



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

IDENTIFICACIÓN DE QUIEBRES DE STOCK EN GÓNDOLA MEDIANTE UN  
MODELO DE MARKOV OCULTO CON INFORMACIÓN PARCIAL DE CLIENTES  
LEALES

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL INDUSTRIAL

ENRIQUE ANDRÉS INFANTE MEDINA

PROFESOR GUÍA:  
RICARDO MONTOYA MOREIRA

MIEMBROS DE LA COMISIÓN:  
ANDRÉS MUSALEM SAID  
MARCELO OLIVARES ACUÑA  
ALEJANDRA PUENTE CHANDÍA

SANTIAGO DE CHILE  
2018

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL  
POR: ENRIQUE ANDRÉS INFANTE MEDINA  
FECHA: 2018  
PROF. GUÍA: SR RICARDO MONTOYA MOREIRA

## IDENTIFICACIÓN DE QUIEBRES DE STOCK EN GÓNDOLA MEDIANTE UN MODELO DE MARKOV OCULTO CON INFORMACIÓN PARCIAL DE CLIENTES LEALES

Un aspecto fundamental en el funcionamiento de un supermercado es la gestión de inventario. Entender la demanda permite determinar cuánto pedir y con qué frecuencia, y a la vez definir el mix de productos a ofrecer. Sin embargo, todo el trabajo de inventario se pierde si no se acompaña de una correcta gestión dentro de la tienda, esto es, que los productos se encuentren disponibles en góndola cuando un cliente desee comprarlos. Los problemas relacionados a esta mala gestión generan pérdidas cercanas a los US\$500 millones para la industria de los supermercados en Chile cada año.

Para alcanzar niveles óptimos en la reposición de productos en góndola es necesaria una correcta identificación de los *quiebres de stock*, fenómeno que se lleva a cabo cuando un producto no se encuentra disponible para su compra. Se utiliza el Faltante de Mercadería en Góndola (FMG) como indicador de disponibilidad de los productos, el cual para Chile es de aproximadamente un 15 %.

El presente trabajo tiene por objetivo refinar mecanismos existentes diseñados para identificar los quiebres de stock en una gran cadena de supermercados, en base a información transaccional. Para esto, se utilizan Modelos de Markov Oculto, por su capacidad de adaptarse a la naturaleza del problema. Se cree que el comportamiento de compra de los clientes puede entregar información valiosa sobre el estado de cada producto en góndola y, por lo tanto, ayudar a identificar los quiebres de stock con mayor precisión. Por esta razón, se estiman modelos de identificación de quiebres para distintos segmentos de clientes agrupados según su *nivel de lealtad*, noción asociada al porcentaje de veces que un cliente compra un producto en un período determinado.

Para estimar los modelos, se utiliza información transaccional agregada para 10 locales y para 58 productos, durante 469 días. También se cuenta con información desagregada por cada transacción que realiza cada cliente en el mismo marco temporal y espacial. Finalmente, se cuenta con mediciones del estado de disponibilidad real en góndola para 14 de estos días.

Evaluando los modelos sobre segmentos de clientes generados en base a distintos criterios, el mejor desempeño es alcanzado considerando la información de los clientes con un nivel de lealtad igual o mayor al 1 %, y utilizando el total de transacciones del supermercado para calcular las probabilidades de ocurrencia de las observaciones. Este modelo logra identificar un 65.42 % de los quiebres ocurridos con un porcentaje de falsas alarmas de 6.15 %, superando los resultados obtenidos para el modelo con información completa.

# Agradecimientos

En primer lugar, agradezco a mi familia. A mi madre, Rosa, por confiar en mis capacidades aún cuando yo no lo hiciera. A mi padre, Enrique, por su apoyo incondicional en cada una de las decisiones tomadas a lo largo de mi proceso universitario. A la Yoyo, por el chaqueteo constructivo y por compartir mi humor. A la Jose y el Nico, por darme el privilegio de ser una fuente de apoyo para ellos.

A mis queridos Engranaje, por su amistad a lo largo de todos estos años. Gracias a ustedes me he sentido siempre acompañado a lo largo de mi vida universitaria, en los buenos y en los malos momentos.

A todas las amistades que formé durante mi estancia en la Facultad. A Wena, por demostrarme desde el inicio la calidad de personas con las que me iba a encontrar. A la Esquina, por entregarme grandes amigos y muy buenos momentos fuera de la sala de clases. A mis queridos ZD, por compartir la locura de Análisis y Medida, incluyendo un viaje durante el paro del 2013. A Gamanchester, por entregarme su amistad en mis últimos años de carrera.

Al Área de Difusión de la FCFM, que me acogió como parte de su familia durante muchos años y donde crecí como profesional, pero por sobre todo como persona. Agradezco a las amistades formadas durante este período, y en especial a la Pili, quien fue una amiga y una consejera, y a la Mane, cuyo amor infinito y alegría siempre recordaré.

Al equipo de la Feria Empresarial, por ayudarme a reencontrarme con mi vocación cuando la pensaba perdida. En especial agradezco a todas las grandes amistades que formé, y que perduran hasta el día de hoy. A Ex-Fábrica, por ser unos tremendos amigos y por estar siempre ahí para mi.

Agradezco a la Coni, por todo su apoyo y por ser mi compañera estos últimos años. Tu amor y cariño hacen que los días malos no lo parezcan tanto.

A todos los compañeros, auxiliares y profesores que aportaron en mi formación, ya sea en una sala de clases o en la terraza un viernes después de clases. A Ricardo Montoya, por su constante apoyo durante todo el proceso y su compromiso con la docencia. A los miembros de la comisión, por el interés mostrado en mi memoria y por sus valiosos comentarios.

Finalmente, agradezco a todos quienes ya no forman parte de mi vida, pero que de una u otra forma me ayudaron a ser quien soy el día de hoy. Gracias.

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes Generales . . . . .	2
1.2. Quiebres de Stock . . . . .	3
1.2.1. Causas y Consecuencias . . . . .	4
1.2.2. Soluciones . . . . .	5
1.3. Trabajos Relacionados . . . . .	6
<b>2. Objetivos</b>	<b>8</b>
2.1. Objetivo General . . . . .	8
2.2. Objetivos Específicos . . . . .	8
2.3. Alcances . . . . .	8
<b>3. Marco Teórico</b>	<b>10</b>
3.1. Modelo de Markov Oculto . . . . .	10
3.1.1. Verosimilitud . . . . .	11
3.1.2. Pertenencia a un Estado . . . . .	11
3.2. Aplicación a Quiebres de Stock . . . . .	12
3.2.1. Consideraciones Generales . . . . .	12
3.3. Estimación Jerárquica Bayesiana . . . . .	15
3.4. Métricas de Desempeño . . . . .	16
<b>4. Clientes Leales</b>	<b>18</b>
4.1. Nivel de Incidencia de Productos . . . . .	19
4.2. Comentarios . . . . .	19
<b>5. Datos</b>	<b>23</b>
5.1. Datos disponibles . . . . .	23
5.1.1. Datos transaccionales . . . . .	23
5.1.2. Auditoría . . . . .	24
5.2. Análisis Exploratorio . . . . .	24
5.2.1. Mediciones de Quiebre . . . . .	27
5.2.2. Segmentos de Clientes . . . . .	27
5.2.3. Días sin Incidencias y Lealtad . . . . .	32
5.2.4. Frecuencia y Amplitud de Compra . . . . .	34
<b>6. Análisis de Resultados</b>	<b>37</b>
6.1. Resultados Iniciales . . . . .	38

6.1.1.	Definición de Modelos a Comparar . . . . .	38
6.1.2.	Análisis de Sensibilidad . . . . .	39
6.2.	Segmentación en base a Nivel de Compra Total . . . . .	41
6.3.	Enfoque Híbrido para Estimación de Modelos . . . . .	42
6.4.	Enfoque de Lealtad en base a Frecuencia y Amplitud . . . . .	43
6.5.	Elección Final de Productos . . . . .	43
6.6.	Robustez de los resultados . . . . .	45
6.7.	Discusión . . . . .	46
6.7.1.	Desempeño de los Modelos . . . . .	47
<b>7.</b>	<b>Conclusiones</b>	<b>49</b>
7.1.	Conclusiones . . . . .	49
7.2.	Trabajo Futuro . . . . .	50
	<b>Bibliografía</b>	<b>52</b>
	<b>A. Grupos de Incidencia por Producto</b>	<b>53</b>
	<b>B. Modelos por Producto para todos los productos</b>	<b>54</b>
	<b>C. Análisis de Resultados para Niveles de Lealtad</b>	<b>58</b>
C.1.	Predicción de los Modelos . . . . .	59
	<b>D. Elección de Modelo Híbrido a utilizar</b>	<b>61</b>
	<b>E. Modelos por Producto para Productos Finales</b>	<b>63</b>
	<b>F. Resultados por Producto del Modelo al 15 % de Lealtad</b>	<b>68</b>

# Índice de Tablas

5.1. Cantidad de días válidos por local . . . . .	24
5.2. Información de Productos en estudio . . . . .	25
5.3. Incidencias de cada producto por local . . . . .	26
5.4. Cantidad de clientes con más de 10 boletas totales según nivel de lealtad para cada producto . . . . .	31
5.5. Lealtad máxima para cada producto . . . . .	33
5.6. Descriptivos en base a frecuencia de compra para los productos sobre los 469 días en estudio . . . . .	35
6.1. Desempeño Modelo Completo para 58 productos . . . . .	38
6.2. Desempeño Modelo con ID para 58 productos . . . . .	39
6.3. Desempeño Modelo al 0 % de Lealtad para 58 productos . . . . .	39
6.4. Desempeño Modelo Híbrido al 1 % de Lealtad para 58 productos . . . . .	42
6.5. Desempeño de los modelos propuestos . . . . .	43
6.6. Desempeño Modelo Completo para Productos Finales . . . . .	44
6.7. Desempeño Modelo con ID para Productos Finales . . . . .	44
6.8. Desempeño Modelo Híbrido al 1 % de Lealtad para Productos Finales . . . . .	44
6.9. Desempeño Modelo al 0 % de Lealtad para Productos Finales . . . . .	44
6.10. Resultados de los submodelos para estimación de Modelo Completo . . . . .	46
A.1. Segmentación de productos según nivel de incidencia . . . . .	53
B.1. Resultado del Modelo Completo para los 58 productos . . . . .	55
B.2. Resultado del Modelo con ID para los 58 productos . . . . .	56
B.3. Resultados del Modelo Híbrido al 1 % de Lealtad para los 58 productos . . . . .	57
C.1. Desempeño Modelo con ID en comparación a Modelo Completo . . . . .	58
C.2. Desempeño del Modelo con ID por producto . . . . .	59
C.3. Resumen con desempeño de modelos estimados . . . . .	59
D.1. Desempeño de Modelos Híbridos . . . . .	61
D.2. Desempeño por Producto para el Modelo Híbrido al 1 % de Lealtad . . . . .	62
E.1. Resultados del Modelo Completo para 47 productos finales . . . . .	64
E.2. Resultados del Modelo con ID para 47 productos finales . . . . .	65
E.3. Resultados del Modelo Híbrido al 1 % de Lealtad para 47 productos finales . . . . .	66
E.4. Resultados del Modelo al 0 % de Lealtad para 47 productos finales . . . . .	67

F.1. Resultados del Modelo al 15% de Lealtad para 14 productos . . . . . 68

# Índice de Ilustraciones

1.1.	Crecimiento de ventas en Supermercados . . . . .	2
1.2.	Quiebres de stock para distintas categorías (Bosch, et al., 2005) . . . . .	3
1.3.	Causas de los quiebres de stock (Gruen et al., 2002) . . . . .	4
1.4.	Consecuencias para el cliente de un quiebre de stock (Gruen et al., 2002) . .	5
3.1.	Matriz de Confusión para evaluar el desempeño del modelo . . . . .	17
4.1.	Incidencia media de productos (González (2016)) . . . . .	20
5.1.	Incidencia promedio para cada producto . . . . .	27
5.2.	Distribución de las observaciones medidas en góndola por producto . . . . .	28
5.3.	Distribución de clientes según boletas totales . . . . .	29
5.4.	Cantidad de clientes según nivel de lealtad para cada producto . . . . .	30
5.5.	Cantidad de clientes con Incidencias totales mayores a 10 según nivel de lealtad para cada producto . . . . .	30
5.6.	Cantidad de clientes según nivel de lealtad para cada categoría . . . . .	31
5.7.	Porcentaje de clientes según nivel de lealtad para cada categoría . . . . .	31
5.8.	Porcentaje de días sin incidencias promedio, para distintos niveles de lealtad	32
5.9.	Porcentaje de días sin incidencias por producto, para distintos niveles de lealtad	33
5.10.	Distribución de clientes según períodos de compra para distintas amplitudes de compra . . . . .	36
6.1.	Cambio en el desempeño del modelo para distintos cortes en la regla de elección del Estado 0 . . . . .	40
6.2.	Comparación de desempeños del Modelo con ID para distintos segmentos de clientes . . . . .	41
6.3.	Comparación de desempeños del Modelo al 5% de Lealtad para distintos seg- mentos de clientes . . . . .	41
6.4.	Comparación de Falsas Alarmas entre modelos para 14 productos . . . . .	47
6.5.	Comparación de Error Tipo I entre modelos para 14 productos . . . . .	47



# Capítulo 1

## Introducción

La cadena de suministros juega un papel fundamental en el funcionamiento de los supermercados, quienes se enfocan en diseñar herramientas que les permitan responder de mejor manera a preguntas como: qué productos pedir y en qué cantidades, qué nivel de inventario mantener o qué rutas utilizar para la compra de productos. La elección del local de compra por parte del cliente se basa, principalmente, en la cercanía del local y en el mix de productos que éste tiene para ofrecer. Así, la disponibilidad de los productos juega un papel clave para las grandes cadenas de supermercados quienes, al ofrecer miles de productos diariamente hacen de esta una tarea compleja.

