencias de ocurrencia de éstas, buscando aquellos intervalos que tuviesen barras más equilibradas en sus histogramas. Los rangos encontrados en la prueba de distintos conjuntos de

intervalos, con frecuencias más balanceadas, son los que se muestran en el histograma a continuación:

Gráfico 6: Histograma de variaciones de precios



Fuente: Elaboración propia

Dado lo anterior, la ramificación de las reglas significa levantar reglas sobre las ya encontradas e identificar la frecuencia con que éstas ocurren con cada una de las variaciones de precios que las ramas representan.

La lógica de la constitución de las reglas ramificadas se muestra a continuación:

Tabla 9: Formato ramificación reglas a nivel sucursal

Sucursal	Rama1	Rama2	Rama3	Rama4	Total Ramas	Frec. Ocurrencia regla original
SAN PABLO	31	21	0	1	53	53
BELLOTO	34	19	0	1	54	54

Fuente: Elaboración propia

Donde, la suma de las ocurrencias de las ramas debe ser igual a la frecuencia de ocurrencia de la regla original.

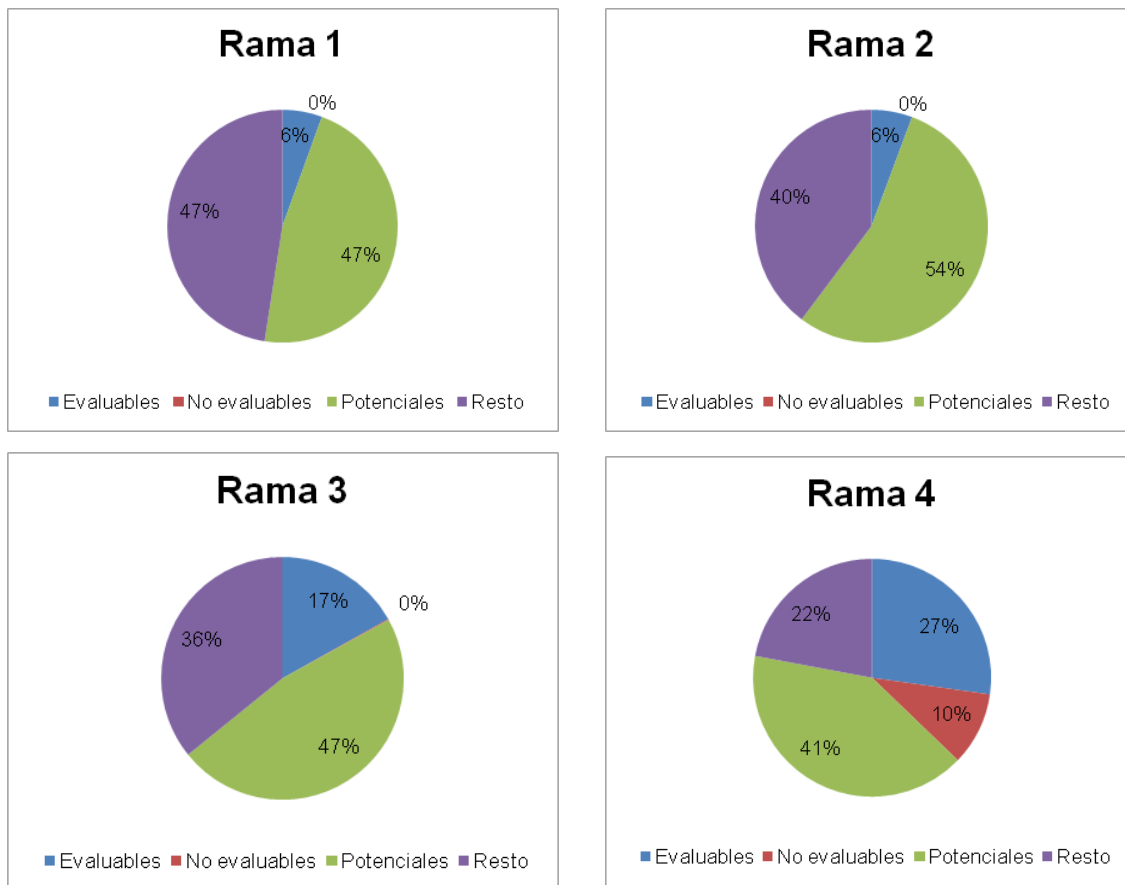
El total de reglas obtenidas en la ramificación fue de 7.369 reglas, distribuidas en los rangos como se muestra a continuación:

Rama	Rango	Total Reglas	% Total Reglas
1	(... , 5%]	2.804	38,05%
2	(5% , 10%]	1.994	27,06%
3	(10% , 20%]	1.587	21,54%
4	(20% , ...)	984	13,35%

Fuente: Elaboración propia

La distribución de los tipos de reglas según Ramificación se muestra en los gráficos a continuación, donde se indica el porcentaje de reglas evaluables, no evaluables, potenciales y el resto de las reglas resultantes.

Gráfico 7: Porcentajes de Tipos de Reglas Ramificación



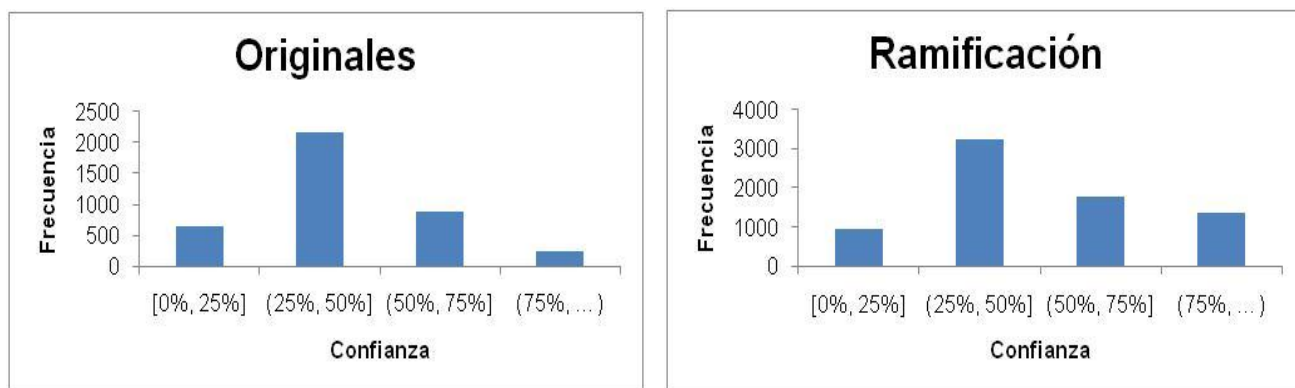
Fuente: Elaboración propia

Como se puede apreciar, el mayor porcentaje lo tienen en todos los casos las reglas potenciales, lo que rectifica lo explicado al inicio de la sección acerca de la disminución en los resultados de los soportes. Le siguen en participación, las reglas que no pertenecen a ningún segmento y por ende no conforman objeto de estudio, con estas reglas se recomienda realizar un estudio posterior con más data, o bien, una vez implementadas las primeras reglas.

Y finalmente en tercer lugar, aparecen las reglas evaluables que son las reglas objetivo para el trabajo, éstas varían entre un 6% y un 27% de los resultados, correspondiendo a un porcentaje bajo que reafirma el hecho de que la fijación de precios, actualmente, no es un proceso ordenado que tenga una estructura lógica entre los productos de una misma categoría.

Por otro lado, las confianzas que se aprecian en los resultados de las reglas ramificadas son efectivamente mejores que los valores obtenidos con las sentencias de “mayor que” y “menor que”. A continuación se muestran los histogramas de las confianzas en las reglas con intervalos grandes (originales) y las con intervalos pequeños (ramificación).

Gráfico 8: Histogramas Confianzas Reglas Originales y Ramificadas



Fuente: Elaboración propia

Tabla 10: Variación Confianzas con Ramificación

Rangos	Frecuencia Originales	%Total Originales	Frecuencia Ramificadas	%Total Ramificadas
[0%, 25%]	649	17%	965	13%
(25%, 50%]	2157	55%	3242	44%
(50%, 75%]	872	22%	1793	24%
(75%, ...)	230	6%	1369	19%
TOTAL	3908	100%	7369	100%

Fuente: Elaboración propia

Como se observa en los histogramas y en la tabla recién presentados, las confianzas de las reglas originales migran a los rangos de mayores valores con la ramificación (aumentando de un 22% a un 24% en el rango de (50%,75%], y de un 6% a un 19% cuando las confianzas son mayores que 75%), entendiéndose esto, como una mejora en los resultados. Sin embargo, es en el capítulo de selección de reglas, donde se identificará cuál de los dos efectos (mejora en confianza o baja en soporte) es más significativo, principalmente en base a la cantidad de reglas que se puedan rescatar, pues si son pocas, significará que la disminución del soporte es más fuerte que la mejora en las confianzas.

4.4.1.2 Reglas a nivel Sucursal Atributadas

Levantar reglas por niveles de atributos, significa levantar reglas sobre las ya encontradas que tengan un soporte mayor al 50%, atributarlas y combinarlas de modo de obtener los indicadores estudiados. La condición sobre el soporte tiene por objeto que no aparezcan las relaciones en ambas direcciones sino sólo la más frecuente de ellas. Es decir, que entre las sentencias 1) producto i es más caro que producto j y 2) producto j es más caro que producto i, sólo se considere aquella que ocurra más veces.

El resultado de las reglas por atributos es de un total de 1.184 reglas en las distintas sucursales, de las cuales 860 son del tipo Evaluable, cantidad que corresponde al 73% de las reglas resultantes. No se encuentran reglas del tipo potencial, ni del tipo No evaluable.

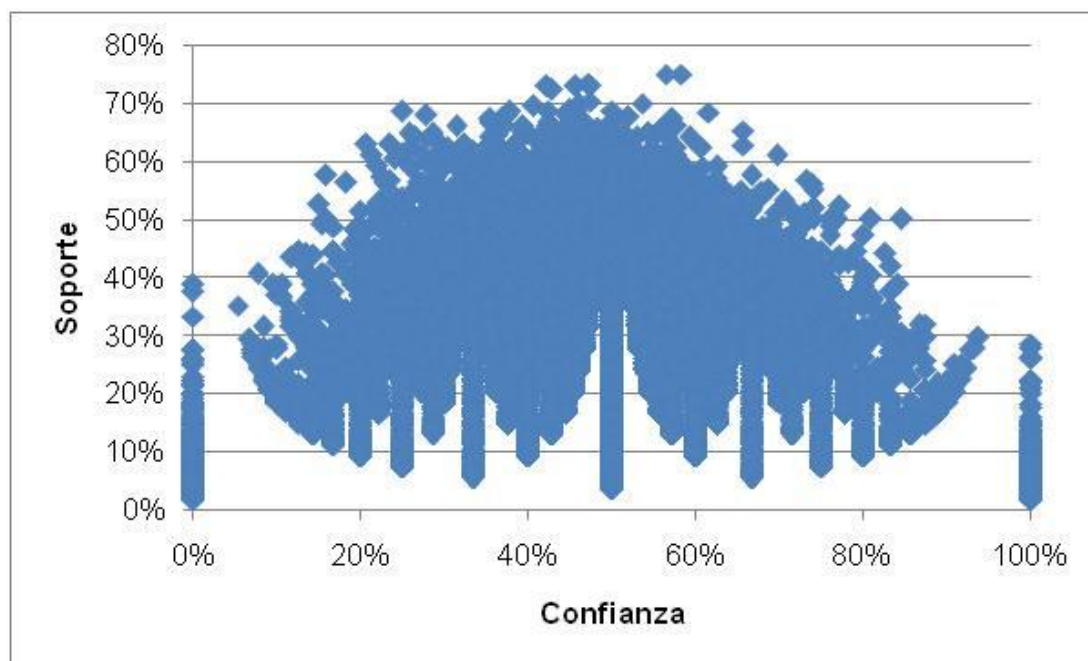
Los resultados que se observan en esta sección, tienen directa relación con los patrones encontrados en las reglas a nivel sucursal, y esto hace pensar que la cantidad de productos que se pueden comparar entre los distintos atributos es baja para una sala. Se espera encontrar mejores resultados al quitar la etiqueta de local en este análisis.

4.4.2 Reglas de Comparación entre Locales

El levantamiento de las reglas a nivel cadena, se realiza tomando todas las combinaciones de sucursales y comparando entre ellas un mismo producto.

El resultado de las reglas a nivel cadena es de un total de 29.986 reglas, de las cuales. Sólo el 4% son reglas evaluables, y el 44% corresponde a regla potenciales. Sus dispersiones de los pares (confianza, soporte) se pueden apreciar en el siguiente gráfico:

Gráfico 9: Dispersión de pares (confianza, soporte) de Reglas de Comparación entre Locales - Productos



Fuente: Elaboración propia

Los valores bajos del indicador soporte se pueden explicar por la manera independiente de funcionar que las sucursales de la cadena tienen. No existe coordinación de los precios a nivel cadena, cada sala pone precios de manera independiente y siguiendo la misma metodología. Esto se manifiesta en los resultados, ya que las reglas encontradas entre sucursales no tienen frecuencias de ocurrencia alta. Es decir, que un producto no está sistemáticamente más caro o más barato en una sucursal que en otra.

Bajo el cuestionamiento de la posibilidad de establecer patrones que relacionen los precios de pares de sucursales y así generar resultados más fáciles de gestionar, se siguió la metodología utilizada para las reglas entre atributos, es decir, se obtuvo indicadores para reglas entre sucursales combinando los indicadores de las reglas obtenidas entre productos de distintas sucursales. Sin embargo, los resultados obtenidos para las reglas son deficientes en sus soportes y confianzas (menores al 40%) y no permiten concluir sobre patrones algunos entre los precios de las sucursales.

4.4.2.1 Reglas de Comparación entre Locales Atributadas

La combinación de las reglas a nivel cadena busca sacar la etiqueta “local” para las reglas por atributos, de modo de tener más casos y poder construir reglas mejor consolidadas.

Los resultados obtenidos son 63 reglas (Ver detalle en ANEXO F: Reglas de comparación entre locales Atributadas), de las cuales el 90% son reglas evaluables. En este caso, no se encuentran reglas potenciales (reglas con soporte bajo 50% y confianza alta), con lo que se puede comprobar que las reglas por atributos, sin la etiqueta local, tienen mejores resultados en su soporte y por ende son más determinantes que las reglas por atributos entre productos. Esto último, se debe a la mayor cantidad de datos que se tiene, generando mayor información en cuanto a la frecuencia de ocurrencia de las reglas (soporte).

En los resultados de esta sección, se encontró patrones contradictorios entre sí con indicadores similares tanto en soporte, como en confianza. Las reglas entre los sabores son las que tienen un mayor porcentaje de contradicciones con respecto al total de reglas obtenidas (en la Tabla 11 se muestran las estadísticas de los resultados).

Tabla 11: Estadísticas de resultados obtenidos en Reglas de Comparación entre Locales Atributadas

	TOTAL REGLAS	%TOTAL REGLAS	TOTAL CONTRADICCIONES	%TOTAL CONTRADICCIONES
Marca	38	60%	5	13%
Tipo	12	19%	3	25%
Sabor	10	16%	3	30%
Packaging	3	5%	0	0%
TOTAL	63	100%	11	17%

Fuente: Elaboración propia

De las contradicciones encontradas, las confianzas son bastantes bajas en la mayoría de los casos, es decir no conforman reglas seleccionables en el trabajo y por lo tanto no representan un conflicto de indicadores (Ver detalle en ANEXO G: Reglas contradictorias encontradas en reglas de comparación entre locales atributadas). Sólo en un caso de sabor, donde se relacionan el sabor frutilla con el sabor chocolate manjar se tienen valores más altos en los indicadores que pudiesen resultar confusos, sin embargo, la confianza de uno es doce puntos

porcentuales mayor que el de su contradicción, no representando tampoco, un problema para el capítulo de selección.

Además se observa en esta sección, que los soportes de las reglas evaluadas no tienen soportes que sumen 100% en sus dos sentencias (“mayor que” y “menor que”), esto se debe a que son reglas creadas en función de reglas a nivel producto, y sus indicadores son combinaciones de los indicadores de las reglas que los conforman. Si bien esto resulta confuso en su definición, representa una forma de agrupar los atributos como tales, bajo la restricción de sólo comparar productos que tengan un atributo distinto. En esta sección se está agregando una nueva dimensión en la comparación que se hace para generar las relaciones entre los precios (los atributos), razón por la cual se presentan estas nuevas problemáticas.

4.4.3 Reglas en Comparación con la Competencia

Los datos utilizados para levantar reglas en comparación con la Competencia, corresponden al segundo trimestre del año 2011, conforman 13 semanas de la 13 a la 26, de las cuales se cuenta sólo con 11. Se comparan en total 9 productos, los cuales se presentan en la tabla a continuación con sus atributos y el total de semanas en los que se tiene registro de cada uno.

Tabla 12: Descripción productos estudiados de la competencia

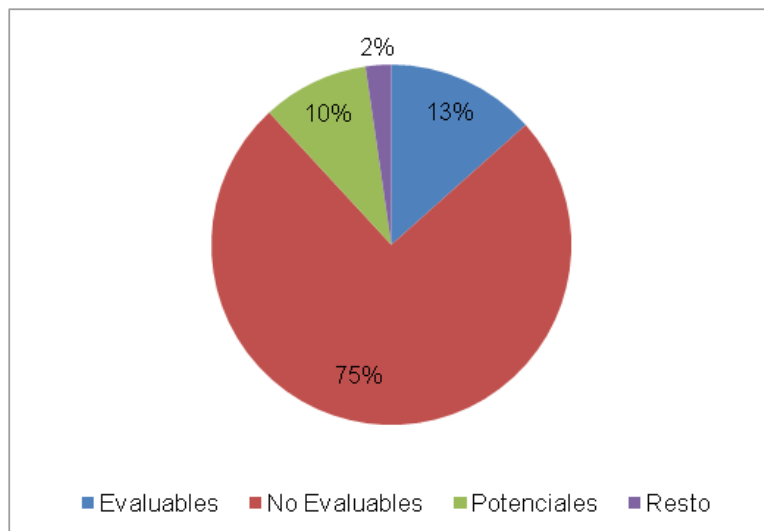
Prod	Marca	Tipo	Sabor	Packaging	umb	Total semanas
12895	CALO	ENTERA	NATURAL	BOLSA	UN	5
10882	COLUN	DESCREMADA	NATURAL	CAJA 1L	UN	6
10883	COLUN	SEMI DESCREMADA	NATURAL	CAJA 1L	UN	6
10887	COLUN	ENTERA	NATURAL	CAJA 1L	UN	6
11178	PARMALAT	ENTERA	NATURAL	CAJA 1L	UN	6
10853	SOPROLE	ENTERA	NATURAL	CAJA 1L	UN	11
10856	SOPROLE	SEMI DESCREMADA	CHOCOLATE	CAJA 1L	UN	9
10888	COLUN	SEMI DESCREMADA	CHOCOLATE	CAJA 1L	UN	6
845	SOPROLE	SEMI DESCREMADA	CHOCOLATE	CAJA 200	DIS	7

Fuente: Elaboración propia

Debido a los datos faltantes, se decide considerar la moda nacional de los datos levantados como un único precio de la competencia y comparar estos con los precios de las distintas sucursales de la cadena.

Como resultado de reglas en comparación con la competencia se obtuvo 261 reglas, de las cuales se pueden evaluar sólo un 13%, encontrándose un alto porcentaje de reglas no evaluables (Ver Gráfico 10).

Gráfico 10: Porcentajes de Tipos de Reglas en Comparación con la Competencia



Fuente: Elaboración propia

Debido a la poca disponibilidad de datos, este resultado parece lógico, debido a que en 11 semanas de datos se pueden encontrar varios eventos casuales que tengan altos indicadores.

4.5 SELECCIÓN DE REGLAS

El criterio para seleccionar una regla será tener indicadores de soporte y confianza mayores a 65%. –es decir, que ocurren al menos un 65% de las veces en que ambos productos son transados, y que al menos el 65% de las veces en que ocurre la contribución es mayor o igual al promedio histórico.

Para la elección del criterio de selección fueron probados 5 cortes partiendo del límite inferior del segmento de reglas evaluables (50%), y aumentando en 5%. Es decir, se probaron los cortes 50%, 55%, 60%, 65% y 70%, ya que con 75% el porcentaje de selección era mínimo casi nulo.

Dichos cortes fueron expuestos a ojo de expertos mostrando las reglas que iban siendo seleccionadas con cada nuevo corte. Lo anterior tenía por objetivo identificar la credibilidad de las reglas que se iban agregando a medida que se disminuía el criterio, y esto es, se relajaba la exigencia en los indicadores.

Para evaluar hasta qué punto las reglas que se seleccionaban hacían sentido fue construido un cuestionario, este fue consultado a expertos en la materia, y los resultados constituyeron el método de definición del criterio de corte, el cual fue de un 65% como criterio único para el soporte y la confianza. Además, para que las reglas seleccionadas sean del tipo Evaluables no pueden tener soporte 100%.

Finalmente, se estableció como condición higiénica que los pesos de las reglas deben ser mayores o iguales al 50%, es decir, que al menos el estudio de cada regla debe considerar 27 semanas de las 54 totales. Lo anterior, debido a que se busca asegurar una mínima significancia de los indicadores con respecto al periodo de estudio.

4.5.1 Reglas Seleccionadas a Nivel Sucursal

Las reglas a nivel sucursal seleccionadas según los criterios explicitados son 20 y corresponden a 11 distintas sucursales. A continuación se presentan las reglas en la dirección el **Producto A** es más caro que el **Producto B**, y el tipo de sucursal será indicado con las letras: (S.P) si corresponde a una sala pequeña, (S.M) si corresponden a una sala mediana, y (S.G) si corresponde tener una sala grande.

Tabla 13: Resultados a Nivel Sucursal

N°	Sucursal	Producto A	Producto B	Peso	Sop.	Conf.
1	TALCA MERCADO(S. P)	LECHE ENTERA COLUN CAJA 1L	LECHE COLUN SEMI DESCREMADA CAJA 1L	98%	70%	68%
2	TEMUCO (S.P)	LECHE ENTERA COLUN CAJA 1L	LECHE COLUN SEMI DESCREMADA CAJA 1L	94%	80%	66%
3	LA PINTANA (S.M)	LECHE ENTERA COLUN CAJA 1L	LECHE COLUN SEMI DESCREMADA CAJA 1L	94%	71%	83%
4	LA PINTANA (S.M)	LECHE ENTERA COLUN CAJA 1L	LECHE COLUN DESCREMADA CAJA 1L	93%	84%	71%
5	EL BOSQUE (S.G)	LECHE ENTERA COLUN CAJA 1L	LECHE PARMALAT ENTERA CAJA 1LT	91%	80%	72%
6	VESPUCIO (S.G)	LECHE ENTERA COLUN CAJA 1L	LECHE PARMALAT ENTERA CAJA 1LT	89%	81%	69%
7	CONCHA Y TORO (S.G)	LECHE ENTERA COLUN CAJA 1L	LECHE PARMALAT ENTERA CAJA 1LT	80%	77%	67%
8	BELLOTO (S.M)	LECHE COLUN VAINILLA CAJA 200ML	LECHE COLUN FRUT CAJA 200ML	100%	78%	69%
9	LINARES (S.P)	LECHE SOPROLE CHOC CAJA 200ML	LECHE COLUN CHOC CAJA 200ML	98%	75%	65%
10	COQUIMBO (S.M)	LECHE SOPROLE CHOC CAJA 200ML	LECHE LONCOLECHE CHOC CAJA 200ML	98%	77%	66%
11	COQUIMBO (S.M)	LECHE SOPROLE CHOC CAJA 200ML	LECHE COLUN CHOC CAJA 200ML	93%	70%	69%
12	BELLOTO (S.M)	LECHE SOPROLE DESCREMADA CAJA 1L	LECHE SOPROLE ENTERA CAJA 1L	98%	79%	79%
13	PUERTO MONTT (S.P)	LECHE SOPROLE DESCREMADA CAJA 1L	LECHE SOPROLE ENTERA CAJA 1L	94%	82%	67%
14	COQUIMBO (S.M)	LECHE SOPROLE ENTERA CAJA 1L	LECHE PARMALAT ENTERA CAJA 1LT	94%	73%	70%
15	TEMUCO (S.P)	LECHE SOPROLE ENTERA CAJA 1L	LECHE PARMALAT ENTERA CAJA 1LT	89%	83%	65%
16	RANCAGUA (S.M)	LECHE SOPROLE ENTERA CAJA 1L	LECHE PARMALAT ENTERA CAJA 1LT	83%	78%	71%
17	BELLOTO (S.M)	LECHE SOPROLE ENTERA CAJA 1L	LECHE PARMALAT ENTERA CAJA 1LT	81%	77%	74%
18	BELLOTO (S.M)	LECHE SOPROLE SEMI DESCREMADA CAJA 1L	LECHE SOPROLE ENTERA CAJA 1L	98%	75%	78%
19	PAJARITOS (S.M)	NESQUIK FRUT CAJA 200ML	BEBIDA LACTEA SUPER CALO FRUT CAJA 200ML	91%	98%	67%
20	PAJARITOS (S.M)	NESQUIK FRUT CAJA 200ML	LECHE LONCOLECHE FRUT CAJA 200ML	87%	98%	65%

Fuente: Elaboración propia

Los atributos comparados en las relaciones son mayormente Marca, Tipo para las 2 marcas líderes, y Sabor en cajas de 200ml saborizadas para niños. Lo anterior deja en evidencia la importancia del atributo Marca por sobre el Tipo en la decisión de compra de Leches Liquidadas de los clientes, ya que los ordenamientos de tipo dependen y/o están sujetos a la Marca que tengan.

De las relaciones establecidas entre marcas se tienen las leches de caja de 1L, que Colun y Soprole son más caras que Parmalat en leche entera que es el producto básico de la categoría, esto hace sentido, y puede deberse al valor de marca que tienen las empresas mencionadas por sobre Parmalat.

En las cajas de 200ml saborizadas, la marca con mayor precio en chocolate es la Soprole por sobre Colun y Loncoleche, lo que puede atribuirse a la altísima demanda que existe por este producto en comparación con la demanda de las otras marcas (como se indica en el Gráfico 2), lo que le da espacio para extraer un poco más de margen. Por otro lado, en las leches de frutilla la marca más cara es Nesquik por sobre Super Calo y Loncoleche, esto último se atribuye a la calidad de las leches mencionadas.

En cuanto a los tipos se tiene que en Soprole las leches semi descremada y descremada son más caras que la leche entera, lo que puede imputarse al costo de los productos. Sin embargo, en Colun la leche entera es más cara que la leche semi descremada y descremada, lo que si se analiza desde el punto de vista de la competencia, se puede atribuir a que la leche entera Soprole es el producto líder de la categoría, y por ende, la marca se preocupa de mantener sus ventas altas teniendo un precio menor al de sus otros productos, en consecuencia, la mejor forma que tiene Colun, que es la marca que le sigue en ventas, de ser competitivo frente a Soprole es potenciar sus otros tipos de leches (descremada y semi descremada) teniéndolas a un menor precio que su leche entera, de modo de transformarlos en sus “caballitos de carrera”² en la categoría.