Para que un producto se encuentre disponible en góndola para el cliente deben llevarse a cabo una serie de procesos, que involucran desde el pedido al proveedor (cuánto y con qué frecuencia) hasta la ubicación de los productos en góndola por parte del supermercado. Para describir la situación en la que un cliente no es capaz de encontrar el producto que desea comprar se define el concepto de Quiebre de Stock (OOS, por sus siglas en inglés “Out of Stock”). Este concepto ha adquirido gran importancia en los últimos años, y ha motivado constantes mejoras por parte de los retailers y distribuidores. Las principales razones asociadas a este comportamiento son:

- Por una parte, los consumidores se están volviendo menos tolerantes ante una situación de OOS. Cada vez existe mayor información disponible y más canales a través de los cuales comprar productos, ya sea en tiendas físicas o virtuales. Esto hace más fácil para el cliente comprar en un supermercado diferente el producto que busca.
- Existen nuevas tecnologías que entregan novedosas formas de solucionar los problemas asociados a los OOS. Esto deja de lado las soluciones tradicionales que utilizan los retailers, como aumentar el personal o tener un mayor nivel de stock de seguridad en los sistemas de inventario.

En la actualidad, existen distintas tecnologías que intentan mejorar los niveles de disponibilidad de los productos (de las cuales se hablará más adelante en este capítulo). Sin embargo, estos avances en las tecnologías no han sido capaces de adaptarse a la complejidad del proceso, como lo son promociones cada vez más diversas y el aumento continuo de los SKUs disponibles en los supermercados. Estos factores hacen del problema materia de estudio

constante, y motiva a buscar soluciones desde múltiples perspectivas.

## 1.1. Antecedentes Generales

La definición de retail corresponde a toda actividad que está directamente relacionada con la venta de bienes y servicios para el consumidor final para uso personal, no comercial (Lamb et al., 2009). En Chile, la industria del retail representa un sector productivo de gran envergadura, generando alrededor de US\$61.500 millones en venta el año 2017 (22.2% del PIB nacional) (Estrategia, 2018).

Los retailers pueden clasificarse, a grandes rasgos, en Tiendas de especialidad (como tiendas de zapatos y marcas exclusivas), Supermercados (incluyendo hipermercados), Farmacias, Tiendas de conveniencia (supermercados de pequeña envergadura), Tiendas de descuento, Restaurantes y en Retail no presencial (e-commerce y automatic vending) (Von Hausen, 2014).

Los supermercados son el canal de distribución más importante del país, representando el 23.9% del total de las ventas del retail en 2017 (Retail, 2018). Como se aprecia en la Figura 1.1, esta industria ha tenido un desarrollo sostenido los últimos años, y a pesar de una reciente baja, alcanzó los US\$14.690 millones de ingresos por ventas en 2017. Existen cuatro grandes cadenas de supermercados, quienes para el año 2016 poseían en conjunto más del 99% de la cuota de mercado, según las memorias anuales de cada una (información de libre acceso más actualizada a la fecha). La primera es Walmart Chile con un 41.1% (Feller Rate, 2016), seguida de Cencosud con un 28.1% (Cencosud, 2017), SMU con un 23.3% (SMU, 2017) y finalmente Tottus con un 7.5% (Tottus, 2017).

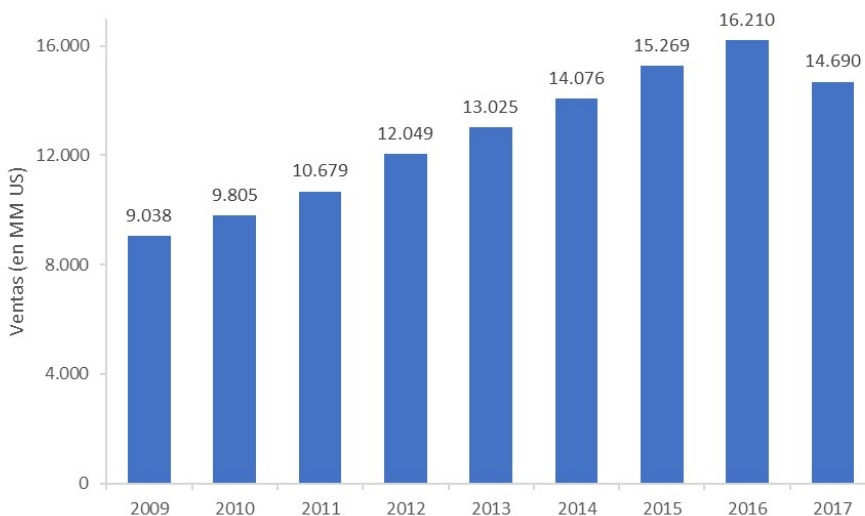


Figura 1.1: Crecimiento de ventas en Supermercados

## 1.2. Quiebres de Stock

Se define el evento de quiebre de stock cuando el retailer pretende tener un ítem para la venta, pero no existe presencia física de ninguna unidad en la góndola. Este evento comienza cuando la última unidad de un SKU es removida de la góndola y termina cuando alguna unidad vendible es repuesta en góndola (Gruen and Corsten, 2007). Sin embargo, existe una definición alternativa que considera un quiebre de stock cuando el consumidor no es capaz de encontrar el producto que busca, ya sea porque efectivamente no está disponible o por la presencia de otros factores que dificultan su visibilidad, tales como exhibiciones especiales de otro producto, superposición de productos, entre otros (Gruen et al., 2002).

Para medir el nivel de quiebre de stock en un supermercado se define el Faltante de Mercadería en Góndola (en adelante FMG) como sigue (Bosch et al., 2005):

$$\text{FMG} = \frac{\text{N}^\circ \text{SKUs en OOS}}{\text{N}^\circ \text{SKUs Totales}}$$

Se estima el promedio de FMG a nivel mundial en un 8.3% (Gruen et al., 2002). En Chile, la Universidad de Chile realiza el Primer Estudio Nacional de FMG en 2004, arrojando un estimado de 15%, muy por encima del promedio (Bosch et al., 2005). Al observar la Figura 1.2, las categorías de Aseo Personal (23.87%) junto con las de Golosinas y Confites (21.12%) son las que tienen mayor FMG, en contraste con las bebidas alcohólicas, que poseen solo un 10.63%.

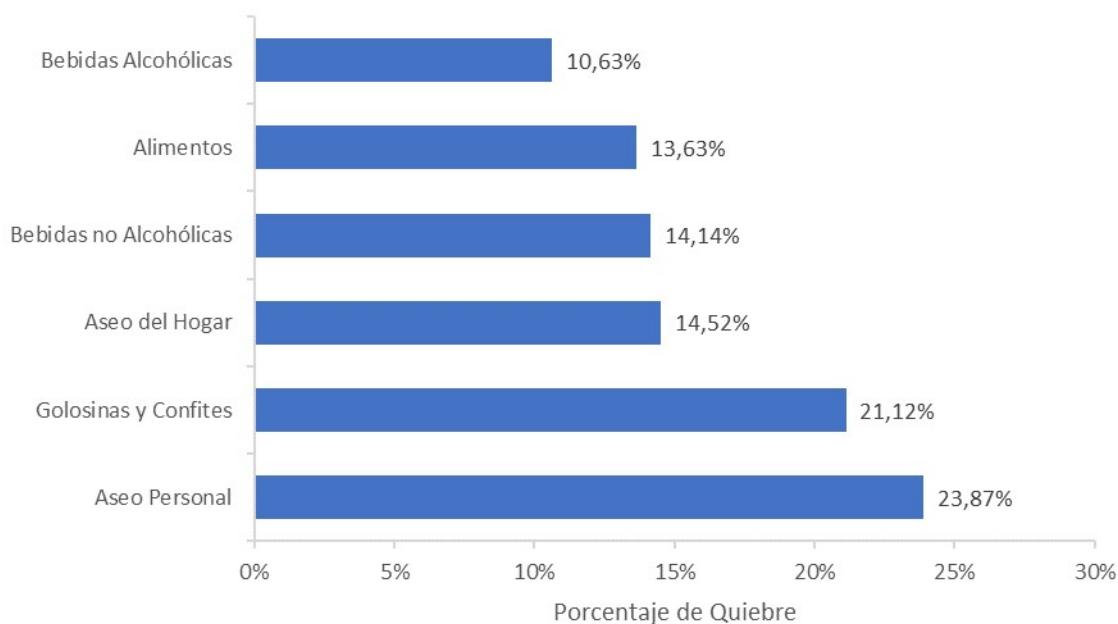


Figura 1.2: Quiebres de stock para distintas categorías (Bosch, et al., 2005)

### 1.2.1. Causas y Consecuencias

Los quiebres de stock pueden ser causados por diversos motivos. Al decidir los niveles de inventario, se busca encontrar un nivel de servicio óptimo evaluando los costos de satisfacer toda la demanda por productos (reflejado en un aumento en el stock de productos, generando excedentes y costos adicionales de bodega) en comparación a no hacerlo (minimizando los costos de bodega, pero no logrando satisfacer la totalidad de la demanda). Si bien los quiebres de stock están asociados al proceso de reposición a nivel de góndola, esta reposición no puede llevarse a cabo si no existe el stock suficiente en inventario para realizar dicha reposición.

Existe, por otro lado, un factor de incertidumbre que es inherente al proceso relacionado a la estimación de demanda por productos, ya que hay un sinnúmero de variables que definen sus niveles y que hacen que los modelos predictivos no puedan describir perfectamente su comportamiento. Por último, los quiebres de stock pueden ser causados por una mala gestión en la cadena de suministros. El manejo de inventario es complejo para la cadena en estudio, considerando que el retailer ofrece más de 9.000 SKUs distintos. Como se observa en la Figura 1.3, las principales razones que causan los quiebres de stock son la subestimación de la demanda hacia el proveedor y un mal sistema de reposición, causando más de un 70 % de los quiebres. A estas causas se les denominan problemas dentro de la tienda (Gatica, 2014). El porcentaje restante se atribuye a problemas río arriba, es decir, problemas relacionados con los proveedores que se asumen extrínsecos.

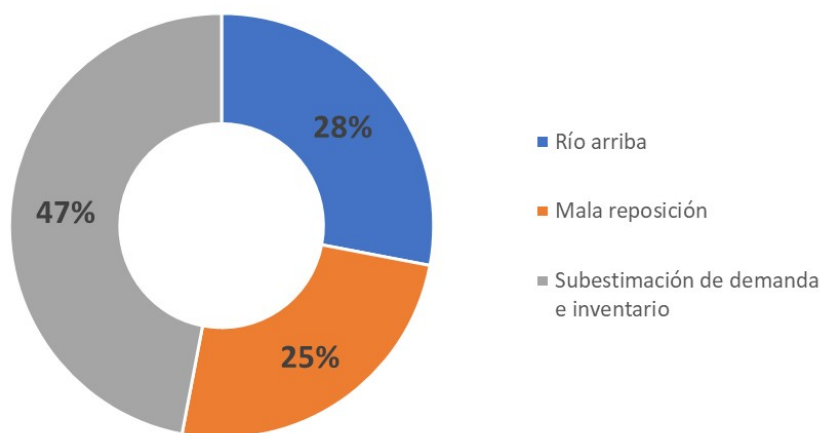


Figura 1.3: Causas de los quiebres de stock (Gruen et al., 2002)

La no disponibilidad de los productos en góndola genera un impacto en el comportamiento de compra del consumidor. Como se observa en la Figura 1.4, las principales reacciones son sustituir el formato del producto (sabor, tamaño, etc) pero manteniendo la marca, sustituir la marca, comprar el producto en otro supermercado, retrasar la compra o finalmente no volver a comprar el producto en el supermercado. Gruen et al. (2002) sugieren que el costo económico de los quiebres de stock en las ventas de un supermercado es de un 4 %. Por otro lado, Gatica (2014) estima este costo en un 3 %. Considerando los ingresos percibidos en 2016 por la industria, los costos alcanzarían entre US\$486.3 millones y US\$648.4 millones al año

en ventas perdidas.

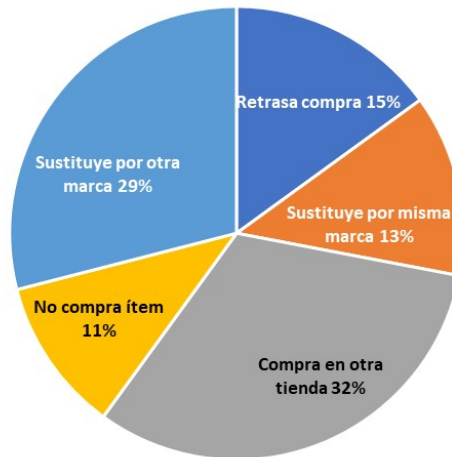


Figura 1.4: Consecuencias para el cliente de un quiebre de stock (Gruen et al., 2002)

### 1.2.2. Soluciones

En la actualidad, y dada la importancia que reviste tener un control eficiente del inventario en góndola, los retailers utilizan distintas técnicas que los ayudan en este propósito. A continuación, se mencionan algunas de ellas:

1. Inspección visual: Se asigna un trabajador que inspeccione periódicamente el estado de góndola (manualmente). Este método es ineficiente, principalmente por los costos asociados al personal encargado de hacer las revisiones y por la limitada cantidad de productos que pueden supervisar.
2. Identificación por Radiofrecuencia (RFID): Se dispone un identificador remoto a cada producto, entregando información básica del estado del producto (georeferencia y SKU). Sus altos costos de implementación, a pesar de su alta eficiencia, obliga a que las cadenas solo lo utilicen en los productos más importantes o a mayores niveles de agregación (por ejemplo, en pallets de productos).
3. Cámaras en góndolas: Permite un monitoreo a tiempo real, pero requiere constante control por parte de un trabajador. Sumado a esto, no entrega visibilidad perfecta debido principalmente a limitaciones técnicas.
4. Robots: Actualmente se están desarrollando tecnologías que permiten monitorear el estado de las góndolas de forma automática. El desarrollo del robot Zippedi permite identificar con gran precisión quiebres de stock, el surtido de productos en góndola y a la vez realiza comprobaciones de precios de los productos (Olivares and Schilkrut, 2018).

Con todos estos avances tecnológicos, la detección de quiebres de stock sigue siendo un problema sin una solución que sea eficiente y a su vez económicamente rentable. La mayoría de las soluciones expuestas tiene baja escalabilidad para el total de productos disponibles en un supermercados, no permitiendo extrapolar los resultados a su totalidad (Hardgrave et al., 2006). Así es como se exploran soluciones alternativas, mediante variados modelos matemáticos.