Sin embargo, al estudiar las estadísticas de los indicadores de estas reglas en todas las sucursales, y contrastarlos con sus inversas se tiene lo siguiente:

Tabla 14: Estadísticas reglas de Tipo, marca Colun

COLUN	REGLA			INVERSA		
	Soporte	Peso	Confianza	Soporte	Peso	Confianza
LECHE ENTERA COLUN CAJA 1L más cara que LECHE COLUN SEMI DESCREMADA CAJA 1L						
MIN	70%	87%	38%	16%	87%	0%
MAX	80%	100%	83%	29%	100%	33%
PROMEDIO	76%	95%	55%	22%	95%	16%
LECHE ENTERA COLUN CAJA 1L más cara que LECHE COLUN DESCREMADA CAJA 1L						
MIN	82%	91%	36%	2%	91%	0%
MAX	98%	100%	71%	14%	98%	100%
PROMEDIO	89%	95%	54%	8%	95%	19%

Fuente: Elaboración propia

En base a la Tabla 14, se puede decir que la regla es efectivamente mejor que su inversa, tanto en su frecuencia de ocurrencia, como en su aporte a la contribución del par de productos. Aún cuando el soporte del segundo caso no sea lo suficientemente alto como para concluir respecto de la inversa, la regla tiene resultados buenos para la cadena completa.

²Se le llama “caballito de carrera” al producto con mayor venta de una marca, y por el cual la marca es reconocida.

Tabla 15: Estadísticas reglas de Tipo, marca Soprole

SOPROLE	REGLA			INVERSA		
	Soporte	Peso	Confianza	Soporte	Peso	Confianza
LECHE SOPROLE SEMI DESCREMADA CAJA 1L más cara que LECHE SOPROLE ENTERA CAJA 1L						
MIN	75%	91%	18%	2%	91%	0%
MAX	93%	100%	78%	13%	100%	100%
PROMEDIO	85%	99%	36%	6%	99%	76%
LECHE SOPROLE DESCREMADA CAJA 1L más cara que LECHE SOPROLE ENTERA CAJA 1L						
MIN	79%	85%	16%	7%	85%	22%
MAX	91%	100%	79%	18%	100%	100%
PROMEDIO	85%	98%	36%	12%	98%	87%

Fuente: Elaboración propia

En base a la Tabla 15, en cambio, no se puede concluir certeramente acerca del comportamiento de la regla y su inversa en toda la cadena, pues se tienen mejores resultados de confianza en la inversa, pero con frecuencias de ocurrencias muy bajas. Lo que se interpreta como que los casos pueden resultar ser eventos casuales, y no representar una medida significativa. Con esto, se distingue el alcance de las reglas, siendo éstas, aplicables exclusivamente en las condiciones en que aparecen.

A raíz del análisis anterior, se recomienda hacer seguimiento de los resultados que se tengan al aplicar las reglas de Tipo para Soprole, pues aún cuando se pueda dar interpretación al resultado, esto sólo es evaluable al momento de implementar.

Por último, se tiene una regla por sabores de productos Colun, donde el precio de la caja de 200ml de vainilla es más cara que la caja de 200ml de frutilla, lo que se justifica en que las ventas (como se puede ver en el Gráfico 2) indican que la frutilla es preferida por sobre la vainilla y como sus ventas no son muy diferentes, respondiendo a la demanda, el precio de la frutilla deberá ser inferior.

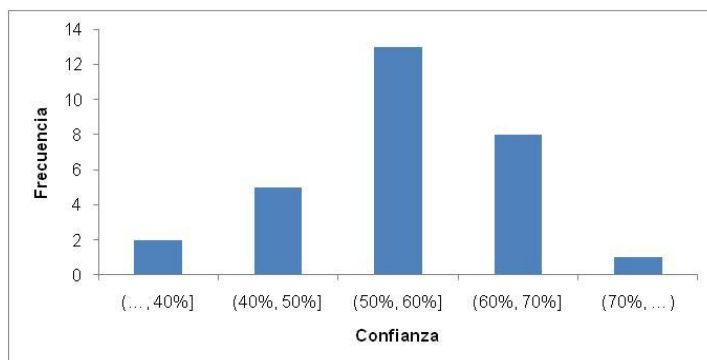
Como se observa en la Tabla 13, hay reglas que relacionan los mismos productos, pero se dan en sucursales diferentes. A estas reglas que tienen repeticiones seleccionadas, se evaluó los indicadores en el resto de las sucursales, con el objeto de identificar cuáles de ellas pueden ser replicables para toda la cadena. En cada caso se concluyó sugiriendo si replicar o no, la regla.

Para estudiar la posibilidad de replicar las reglas repetidas se estudian los mínimos y máximos de cada indicador en todas las sucursales de la cadena, y se considera que el indicador confianza es el importante para el cliente, debido a que abarca la contribución en su cálculo, se construye un histograma con éstas, de modo de poder concluir sobre su comportamiento en el resto de las sucursales, y si es una regla posible de replicar.

De las reglas N°1, N°2, y N°3 se obtienen los siguientes resultados en el análisis de los indicadores de todas las sucursales de la cadena:

Gráfico 11: Estudio para replicar reglas N°1, N°2, y N°3 (LECHE ENTERA COLUN CAJA 1L más cara que LECHE COLUN SEMI DESCREMADA CAJA 1L).

	Soporte	Peso	Confianza
Mínimo	70%	87%	38%
Máximo	80%	100%	83%



Fuente: Elaboración propia

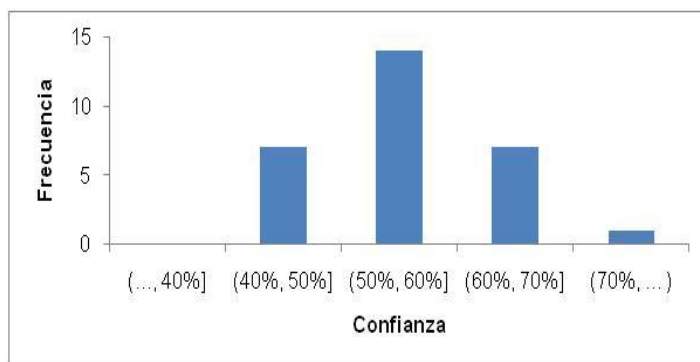
A partir del gráfico anterior se puede identificar que las confianzas de esta regla en el resto de las sucursales de la cadena se concentran en las clases entre 50% y 70%, lo que representa un buen nivel para este indicador.

En base a lo anterior, considerando que el mínimo soporte que presenta esta regla en toda la cadena es de un 70% y dado que los soportes están consolidados con un peso mínimo del 87% es posible afirmar que esta regla es de alta frecuencia de ocurrencia, y que además, más del 50% de los casos en que ocurre esta regla representa una mejora a la contribución del par de productos, es decir, se recomienda replicar.

De las reglas N°5, N°6, y N°7 se obtienen los siguientes resultados en el análisis de los indicadores de todas las sucursales de la cadena:

Gráfico 12: Estudio para replicar reglas N°5, N°6, y N°7 (LECHE ENTERA COLUN CAJA 1L más cara que LECHE PARMALAT ENTERA CAJA 1LT)

	Soporte	Peso	Confianza
Mínimo	77%	74%	44%
Máximo	90%	100%	72%



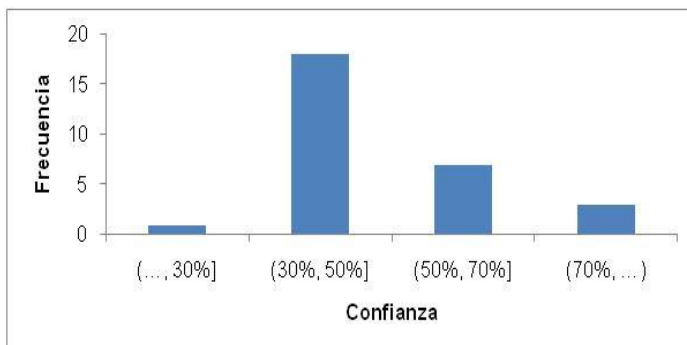
Fuente: Elaboración propia

En este caso se observa que la regla efectivamente es frecuente también en las otras sucursales, y el peso mínimo respalda esto ya que al menos se están evaluando un 74% de las semanas estudiadas. Por otro lado, se tiene que las confianzas de esta relación entre productos se concentran sobre el 60%, lo que permite recomendar replicarla al resto de las sucursales de la cadena.

De las reglas N°14, N°15, N°16, y N°17 se obtienen los siguientes resultados en el análisis de los indicadores de todas las sucursales de la cadena:

Gráfico 13: Estudio para replicar reglas N°14, N°15, N°16, y N°17 (LECHE SOPROLE ENTERA CAJA 1L más cara que LECHE PARMALAT ENTERA CAJA 1LT)

	Soporte	Peso	Confianza
Mínimo	68%	81%	28%
Máximo	89%	100%	74%



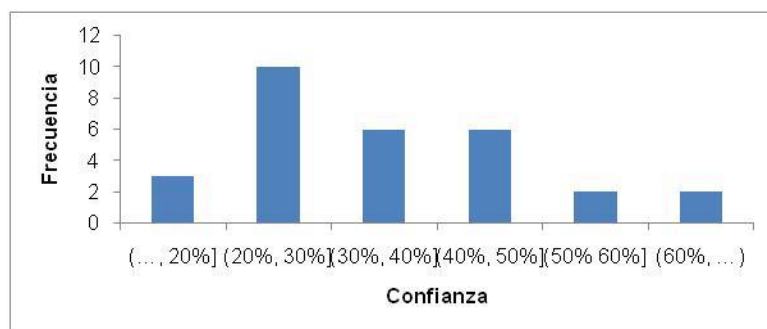
Fuente: Elaboración propia

De la tabla se concluye que la regla ocurre con alta frecuencia, sin embargo, en el gráfico se puede observar que la concentración de confianzas se ubica en la clase entre 30% y 50%, lo que no es suficiente y por lo tanto no se recomienda replicar la regla.

De las reglas N°12 y N°13 se obtienen los siguientes resultados en el análisis de los indicadores de todas las sucursales de la cadena:

Gráfico 14: Estudio para replicar reglas N°12 y N°13 (LECHE SOPROLE DESCREMADA CAJA 1L más cara que LECHE SOPROLE ENTERA CAJA 1L)

	Soporte	Peso	Confianza
Mínimo	79%	85%	16%
Máximo	91%	100%	79%



Fuente: Elaboración propia

Finalmente, las reglas N°12 y N°13 ocurren frecuentemente en el resto de las sucursales como indican los mínimos de soporte y peso, pero sus confianzas tienen apariciones dispersas en las clases de valores, y más bien se concentran bajo el 50%, lo que no hace recomendable que esta regla sea replicada a la cadena completa pues no se puede concluir claramente acerca del efecto que tendría en las contribuciones de los productos de las otras salas.

Además de las reglas que puedan replicarse, se pueden analizar las reglas que podrían generar buenos resultados pero que no son una práctica frecuente, es decir, evaluar las reglas potenciales.

En la evaluación de las reglas potenciales, es decir, las que son del tipo Evaluable pero que tienen soporte bajo al criterio de selección y que se destacan por tener confianzas altas se puede encontrar las siguientes reglas:

N°	Sucursal	Producto A	Producto B	Peso	Sop.	Conf.
1	TALCA (S.P)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE SOPROLE FRUTILLA CAJA 200ML	98%	55%	72%
2	TALCA MERCADO (S.P)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE SOPROLE FRUTILLA CAJA 200ML	100%	54%	72%
3	CHILLAN (S.P)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE SOPROLE FRUTILLA CAJA 200ML	100%	52%	71%
4	MAIPU (S.G)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE LONCO CHOCOLATE CAJA 200ML	100%	63%	79%
5	RENCA (S.M)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE LONCO CHOCOLATE CAJA 200ML	91%	55%	74%
6	QUILICURA (S.M)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE LONCO CHOCOLATE CAJA 200ML	100%	50%	74%
7	DEPARTAMENTAL (S.G)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE LONCO CHOCOLATE CAJA 200ML	87%	51%	71%
8	SAN PABLO (S.M)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE LONCO CHOCOLATE CAJA 200ML	100%	50%	70%
9	INDEPENDENCIA (S.G)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE LONCO CHOCOLATE CAJA 200ML	98%	57%	70%
10	RECOLETA (S.G)	LECHE SOPROLE CHOCOLATE CAJA 200ML	LECHE LONCO CHOCOLATE CAJA 200ML	100%	56%	70%
11	PAJARITOS (S.M)	LECHE NESQUIK CHOCOLATE CAJA 200ML	LECHE MIMUN CHOCOLATE CAJA 200ML	56%	63%	89%
12	EL BOSQUE (S.G)	LECHE COLUN FRUTILLA CAJA 200ML	LECHE SOPROLE FRUTILLA CAJA 200ML	93%	58%	72%
13	TALCA MERCADO (S.P)	LECHE COLUN CHOCOLATE CAJA 1L	LECHE NESTLE CHOCOLATE CAJA 1L	98%	64%	74%

Fuente: Elaboración propia

De las reglas potenciales se destaca la aparición de la regla seleccionada N°10 (ver Tabla 13) en más sucursales de la cadena, la aparición de otra relación de precio entre los sabores de caja 200ml de Soprole (reglas N°1, N°2, y N°3) lo que tiene sentido ya que como se puede ver en el Gráfico 1 este producto corresponde al tercero más vendido, lo que le da espacio para buscar generar más margen.

La superioridad en precio de Colun sobre Soprole en frutilla caja 200ml, lo que repite la idea de que Colun baje sus precios en los productos segundos en preferencias mejora las contribuciones para la categoría.