Distintos enfoques se han desarrollado, principalmente utilizando información transaccional en los puntos de venta. Esta información corresponde a los registros de venta de productos en las cajas de los supermercados, al momento en que el cliente debe pagarlos. Los datos transaccionales no tienen la capacidad de describir exactamente lo que ocurre en góndola, pero los distintos niveles de venta de los productos guardan relación con su disponibilidad. Por la naturaleza de la transacción, esta se registra de forma automática, lo que permite utilizarla con bastante facilidad. Esta elección tiene varias ventajas: en primer lugar, tiene relativamente bajo costo de implementación, en comparación con las soluciones usuales. Sumado a esto, tiene una alta escalabilidad debido a la naturaleza de la información, pues permite una sincronización con información de inventario o distribución, por ejemplo.

### 1.3. Trabajos Relacionados

En el año 2014, el Centro de Estudio del Retail (CERET) realiza una auditoría a la cadena estudiada, con el objetivo de evaluar el sistema actual y desarrollar un nuevo modelo de predicción mediante Límites de Control (CERET, 2014). Para esto, se realizan mediciones diarias del real estado de la góndola, durante 14 días. Un primer resultado arroja un FMG de 13.9 %, bastante similar a los resultados del estudio realizado por Bosch et al. (2005).

En general, se presentan en la literatura dos familias de modelos utilizados para la detección de quiebres, los *modelos supervisados* y los *no supervisados*. El primer grupo corresponde a modelos donde, al menos en su período de calibración (o bien durante todo lo que dure el estudio), se conoce el estado real de la góndola (por alguno de los métodos de revisión descritos en la sección anterior). Figueroa (2009) desarrolla un modelo de clasificación supervisado de predicción de quiebres de stock mediante árboles de decisión para distintas categorías, logrando detectar los quiebres de stock en un 56 % de los casos. Sin embargo, en este trabajo no se consideran los efectos de la estacionalidad en las ventas para discernir si un producto se encuentra disponible o no en la góndola.

Por otra parte, en los modelos no supervisados no se conoce el estado de la góndola, por lo que mediante las características de las observaciones se generan clasificaciones. Chen (2014) propone un Modelo de Markov Oculto para determinar problemas relacionados con la ejecución de la reposición de inventario y con los quiebres ocasionados por la *demanda fantasma* (termino acuñado por Fisher and Raman (2010)) y que se describe como “desajustes en el control de inventario, causados por robo y por el deterioro de los productos, provocando que el sistema contabilice un mayor stock que el real, congelando la reposición” (Kang and Gershwin, 2005)).

Bhan (2015) utiliza Modelos de Markov Oculto para estimar la evolución en los niveles de inventario en una tienda de retail. Para esto, intenta determinar el momento en que un quiebre de stock ocurre utilizando la información de punto de venta como una serie de tiempo, lo que se asemeja al enfoque utilizado en el presente Trabajo. Luego, estima su modelo mediante la técnica de muestreo Gibb sampling.

Arriagada (2018) hace uso de distintos métodos estadísticos para predecir quiebres de stock en una cadena de supermercados. Dentro de sus hipótesis, realiza una segmentación de los clientes poseedores de la tarjeta de lealtad del supermercado según preferencias y comportamientos de compra de una forma similar a la planteada en el presente Trabajo. Con esto, propone mejorar los tiempos de detección de un modelo de control estadístico de proceso, y logra mejorar los indicadores de desempeño. En un trabajo reciente, Montoya and González (2018) hacen uso de un Modelo de Markov Oculto para identificar los quiebres de stock en la cadena en estudio. En su trabajo, estiman los modelos utilizando la información completa de todos los clientes, logrando predecir más del 60% de los quiebres ocurridos.

Estos buenos resultados sirven de base para el desarrollo del presente Trabajo, sin embargo, este último se centrará en segmentar a los clientes, entregando nociones de lealtad y estudiando la capacidad que estas segmentaciones tengan para mejorar los niveles de predicción de quiebres. Se espera que el presente Trabajo aporte al entendimiento del papel que juegan los clientes y su comportamiento de compra en la determinación del estado de disponibilidad de los productos en góndola.

# Capítulo 2

## Objetivos

### 2.1. Objetivo General

Diseñar un método de identificación de quiebres de stock en góndola utilizando Modelos de Markov Ocultos incorporando la información parcial de clientes leales a cada producto.

### 2.2. Objetivos Específicos

- Construir un modelo de identificación para cada producto, limitado a la información de los clientes leales, para distintos niveles de lealtad.
- Encontrar el nivel de lealtad que mejor se ajuste a cada producto.
- Evaluar el desempeño de los modelos para cada producto en comparación al modelo completo, y discernir sobre el valor de los modelos segmentados.

### 2.3. Alcances

- El presente Trabajo de Título utiliza la información recolectada durante la auditoría realizada en el año 2014 a una cadena de supermercados presente en Chile junto con información transaccional de 10 salas de la cadena, comprendida entre el 2 de febrero de 2012 y el 1 de junio de 2014.
- Se cuenta con una medición del real estado de la góndola para 58 productos los últimos 14 días de estudio. Con esta información, se testearán los modelos y se validarán los supuestos.
- El resultado esperado del trabajo es un modelo de identificación, que se evaluará en comparación a los resultados obtenidos por Montoya and González (2018).



- Este trabajo no contempla la recolección de información adicional, ya que los datos disponibles se consideran suficientes para su desarrollo.
- Por la naturaleza de la información y por el enfoque postulado el estudio se limitará al análisis de la actividad en góndola y en las cajas. Se asumirá en todo momento la existencia de stock en bodega, y no se analizarán problemas relacionados a inventario.

# Capítulo 3

## Marco Teórico

### 3.1. Modelo de Markov Oculto

Los Modelos de Markov Ocultos son una forma particular dentro de los procesos Markovianos. Estos últimos son procesos estocásticos caracterizados por tener la propiedad de Markov, esto es, que la probabilidad de que ocurra un evento depende exclusivamente del evento inmediatamente anterior. Para el caso de los Modelos de Markov Ocultos, los estados son no observables (puede pensarse que son desconocidos), por lo que requieren ser observados indirectamente y a través a algún parámetro que sea observable.

Las variables necesarias para definir un Modelo de Markov Oculto son (Zucchini and MacDonald, 2009):

- Estados: Variable oculta del modelo y cuyas probabilidades se desean obtener. Se identifica como  $S_t$ , donde  $t$  es el período en que se analiza la variable. Se define la colección  $\mathbb{S} = (S_i)_{i=1}^T$  como el conjunto de estados posibles, donde  $T$  es el último período en estudio.
- Observaciones: Variable observable del modelo. Se identifica como  $n_t$ , donde  $t$  es el período donde se observa la variable.  $\mathbb{N} = (n_i)_{i=1}^T$  se define como el conjunto de observaciones, que puede ser un conjunto discreto o continuo.
- Probabilidades Iniciales: Probabilidad de que una secuencia de observaciones comience en cierto estado. Se identifica como  $\Pi = (\pi_s)_{s \in \mathbb{S}}$ , con  $\sum_{i=1}^T \pi_{s_i} = 1$ .
- Matriz de Transición: Matriz cuadrada  $Q$  de dimensiones  $\text{card}(\mathbb{S})^2$ , donde cada componente  $q_{ij}$  indica la probabilidad de pasar del estado  $i$  al estado  $j$ . Esta matriz puede ser estática (es decir,  $Q(t) = Q \forall t$ ) o bien dinámica, con  $Q = Q_t$  variable en el tiempo.
- Probabilidad de Ocurrencia: Probabilidad de realización de una observación, dada la pertenencia a un estado  $s \in \mathbb{S}$ . Se define como  $f(n_t | S_t = s) \doteq f_s(n_t)$ .

### 3.1.1. Verosimilitud

Es estadísticas, se define la verosimilitud de un parámetro  $\theta$  sobre una muestra  $(x_i)_{i=1}^n$  como sigue:

$$L(\theta) = L(\theta|x_1, \dots, x_n) \doteq \prod_{i=1}^n f_{\theta}(x_i)$$

Esto se interpreta como la probabilidad de observar las realizaciones de la muestra, dada la distribución de probabilidad  $f$  con parámetro  $\theta$ . En otras palabras, mientras mayor sea la verosimilitud de una distribución de probabilidad encontrada, mejor se ajusta a los datos disponibles. En el contexto de Cadenas de Markov Ocultas, se define la matriz de probabilidad de ocurrencia como la siguiente matriz diagonal:

$$M(x_t) \doteq \begin{pmatrix} f_1(x_t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_S(x_t) \end{pmatrix}$$

Con esto, se define la verosimilitud del Modelo como sigue (Zucchini and MacDonald, 2009):

$$L_T \doteq \Pi' M(x_1) Q M(x_2) \cdots Q M(x_T) 1$$

Si bien esta fórmula entrega una función de verosimilitud, en la práctica se trabaja con su logaritmo natural, para evitar problemas de convergencia al tratar con un gran número de observaciones. Realizar la transformación requiere de un tratamiento previo de los datos, hasta conseguir lo que se conoce como LogVerosimilitud Reescalada (Zucchini and MacDonald, 2009). Los pasos se detallan a continuación:

1. Se define  $L_1 \doteq \Pi' M(x_1)$  y  $\Delta_1 \doteq \frac{L_1 1}{\mathbb{S}}$ , con lo que la nueva verosimilitud resulta  $L_1^* = \frac{L_1}{\Delta_1}$ . Sumado a esto, se define  $LL_{temp} = \ln \Delta_1$ .
2. Se define  $\forall t \in \{2, \dots, T-1\}$   $L_t = L_{t-1} Q M(x_t)$  y  $\Delta_t = \frac{L_t 1}{\mathbb{S}}$ , con lo que la nueva verosimilitud resulta  $L_t^* = \frac{L_t}{\Delta_t}$ . Sumado a esto, se actualiza  $LL_{temp} = LL_{temp} + \ln \Delta_t$ .
3. Se define finalmente la LogVerosimilitud como sigue:  $LL_{final} = LL_{temp} + \ln L_T^* 1$ . Esto corresponde a la suma de los logaritmos de las verosimilitudes de cada estado.

### 3.1.2. Pertenencia a un Estado

Calcular la probabilidad pertenencia de una observación de POS a cada estado es fundamental para el desarrollo del modelo. Con este fin, se define la verosimilitud hasta el período  $t$ :

$$L'_t \doteq \Pi' M(x_1) Q M(x_2) \cdots Q M(x_t) = [L'_{t1} \cdots L'_{tS}]$$

Donde  $L'_{ts}$  corresponde a la verosimilitud de pertenecer al estado  $s \in \mathbb{S}$  en el período  $t$ . Para facilitar la interpretación del conjunto de verosimilitudes como probabilidades de pertenencia al estado, se realiza la siguiente distinción:

$$P_t = \frac{L'_t}{\sum_{i=1}^S L'_{ti}} = [P_{t1} \dots P_{tS}]$$

Donde  $P_{ts}$  se puede interpretar como la probabilidad de estar en el estado  $s$  en el período  $t$ . Finalmente, en cada período  $t$  se escoge el estado que tenga asociado la mayor probabilidad  $P_{ts}$ .

## 3.2. Aplicación a Quiebres de Stock

Los Modelos de Markov Oculto introducidos en la Sección 3.1 tienen variadas aplicaciones según sus especificaciones. En esta sección se describe la topología de estos modelos en el marco del estudio de Detección de Quiebres de Stock. En particular, se describen sus principales componentes y parámetros, junto con las técnicas para estimarlos.

### 3.2.1. Consideraciones Generales

#### Estados y Observaciones

Los *estados* se definen como los posibles niveles de demanda por un producto. La mejor identificación se realiza al existir tres estados, los que se denominarán estado de Demanda Alta, Demanda Media y Quiebre (llamado en adelante Estado 0) (González, 2016).

Por otra parte, las *observaciones* se definen como las variables observables del modelo. Se identifican como las boletas que se transan durante un día, donde  $n_{ijt}$  representa el número de boletas observadas para el producto  $i$  en el local  $j$  durante el día  $t$ .

#### Probabilidades Iniciales

La pertenencia a cada uno de los estados al inicio de la medición se puede establecer de antemano o bien puede ser parte de los parámetros a estimar. La literatura sugiere que los resultados entregados por las diferentes especificaciones no varía de forma significativa (Montoya and González, 2018), por lo tanto la pertenencia inicial se asume equiprobable, es decir,  $\pi_{ij} = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$ .

## Probabilidades de Ocurrencia

Para modelar la probabilidad de ocurrencia de observación de boletas para el día  $t$  y el estado  $s$ , se asume una Distribución Binomial con parámetros  $N_{jt}$  y  $p_{ijts}$ , lo que se traduce en la siguiente estructura para la función de probabilidad:

$$P_{ijt}(n_{ijt}|X_{ijt} = s, z_{ijt}) = \binom{N_{jt}}{n_{ijt}} p_{ijts}^{n_{ijt}} (1 - p_{ijts})^{N_{jt} - n_{ijt}}$$

Donde  $N_{jt}$  corresponde a las boletas totales observadas en el local  $j$  durante el período  $t$ ,  $X_{ijt}$  corresponde al estado en que se encuentra el producto  $i$  en el local  $j$  y durante el período  $t$ ,  $z_{ijt}$  corresponde a una colección de variables observables que afectan la demanda, y  $p_{ijts}$  corresponde a la probabilidad de que una boleta del producto  $i$  en el local  $j$  y en el período  $t$  contenga al producto estudiado, condicionado a que el sistema se encuentre en el estado  $s \in \mathbb{S}$ .

## Probabilidad de Observación de un Producto

La estimación de esta probabilidad se realiza agregando heterogeneidad observable obtenida de los datos, de la siguiente manera:

$$p_{ijts} = \frac{1}{1 + \exp(-(\hat{\beta}_{0,ij s} - \exp(\beta_{\text{precio},ij s}) * \text{Precio}_{ijt}^* + z_{ijt}\beta_{ij s}))}$$

La variable  $\text{Precio}_{ijt}^*$  es una transformación del precio original. Con el objetivo de que las variaciones de los precios se hagan comparables con las demás variables y que solo se perciban aumentos o disminuciones en su valor, se realiza el procedimiento llamado mean center, el cual consiste en calcular el precio promedio por producto para cada local para luego restarlo del valor de cada observación. Sumado a esto, se divide el resultado por 1000. Esto es:

$$\text{Precio}_{ijt}^* = \frac{\text{Precio}_{ijt} - \text{Precio}_{i,t}}{1000}$$

Sumado a esto, y para adaptarse a la intuición del problema, se realiza una transformación que tiene por objetivo imponer un orden no decreciente en las probabilidades de observación de boletas en cada uno de los estados ( $p_{ijt1} \leq p_{ijt2} \leq p_{ijt3}$ ). Para esto, sean  $\beta_{0,ij1}$ ,  $\beta_{0,ij2}$  y  $\beta_{0,ij3}$  los interceptos estimados para los tres estados definidos por el modelo. Los parámetros definitivos asociados a cada estado quedan definidos por la siguiente recurrencia:

$$\begin{aligned}\hat{\beta}_{0,ij1} &= \beta_{0,ij1} \\ \hat{\beta}_{0,ij2} &= \hat{\beta}_{0,ij1} + e^{\beta_{0,ij2}} \\ \hat{\beta}_{0,ij3} &= \hat{\beta}_{0,ij2} + e^{\beta_{0,ij3}}\end{aligned}$$

Con esta transformación se facilita la interpretación de lo que representa cada estado. Como se tiene una desigualdad estricta entre la probabilidad de observación de cada estado, esto induce un mayor nivel de ventas mientras se avanza en los estados posibles. Así, se fija el estado de quiebre como el primer estado del modelo.

## Matriz de Transición

La matriz  $Q$  describe las transiciones entre los estados después del primer período. Sea  $X_{ijt} \in \{1, \dots, S\}$  el nivel de demanda por el producto  $i$  en el local  $j$  para el día  $t$ . Luego, cada elemento de la matriz se puede escribir como:

$$q_{ijt}^{s's} = P(X_{ijt} = s | X_{ijt-1} = s', \text{precio}_{ijt})$$

Donde  $q_{ijt}^{s's} \geq 0$ ,  $\sum_{s=1}^S q_{ijt}^{s's} = 1$  y  $\text{precio}_{ijt}$  corresponde al precio del producto  $i$  en el local  $j$  para el día  $t$  (luego de aplicar la transformación de la variable). Esta estimación requiere una parametrización adicional, ya que se exige que la suma de las componentes de cada fila sume 1. Por esto, se estiman  $S - 1$  parámetros por fila (2 parámetros para efectos del presente trabajo), lo que resulta en 6 parámetros por determinar. La parametrización utilizada es la siguiente (simplificando para el caso en estudio):

$$q_{ijt}^{s's} = \frac{e^{\gamma_{ij}^{s's}}}{1 + e^{\gamma_{i1}^{s'1}} + e^{\gamma_{i2}^{s'2}}} \quad s < 3$$

$$q_{ijt}^{s3} = 1 - q_{ijt}^{s2} - q_{ijt}^{s1}$$

Donde  $\gamma_{ij}$  son los parámetros que el modelo debe estimar. Esto es válido para cada fila  $i \in \{1, 2, 3\}$ . Con esta transformación se garantiza que las componentes tomen valores entre 0 y 1, y que sumen 1.

## Estado 0 o de Quiebre

Como se explicó anteriormente, para la construcción del modelo se restringe un estado para que represente la condición de quiebre de stock (este estado se denota por  $s = 0$ ). La intuición detrás de esto es que se espera que un quiebre de stock genere un gran impacto en el nivel de venta de los productos, disminuyendo sus niveles de venta normales. Sin embargo, debido a que es posible observar ventas del producto a pesar de que esté en estado de quiebre, la probabilidad de observar un producto no es necesariamente cero. Así,

$$P(n_{ijt} | X_{ijt} = 0, z_{ijt}) = \binom{N_{jt}}{n_{ijt}} \varepsilon^{n_{ijt}} (1 - \varepsilon)^{N_{jt} - n_{ijt}}$$

Donde  $\varepsilon = 10^{-5}$  se define como la probabilidad de observación dado que el producto  $i$  en el local  $j$  y en el período  $t$  se encuentra en estado de quiebre. Para determinar este valor, González (2016) obtiene el desempeño promedio del modelo para los valores  $10^{-4}$ ,  $10^{-5}$  y  $10^{-6}$  (las métricas de desempeño se describen más adelante en este capítulo). No se consideran valores de probabilidad mayores, ya que esta no debe ser mayor a las probabilidades estimadas para el resto de los estados, ya que en caso contrario se produciría un intercambio en las etiquetas o interpretación de los estados, lo que afectaría de manera importante a la clasificación del modelo y la interpretación de sus resultados.

Finalmente, se obtiene que para el valor  $10^{-5}$  el modelo tiene el desempeño óptimo, en cuanto minimiza los errores y alcanza altos valores de predicción de quiebres.

### 3.3. Estimación Jerárquica Bayesiana

Para estimar los parámetros asociados a la probabilidad de pertenencia a cada estado en el modelo, un primer enfoque sugiere emplear el método de Máxima Verosimilitud, a través del cual se encuentran los parámetros que con una mayor probabilidad predigan las observaciones condicionadas. Sin embargo, esta metodología no resulta del todo apropiada para el problema en estudio. Las principales causas son:

- La cantidad de datos provenientes de los productos, locales, días y estados posibles hace necesario estimar una enorme cantidad de parámetros, lo que dificulta la resolución del problema de optimización.
- La estimación mediante Máxima Verosimilitud entrega parámetros desagregados y estáticos, lo que dificulta realizar un análisis dinámico efectivo de la información.

Por esta razón, los parámetros se determinan utilizando Estimación Bayesiana. Esta corriente está sustentada en el Teorema de Bayes, que relaciona las creencias iniciales sobre los valores de un parámetro con su refinamiento a través de la información entregada. Formalmente, para un conjunto de observaciones  $x$  y un parámetro  $\theta$ , por determinar, el teorema muestra que:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Tomando en cuenta que  $p(x)$  se mantiene constante al moverse en los parámetros, y que  $p(x|\theta) = L_T(\theta|x)p(\theta)$  (ver Sección 3.1.1), resulta más útil trabajar con la siguiente relación:

$$p(\theta|x) \propto L_T(\theta|x)p(\theta)$$

Sea  $(n_{ijt})_{t=1}^T$  una secuencia de observaciones del producto  $i$  para el local  $j \in \{1, \dots, J\}$ . La función de verosimilitud puede escribirse de la siguiente manera:

$$L_i = \prod_{j=1}^J P(n_{ij1}, \dots, n_{ijT}) = \prod_{j=1}^J \pi'_{ij} M_{ij1} \prod_{t=1}^T Q_{ijt} M_{ijt} 1$$

Es importante notar que todos los productos medidos se encuentran en todos los locales en estudio, por lo que se cuenta con información de cada uno. Sin embargo, cada producto tiene distintos niveles de demanda por parte de los consumidores, y un mismo producto puede tener distinta demanda según tipo de local (tamaño, ubicación, etc.). Estas diferencias en el surtido deben reflejarse en el tratamiento de los datos. Por esta razón, se debe incluir un nivel de heterogeneidad en los parámetros a estimar. Cabe destacar que la información utilizada en la estimación es propia de cada producto, por lo que se estiman modelos independientes para cada uno de ellos. Para hacerlo, se recurre a una estimación con enfoque bayesiano jerárquico, donde se asume que la distribución de los parámetros para cada local se deriva de una distribución poblacional. En particular, se utilizan métodos jerárquicos Bayesianos de Monte Carlo para Cadenas de Markov (MCMC) para estimar estos parámetros.

Específicamente, sea  $\Phi_{ij} = \{\gamma_{ij}^{ss'}, \beta_{0,ij}, \beta_{precio,ij}, \beta_{ij}\} \forall s, s'$  el conjunto de parámetros a estimar para un producto  $i$  en un local  $j$ . Para capturar la heterogeneidad no observable se asume que los parámetros distribuyen normal multivariado, con media  $\Phi_{i0}$  y varianza  $V_{i\Phi}$ . Así,

$$\Phi_{ij} = \Phi_{i0} + \eta_{ij}, \quad \eta_{ij} \sim N(0, V_{i\Phi})$$

Los hiperparámetros  $\Phi_{i0}$  y  $V_{i\Phi}$  distribuyen como normal multivariada e invertida de Wishart, respectivamente. Finalmente, mediante la función de verosimilitud se estima la distribución condicional y la distribución a priori de los parámetros. La distribución condicional no tiene forma cerrada, por lo que se utiliza el algoritmo de Metropolis-Hasting para derivar la distribución a posterior.

El algoritmo de Metropolis-Hasting consiste en generar nuevos candidatos a los parámetros durante cada iteración, y según criterios de aceptación, se actualizan o nos los datos. Naturalmente, a medida que aumentan las iteraciones, mejor se ajusta la estimación a la distribución real de los parámetros. Formalmente, el algoritmo tiene la siguiente secuencia (Yildirim, 2012):

1. Se escoge arbitrariamente un set de parámetros iniciales y una distribución de probabilidad.
2. Se define una función a aplicar a la muestra de parámetros, que entrega los nuevos candidatos a parámetros. Esta función debe ser proporcional a la función de distribución buscada.
3. En cada iteración, se actualizan los parámetros de la distribución.
4. Luego, se genera un nuevo valor a partir del condicionamiento en el valor anterior. Este elemento se denomina candidato.
5. Se evalúa la función en el candidato y se determina una probabilidad de aceptación, en base al siguiente ratio:

$$a = \frac{f(\theta_c)\mathbb{P}(\theta_c|\bar{\theta})}{f(\theta_{t-1})\mathbb{P}(\theta_{t-1}|\bar{\theta})}$$

Siendo  $\theta_{t-1}$  el último valor aceptado,  $\theta_c$  el candidato y  $\bar{\theta}$  la media estimada de la distribución. Así, el candidato se acepta con probabilidad  $a$  (si es mayor a 1, se acepta con probabilidad 1). Para evaluar los candidatos se utiliza la función de verosimilitud.

6. Si se acepta el candidato, se guarda su valor y se actualiza la distribución en la iteración siguiente. En caso contrario, se mantiene el valor anterior.

### 3.4. Métricas de Desempeño

Para medir el desempeño de los modelos estimados se utilizará una Matriz de Confusión. Esta matriz se utiliza para modelos de clasificación tanto supervisados como no supervisados.



		Tienda		Total
		Quiebre	No Quiebre	
Modelo	Alerta	Verdaderos Positivos	Falsos Positivos	P'
	No Alerta	Falsos Negativos	Verdaderos Negativos	N'
Total		P	N	

Figura 3.1: Matriz de Confusión para evaluar el desempeño del modelo

En base a esta matriz, se definen las métricas más importantes utilizadas:

- Poder de Detección: Porcentaje de quiebres de stock correctamente identificados.

$$\frac{\text{Verdaderos positivos}}{P}$$

- Falsas Alarmas: Porcentaje de veces que alertas no eran realmente quiebres de stock.

$$\frac{\text{Falsos positivos}}{P'}$$

- Error Tipo I: Porcentaje de veces que se clasifica como quiebre de stock cuando no lo era.

$$\frac{\text{Verdaderos positivos}}{N}$$

# Capítulo 4

## Cientes Leales

Parte importante del presente Trabajo de Título se centra en analizar el comportamiento de compra de los clientes en los supermercados. Cada cliente tiene preferencias personales a la hora de escoger el mix de productos que compra, lo que puede estar determinado por distintos factores no observables, tales como el número de su grupo familiar, su nivel socioeconómico, si vive en pareja o con sus padres, etc. Distintos tipos de personas poseen características de compra que los diferencian entre sí, y dentro de un mismo tipo de persona existen quienes poseen un comportamiento de compra estable a lo largo del tiempo.

Las personas que compran de forma estable y que además tienden a repetir su canasta con cierta frecuencia, como lo pueden ser la harina para un cocinero o los platos instantáneos para un joven que vive solo, son de especial interés, pues en el comportamiento de estas personas pueden encontrarse potentes indicadores sobre el estado de disponibilidad de un producto. Esta discusión lleva a definir el concepto de “lealtad” hacia un producto, desde el punto de vista de su comportamiento de compra, como sigue:

**Definición 4.1** *Un cliente tiene un nivel de lealtad  $\lambda$  a un producto  $i$  si:*

$$\frac{N^\circ \text{ de compras del cliente que contengan al producto } i}{N^\circ \text{ de compras totales del cliente}} \geq \lambda$$

Donde el cociente se calcula sobre todas las compras realizadas por el cliente durante los 469 días que dura el estudio, en cualquiera de los 10 locales. Es fácil ver que  $\lambda \in [0, 1]$ .

Esta definición entrega un indicador de la frecuencia de compra de un cliente a un SKU en particular, aludiendo a una lealtad a nivel de producto. Se podría utilizar otra medida de lealtad, asociada a la frecuencia de compra de un producto dentro del total de compras en la categoría a la que pertenece (en cuyo caso se hablaría de lealtad a nivel de categoría). Este enfoque ofrece una mirada interesante, ya que se podrían realizar análisis, por ejemplo, para jerarquizar a cada producto dentro de una categoría.

Sin embargo, bajo este enfoque se pierde parte de la noción de lealtad que se pretende entregar. Esto, ya que pueden existir casos de clientes que solo realizando un par de compras

de una categoría tengan un 100 % de lealtad a un producto dentro de ella (basta con que las veces que realicen dichas compras elijan llevar el producto).

## 4.1. Nivel de Incidencia de Productos

En base a las definiciones previas, se pretende entregar una clasificación que asocie la lealtad de un cliente a un producto, esperando que esta ayude en la identificación de los quiebres de stock. Se definen los Niveles de Incidencia de cada producto como sigue:

- Nivel de Incidencia Alta: Productos presentes, en promedio, en más del 0.6 % de todas las boletas en un supermercado.
- Nivel de Incidencia Media: Productos presentes entre el 0.3 % y el 0.6 % de todas las boletas en un supermercado.
- Nivel de Incidencia Baja: Productos presentes en menos del 0.3 % de todas las boletas en un supermercado.