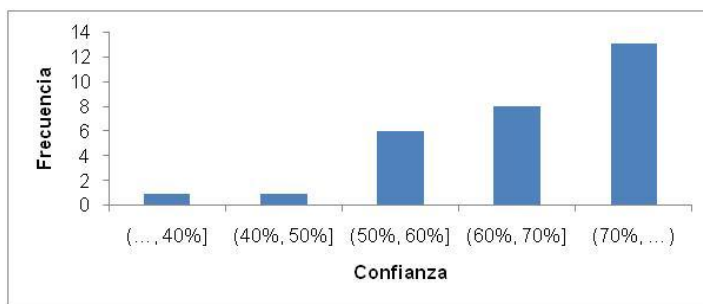
Y finalmente, la superioridad en precio de Colun sobre Nestlé en chocolate caja 1L, y la superioridad en precio de Nesquik chocolate sobre Mimun chocolate en cajas de 200ml, que comparan calidades y propuestas de valor de los productos.

A continuación, del mismo modo que en la etapa anterior, se estudia la posibilidad de replicar en el resto de la cadena las reglas que aparecen repetidas en más de una sucursal.

De las reglas N°1, N°2 y N°3 se obtienen los siguientes resultados en el análisis de los indicadores de todas las sucursales de la cadena:

Gráfico 15: Estudio para replicar reglas potenciales N°1, N°2, y N°3 (LECHE SOPROLE CHOCOLATE CAJA 200ML más cara que LECHE SOPROLE FRUTILLA CAJA 200ML)

	Soporte	Peso	Confianza
Mínimo	17%	72%	38%
Máximo	56%	100%	92%



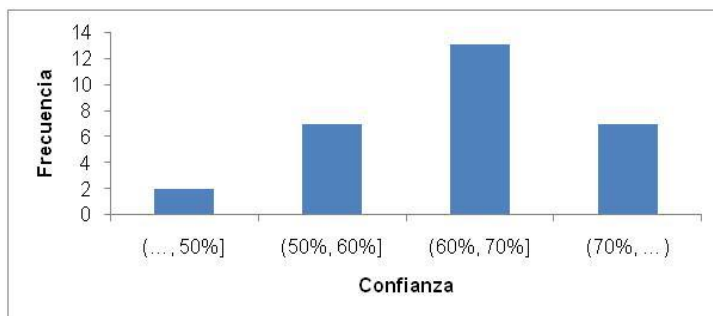
Fuente: Elaboración propia

En este caso, si bien las confianzas parecen apuntar a que las contribuciones experimentan una mejora la mayoría de las veces que la regla ocurre, los soportes de la regla se ubican entre un 17% y un 56%, lo que no es suficiente para recomendar replicarla.

De las reglas N°4, N°5, N°6, N°7, N°8, N°9, y N°10 se obtienen los siguientes resultados en el análisis de los de indicadores de todas las sucursales de la cadena:

Gráfico 16: Estudio para replicar reglas potenciales N°4, N°5, N°6, N°7, N°8, N°9, y N°10 (LECHE SOPROLE CHOCOLATE CAJA 200ML más cara que LECHE LONCO CHOCOLATE CAJA 200ML)

	Soporte	Peso	Confianza
Mínimo	43%	59%	47%
Máximo	85%	100%	79%



Fuente: Elaboración propia

Esta regla, según lo que se puede apreciar en la tabla no ocurre muy frecuentemente (lo que responde a su carácter de potencial), sin embargo, sus confianzas en todas las sucursales de la cadena se concentran entre el 60% y el 70%, lo que la convierte en una regla candidata a ser probada y recomendada para replicar en todas las sucursales de la cadena.

De la ramificación de las reglas a nivel sucursal se obtienen las siguientes reglas, donde se muestran los indicadores de la regla original, y a continuación, marcados con "r4" los de la ramificación (sólo en la rama 4 se encontró una regla que en su versión original también superara los criterios de corte).

Tabla 16: Resultados Ramificación a Nivel Sucursal

N°	Sucursal	Producto A	Producto B	Peso	Sop.	Conf.	Peso_r4	Sop_r4	Conf_r4
1	PAJARITOS (S.M)	NESQUIK FRUTILLA CAJA 200ML	BEBIDA LACTEA SUPER CALO FRUT CAJA 200ML	91%	98%	67%	83%	92%	69%

Fuente: Elaboración propia

Como se observa la regla que es seleccionada de la ramificación corresponde a la regla N° 19 de la Tabla 13, lo que se interpreta como que no sólo Nesquik es más caro que Super Calo, si no que es más de un 20% más caro.

De los resultados de las reglas a nivel sucursal atribuidas, se obtienen reglas que están compuestas por los mismos productos que las reglas seleccionadas en la primera etapa, por lo que no se considera que tengan una validez aceptable. Además, estas reglas están conformadas con pocos productos transados debido a que están bajo la etiqueta de la sucursal. (VerANEXO: Reglas atribuidas seleccionadas a nivel sucursal).

4.5.2 Reglas Seleccionadas de Comparación entre Locales

En la selección de las reglas de comparación entre locales se obtienen los niveles más bajos en indicadores de todo el proyecto, por lo que se decide relajar el criterio de selección definido y cortar en un 60%, de modo que sean visibles reglas en esta etapa.

Según el nuevo criterio definido las reglas seleccionadas son cuatro y se muestran en la siguiente tabla, donde la dirección de la sentencia es el **Producto X** es más caro en la **Sucursal A** que en la **Sucursal B**, y al igual que en la sección anterior, se reconocerá los tamaños de las salas indicando con (S.P) si corresponde a una sala pequeña, (S.M) si corresponden a una sala mediana, y (S.G) si corresponde tener una sala grande.

Tabla 17: Resultados Reglas de Comparación entre Locales

N°	Producto	Sucursal A	Sucursal B	Peso	Sop.	Conf.
1	LECHE COLUN ENTERA CAJA 1L	J.J. PEREZ (S.G)	TEMUCO (S.P)	98%	62%	61%
2	LECHE PARMALAT CHOCOLATE CAJA 1L	MAIPU (S.G)	BELLOTO (S.M)	100%	61%	70%
3	LECHE NESTLE CHOCOLATE CAJA 1L	RANCAGUA (S.M)	BELLOTO (S.M)	94%	63%	66%
4	LECHE COLUN CHOCOLATE CAJA 1L	LAS CONDES (S.G)	LINARES (S.P)	91%	65%	66%

Fuente: Elaboración propia

Dado que los niveles de precios entre las distintas sucursales no tienen un ordenamiento intersucursales, y responden al mismo análisis y metodología de fijación de precios, los patrones de relaciones no son sistemáticos, y por ende, no tienen indicadores altos que resulten convincentes para su aplicación.

De las reglas atribuidas de comparación entre locales, se obtienen soportes en promedio de un 83,85%, sin embargo las confianzas presentan valores más bajos con un promedio de un 48%. Como los resultados tienen indicadores con valores bastante bajos, se optó por relajar el criterio de corte y seleccionar solo las mejores reglas según sus pares (soporte, confianza).

Tabla 18: Resultados Reglas de Comparación entre Locales Atributadas

N°	Atributo A	Atributo B	Soporte	Confianza
1	CHOCOLATE	FRUTILLA	54%	67%
2	NESQUIK	PARMALAT	98%	62%
3	LONCOLECHE	PARMALAT	80%	59%
4	CHOCOLATE	CHOCOLATE MANJAR	59%	59%
5	SOPROLE	LONCOLECHE	63%	58%

Fuente: Elaboración propia

Las reglas N°2, y N°3 hacen sentido considerando la calidad de los productos y el valor de marca de estos y en el caso de Loncoleche, por la propuesta de valor que hace en el packaging de sus productos.

Las reglas N°1, y N°4 por otro lado respetan el grado de libertad que da el sabor preferido entre los consumidores, pudiendo aumentar el precio de este por sobre los otros sabores.

La regla N°5 no resulta muy intuitiva, y la credibilidad sobre las mejoras que pueda traer esta regla a las contribuciones de los productos se ve disminuida por la amplia gama de productos que tienen ambas marcas lo que hace difícil poder compararlas, debido a esto, la construcción de la reglas podría estar distorsionada y se prefiere no considerarla como un resultado aplicable.

Finalmente, cabe destacar que los soportes y confianzas de estas reglas no tienen valores superiores al 60%, a excepción de la regla N°2 que podría ser la más segura en sus resultados.

Las reglas que comparan atributos resultan poco rigurosas debido a la dificultad de separar los efectos del resto de los atributos en los productos en estudio. Sin embargo, se recomienda aplicar y evaluar prácticamente su desempeño, ya que si los resultados son positivos, estas reglas constituyen información más gestionable por su nivel de agregación, que la información que se tenga a nivel producto.

Por otro lado, en este ítem se pueden rescatar reglas que si bien no signifiquen una aporte a la contribución de los productos, son básicas para el ordenamiento de los formatos de los productos, es decir, las reglas del atributo packaging.

Tabla 19: Reglas de Orden

N°	Atributo A	Atributo B	Soporte	Confianza
1	CAJA 200ML	CAJA 1L	100%	46%
2	CAJA 200ML	SQUARE	100%	43%
3	CAJA 200ML	BOLSA 500ML	100%	37%
4	BOLSA 500ML	CAJA 1L	100%	55%

Fuente: Elaboración propia

Las reglas de orden, como se menciona anteriormente no significan un aporte a la contribución, lo que se puede verificar con los valores de las confianzas, pero son utilizadas un 100% de las veces como indican los soportes.

Estas reglas constituyen un caso especial, y a pesar de sus bajas confianzas se recomienda seguir cumpliéndolas, puesto que si los productos en formatos individuales resultan más económicos que los formatos familiares o a granel, será tentador para el cliente comprar los volúmenes de productos que busca, en formatos pequeños. Si esto ocurre, se canibalizará fuertemente la venta de los formatos más grandes, y se arriesgarán quiebres de stock. Además, independiente de que los distintos formatos puedan representar ocasiones de consumo diferentes, el cliente siempre decidirá su compra en base a lo que le resulte más económico.

4.5.3 Reglas Seleccionadas en Comparación con la Competencia

De las reglas resultantes del análisis de la comparación de precios de productos en las distintas sucursales de la cadena, con el precio moda nacional de la competencia, se pueden seleccionar 40 reglas, las cuales son todas de tipo No Evaluable, es decir, tienen todas un soporte de 100%. (Ver detalle en ANEXO I: Reglas seleccionadas en comparación con la competencia)

Las reglas comparan 7 productos, que en su mayoría son de altos niveles de ventas, y lo hacen en 23 sucursales distintas. Con esto se observa la estrategia de estar bajo la competencia en los productos importantes de la categoría, que análogamente a los roles de éstas, podrían tener un carácter de destino para los clientes que compran esta línea de productos en la cadena. Los 7 productos y sus posiciones en ventas se listan en la tabla a continuación:

Producto	Posición en ranking de ventas
LECHE SOPROLE ENTERA CAJA 1L	1°
LECHE COLUN ENTERA CAJA 1L	8°
LECHE COLUN DESCREMADA CAJA 1L	7°
LECHE COLUN CHOCOLATE CAJA 1L	12°
LECHE SOPROLE CHOCOLATE CAJA 1L	2°
LECHE COLUN SEMI DESCREMADA CAJA 1L	15°
LECHE PARMALAT ENTERA CAJA 1L	4°

Fuente: Elaboración propia

De las reglas potenciales se encuentran 3 reglas más, que también sugieren estar debajo del precio de la competencia en el producto leche Soprole chocolate caja 200ml que corresponde al tercer más vendido de la categoría. Las sucursales en las que aparece la regla son La Pintana, Recoleta y Los Ángeles, y el soporte y la confianza son, en los tres casos, de 57% y 75% respectivamente.

Los resultados obtenidos en este subcapítulo se ven fuertemente influenciados por la estrategia de la empresa, que apunta a tener los precios más bajos que la competencia. Sin embargo, se observa que las reglas seleccionadas corresponden sólo a los productos más vendidos, dejando entre ver la posibilidad de que para el resto de los productos de la categoría, esta no sea la mejor opción.

4.6 Evaluación del Modelo

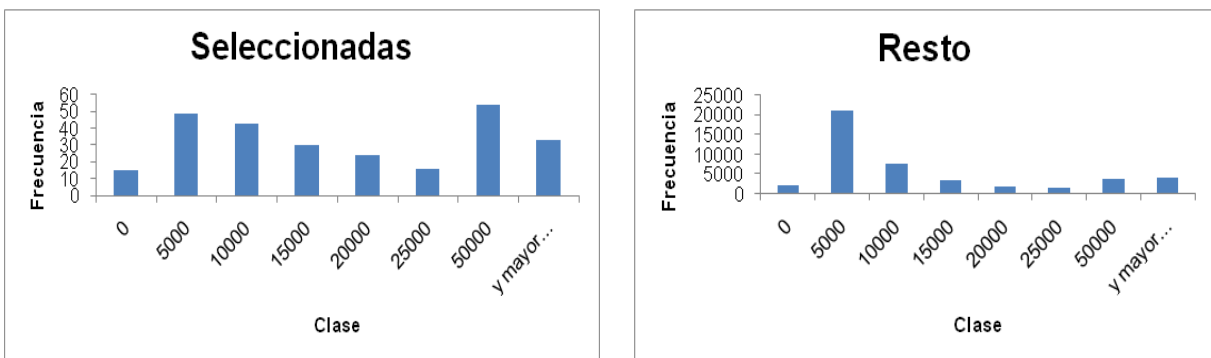
Para evaluar la capacidad de discriminación de reglas del modelo se usaron los datos comprendidos en el periodo que se extiende desde Julio de 2011 a Septiembre de 2011, y corresponde al tercer trimestre de dicho año. En total se utilizan para esta etapa, 13 semanas de transacciones distintas a las semanas utilizados para levantar las reglas.