La literatura sugiere que la identificación de quiebres de stock se puede realizar con buenos indicadores de desempeño para productos de Incidencia Alta. Esto se debe a que, al tener mayor presencia en las boletas, es menos probable que pasen tiempos prolongados de tiempo sin ventas. Así, al momento de detectar una ausencia prolongada en sus ventas, se activa la alarma que da aviso de la existencia de un quiebre de stock (que, en general, se activa correctamente). No ocurre de la misma manera para productos de baja incidencia, donde es normal que haya periodos prolongados sin ventas. Sin embargo, como se aprecia en la Figura 4.1, menos de un 2 % de todos los productos disponibles en góndola corresponden a esta categoría, y más del 95 % tiene una incidencia menor al 0.3 %.

## 4.2. Comentarios

Bajo la definición de Cliente Leal ya expuesta, en el caso ideal se pensaría que aumentando el nivel de lealtad, los clientes elegidos tendrían un mayor poder explicativo a la hora de determinar la rotación y disponibilidad de un producto. Esto se explica con el siguiente ejemplo: en un supermercados existe un producto que tiene presencia en el 0.1 % de las boletas. Sin embargo, existe un subconjunto de clientes que al menos el 50 % de las veces que acude al supermercados compra dicho producto. Si un día particular en que no se registran compras del producto se observa que ninguno de estos clientes acudió al supermercados, es difícil concluir si el producto se quebró o simplemente no hubo demanda por parte de los clientes que acudieron al supermercados (ya que estos usualmente compran el producto en menos del 50 % de sus compras totales). Si se observa, en cambio, que ese mismo día acudieron clientes del subconjunto y aún así no se registran compras del producto, es más probable que el producto se encuentre, efectivamente, quebrado.

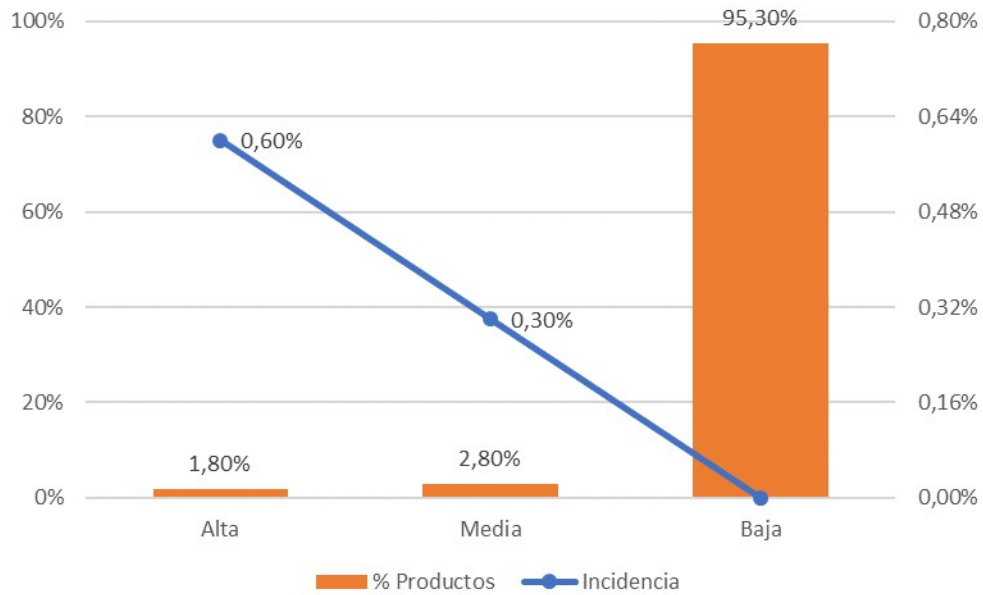


Figura 4.1: Incidencia media de productos (González (2016))

Sin embargo, el efecto real que estos subconjuntos de clientes ejercen sobre el comportamiento de los productos no es fácil de determinar, debido a todas las variables que hay involucradas tanto en las especificaciones del modelo como en los factores intrínsecos de los clientes en cuestión.

A pesar de estas limitaciones, en base al planteamiento del problema y al modelo escogido se pueden realizar ciertas observaciones sobre su posible comportamiento al segmentar a los clientes. Se observa en la Figura 4.1 que la gran mayoría de los productos posee una incidencia media de menos de un 0.5 %, es decir, son productos que aparecen con una frecuencia menor a una de cada 200 boletas. Con esta información, es de esperar que la probabilidad media de observar un producto cualquiera en una boleta oscile entre  $10^{-3}$  y  $10^{-4}$ .

Sin embargo, estas probabilidades cambian al segmentar a los clientes. Si ahora se considera la información de los clientes que contengan, dentro del total de sus compras, al menos un porcentaje  $\lambda \in (0, 1)$  de compras de un producto  $i$  (es decir, si se considera a los clientes con un nivel de lealtad  $\lambda$ ) es esperable que la probabilidad de observación de cada producto aumente. Bajo el mismo argumento, para estos clientes la frecuencia media de aparición de cada producto en las boletas es a lo menos  $\lambda$  (ya que los clientes que tengan menor frecuencia no son considerados). Así, se espera que la probabilidad de observar un producto en una boleta se acerque a este valor, y mientras mayor sea el valor de  $\lambda$ , mayor se espera que sea la probabilidad de observar un producto particular en la boleta.

Los modelos planteados se ajustan de tal forma que puedan decidir cada día si un producto se encuentra o no quebrado según el criterio de máxima verosimilitud (en base a las probabilidades de ocurrencia y a las probabilidades de transición entre estados). Estos modelos toman especial relevancia al analizar días en que no se registran ventas del producto, ya que en estos casos se debe decidir si el producto se encuentra quebrado realmente o si, por otro lado, el producto se encuentra disponible pero ningún cliente decide comprarlo.

Por esta razón, es lógico preguntarse qué ocurre con la verosimilitud de cada estado al segmentar a los clientes. Para que la verosimilitud de pertenecer al estado  $s$  (distinto del Estado 0 o de Quiebre) aumente un día sin ventas del producto, o bien su probabilidad de observar venta cero o bien las probabilidades de transiciones al estado debiesen aumentar. Si la verosimilitud del estado  $s$  aumenta, entonces hay una mayor probabilidad de que al finalizar el día se escoja este estado por sobre el Estado de Quiebre, lo que disminuiría la cantidad total de alertas de quiebres.

La estimación de las probabilidades de transición es compleja y en ella intervienen muchos factores. Sin embargo, se puede analizar con mayor facilidad el comportamiento de la probabilidad de ocurrencia en estos casos. Para un producto  $i$  en un día en particular  $t$  y en un local  $j$ , si no se registran ventas del producto se tiene que:

$$f_s(n_{ijt} = 0) = (1 - p_s)^{N_{jt}}$$

Donde  $p_s$  es la probabilidad de observar al producto en una boleta dado que se encuentra en el estado  $s$  y  $N_{jt}$  la cantidad de boletas totales para el día  $t$  en el local  $j$ . Es importante notar que si  $p_{\hat{s}} > p_s$ , entonces  $f_{\hat{s}}(n_{ijt} = 0) < f_s(n_{ijt} = 0)$ . Es decir, mientras mayor sea la probabilidad de observar al producto en una boleta, menor será la probabilidad de observar cero ventas del producto.

Al segmentar a los clientes que tengan un nivel de lealtad  $\lambda$ , el valor  $N_{jt}$  disminuye en comparación al análisis del total de clientes, ya que se observan menos transacciones en total. Así, para cualquier nivel de lealtad, se tiene que  $N_{jt}^\lambda \leq N_{jt}$ , lo que a su vez implica:

$$(1 - p_s)^{N_{jt}^\lambda} \geq (1 - p_s)^{N_{jt}}$$

Esto para cualquier estado  $s$ . En otras palabras, a medida que aumenta el nivel de lealtad, mayor será la probabilidad de observar cero ventas del producto, independiente del estado en que se encuentre (hasta el momento asumiendo que la probabilidad  $p_s$  se mantiene constante). Una información útil sería saber si al segmentar a los clientes su probabilidad de observar venta cero aumenta o disminuye, pero los resultados descritos parecen contradictorios. Por un lado, esta probabilidad aumenta al disminuir la cantidad de boletas totales observadas, pero por el otro disminuye al crecer la probabilidad  $p_s$  de observar al producto en una boleta cualquiera. Dependiendo los valores que tomen  $N_{jt}$ ,  $N_{jt}^\lambda$ ,  $p_s$  y  $p_{\hat{s}}$  la probabilidad de ocurrencia podría aumentar o disminuir.

A modo de ejemplo, tomando un producto  $i$ , con  $N_{jt} = 5.000$  y  $p_s = 0.15\%$  (valor razonable considerando las incidencias medias descritas en la Figura 4.1), se tiene que:

$$f_s(n_{ijt} = 0) = (1 - p_s)^{N_{jt}} = 0,055\%$$

Si al segmentar a un nivel de lealtad  $\lambda = 0.01$  se obtiene que  $N_{jt}^\lambda = 500$  y  $p_s^\lambda = 1.2\%$  (valor razonable considerando el nivel de lealtad), la nueva probabilidad de ocurrencia sería:

$$f_s^\lambda(n_{ijt} = 0) = 0,24\%$$

Con esto, se tendría que la probabilidad de ocurrencia al segmentar a los clientes aumenta. A pesar de esto, no se puede concluir sobre el cambio de la verosimilitud del estado  $s$ , ya que no se sabe con certeza lo que pueda ocurrir con las variaciones en las probabilidades de transición entre estados.

# Capítulo 5

## Datos

### 5.1. Datos disponibles

#### 5.1.1. Datos transaccionales

Para efectos del presente Trabajo, se trabajará con información transaccional de una gran cadena de supermercados. La base de datos disponible cuenta con información transaccional entre los días 1 de junio de 2012 y 1 de junio de 2014, y cuenta con las siguientes columnas:

- Código del local donde se realiza la transacción.
- Número de la boleta de compra.
- Identificador del cliente que compra.
- Fecha de la transacción.
- Hora de la transacción.
- Monto total de la boleta de compra.
- Código de la caja donde se realiza la transacción.
- Código del producto que se está comprando.
- Cantidad de unidades del producto que se están comprando.
- Precio unitario del producto (en CLP).
- Precio total pagado por el producto (en CLP).

Esta base cuenta con 477.915.117 entradas, correspondiente a la información de 18 locales a lo largo de 731 días y de 97.629 productos para 1.289.616 clientes distintos. Sin embargo, es necesario transformar la forma en que se presentan estos datos para la correcta estimación de los modelos predictivos. Al modificar la base para corresponder a los criterios estipulados, la data resultante cuenta con 241.772.131 entradas, donde se identifican 76.908 productos y 829.873 clientes distintos. De la totalidad de entradas, 179.783.114 cuentan con un cliente asociado, correspondiendo al 74.36% de los datos.

### 5.1.2. Auditoría

Esta auditoría realiza análisis de los mismos datos provistos anteriormente. Dentro del estudio, se inspeccionan las góndolas de los 10 locales dos veces al día para 58 productos. La primera medición se realiza a las 12:00 horas, y acumula la información de ventas hasta ese momento, mientras que la segunda medición se realiza a las 18:00 horas, y contabiliza las ventas entre mediciones.

Para definir si un producto se encuentra o no en quiebre de stock, el criterio a utilizar requiere que ambas mediciones diarias arrojen el mismo resultado, ya sea que las dos arrojen disponibilidad para el producto o quiebre de stock. Las mediciones que no coincidan no se contabilizan para la validación.

## 5.2. Análisis Exploratorio

La información con la que se trabaja inicialmente sirve de comparación al estudio anterior ya mencionado. Este cuenta con la información de 469 días para cada local, entre el 18 de febrero de 2013 y el 1 de junio de 2014. Una primera observación que hacer respecto a estos datos, es que existen días en los que se registra una cantidad total de boletas igual a 0, lo que se considera un error, ya que esto representaría que no se vendió ningún producto en todo el local (algo que no es razonable considerar).

Estos días varían según el local, pero existen 7 días en que coincide el comportamiento de todos, los cuales son: 1 de mayo de 2013, 18 de septiembre de 2013, 19 de septiembre de 2013, 25 de diciembre de 2013, 1 de enero de 2014, 1 de mayo de 2014 y 20 de mayo de 2014. Los días que presentan este comportamiento no serán considerados en el análisis posterior, por lo que toma especial relevancia asegurar la existencia de suficientes días válidos en cada local. Como se observa en la Tabla 5.1, el 93.11 % de los días son válidos, lo que permite utilizar los locales para el análisis.

Local	Días válidos
1	444
2	451
3	451
4	459
5	462
6	447
7	462
8	451
9	451
10	289

Tabla 5.1: Cantidad de días válidos por local

Como se explicó anteriormente, se cuenta con información real del estado de góndola para



58 productos. El detalle de los productos se aprecia en la Tabla 5.2. La numeración sirve como referencia para los productos, y los productos 1 al 14 coinciden con los 14 estudiados por Montoya and González (2018).

Numeración	Producto	Marca	Numeración	Producto	Marca
1	Atún	Angelmó	30	Crackeler	Costa
2	Azúcar	IANSÁ	31	Gaseosa Cola	Coca Cola
3	Camarón	El Golf	32	Gaseosa Cola Light	Coca Cola
4	Chocolate Trencito	Nestlé	33	Leche en Polvo	Calo
5	Gaseosa Naranja	Fanta	34	Leche Entera	Marca Propia
6	Jugo en Polvo	Livean	35	Nidal 1	Nidal
7	Jurel	Colorado	36	Leche sin Lactosa	Loncoleche
8	Lasaña	Luchetti	37	Maizena	Dropa
9	Lavalozas	Quix	38	Margarina	Sadia
10	Leche Chocolate	Soprole	39	Mayonesa	Marca Propia
11	Toalla de Papel	Nova	40	Mayonesa	Hellmanns
12	Papas Duquesas	Minuto Verde	41	Néctar	Watts
13	Quesillo	Colún	42	Nescafé	Nescafé
14	Té	Supremo	43	Nido Polvo	Nido
15	Aceite Soya	Marca Propia	44	Palmitos	Marca Propia
16	Aceite Oliva	Banquete	45	Pañal	Huggies
17	Alimento para Perro	Guau Forte	46	Pañal Activsec	Huggies
18	Avena	Quaker	47	Papas Stax	Lay's
19	Café	Dolca	48	Papel Higiénico	Favorita
20	Caluga Suny	Calaf	49	Pisco	Mistral
21	Chocapic	Nestlé	50	Puré de Papas	Maggi
22	Cerveza Escudo	Escudo	51	Queso Fresco	Quillayes
23	Cerveza S/Alcohol	Cristal	52	Queso Gauda	Calo
24	Cerveza Stella Artois	Stella Artois	53	Salchichas	Llanquihue
25	Champagne	Valdivieso	54	Salmón	Super
26	Chorizo	Marca Propia	55	Shampoo	Elvive
27	Detergente Líquido	Drive	56	Tortilla	Pancho Villa
28	Fideo	Lucchetti	57	Vienesas	JK
29	Fósforos	Copihue	58	Vienesas de Ave	Montina

Tabla 5.2: Información de Productos en estudio

Un primer análisis interesante consiste en calcular la incidencia de cada producto. Este indicador corresponde al porcentaje promedio de veces que aparece el producto en una boleta sobre el total de boletas diarias, a lo largo del horizonte de estudio.

Como se observa en la Tabla 5.3, todos los productos tienen, en el promedio de los locales, incidencias comprendidas entre 0.02% para el caso del Pañal y 1.81% para el caso de la Leche Entera, que presenta su máximo valor en el local 6. Destaca el caso del Pañal, que posee una incidencia de tan solo el 0.0102% en el local 10 (lo que se interpreta como aparecer en 1 de cada 1.000 boletas), siendo la menor incidencia entre los productos. En la Figura 5.1 se aprecia la incidencia promedio de cada producto considerando todos los locales, donde se mantienen las tendencias.

Tomando en cuenta la categorización de productos según su nivel de incidencias en las ventas de un supermercado entregada en la Sección 4.1, los productos pueden segmentarse en tres categorías, Incidencia Alta, Media o Baja. En el universo de productos se encuentran

Producto	Local										Total
	1	2	3	4	5	6	7	8	9	10	
Atún	0.29%	0.34%	0.30%	0.27%	0.34%	0.59%	0.47%	0.32%	0.32%	0.34%	0.35%
Azúcar	0.41%	0.33%	0.25%	0.15%	0.48%	0.61%	0.36%	0.30%	0.52%	0.30%	0.37%
Camarón	0.46%	0.46%	0.26%	0.21%	0.43%	0.36%	0.32%	0.47%	0.26%	0.27%	0.37%
Chocolate Trencito	0.29%	0.43%	0.18%	0.20%	0.14%	0.38%	0.46%	0.52%	0.30%	0.28%	0.33%
Gaseosa Naranja	0.44%	0.62%	0.23%	0.18%	0.45%	0.25%	0.51%	0.55%	0.45%	0.37%	0.41%
Jugo en Polvo	0.64%	0.67%	0.34%	0.18%	0.40%	0.44%	0.42%	1.04%	0.27%	0.46%	0.51%
Jurel	0.29%	0.29%	0.20%	0.15%	0.44%	0.46%	0.32%	0.30%	0.12%	0.23%	0.29%
Lasaña	0.18%	0.27%	0.12%	0.15%	0.17%	0.40%	0.26%	0.29%	0.16%	0.16%	0.22%
Lavalozas	1.12%	1.30%	0.48%	0.45%	0.97%	1.74%	0.97%	1.37%	0.94%	0.52%	1.04%
Leche Chocolate	0.97%	0.89%	0.59%	0.43%	0.64%	1.02%	0.81%	1.00%	0.93%	0.77%	0.81%
Toalla de Papel	1.70%	1.57%	0.94%	0.70%	0.76%	1.71%	1.38%	2.08%	0.68%	0.89%	1.32%
Papas Duquesas	0.25%	0.28%	0.10%	0.11%	0.16%	0.29%	0.20%	0.27%	0.19%	0.17%	0.21%
Quesillo	0.26%	0.51%	0.22%	0.26%	0.32%	0.54%	0.36%	0.65%	0.25%	0.44%	0.39%
Té	1.00%	1.28%	0.66%	0.52%	0.99%	1.41%	0.66%	0.96%	0.86%	0.68%	0.93%
Aceite Soya	0.26%	0.17%	0.10%	0.11%	0.21%	0.20%	0.13%	0.17%	0.14%	0.08%	0.17%
Aceite Oliva	0.03%	0.04%	0.02%	0.01%	0.02%	0.04%	0.03%	0.06%	0.02%	0.03%	0.03%
Alimento para Perro	0.09%	0.18%	0.16%	0.02%	0.09%	0.25%	0.17%	0.13%	0.22%	0.07%	0.14%
Avena	0.78%	0.82%	0.51%	0.50%	0.80%	1.02%	0.69%	0.85%	0.73%	0.63%	0.75%
Café	0.05%	0.13%	0.12%	0.07%	0.14%	0.11%	0.15%	0.11%	0.15%	0.10%	0.11%
Caluga Suny	0.16%	0.14%	0.10%	0.16%	0.17%	0.18%	0.13%	0.14%	0.18%	0.18%	0.15%
Chocapic	0.16%	0.25%	0.07%	0.13%	0.16%	0.38%	0.19%	0.29%	0.14%	0.09%	0.20%
Cerveza Escudo	0.43%	0.83%	0.49%	0.14%	0.45%	0.94%	0.65%	0.54%	0.67%	0.54%	0.56%
Cerveza S/Alcohol	0.09%	0.14%	0.04%	0.04%	0.10%	0.14%	0.18%	0.14%	0.07%	0.15%	0.11%
Cerveza Stella Artois	0.08%	0.11%	0.03%	0.06%	0.07%	0.11%	0.07%	0.08%	0.07%	0.05%	0.08%
Champagne	0.10%	0.11%	0.06%	0.06%	0.10%	0.09%	0.10%	0.14%	0.10%	0.17%	0.10%
Chorizo	0.38%	0.38%	0.18%	0.11%	0.37%	0.65%	0.54%	0.26%	0.50%	0.35%	0.36%
Detergente Líquido	0.07%	0.08%	0.03%	0.03%	0.03%	0.04%	0.03%	0.11%	0.02%	0.02%	0.05%
Fideo	0.81%	1.16%	0.55%	0.36%	0.69%	1.44%	0.64%	0.68%	0.83%	0.64%	0.79%
Fósforos	1.17%	1.44%	0.50%	0.32%	1.01%	1.18%	1.15%	1.05%	0.99%	0.36%	0.96%
Crackeler	1.12%	1.38%	0.84%	0.88%	1.23%	1.12%	1.03%	1.81%	0.62%	1.67%	1.17%
Gaseosa Cola	1.40%	2.12%	1.09%	0.97%	1.69%	1.59%	1.81%	2.32%	1.52%	1.81%	1.63%
Gaseosa Cola Light	0.24%	0.17%	0.08%	0.04%	0.10%	0.14%	0.19%	0.39%	0.03%	0.20%	0.16%
Leche en Polvo	0.06%	0.10%	0.03%	0.03%	0.05%	0.13%	0.06%	0.10%	0.10%	0.04%	0.07%
Leche Entera	1.43%	2.29%	1.37%	1.07%	2.40%	2.60%	2.17%	1.37%	1.93%	1.62%	1.81%
Nidal 1	0.03%	0.04%	0.01%	0.02%	0.03%	0.04%	0.03%	0.03%	0.03%	0.04%	0.03%
Leche sin Lactosa	0.36%	0.49%	0.30%	0.27%	0.34%	0.34%	0.35%	0.53%	0.22%	0.42%	0.37%
Maizena	0.18%	0.14%	0.08%	0.08%	0.09%	0.14%	0.13%	0.20%	0.08%	0.09%	0.13%
Margarina	0.58%	0.49%	0.45%	0.41%	0.59%	1.16%	0.54%	0.55%	0.56%	0.44%	0.59%
Mayonesa Marca Propia	0.11%	0.18%	0.12%	0.07%	0.12%	0.13%	0.10%	0.11%	0.12%	0.07%	0.11%
Mayonesa Hellmanns	0.69%	0.47%	0.34%	0.27%	0.53%	0.76%	0.70%	0.58%	0.79%	0.37%	0.56%
Néctar	0.79%	0.92%	0.64%	0.50%	0.76%	0.97%	0.60%	1.09%	0.68%	0.78%	0.79%
Nescafé	0.70%	0.95%	0.54%	0.44%	0.83%	1.15%	0.96%	0.90%	0.69%	0.64%	0.79%
Nido Polvo	0.15%	0.15%	0.07%	0.11%	0.13%	0.21%	0.14%	0.25%	0.10%	0.21%	0.15%
Palmitos	0.55%	1.09%	0.47%	0.39%	0.82%	1.09%	0.55%	0.98%	0.25%	0.78%	0.72%
Pañal	0.02%	0.02%	0.01%	0.02%	0.02%	0.03%	0.01%	0.02%	0.03%	0.01%	0.02%
Pañal Activsec	0.02%	0.01%	0.02%	0.01%	0.02%	0.02%	0.01%	0.02%	0.03%	0.01%	0.02%
Papas Stax	0.32%	0.30%	0.21%	0.25%	0.33%	0.35%	0.28%	0.40%	0.30%	0.39%	0.31%
Papel Higiénico	0.14%	0.12%	0.06%	0.04%	0.11%	0.20%	0.10%	0.33%	0.05%	0.11%	0.14%
Pisco	0.14%	0.20%	0.10%	0.07%	0.12%	0.17%	0.16%	0.27%	0.14%	0.11%	0.15%
Puré de Papas	0.43%	0.80%	0.18%	0.39%	0.54%	0.49%	0.64%	0.90%	0.45%	0.57%	0.55%
Queso Fresco	0.22%	0.13%	0.13%	0.17%	0.11%	0.12%	0.09%	0.17%	0.05%	0.03%	0.13%
Queso Gauda	0.53%	0.72%	0.22%	0.41%	0.26%	0.85%	0.66%	0.46%	0.31%	0.64%	0.51%
Salchichas	0.12%	0.10%	0.04%	0.05%	0.07%	0.08%	0.06%	0.17%	0.04%	0.10%	0.09%
Salmón	0.24%	0.22%	0.10%	0.15%	0.24%	0.13%	0.19%	0.30%	0.08%	0.20%	0.19%
Shampoo	0.14%	0.14%	0.05%	0.07%	0.09%	0.15%	0.10%	0.16%	0.06%	0.09%	0.11%
Tortilla	0.07%	0.14%	0.04%	0.09%	0.10%	0.08%	0.08%	0.39%	0.05%	0.10%	0.13%
Vienesas	0.31%	0.41%	0.27%	0.24%	0.42%	0.63%	0.52%	0.21%	0.58%	0.28%	0.38%
Vienesas de Ave	0.68%	0.59%	0.56%	0.43%	0.65%	1.10%	0.77%	0.32%	1.04%	0.69%	0.66%
<b>Promedio</b>	<b>0.43%</b>	<b>0.52%</b>	<b>0.28%</b>	<b>0.24%</b>	<b>0.41%</b>	<b>0.57%</b>	<b>0.44%</b>	<b>0.52%</b>	<b>0.38%</b>	<b>0.38%</b>	<b>0.42%</b>

Tabla 5.3: Incidencias de cada producto por local

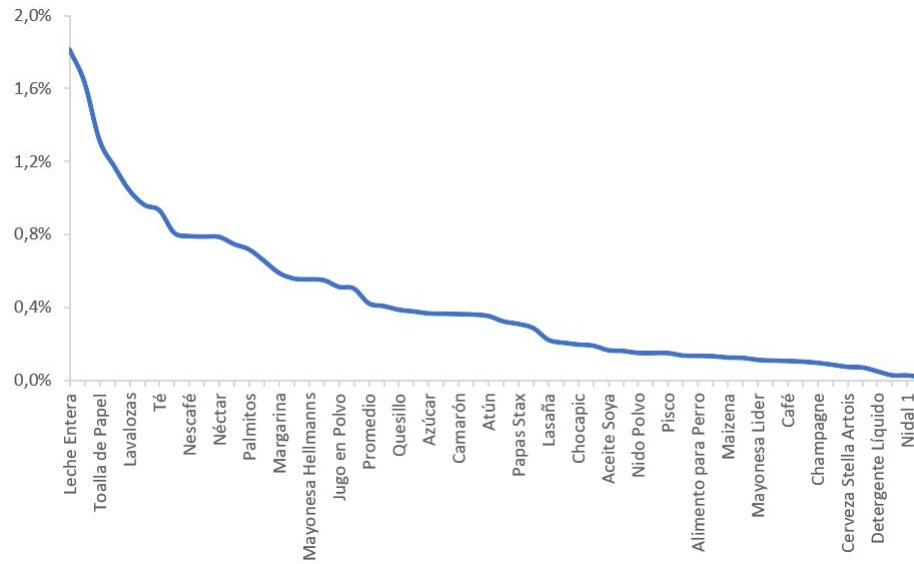


Figura 5.1: Incidencia promedio para cada producto

14 de Incidencia Alta, 16 de Incidencia Media y 28 de Incidencia Baja (ir al Apéndice A para ver detalle por producto).

### 5.2.1. Mediciones de Quiebre

El fenómeno que este Trabajo de Título pretende analizar y desarrollar es el de quiebres de stock ocurrido en góndola. Para lograr esto, se analizan los niveles de quiebre para cada uno de los 58 productos, según los criterios definidos en la Sección 5.1.2. Como se observa en la Figura 5.2, el promedio de los productos presenta quiebres solo en el 7.3% de los días, valor menor en comparación al 12.6% que alcanzan los días que no se consideran en el análisis. Dentro del total de productos, 12 no presentan ningún quiebre en el marco de análisis, y 9 presentan más de 10 quiebres. Destacan el Pañal Activsec y las Papas Stax, con 81 y 135 quiebres medidos, respectivamente.