Además, para el grupo de reglas que no habían sido seleccionadas se escogió de manera aleatoria uno de los dos posibles sentidos¹⁾ producto *i* es más caro que producto *j* y 2) producto *j* es más caro que producto *i*.

El objetivo en este punto es contrastar las contribuciones resultantes en el grupo seleccionado con las contribuciones del resto de las reglas y así evaluar si el modelo es capaz de indicar reglas que significan efectivamente una mejora para las contribuciones de los productos.

Del análisis de las distribuciones de las contribuciones semanales, de las reglas seleccionadas y del resto de las reglas, se obtienen medianas de \$13.808 y \$4.972 respectivamente. A continuación se presentan los histogramas de las contribuciones de cada grupo, donde se observa que las reglas seleccionadas tienen su peak de contribuciones entre los \$25.000 y los \$50.000, y el resto de las reglas se concentra entre \$0 y \$5.000.

Gráfico 17: Histograma de contribuciones reglas seleccionadas y resto



Fuente: Elaboración propia

Con lo anterior se puede concluir sobre el funcionamiento del modelo, ya que las medianas de las contribuciones de las reglas seleccionadas y del resto de las reglas se diferencian claramente, y al ser la primera casi el triple de la segunda se puede interpretar que el modelo está seleccionando, en base a sus indicadores, las reglas que efectivamente significan un real aporte a los beneficios de la compañía.

Como una segunda prueba, la evaluación del modelo considera la observación de los soportes de las reglas en el periodo de testeo antes explicitado, con el fin de identificar la consistencia de las reglas en periodos distintos.

En base a los resultados encontrados se puede decir que las reglas encontradas son consistentes en el tiempo, ya que en su mayoría mantienen o aumentan la frecuencia de su ocurrencia. Sin embargo, estos valores no son absolutos (soportes 100%), lo que significa que se siguen equivocando al romper estas reglas, y continúan así dejando dinero sobre la mesa.

En la tabla a continuación, se muestran los porcentajes de reglas que en el periodo de testeo tienen soportes que pasen los criterios de selección, y el promedio de las confianzas obtenidas en cada sección:

Tabla 20: Porcentajes de reglas que pasan testeo

	%Reglas que siguen teniendo un soporte alto (supera criterio de selección).	Promedio Confianzas
Nivel Sucursal	55%	69,77%
Ramificación Sucursal	25%	30,30%
Nivel Cadena	25%	30,26%
Comparación Competencia	100%	10,38%

Fuente: Elaboración propia

Que se sigan rompiendo estas reglas confirma que los precios se desordenan, y no tienen un sistema que alerte de esto. El desorden podría atribuirse a la influencia de los proveedores en los tomadores de decisiones, ya que éstos, con promociones u otras técnicas buscan potenciar las ventas de sus productos incentivando al desequilibrio de los precios del resto de los productos de la categoría.

Por otro lado, los promedios de las confianzas obtenidas son bajos, lo que podría estar influenciado por decrecimientos orgánicos de las ventas de la cadena y/o por el mayor efecto porcentual que tiene una semana en la menor cantidad de datos que se están evaluando. Al estar evaluando un 24% de semanas con respecto al periodo de calibración, el peso relativo de las semanas del periodo de testeo es mayor, y por ende, una semana menos de contribución que supera o iguala al promedio histórico (semana favorable) puede significar una confianza mucho menor.

Para ejemplificar sobre la pérdida que conlleva quebrar las reglas, se presenta a continuación una evaluación económica de una regla a nivel sucursal que haya resultado replicable, y el cálculo se hace estimando el valor de la pérdida.

La evaluación se realizó sobre las reglas a nivel sucursal N°14, N°15, N°16, y N°17 que se muestran en la Tabla 13. Dicha regla, establece que la leche Soprole entera de caja 1L debe ser más cara que la leche Parmalat entera de caja 1L, en las sucursales Coquimbo, Temuco, Rancagua y Belloto.

Como medida aproximada de pérdida, se calcula la diferencia entre el promedio de contribuciones semanales en que se cumple la regla y en que no se cumple la regla, esto multiplicado por las veces en que no se cumple la regla. Es decir, se estima la diferencia entre los estados con regla y sin regla como lo que se está dejando de ganar y esto se multiplica la cantidad de veces en que se equivocan y rompen la regla.

Tabla 21: Estimación de pérdida en regla ejemplo

Sucursal	Prom. con regla	Prom. sin regla	Diferencia	Veces con regla	Total transados	Veces sin regla	Pérdida
COQUIMBO	\$ 169.276	\$ 95.484	\$ 73.793	37	51	14	\$ 1.033.100

Sucursal	Prom. con regla	Prom. sin regla	Diferencia	Veces con regla	Total transados	Veces sin regla	Pérdida
TEMUCO	\$ 34.215	\$ 19.217	\$ 14.998	40	48	8	\$ 119.985
RANCAGUA	\$ 105.693	\$ 70.595	\$ 35.098	35	45	10	\$ 350.980
BELLOTO	\$ 63.000	-\$ 19.948	\$ 82.947	34	44	10	\$ 829.474

Fuente: Elaboración propia

Con el objetivo de visualizar lo que significa la pérdida, a continuación se contrastan los valores de las pérdidas con los valores de las contribuciones totales que suman los pares de productos en el año estudiado, para establecer el porcentaje de la ganancia anual, que la pérdida significa.

Tabla 22: Porcentaje de pérdida estimado en regla ejemplo

Sucursal	Contribución Pérdida	Contribución Total	% Pérdida
COQUIMBO	\$ 1.033.100	\$ 7.599.998	14%
TEMUCO	\$ 119.985	\$ 1.522.336	8%
RANCAGUA	\$ 350.980	\$ 4.405.214	8%
BELLOTO	\$ 829.474	\$ 1.942.518	43%

Fuente: Elaboración propia

Los valores se mueven entre un 8% y un 43% para la regla ejemplo, lo que implicaría una alta incidencia en los resultados de la compañía. Si las pérdidas que significan las distintas reglas en el trabajo presentadas no fueran dejadas sobre la mesa los beneficios de la empresa mejorarían significativamente, y si la información sobre las relaciones entre precios de los productos de las distintas categorías los montos podrían aumentar considerablemente, sobre todo en categorías de conveniencia y no de destino como es la categoría estudiada.

4.7 Aplicación

Los resultados del presente trabajo corresponden a información valiosa en la fijación de los precios de una compañía, y constituyen un pre pricing que como su objetivo lo indica, pueden constituir un aporte clave en la realización del pricing.

La información levantada de relaciones entre los precios se puede generar filtros o reglas que revisen la estructura de los precios que se están poniendo, en el caso en que este procedimiento se realice en base a políticas de precio y/o márgenes de utilidades.

Por otro lado, si el pricing se realiza optimizando en base a estimaciones de demanda, se permite apoyar los conocimientos de los tomadores de decisión y reducir el espacio de soluciones del problema, en este caso, las reglas podrían acotar el problema y convertirse en un input o dato inicial.

Cualquiera sea la metodología para la fijación de los precios, la información levantada con las reglas corresponde a información no necesariamente conocida por los managers, debido a que no solo rescata las sensibilidades de los consumidores (al estar trabajando con las contribuciones y no las ventas), sino, también las relaciones de precios que más beneficios

traen. Constituyéndose como un paso fundamental en el proceso de fijación de precios, que seguro tendrá un pago altísimo en los resultados de la empresa.

En el ANEXO J: Set resumen de reglas a aplicar, se presenta el resumen de los resultados obtenidos, indicando cuales reglas se deben aplicar a la fijación de precios de la cadena.

5 CONCLUSIONES

5.1 Conclusiones del Trabajo

Como se ha comentado, de entre las variables claves para mejorar los resultados de una compañía (unidades vendidas, precio, costos fijos y costos variables) el instrumento más poderoso es la fijación de precios, ya que una buena estrategia en este ámbito, asegura traducirse en resultados con un mayor porcentaje de impacto.

Es por la importancia de la correcta fijación de precios, que en el presente trabajo se desarrolla una metodología alternativa que busca disminuir la dificultad del problema, generando filtros si los precios son fijados con políticas de la empresa, o bien constituyendo restricciones que reduzcan el tamaño del problema si los precios son optimizados en base a estimaciones de demanda.

El trabajo cumple el objetivo de desarrollar una metodología genérica para todas las categorías, en cuanto se presenta un método que es capaz de identificar patrones de comportamiento de precios, que generan un aporte a la contribución de la categoría. Sin embargo es importante hacer énfasis en que los resultados obtenidos pueden no ser iguales, más aún puede no encontrarse resultados que superen los cortes de los indicadores, en la aplicación de la metodología a otra categoría. Es decir, si bien la metodología resulta genérica en su construcción, esta depende de los datos, en especial del orden de los precios, para generar resultados satisfactorios.

En el proyecto, la metodología fue aplicada en la categoría Leches Liquidas de una cadena de supermercados mayorista, identificando un total de 40 reglas de los precios tanto a nivel producto, a nivel de atributos y en comportamiento frente a la competencia. Sin embargo, debido al desorden actual de los precios que se utilizaron en el estudio, no fue posible encontrar patrones de comportamiento para todos los productos de la categoría y en consecuencia no es posible generar una estructura que los ordene a todos en una línea decreciente de precios.

A partir del trabajo realizado se puede concluir entonces que las relaciones entre los precios de los productos pueden ser correctamente evaluadas con indicadores que midan su frecuencia de ocurrencia (y su respectiva significancia que valide este resultado) y sus casos favorables para la contribución. Ya que estos indicadores poseen capacidad de discriminación de reglas efectivamente mejores para la contribución de los productos, y son válidos en periodos diferentes de datos, en cuanto que, las reglas seleccionadas siguen ocurriendo y constituyen un patrón de comportamiento.

En base a la evaluación de una regla replicable (tomada como ejemplo), se obtuvo que el romper una regla significaría una pérdida de entre un 8% y un 43% de la contribución anual del par de productos. Llevado a dinero, se tiene que en promedio en cada sucursal se pierden \$583.385, lo que proyectado en las 29 sucursales suma \$16.918.158 anuales.

Además, se concluye que el análisis de las reglas entrega información importante que conforma el pre pricing. Esta información puede aplicarse como filtros que revisen la estructura de los precios que se están fijando, en el caso en que el pricing se realice en base a políticas de precio y/o márgenes de utilidades, o bien, puede reducir el espacio de soluciones del problema, si la fijación de precios se hace con modelos de optimización.

5.2 Discusión de Resultados

A partir de los resultados obtenidos en la aplicación de la metodología desarrollada, surgen diversas aristas de cuestionamiento y análisis respecto de la generación de reglas para la fijación de precios y la creación de indicadores capaces de evaluarlas.

En primera instancia, se ha obtenido un set de reglas aplicables como filtros reguladores en distintos niveles de agregación. De estos, destaca las relaciones a nivel producto, ya que poseen mejores frecuencias de ocurrencia y representan, en más casos, una mejora para la contribución. Sin embargo, dicho nivel, resulta poco gestionable debido al detalle que posee, lo cual lo presenta una especificidad tal, que dificulta su implementación en la praxis.

Para simplificar los resultados en cuanto a la cantidad de reglas y a su aplicabilidad, se desarrollan dos agregaciones de lo obtenido en la etapa anterior. Estos son: la ramificación y el atribuir las reglas. En cuanto a la ramificación de las reglas, se puede decir que dado el desorden actual de los precios, la confianza no aumenta en intervalos más pequeños como se esperaba, resultando sólo una regla que supera los criterios de corte. Para mejorar los resultados obtenidos en esta sección, se sugiere realizar nuevamente este paso una vez implementadas las reglas a nivel producto; ya que datos más ordenados podrían tener frecuencias de ocurrencia más altas, bajo reglas más estrictas; como sucede, en la ramificación. Por otro lado, al atribuir las reglas se obtienen relaciones de precios más claras y manejables para quienes toman las decisiones.

En la sección de comparación con la competencia, se obtienen sólo reglas que ocurren un 100% de las veces, y por ende no existen semanas con que contrastar los desempeños. Sería recomendable probar que sucede si alguna vez no se está bajo la competencia para poder evaluarlo y dilucidar si efectivamente representan una mejora para la rentabilidad de la compañía.

En cuanto a la compatibilidad de las reglas, se puede decir que éstas no se contradicen entre sí, facilitando la implementación propuesta en el capítulo de aplicación. Si hubiese ocurrido el caso contrario, se debería definir la prioridad de las reglas en base a la estrategia de la compañía.

Como principal limitación de la metodología, se puede mencionar que la confianza no cubre el espacio de posibilidades que se da en una unidad de período de tiempo. Las cuales pueden ser:

	Se cumple la regla	No se cumple la regla
Sem. favorable		
Sem. No favorable		

Ésta sólo evalúa el caso en que se cumple la regla y es semana favorable y el resto de las posibilidades, quedan fuera del análisis.

Un método alternativo a la confianza, para evaluar la bondad de una regla, puede ser generar un estadístico de la contribución en los casos con reglay sin regla, y probar la significancia de la hipótesis nula “en ambos casos las contribuciones son iguales”. De esta forma, se estaría abarcando también el caso en que no se cumple la regla (al considerar las semanas sin regla).

Con la etapa de evaluación del modelo se concluye que éste es capaz de discriminar entre reglas efectivamente mejores para las contribuciones de los productos, ya que existe diferencia considerable entre las medianas de las series de datos de cada caso. Además, los soportes de las reglas seleccionadas son sustentables en el tiempo, pues se validan en su ocurrencia con un periodo de datos distinto al de levantamiento de las reglas. Sin embargo, se siguen teniendo casos en que las relaciones entre los productos no son respetadas y se continúa perdiendo dinero.

Finalmente, se plantea la discusión de la metodología en otro tipo de retailers, la cual sería aplicable en caso que: se tengan categorías bien definidas, que los productos sean atributables, y que las áreas de interacción de los precios sean conocidas y abordables. En caso contrario, la metodología no sería útil.

5.3 Trabajos Futuros

Se propone la implementación de la propuesta de reglas desarrollada en la categoría Leches Líquidas de forma de evaluar su impacto real en los resultados económicos de la cadena. Midiendo tanto las mejoras en los beneficios como la mejora en las herramientas de negociación con proveedores.

Se propone plantear la búsqueda de reglas en forma inversa, identificando mediante experimentos con las reglas que tengan nivel alto en soporte pero bajo en confianza, aquellos comportamientos que no deben repetirse porque representan malas prácticas para la compañía.

Como trabajos futuros se sugiere también, profundizar en las relaciones que se puedan encontrar con los precios de la competencia, ya sea diferenciando los locales que son competencia para cada sucursal, o bien, ramificando las reglas por sus rangos de diferencia de precios. Éste sería un aporte importante al dinamismo que alcance el post pricing de la compañía, y le daría ventajas competitivas al poder reaccionar más rápido que sus competidores a los cambios del entorno.

Debido a la poca información que se tiene sobre los movimientos de los precios de los competidores, en este tema se abre un amplio abanico de posibilidades al conocer un poco más sobre la respuesta de los consumidores cuando por ejemplo la competencia lanza promociones.

6 REFERENCIAS

- [1] Informe Retail 2011. Reportajes y Entrevistas, Revista Capital, Mayo 2011.
- [2] Neculman, Gonzalo. Retail en Chile Evolución, Crecimiento y Riesgos, Humphreys, clasificadora de riesgo, Agosto 2010.
- [3] Informe Mensual Sectores Económicos, INE Chile, Febrero 2012.
- [4] Ventas de Comercio en Chile se Moderarían durante 2012. Reuters América Latina. <<http://lta.reuters.com/article/domesticNews/idLTASIE80P06Q20120126>> [Consulta en línea: Marzo 2012]
- [5] Michael V. Marn, Eric V. Roegner, and Craig C. Zawada. *The power of pricing: Transaction pricing is the key to surviving the current downturn—and to flourishing when conditions improve*. Mc Kinsey&Company, February 2003.
- [6] Marn, Michael, and Rosiello, Robert. *Managing Price, Gaining Profit*, Harvard Business Review, September-October 1992.
- [7] Andersen, Arthur. *At what price?*, 2000.
- [8] Pizarro, Claudio. 2010. Apuntes IN547: Gestión de Retail [Clases]. Facultad de Ciencias Físicas y Matemáticas. Departamento de Ingeniería Industrial. Universidad de Chile.
- [9] Montgomery, Alan L, Marcel Goic. *Making Better Pricing Decisions with Implied Priors*, working paper, Carnegie Mellon University, 2011.
- [10] Troncoso, Catalina. *Determinación de precios óptimos para una cadena de supermercado utilizando modelos jerárquicos Bayesianos*, Universidad de Chile, Septiembre 2010.
- [11] Cruz, Gustavo. *Determinación de precios óptimos de una categoría para una cadena de supermercado*, Universidad de Chile, Octubre 2008.
- [12] Schwartz, Daniel. *Modelo de Análisis de Canasta y su aplicación a una empresa de Home Improvement*, Universidad de Chile, Enero 2004.
- [13] Montgomery, Alan L. The Implementation Challenge of Pricing Decision Support System for Retail Managers, March 2004.
- [14] Rajaraman A., Ullman J. *Minig of Massive Datasets*, Cambridge University Press, 2010.
- [15] Aburto, Luis. 2010. Apuntes IN58B: Ingeniería de Marketing [Clases]. Facultad de Ciencias Físicas y Matemáticas. Departamento de Ingeniería Industrial. Universidad de Chile.
- [16] Hyde, V., *Train, Validate, and Test for Data Mining in JMP*, JMP® Blog, JMP Statistical Discovery Software from SAS, Tuesday, July 6, 2011.
- [17] Agrawal, Rakesh, and Srikant, Ramakrishnan. *Fast Algorithms for Mining Association Rules*. IBM Almaden Research Center. VLDB Conference, Santiago, Chile 1994.
- [18] Härdle, Wolfgang and Symar, Léopold. *Applied Multivariate Statistical Analysis*, Method and Date Technologies, April 29, 2003.
- [19] Srikant, Ramakrishnan and Agrawal, Rakesh. *Mining Quantitative Association Rules in Large Relational*, IBM Almaden Research Center, Canada, June 1996.
- [20] Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., Verkamo, I. *Finding Interesting Rules from Large Sets of Discovered Association Rules*, Department of Computer Science, University of Helsinki, Finland, November 1994.

7 ANEXOS

7.1 ANEXO A: Maestro de Productos Categoría Leches Líquidas

prod_id	ml_un	ml_dis	marca	tipo	sabor	packaging	umb	fvf
12833	500	0	CALO	SEMI DESCREMADA	CHOCOLATE	BOLSA	UN	6
12895	500	0	CALO	ENTERA	NATURAL	BOLSA	UN	6
12835	500	0	LONCOLECHE	SEMI DESCREMADA	CHOCOLATE	BOLSA	UN	6
13071	190	0	BAMBINO LAC	SEMI DESCREMADA	CHOCOLATE	TETRA WEDGE	UN	6
13072	190	0	BAMBINO LAC	SEMI DESCREMADA	CHOCOLATE MANJAR	TETRA WEDGE	UN	6
10866	1000	0	CALO	DESCREMADA	NATURAL	CAJA	UN	3
10882	1000	0	COLUN	DESCREMADA	NATURAL	CAJA	UN	3
10883	1000	0	COLUN	SEMI DESCREMADA	NATURAL	CAJA	UN	3
10854	1000	0	SOPROLE	DESCREMADA	NATURAL	CAJA	UN	3
11179	1000	0	PARMALAT	SEMI DESCREMADA	CHOCOLATE	CAJA	UN	3
14356	1000	0	PARMALAT	SEMI DESCREMADA	FRUTILLA	CAJA	UN	3
10887	1000	0	COLUN	ENTERA	NATURAL	CAJA	UN	3
11178	1000	0	PARMALAT	ENTERA	NATURAL	CAJA	UN	3
10853	1000	0	SOPROLE	ENTERA	NATURAL	CAJA	UN	3
10852	1000	0	SOPROLE	SEMI DESCREMADA	NATURAL	CAJA	UN	3
10913	1000	0	SVELTY	DESCREMADA	NATURAL	CAJA	UN	3
11468	1000	0	PARMALAT	LIGHT	CHOCOLATE	CAJA	UN	3
1028	1000	0	NESTLE	SEMI DESCREMADA	CHOCOLATE	CAJA	UN	3
10856	1000	0	SOPROLE	SEMI DESCREMADA	CHOCOLATE	CAJA	UN	3
10855	1000	0	SOPROLE	SEMI DESCREMADA	FRUTILLA	CAJA	UN	3
10888	1000	0	COLUN	SEMI DESCREMADA	CHOCOLATE	CAJA	UN	3
11418	1000	0	LONCOLECHE	DESCREMADA EXTRA CALCIO	NATURAL	SQUARE	UN	3
11423	1000	0	LONCOLECHE	SEMI DESCREMADA	CHOCOLATE	SQUARE	UN	3
11421	1000	0	LONCOLECHE	SEMI DESCREMADA	VAINILLA	SQUARE	UN	2
11420	1000	0	LONCOLECHE	SEMI DESCREMADA	PLATANO	SQUARE	UN	2
11422	1000	0	LONCOLECHE	SEMI DESCREMADA	FRUTILLA	SQUARE	UN	2
9747	1000	0	PARMALAT	CULTIVADA DIET	ARANDANO	BOTELLA	UN	1
9857	1000	0	PARMALAT	CULTIVADA DIET	VAINILLA	BOTELLA	UN	1
9859	1000	0	PARMALAT	CULTIVADA DIET	CIRUELA	BOTELLA	UN	1
9856	1000	0	PARMALAT	CULTIVADA DIET	FRUTILLA	BOTELLA	UN	1
9860	1000	0	PARMALAT	CULTIVADA DIET	FRAMBUESA	BOTELLA	UN	1
6485	1000	0	SOPROLE	S/LACTOSA DESCREMADA	NATURAL	SQUARE	UN	1

prod_id	ml_un	ml_dis	marca	tipo	sabor	packaging	umb	fv
15528	1000	0	LONCOLECHE	BIO OMEGA	NATURAL	SQUARE	UN	1
15205	1000	0	LONCOLECHE	BIO FIBRAS	NATURAL	SQUARE	UN	1
11417	1000	0	LONCOLECHE	S/LACTOSA	NATURAL	SQUARE	UN	1
6750	1000	0	LONCOLECHE	S/LACTOSA ENTERA	NATURAL	SQUARE	UN	1
6484	1000	0	SOPROLE	S/LACTOSA SEMI DESCREMADA	NATURAL	SQUARE	UN	1
11424	1000	0	LONCOLECHE	S/LACTOSA SEMI DESCREMADA	CHOCOLATE	BOLSA	UN	1
5838	200	600	CALO	SEMI DESCREMADA	FRUTILLA	CAJITA	DIS	1
6749	1000	0	LONCOLECHE	S/LACTOSA DESCREMADA	NATURAL	SQUARE	UN	1
3187	200	1200	PARMALAT	SEMI DESCREMADA	FRUTILLA	CAJITA	DIS	1
3188	200	1200	PARMALAT	SEMI DESCREMADA	CHOCOLATE	CAJITA	DIS	1
859	1000	0	SOPROLE	CULTIVADA	CHIRIMOYA	BOTELLA	UN	1
858	1000	0	SOPROLE	CULTIVADA	DURAZNO	BOTELLA	UN	1
856	1000	0	SOPROLE	CULTIVADA	FRUTILLA	BOTELLA	UN	1
967	200	1200	LONCOLECHE	SEMI DESCREMADA	CHOCOLATE	CAJITA	DIS	1
938	200	1200	LONCOLECHE	SEMI DESCREMADA	PLATANO	CAJITA	DIS	1
925	200	1200	LONCOLECHE	SEMI DESCREMADA	VAINILLA	CAJITA	DIS	1
968	200	1200	LONCOLECHE	SEMI DESCREMADA	FRUTILLA	CAJITA	DIS	1
2864	200	1200	COLUN	SEMI DESCREMADA	CHOCOLATE	CAJITA	DIS	1
2866	200	1200	COLUN	SEMI DESCREMADA	FRUTILLA	CAJITA	DIS	1
10130	200	1200	SOPROLE	SEMI DESCREMADA	CHOCOLATE MANJAR	CAJITA	DIS	1
845	200	1200	SOPROLE	SEMI DESCREMADA	CHOCOLATE	CAJITA	DIS	1
6480	200	1200	MIMUN	SEMI DESCREMADA	FRUTILLA	CAJITA	DIS	1
844	200	1200	SOPROLE	SEMI DESCREMADA	FRUTILLA	CAJITA	DIS	1
4060	200	1200	COLUN	SEMI DESCREMADA	VAINILLA	CAJITA	DIS	1
1033	200	1200	NESQUIK	SEMI DESCREMADA	CHOCOLATE	CAJITA	DIS	1
1034	200	1200	NESQUIK	SEMI DESCREMADA	FRUTILLA	CAJITA	DIS	1
966	200	1000	LONCOLECHE	ENTERA	NATURAL	CAJITA	DIS	1
841	200	1200	HUESITOS	SEMI DESCREMADA	CHOCOLATE	CAJITA	DIS	1
1030	200	1200	MILO	SEMI DESCREMADA	CHOCOLATE	CAJITA	DIS	1
1031	200	1200	TRENCITO	ENTERA	CHOCOLATE	CAJITA	DIS	1
12830	500	0	SOPROLE	ENTERA	NATURAL	BOLSA	UN	6
11623	190	0	NEXT	CULTIVADA	FRUTILLA	BOTELLITA	UN	3
6482	200	1200	MIMUN	SEMI DESCREMADA	CHOCOLATE	CAJITA	DIS	1

7.2 ANEXO B: Regresión Lineal de Niveles de Atributos.

	Coeficientes no Estandarizados		Coeficientes Estandarizados	t	Sig.
	B	Estad. Error	Beta		
(Constante)	,512	,023		21,872	,000
calo	-,054	,032	-,080	-1,721	,094
colun	-,002	,024	-,004	-,093	,926
loncoche	,010	,024	,023	,392	,698
nesquik	,091	,039	,097	2,362	,024
parmalat	-,090	,027	-,207	-3,334	,002
otras marcas	,071	,028	,135	2,517	,017
cultivada	,244	,044	,361	5,597	,000
cultivadadiet	,180	,039	,296	4,591	,000
descremada	,007	,037	,011	,194	,847
entera	,089	,030	,170	2,932	,006
slactosadescremada	,111	,048	,118	2,295	,028
slactosasemidescremada	,096	,041	,102	2,308	,027
otros sabores	,096	,039	,124	2,459	,019
chocolatemanjar	-,014	,041	-,015	-,335	,740
frutilla	-,010	,022	-,023	-,450	,656
natural	-,084	,030	-,224	-2,779	,009
platano	-,037	,043	-,040	-,860	,396
vainilla	-,004	,032	-,006	-,122	,903
otros tipos	-,004	,041	-,007	-,105	,917
bolsa	,063	,033	,104	1,900	,066
botellita	,328	,066	,249	5,001	,000
cajita	,306	,021	,854	14,491	,000
square	,110	,037	,209	2,953	,006
tetrawedge	,236	,043	,252	5,470	,000

(*) Todos los niveles que tienen un porcentaje de los productos menor al 2% fueron agrupados en 'Otros' respectivamente.