### 5.2.2. Segmentos de Clientes

Parte importante del desarrollo del presente Trabajo radica en encontrar segmentos de clientes para cada producto que puedan entregar información valiosa al momento de predecir los quiebres de stock, por lo que es de vital importancia tener nociones de su comportamiento de compra. Como se muestra en la Figura 5.3, del total de clientes identificables (es decir, que poseen un ID asociado) presentes en la base de datos, 530 mil (correspondientes al 63.91% del total) registran entre 1 y 10 compras considerando todos los días y todos los locales. Es razonable pensar que, mientras más boletas tenga registradas un cliente, más información entregará al modelo. Por esta razón, interesa ver el comportamiento de clientes que tengan más de 10 boletas registradas.

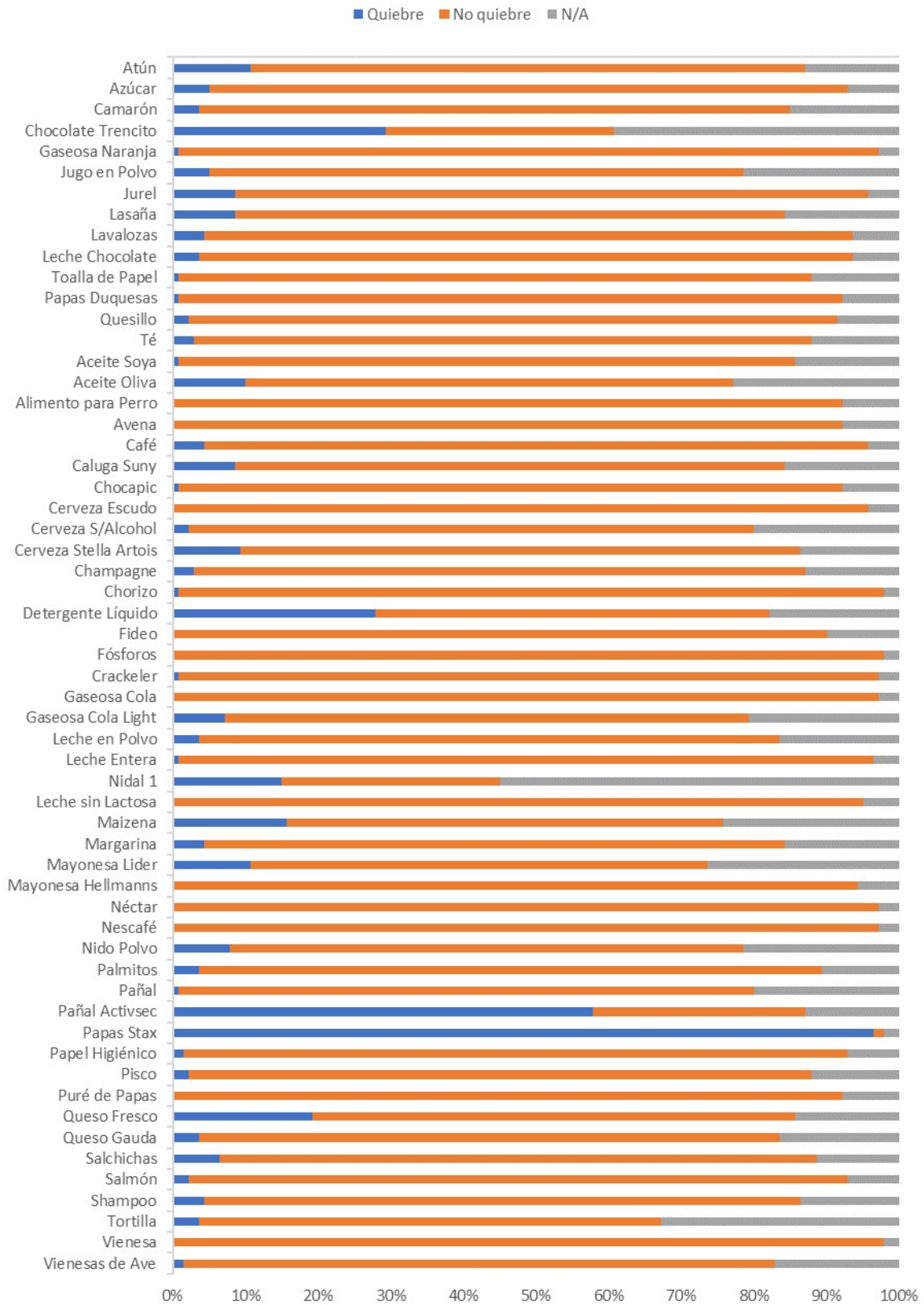


Figura 5.2: Distribución de las observaciones medidas en góndola por producto

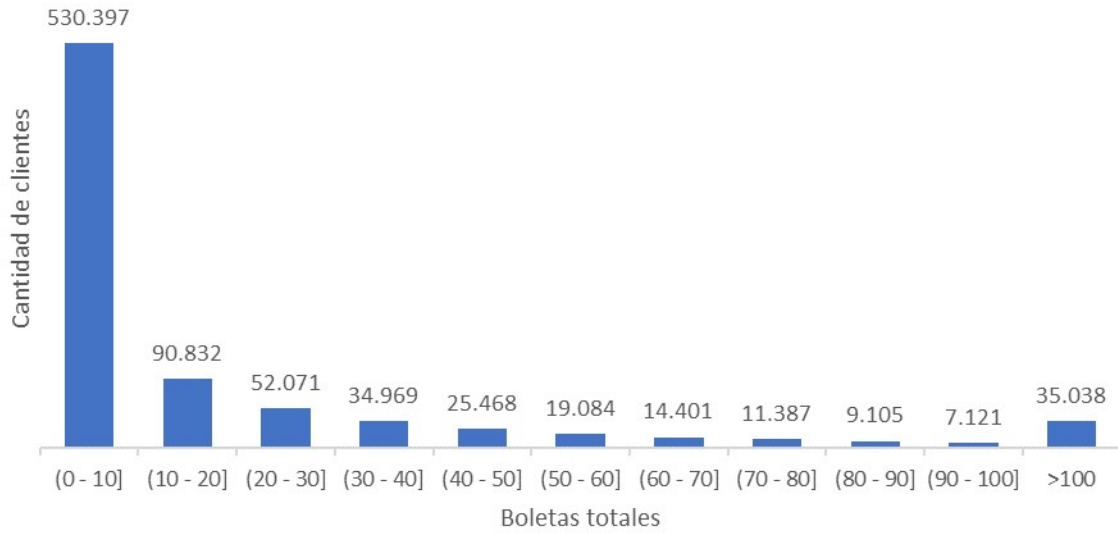


Figura 5.3: Distribución de clientes según boletas totales

Otro aspecto interesante de analizar es ver cómo varía la cantidad de clientes a medida que aumenta el porcentaje de lealtad que tengan a cada producto. En otras palabras, se pretende estudiar cuántas personas tienen, por ejemplo, más de un 50% de presencia de un producto determinado en la totalidad de sus boletas. En las Figuras 5.4 y 5.5 se aprecia que, a medida que aumenta el nivel de lealtad, la cantidad de clientes decae de forma exponencial, tanto para los clientes en su totalidad como para los que presentan más de 10 boletas registradas en el marco temporal estudiado. Para facilitar el análisis descriptivo posterior, se tomarán como referencia los primeros 14 productos en estudio (que coinciden con los productos utilizados por Montoya and González (2018)).

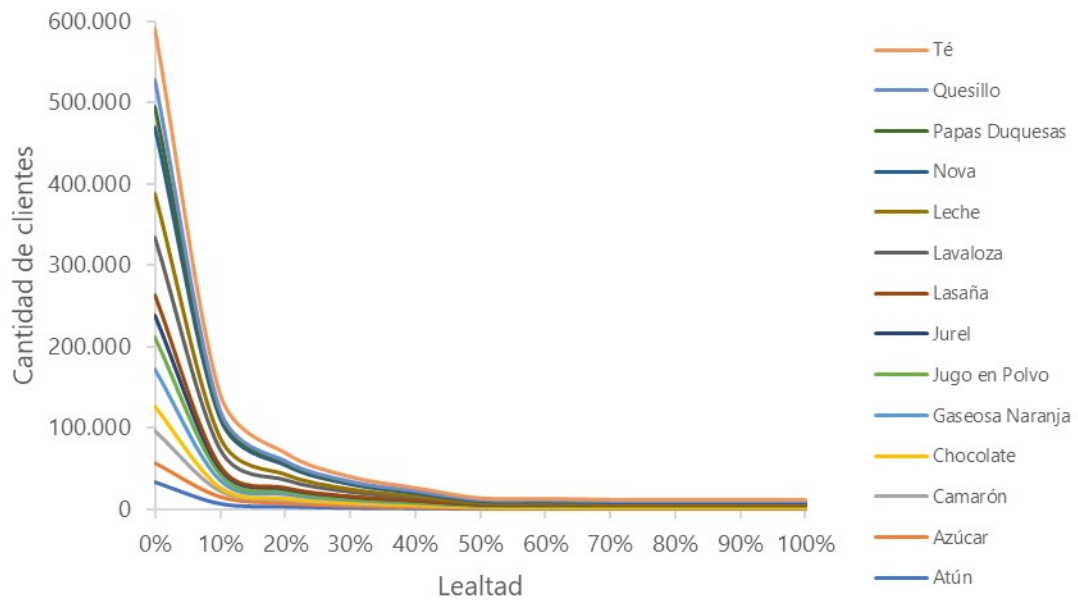


Figura 5.4: Cantidad de clientes según nivel de lealtad para cada producto

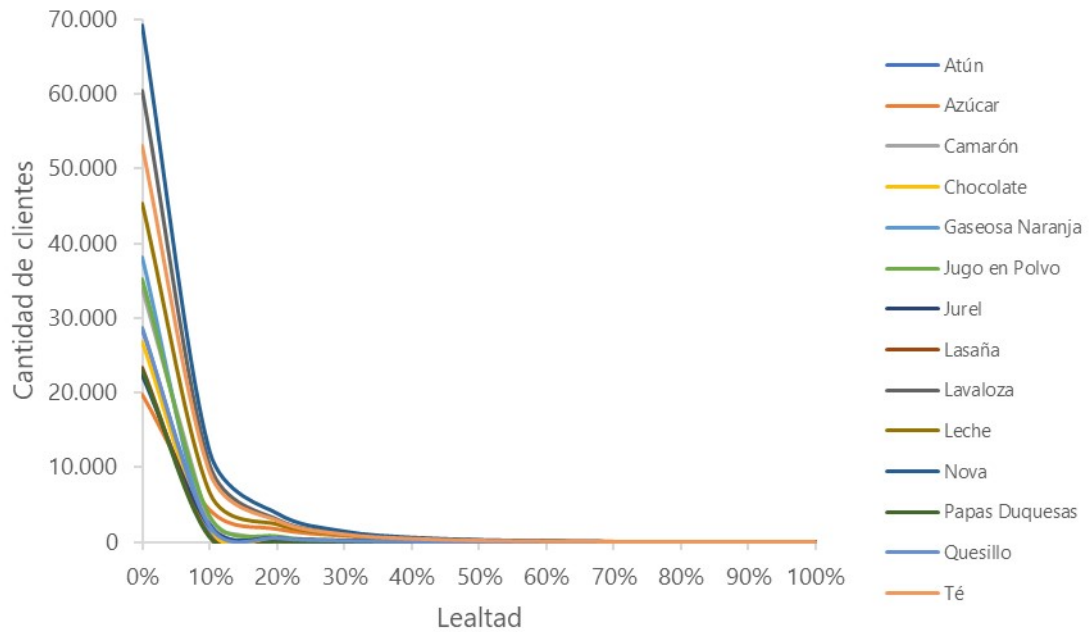


Figura 5.5: Cantidad de clientes con Incidencias totales mayores a 10 según nivel de lealtad para cada producto

Es importante notar que estos resultados definen criterios de búsqueda en cuanto a los niveles de lealtad. Tomando en cuenta el universo de clientes con más de 10 boletas registradas, en promedio solo hay 4.216 personas por producto fijando una lealtad del 10% (lo que corresponde a menos del 12% de los clientes que alguna vez compraron el producto). Con

esta cantidad tan acotada de clientes, es esperable que al llegar a niveles de lealtad mayores se encuentre un elevado número de días en que ninguno de ellos presente compras en ninguno de los locales, lo que se refleja en importantes pérdidas de información. En la Tabla 5.4 se aprecian las diferencias entre productos, destacando Toalla de Papel Nova y Lavalozza (ambas pertenecientes al nivel de incidencia alta), que al 10 % de lealtad cuentan con más de 10.000 clientes.

	Lealtad										
	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Atún	28,548	24,643	18,408	13,764	10,712	8,390	6,400	5,115	4,003	3,283	2,601
Azúcar	19,773	18,231	15,385	12,959	10,977	9,409	7,915	6,822	5,782	4,987	4,319
Camarón	34,007	28,397	19,776	13,795	10,010	7,477	5,342	4,064	2,988	2,223	1,539
Chocolate	26,850	22,169	15,385	10,874	8,086	6,194	4,550	3,519	2,613	2,001	1,485
Gaseosa Naranja	38,214	31,847	21,783	14,948	10,832	7,862	5,418	4,006	2,768	1,956	1,251
Jugo en Polvo	35,205	30,812	23,497	17,862	14,088	11,191	8,591	6,817	5,376	4,317	3,449
Jurel	22,430	19,324	14,518	10,910	8,471	6,637	5,128	4,158	3,276	2,650	2,133
Lasaña	23,382	19,417	13,430	9,236	6,720	4,919	3,441	2,554	1,785	1,295	853
Lavalozza	60,517	56,006	46,648	38,158	31,608	26,087	21,328	17,807	14,822	12,399	10,372
Leche	45,349	40,341	31,879	940	20,643	16,904	13,658	11,433	9,486	7,971	6,658
Nova	69,220	62,992	51,647	41,955	34,868	29,046	23,953	20,137	16,932	14,396	12,161
Papas Duquesas	23,023	18,802	12,528	8,408	6,013	4,282	2,916	2,097	1,426	1,029	650
Quesillo	28,762	24,126	17,290	12,564	9,619	7,534	5,680	4,548	3,514	2,792	2,169
Té	53,127	48,698	40,442	33,116	27,669	23,004	18,747	15,815	13,167	11,162	9,389
<b>Promedio</b>	<b>36,315</b>	<b>31,843</b>	<b>24,473</b>	<b>17,106</b>	<b>15,023</b>	<b>12,067</b>	<b>9,505</b>	<b>7,778</b>	<b>6,281</b>	<b>5,176</b>	<b>4,216</b>

Tabla 5.4: Cantidad de clientes con más de 10 boletas totales según nivel de lealtad para cada producto

A grandes rasgos, el comportamiento mediante se avanza en el nivel de lealtad es similar entre los productos (aparte del hecho que, en cantidad, hay una predominancia de los productos con mayor nivel de incidencia). Analizando a nivel de categorías de incidencia, como es esperable, el comportamiento es similar entre los distintos grupos, como se observa en la Figura 5.6. Las diferencias que se aprecian se relacionan, al igual que a nivel de producto, con la incidencia total de los productos pertenecientes a cada grupo. Al analizar este comportamiento de forma porcentual sobre la totalidad de los clientes que presentan incidencias en cada categoría, es más fácil ver las similitudes entre los grupos, como se aprecia en la Figura 5.7.