7.3 ANEXO C: Ranking de Sucursales

En la siguiente tabla se presentan todas las sucursales de la cadena ordenadas en forma decreciente por su monto vendido en el periodo que comprende el estudio, esto es Julio 2010 a Junio 2011. Además se indican suticket promedio de compra.

Sucursal	Monto	Ticket Promedio
J.J. PEREZ	\$ 543.454.589	\$ 6.057
CONCHA Y TORO	\$ 493.851.617	\$ 5.783
VESPUCIO	\$ 438.759.702	\$ 5.328
RECOLETA	\$ 393.006.679	\$ 5.891
MAIPU	\$ 392.908.055	\$ 5.028
EL BOSQUE	\$ 365.271.384	\$ 5.219
COQUIMBO	\$ 353.029.213	\$ 4.107
QUILICURA	\$ 351.735.210	\$ 4.332
INDEPENDENCIA	\$ 345.035.940	\$ 5.517
DEPARTAMENTAL	\$ 336.489.072	\$ 6.106
RANCAGUA	\$ 325.005.712	\$ 4.549
SAN PABLO	\$ 283.968.930	\$ 4.116
PAJARITOS	\$ 273.559.349	\$ 4.464
LAS CONDES	\$ 266.181.460	\$ 6.683
PUENTE ALTO	\$ 246.615.708	\$ 4.861
LA PINTANA	\$ 209.721.866	\$ 4.222
VALPARAISO	\$ 198.056.198	\$ 3.617
PENALOEN	\$ 197.000.235	\$ 4.169
BELLOTO	\$ 179.523.473	\$ 4.598
LO PRADO	\$ 178.096.123	\$ 3.992
CONCEPCION	\$174.138.110	\$ 3.566
RENCA	\$ 147.861.482	\$ 4.929
TALCA	\$ 142.353.073	\$ 2.651
CHILLAN	\$ 139.201.975	\$ 3.668
LOS ANGELES	\$ 138.152.125	\$ 2.979
LINARES	\$ 126.625.072	\$ 3.385
PUERTO MONTT	\$ 111.546.324	\$ 3.320
TEMUCO	\$ 91.703.599	\$ 2.977
TALCA MERCADO	\$ 81.694.252	\$ 3.454

Fuente: Elaboración propia.

7.4 ANEXO D: Estudio de Outliers

Producto	Media Contribución	Media sin máx.	Media sin min	%Variación de media sin máx.	%Variación de media sin máx.
11178	\$ 26.838	\$ 26.744	\$ 26.838	0%	0%
845	\$ 48.309	\$ 48.161	\$ 48.332	0%	0%
10853	\$ 68.514	\$ 68.102	\$ 68.743	-1%	0%
6484	\$ 1.058	\$ 1.052	\$ 1.058	-1%	0%
6482	\$ 699	\$ 668	\$ 698	-5%	0%
844	\$ 16.072	\$ 16.022	\$ 16.080	0%	0%
10856	\$ 45.830	\$ 45.605	\$ 45.908	0%	0%
10854	\$ 7.442	\$ 7.396	\$ 7.448	-1%	0%
967	\$ 2.742	\$ 2.727	\$ 2.744	-1%	0%
6749	\$ 985	\$ 975	\$ 985	-1%	0%
12895	\$ 25.459	\$ 25.351	\$ 25.459	0%	0%
10882	\$ 13.742	\$ 13.651	\$ 13.744	-1%	0%
1030	\$ 2.318	\$ 2.301	\$ 2.364	-1%	2%
9860	\$ 292	\$ 290	\$ 292	-1%	0%
12830	\$ 20.381	\$ 20.236	\$ 20.417	-1%	0%
11424	\$ 571	\$ 548	\$ 571	-4%	0%
10913	\$ 5.634	\$ 5.606	\$ 5.663	-1%	1%
856	\$ 901	\$ 897	\$ 901	0%	0%
841	\$ 3.440	\$ 3.395	\$ 3.442	-1%	0%
5838	\$ 2.171	\$ 2.162	\$ 2.171	0%	0%
10855	\$ 8.365	\$ 8.332	\$ 8.365	0%	0%
12833	\$ 5.386	\$ 5.356	\$ 5.390	-1%	0%
966	\$ 2.364	\$ 2.346	\$ 2.364	-1%	0%
1033	\$ 1.469	\$ 1.457	\$ 1.498	-1%	2%
10852	\$ 5.978	\$ 5.944	\$ 5.985	-1%	0%
1034	\$ 854	\$ 844	\$ 875	-1%	2%
9857	\$ 299	\$ 297	\$ 299	-1%	0%
12835	\$ 3.911	\$ 3.861	\$ 3.911	-1%	0%
11179	\$ 2.691	\$ 2.676	\$ 2.691	-1%	0%
1028	\$ 1.395	\$ 1.376	\$ 1.400	-1%	0%
11468	\$ 987	\$ 977	\$ 987	-1%	0%
925	\$ 1.804	\$ 1.795	\$ 1.804	-1%	0%
10888	\$ 4.813	\$ 4.776	\$ 4.837	-1%	0%
938	\$ 2.227	\$ 2.217	\$ 2.227	0%	0%
6750	\$ 762	\$ 757	\$ 762	-1%	0%
2864	\$ 5.989	\$ 5.941	\$ 5.991	-1%	0%
9856	\$ 326	\$ 324	\$ 326	-1%	0%
9859	\$ 258	\$ 257	\$ 258	-1%	0%
10883	\$ 6.251	\$ 6.206	\$ 6.251	-1%	0%
968	\$ 1.553	\$ 1.545	\$ 1.553	-1%	0%
11417	\$ 1.651	\$ 1.641	\$ 1.651	-1%	0%

Producto	Media Contribución	Media sin máx.	Media sin min	%Variación de media sin máx.	%Variación de media sin máx.
9747	\$ 276	\$ 273	\$ 276	-1%	0%
6485	\$ 1.194	\$ 1.182	\$ 1.195	-1%	0%
10887	\$ 8.930	\$ 8.895	\$ 8.975	0%	1%
1031	\$ 1.088	\$ 1.074	\$ 1.088	-1%	0%
859	\$ 788	\$ 784	\$ 788	0%	0%
11623	\$ 1.973	\$ 1.934	\$ 1.973	-2%	0%
858	\$ 719	\$ 716	\$ 719	0%	0%
10866	\$ 1.108	\$ 1.076	\$ 1.117	-3%	1%
2866	\$ 1.322	\$ 1.316	\$ 1.330	-1%	1%
4060	\$ 1.571	\$ 1.564	\$ 1.571	0%	0%
11420	\$ 713	\$ 710	\$ 713	0%	0%
11421	\$ 566	\$ 561	\$ 566	-1%	0%
13072	\$ 75	\$ 64	\$ 78	-17%	5%
14356	\$ 869	\$ 860	\$ 870	-1%	0%
13071	\$ 712	\$ 703	\$ 719	-1%	1%
11423	\$ 435	\$ 429	\$ 436	-2%	0%
10130	\$ 3.549	\$ 3.526	\$ 3.553	-1%	0%
11422	\$ 355	\$ 353	\$ 355	-1%	0%
3187	\$ 917	\$ 900	\$ 971	-2%	6%
3188	\$ 2.249	\$ 2.186	\$ 2.330	-3%	4%
11418	\$ 274	\$ 270	\$ 275	-2%	0%
15205	\$ 197	\$ 194	\$ 197	-1%	0%
15528	\$ 197	\$ 189	\$ 196	-4%	0%
6480	\$ 612	\$ 598	\$ 612	-2%	0%

Resumen de productos que caen en el criterio definido para detectar outliers:

Producto	Media Contribución	Media sin máx.	Media sin min	%Variación de media sin máx.	%Variación de media sin máx.
6482	\$ 699	\$ 668	\$ 698	-5%	0%
11424	\$ 571	\$ 548	\$ 571	-4%	0%
13072	\$ 75	\$ 64	\$ 78	-17%	5%
15528	\$ 197	\$ 189	\$ 196	-4%	0%
3187	\$ 917	\$ 900	\$ 971	-2%	6%
3188	\$ 2.249	\$ 2.186	\$ 2.330	-3%	4%

7.5 ANEXO E: Tabla Resumen Reglas a nivel Sucursal

SUCURSAL	TOTAL REGLAS	REGLAS EVALUABLES	REGLAS NO EVALUABLES	REGLAS POTENCIALES
1	150	47	18	11
2	144	43	19	12
3	168	46	27	13
4	118	38	17	6
5	122	37	22	7
6	104	31	20	6
7	142	45	18	9
8	135	38	19	6
9	172	50	26	17
10	143	44	17	11
11	178	51	27	8
12	136	42	13	17
13	122	42	13	8
14	123	41	21	8
15	101	36	17	4
16	131	49	8	17
17	123	41	15	12
18	115	35	17	5
19	121	44	18	7
20	133	39	17	6
21	149	37	24	11
22	128	39	20	10
23	180	50	30	12
24	122	40	20	3
25	140	42	19	14
26	148	40	19	6
38	122	40	18	9
39	119	38	21	10
40	119	36	17	9
TOTAL	3908	1201	557	274

Fuente: Elaboración propia

7.6 ANEXO F: Reglas de comparación entre locales Atributadas

El detalle de las reglas encontradas en este subcapítulo se muestra en la siguiente tabla. Donde se presentan las reglas de la forma: “el **Atributo A** es más caro que el **Atributo B**”.

N°	Atributo A	Atributo B	Soporte	Confianza
1	MILO	PARMALAT	100%	54%
2	MIMUN	SOPROLE	100%	47%
3	MIMUN	LONCOLECHE	100%	45%
4	HUESITOS	LONCOLECHE	100%	43%
5	MILO	MIMUN	100%	37%
6	FRUTILLA	NATURAL	100%	49%
7	CAJITA	BOLSA	100%	37%
8	COLUN	CALO	100%	45%
9	CAJITA	SQUARE	100%	43%
10	CHOCOLATE	NATURAL	100%	43%
11	CAJITA	CAJA	100%	46%
12	LIGHT	SEMI DESCREMADA	100%	41%
13	MILO	SOPROLE	100%	46%
14	MILO	LONCOLECHE	100%	44%
15	NESQUIK	CALO	99%	46%
16	HUESITOS	SOPROLE	99%	46%
17	NESTLE	PARMALAT	99%	44%
18	NESQUIK	LONCOLECHE	99%	45%
19	LONCOLECHE	CALO	99%	41%
20	NESQUIK	PARMALAT	98%	62%
21	S/LACTOSA DESCREMADA	S/LACTOSA	98%	45%
22	SOPROLE	CALO	98%	49%
23	S/LACTOSA DESCREMADA	S/LACTOSA ENTERA	97%	45%
24	MILO	COLUN	97%	44%
25	MILO	HUESITOS	96%	42%
26	SVELTY	COLUN	96%	50%
27	S/LACTOSA SEMI DESCREMADA	SEMI DESCREMADA	95%	36%
28	NESQUIK	SOPROLE	95%	47%
29	NESQUIK	COLUN	94%	48%
30	HUESITOS	COLUN	93%	47%
31	COLUN	PARMALAT	93%	52%
32	SOPROLE	PARMALAT	91%	45%
33	COLUN	LONCOLECHE	91%	46%
34	ENTERA	DESCREMADA	89%	54%
35	SEMI DESCREMADA	DESCREMADA	86%	48%
36	VAINILLA	CHOCOLATE	85%	49%
37	SEMI DESCREMADA	ENTERA	85%	36%
38	DESCREMADA	ENTERA	85%	35%

N°	Atributo A	Atributo B	Soporte	Confianza
39	VAINILLA	FRUTILLA	83%	53%
40	SVELTY	CALO	81%	58%
41	LONCOLECHE	PARMALAT	80%	59%
42	LONCOLECHE	SOPROLE	79%	47%
43	FRUTILLA	CHOCOLATE	78%	39%
44	SVELTY	SOPROLE	77%	48%
45	CALO	COLUN	76%	47%
46	SOPROLE	COLUN	76%	48%
47	ENTERA	SEMI DESCREMADA	76%	55%
48	MILO	NESQUIK	75%	46%
49	COLUN	NESTLE	68%	47%
50	SOPROLE	LONCOLECHE	63%	58%
51	NESQUIK	HUESITOS	61%	53%
52	COLUN	SOPROLE	60%	39%
53	CHOCOLATE	CHOCOLATE MANJAR	59%	59%
54	CHOCOLATE MANJAR	FRUTILLA	59%	50%
55	SOPROLE	NESTLE	58%	36%
56	S/LACTOSA SEMI DESCREMADA	S/LACTOSA DESCREMADA	58%	38%
57	CHOCOLATE MANJAR	CHOCOLATE	56%	44%
58	CHOCOLATE	FRUTILLA	54%	67%
59	NESTLE	SOPROLE	54%	30%
60	LONCOLECHE	COLUN	54%	37%
61	S/LACTOSA	S/LACTOSA ENTERA	54%	41%
62	S/LACTOSA ENTERA	S/LACTOSA	54%	47%
63	FRUTILLA	CHOCOLATE MANJAR	52%	62%

Fuente: Elaboración propia

7.7 ANEXO G: Reglas contradictorias encontradas en reglas de comparación entre locales atributadas

CONTRADICCIONES EN MARCA:

Atributo A	Atributo B	Soporte	Confianza
COLUN	CALO	100%	45%
CALO	COLUN	76%	47%
COLUN	LONCOLECHE	91%	46%
LONCOLECHE	COLUN	54%	37%
LONCOLECHE	SOPROLE	79%	47%
SOPROLE	LONCOLECHE	63%	58%
SOPROLE	COLUN	76%	48%
COLUN	SOPROLE	60%	39%
SOPROLE	NESTLE	58%	36%
NESTLE	SOPROLE	54%	30%

CONTRADICCIONES EN TIPO:

Atributo A	Atributo B	Soporte	Confianza
ENTERA	DESCREMADA	89%	54%
DESCREMADA	ENTERA	85%	35%
SEMI DESCREMADA	ENTERA	85%	36%
ENTERA	SEMI DESCREMADA	76%	55%
S/LACTOSA	S/LACTOSA ENTERA	54%	41%
S/LACTOSA ENTERA	S/LACTOSA	54%	47%

CONTRADICCIONES EN SABOR:

Atributo A	Atributo B	Soporte	Confianza
FRUTILLA	CHOCOLATE	78%	39%
CHOCOLATE	FRUTILLA	54%	67%
CHOCOLATE	CHOCOLATE MANJAR	59%	59%
CHOCOLATE MANJAR	CHOCOLATE	56%	44%
CHOCOLATE MANJAR	FRUTILLA	59%	50%
FRUTILLA	CHOCOLATE MANJAR	52%	62%

7.8 ANEXO H: Reglas atributadas seleccionadas a nivel sucursal

Las reglas que se presentan a continuación, responden a una derivada de las reglas a nivel sucursal, y como consecuencia de esto y de los pocos pares de productos que se pueden comparar en cada sucursal por cada nivel de atributo se encuentran reglas que son las mismas reglas seleccionadas en las comparaciones de productos.

Si bien los indicadores resultan satisfactorios, estos no responden a una comparación entre atributos, es por esto, que en esta etapa no se seleccionan reglas y el resultado es insuficiente para el proyecto.

N°	Sucursal	Atributo A	Atributo B	Soporte	Confianza	Intersección
1	LA PINTANA	ENTERA	DESCREMADA	84%	71%	50%
2	LA PINTANA	ENTERA	SEMI DESCREMADA	71%	83%	50%
3	CONCHA Y TORO	LONCOLECHE	PARMALAT	85%	85%	100%
4	CONCHA Y TORO	NESQUIK	PARMALAT	98%	90%	100%
5	CONCHA Y TORO	COLUN	PARMALAT	92%	67%	100%
6	CONCHA Y TORO	PARMALAT	CALO	68%	93%	100%
7	BELLOTO	VAINILLA	FRUTILLA	78%	69%	25%
8	BELLOTO	SEMI DESCREMADA	ENTERA	75%	78%	50%
9	BELLOTO	DESCREMADA	ENTERA	79%	79%	50%
10	TEMUCO	ENTERA	SEMI DESCREMADA	80%	66%	50%
11	PUERTO MONTT	DESCREMADA	ENTERA	82%	67%	50%
12	COQUIMBO	SOPROLE	LONCOLECHE	77%	66%	33%
13	PAJARITOS	NESQUIK	PARMALAT	97%	94%	100%
14	PAJARITOS	NESQUIK	CALO	98%	67%	100%
15	TALCA MERCADO	ENTERA	SEMI DESCREMADA	70%	68%	50%

Fuente: Elaboración propia

7.9 ANEXO I: Reglas seleccionadas en comparación con la competencia

Las reglas que se muestran a continuación son el detalle de las reglas en comparación con la competencia que cumplen con los criterios de selección, sin embargo, éstas son todas del tipo No Evaluables y por lo tanto no se puede saber si representarán un aporte a las contribuciones.

En este caso, cuando la confianza se construye sobre un soporte de 100%, la confianza no evaluará las contribuciones en los casos con regla con el promedio histórico como referencia para las contribuciones (ya que no hay casos sin regla), si no, se definirá en base a la diferencia que exista entre la media y la mediana de los datos de las contribuciones semanales de los productos y según esto estará un por sobre o bajo el 50%.

Las reglas se presentan de la forma: en la **Sucursal X** el **Producto Y** es más barato que en la competencia. (Las reglas son todas de tipo 2, es decir, la sentencia es que es más barato).

N°	Sucursal	Tipo regla	Producto	Peso	Soporte	Confianza
1	PUENTE ALTO	2	LECHE COLUN CHOCOLATE CAJA 1L	55%	100%	67%
2	QUILICURA	2	LECHE COLUN CHOCOLATE CAJA 1L	55%	100%	67%
3	PUENTE ALTO	2	LECHE COLUN DESCREMADA CAJA 1L	55%	100%	67%
4	LA PINTANA	2	LECHE COLUN DESCREMADA CAJA 1L	55%	100%	67%
5	COQUIMBO	2	LECHE COLUN DESCREMADA CAJA 1L	55%	100%	67%
6	LOS ANGELES	2	LECHE COLUN DESCREMADA CAJA 1L	55%	100%	67%
7	LAS CONDES	2	LECHE COLUN DESCREMADA CAJA 1L	55%	100%	67%
8	VALPARAISO	2	LECHE COLUN DESCREMADA CAJA 1L	55%	100%	67%
9	RENCA	2	LECHE COLUN DESCREMADA CAJA 1L	55%	100%	67%
10	VESPUCIO	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
11	PUENTE ALTO	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
12	J.J. PEREZ	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
13	LA PINTANA	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	83%
14	EL BOSQUE	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	83%
15	RECOLETA	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
16	PENALOEN	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	83%
17	BELLOTO	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
18	TEMUCO	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
19	RANCAGUA	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
20	TALCA	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
21	CHILLAN	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
22	LOS ANGELES	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
23	LO PRADO	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
24	TALCA MERCADO	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
25	VALPARAISO	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
26	RENCA	2	LECHE COLUN ENTERA CAJA 1L	55%	100%	67%
27	LA PINTANA	2	LECHE COLUN SEMIDESCREMADA CAJA 1L	55%	100%	67%
28	LINARES	2	LECHE COLUN SEMIDESCREMADA CAJA 1L	55%	100%	67%
29	COQUIMBO	2	LECHE COLUN SEMIDESCREMADA CAJA 1L	55%	100%	67%
30	RENCA	2	LECHE COLUN SEMIDESCREMADA CAJA 1L	55%	100%	67%
31	RECOLETA	2	LECHE PARMALAT ENTERA CAJA 1LT	55%	100%	83%
32	PUERTO MONTT	2	LECHE PARMALAT ENTERA CAJA 1LT	55%	100%	67%
33	J.J. PEREZ	2	LECHE SOPROLE CHOCOLATE CAJA 1L	82%	100%	67%
34	BELLOTO	2	LECHE SOPROLE CHOCOLATE CAJA 1L	82%	100%	67%
35	RANCAGUA	2	LECHE SOPROLE CHOCOLATE CAJA 1L	82%	100%	67%
36	COQUIMBO	2	LECHE SOPROLE CHOCOLATE CAJA 1L	82%	100%	67%
37	VESPUCIO	2	LECHE SOPROLE ENTERA CAJA 1L	100%	100%	82%
38	J.J. PEREZ	2	LECHE SOPROLE ENTERA CAJA 1L	100%	100%	73%
39	EL BOSQUE	2	LECHE SOPROLE ENTERA CAJA 1L	100%	100%	82%
40	MAIPU	2	LECHE SOPROLE ENTERA CAJA 1L	100%	100%	73%

Fuente: Elaboración propia

7.10 ANEXO J: Set resumen de reglas a aplicar

A continuación se presenta un set resumen de los resultados obtenidos, indicando cuales reglas se deben aplicar a la fijación de precios de la cadena.

Sucursal (es)	A		B	Prom. Peso	Prom. Sop.	Prom. Conf.
TODA LA CADENA	LECHE ENTERA COLUN CAJA 1L	más cara que	LECHE COLUN SEMI DESCREMADA CAJA 1L	95%	76%	55%
LA PINTANA	LECHE ENTERA COLUN CAJA 1L	más cara que	LECHE COLUN DESCREMADA CAJA 1L	93%	84%	71%
TODA LA CADENA	LECHE ENTERA COLUN CAJA 1L	más cara que	LECHE PARMALAT ENTERA CAJA 1LT	89%	84%	57%
BELLOTO	LECHE COLUN VAINILLA CAJA 200ML	más cara que	LECHE COLUN FRUT CAJA 200ML	100%	78%	69%
LINARES	LECHE SOPROLE CHOC CAJA 200ML	más cara que	LECHE COLUN CHOC CAJA 200CC	98%	75%	65%
COQUIMBO	LECHE SOPROLE CHOC CAJA 200ML	más cara que	LECHE LONCOLECHE CHOC CAJA 200ML	98%	77%	66%
COQUIMBO	LECHE SOPROLE CHOC CAJA 200ML	más cara que	LECHE COLUN CHOC CAJA 200CC	93%	70%	69%
BELLOTO	LECHE SOPROLE DESCREMADA CAJA 1L	más cara que	LECHE SOPROLE ENTERA CAJA 1L	98%	79%	79%
PUERTO MONTT	LECHE SOPROLE DESCREMADA CAJA 1L	más cara que	LECHE SOPROLE ENTERA CAJA 1L	94%	82%	67%
COQUIMBO	LECHE SOPROLE ENTERA CAJA 1L	más cara que	LECHE PARMALAT ENTERA CAJA 1LT	94%	73%	70%
TEMUCO	LECHE SOPROLE ENTERA CAJA 1L	más cara que	LECHE PARMALAT ENTERA CAJA 1LT	89%	83%	65%
RANCAGUA	LECHE SOPROLE ENTERA CAJA 1L	más cara que	LECHE PARMALAT ENTERA CAJA 1LT	83%	78%	71%
BELLOTO	LECHE SOPROLE ENTERA CAJA 1L	más cara que	LECHE PARMALAT ENTERA CAJA 1LT	81%	77%	74%
BELLOTO	LECHE SOPROLE SEMI DESCREMADA CAJA 1L	más cara que	LECHE SOPROLE ENTERA CAJA 1L	98%	75%	78%
PAJARITOS	NESQUIK FRUT CAJA 200ML	20% más cara que	BEBIDA LACTEA SUPER CALO FRUT CAJA 200ML	91%	98%	67%
PAJARITOS	NESQUIK FRUT CAJA 200ML	más cara que	LECHE LONCOLECHE FRUT CAJA 200ML	87%	98%	65%
PAJARITOS	LECHE SOPROLE FRUTILLA CAJA 200ML	más cara que	LECHE SOPROLE CHOCOLATE MANJAR CAJA 200ML	50%	52%	71%
TALCA	LECHE SOPROLE CHOCOLATE CAJA 200ML	más cara que	LECHE SOPROLE FRUTILLA CAJA 200ML	98%	55%	72%
TALCA MERCADO	LECHE SOPROLE CHOCOLATE CAJA 200ML	más cara que	LECHE SOPROLE FRUTILLA CAJA 200ML	100%	54%	72%
TODA LA CADENA	LECHE SOPROLE CHOCOLATE CAJA 200ML	más cara que	LECHE SOPROLE FRUTILLA CAJA 200ML	96%	65%	64%
RECOLETA	LECHE SOPROLE CHOCOLATE CAJA 200ML	más cara que	LECHE LONCO CHOCOLATE CAJA 200ML	100%	56%	70%
PAJARITOS	LECHE NESQUIK CHOCOLATE CAJA 200ML	más cara que	LECHE MIMUN CHOCOLATE CAJA 200ML.	56%	63%	89%
EL BOSQUE	LECHE COLUN FRUTILLA CAJA 200ML	más cara que	LECHE SOPROLE FRUTILLA CAJA 200ML	93%	58%	72%
TALCA MERCADO	LECHE COLUN CHOCOLATE CAJA 1L	más cara que	LECHE NESTLE CHOCOLATE 12X1L.	98%	64%	74%
TODA LA CADENA	CHOCOLATE	más cara que	FRUTILLA	-	54%	67%
TODA LA CADENA	NESQUIK	más cara que	PARMALAT	-	98%	62%
TODA LA CADENA	LONCOLECHE	más cara que	PARMALAT	-	80%	59%

Sucursal (es)	A		B	Prom. Peso	Prom. Sop.	Prom. Conf.
TODA LA CADENA	CHOCOLATE	más cara que	CHOCOLATE MANJAR	-	59%	59%
TODA LA CADENA	CAJA 200ML	más cara que	CAJA 1L	-	100%	46%
TODA LA CADENA	CAJA 200ML	más cara que	SQUARE	-	100%	43%
TODA LA CADENA	CAJA 200ML	más cara que	BOLSA 500ML	-	100%	37%
TODA LA CADENA	BOLSA 500ML	más cara que	CAJA 1L	-	100%	55%
TODA LA CADENA	COMPETENCIA	más cara que	LECHE COLUN CHOCOLATE CAJA 1L	55%	100%	67%
TODA LA CADENA	COMPETENCIA	más cara que	LECHE COLUN DESCREMADA CAJA 1L	55%	100%	67%
TODA LA CADENA	COMPETENCIA	más cara que	LECHE COLUN ENTERA CAJA 1L	55%	100%	70%
TODA LA CADENA	COMPETENCIA	más cara que	LECHE COLUN SEMIDESCREMADA CAJA 1L	55%	100%	67%
TODA LA CADENA	COMPETENCIA	más cara que	LECHE PARMALAT ENTERA CAJA 1LT	55%	100%	75%
TODA LA CADENA	COMPETENCIA	más cara que	LECHE SOPROLE CHOCOLATE CAJA 1L	82%	100%	67%
TODA LA CADENA	COMPETENCIA	más cara que	LECHE SOPROLE ENTERA CAJA 1L	100%	100%	78%

Fuente: Elaboración propia