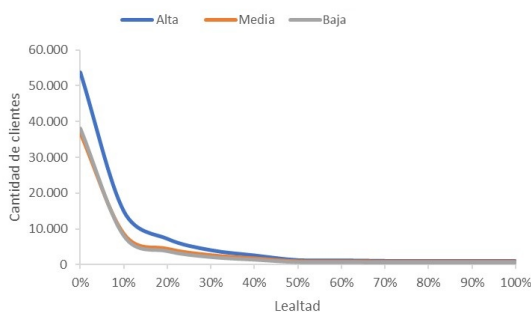


Figura 5.6: Cantidad de clientes según nivel de lealtad para cada categoría

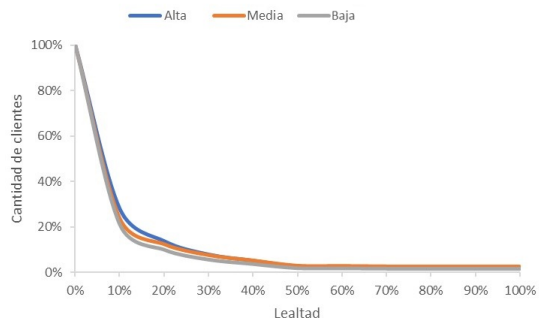


Figura 5.7: Porcentaje de clientes según nivel de lealtad para cada categoría

### 5.2.3. Días sin Incidencias y Lealtad

A medida que aumenta el nivel de lealtad exigido disminuyen los clientes seleccionados. Así, mientras menos son los clientes, se espera que más sean los días en que ninguno de ellos va al supermercado a comprar. Al observar la Figura 5.8, para niveles bajos de lealtad la cantidad de días sin incidencias totales no presenta mayores alzas (a un nivel de lealtad del 15 % hay solo un 3.4 % más de días sin incidencias que en el modelo completo). Sin embargo, se genera un aumento pronunciado al seguir aumentando, llegando a más de un 56 % de días sin incidencias en promedio al exigir un nivel de 50 % de lealtad.

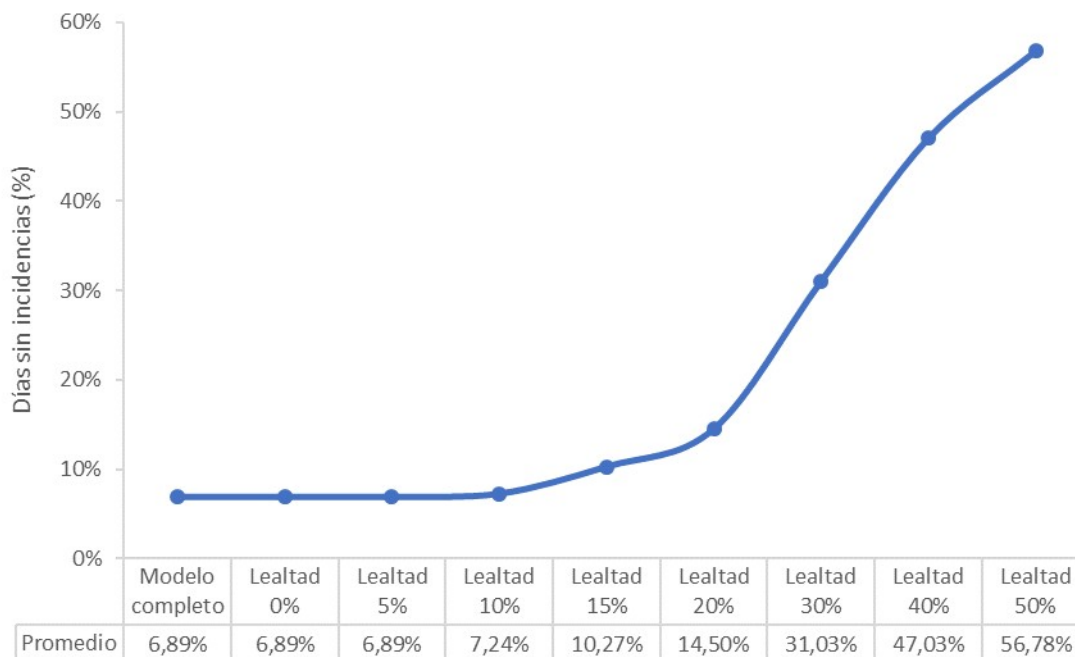


Figura 5.8: Porcentaje de días sin incidencias promedio, para distintos niveles de lealtad

Al analizar este resultado a nivel de producto, como se observa en la Figura 5.9, se encuentran diferencias significativas en la pérdida de información para los distintos productos. Esto se debe principalmente a la cantidad de clientes restantes para cada nivel de lealtad, y que estos varían según el producto (ver Tabla 5.3). Productos de baja incidencia como las Papas Duquesas y la Lasaña son las que presentan mayor pérdida de información. Sin embargo, se observa una disminución significativa en la pérdida de información para el Azúcar, Lavalozza, Té, Leche y Toallas de Papel, quienes a una lealtad del 50 % aún poseen casi un 70 % de la información disponible para ser analizada.



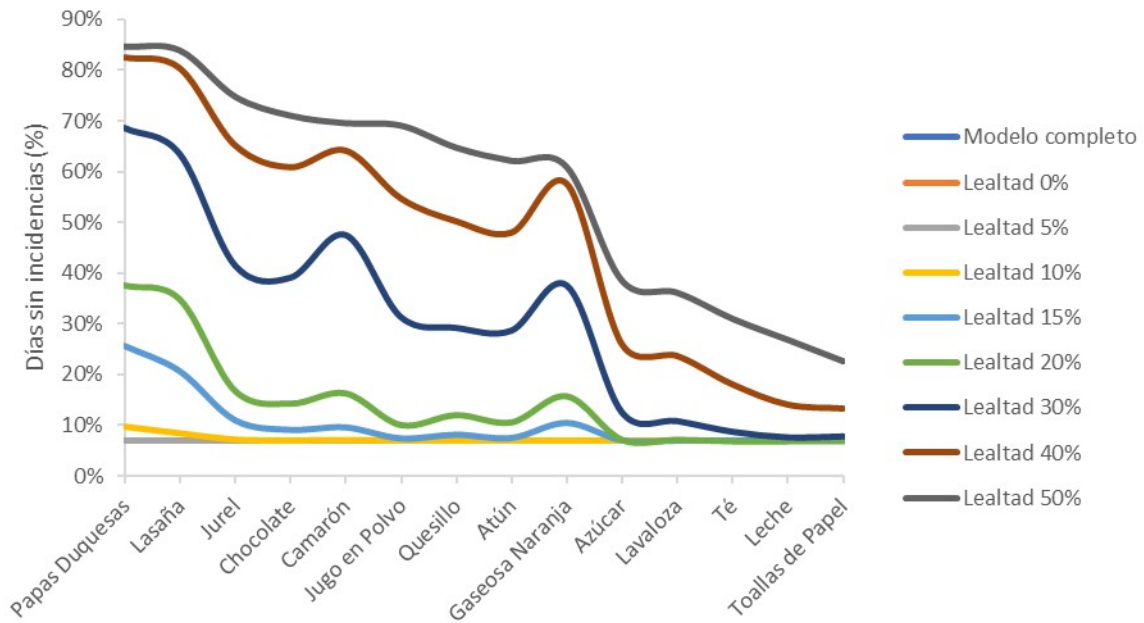


Figura 5.9: Porcentaje de días sin incidencias por producto, para distintos niveles de lealtad

Este análisis entrega información que puede ser utilizada para definir niveles de lealtad máximo a exigir por producto, fijando una cota de tolerancia para la cantidad de pérdida de información permitida. Si se exige a todos los productos que la pérdida de información no supere un 20 %, esto define un rango porcentual máximo para el nivel de lealtad exigido. En la Tabla 5.5 se resumen los rangos para cada producto:

Producto	Grupo de Incidencia	Lealtad máxima
Papas Duquesas	Baja	
Lasaña	Baja	[15 % - 20 %]
Jurel	Baja	
Chocolate	Media	
Camarón	Media	[20 % - 30 %]
Jugo en Polvo	Media	
Gaseosa Naranja	Media	
Quesillo	Media	30 %
Atún	Media	
Azúcar	Media	
Lavalozza	Alta	[40 % - 50 %]
Té	Alta	
Leche Chocolate	Alta	50 %
Toallas de Papel	Alta	

Tabla 5.5: Lealtad máxima para cada producto

## 5.2.4. Frecuencia y Amplitud de Compra

Hasta el momento, se ha utilizado como única métrica de lealtad la incidencia media de compra sobre cada uno de los productos. Si bien esta medida es útil a la hora de seleccionar ocurrencia de compra y mediante esto filtrar a clientes que no tengan suficiente presencia del producto en su canasta de compra, esta nose encuentra exenta de limitaciones.

Cuando se realiza la búsqueda de métricas que discriminen lealtad en los clientes, no solo basta considerar aquellos que hayan realizado compras de un producto, ya que se pueden dar muchos casos que cumplen dicho criterio y que, sin embargo, no siguen la regla de lealtad (en cuanto a nociones se refiere):

- Si se exige un nivel de lealtad muy bajo (cercano al 0%), existe una gran cantidad de clientes que solo compraron una vez el producto durante todo el marco temporal.
- Pueden existir clientes que compraron más de una vez el producto, pero dichas compras estén separadas de varios períodos de tiempo (incluso varios meses). Considerar a estos clientes no entrega información valiosa, ya que su frecuencia es demasiado baja.
- Pueden existir clientes que compraron varias veces el producto, pero concentrando las compras en un marco temporal muy acotado (por ejemplo, en un mes). Considerar a estos clientes puede ser una desventaja, ya que los períodos de tiempo en los que estos no realizan compras serán mal etiquetados como quiebres.

Para mejorar la caracterización de lealtad, poniendo el foco en evitar estos casos “patológicos”, se introducen los conceptos de período (asociado a la frecuencia de compra) y amplitud de compra. El período promedio de compra de un producto se define, para efectos de este Trabajo de Título, como el promedio de la distancia entre las compras del producto por parte de un cliente, y se interpreta como el tiempo promedio que transcurre entre dos compras sucesivas del mismo producto. El concepto de amplitud se define como la distancia entre la primera y la última compra de un producto por parte de un cliente.

En la Tabla 5.6 se realiza un primer análisis donde, en primer lugar, se estiman los períodos promedio de compra de cada producto, analizando sobre el universo de clientes que haya realizado al menos una compra del producto (o lo que es equivalente, clientes con 0% de lealtad) sobre los 469 días en estudio. Se encuentra que el período medio de tiempo que transcurre entre dos compras consecutivas de un producto es de 94.31 días, algo más que 3 meses.

Producto (Incidencia)	Frecuencia Promedio	Compras promedio (mensuales)	Compra máxima	Desviación estándar
Atún (0.35 %)	89.47	0.23	76	0.21
Azúcar (0.37 %)	88.61	0.29	164	0.25
Camarón (0.37 %)	104.7	0.20	54	0.16
Chocolate (0.33 %)	95.09	0.24	88	0.24
Gaseosa naranja (0.41 %)	100.32	0.20	89	0.17
Jugo en polvo (0.51 %)	87.71	0.26	75	0.22
Jurel (0.29 %)	96.52	0.24	42	0.18
Lasaña (0.22 %)	119.23	0.18	26	0.10
Lavalozas (1.04 %)	103.5	0.26	59	0.19
Leche Chocolate (0.81 %)	74.85	0.32	154	0.38
Toallas de papel (1.32 %)	92.23	0.31	94	0.30
Papas duquesas (0.21 %)	91.84	0.18	33	0.11
Quesillo (0.39 %)	82.18	0.27	201	0.32
Té (0.93 %)	94.05	0.27	202	0.22
<b>Promedio</b>	<b>94.31</b>	<b>0.25</b>	<b>96.93</b>	<b>0.22</b>

Tabla 5.6: Descriptivos en base a frecuencia de compra para los productos sobre los 469 días en estudio

También se analiza la cantidad de compras que realiza en promedio un cliente para cada producto. Es esperable que los productos con mayor nivel de incidencia, tales como las Toallas de Papel y el Lavalozas, presenten los mayores niveles de compras promedio. No obstante, productos como el Azúcar y el Quesillo alcanzan altos promedios de compra (0.29 y 0.27 compras mensuales promedio, respectivamente) a pesar de encontrarse en el grupo de incidencia media.

Al repetir este ejercicio, pero solo considerando clientes que, además, tengan una amplitud de compra de al menos 180 días (equivalente a 6 meses), el período promedio entre compras consecutivas aumenta a 130 días. También se ve afectado el número promedio de compras mensuales, pasando de 0.25 a 0.33 para el total de los productos. Sin embargo, y como es de esperar, el número de clientes promedio disminuye de 15.861 para el caso irrestricto a 8.680 imponiendo restricción sobre la amplitud.

Por otra parte, la Figura 5.10 muestra las diferencias entre las distribuciones de clientes por tramos de períodos medios de compra, y para dos amplitudes de compra distintas. Se observa que, sin restringir la amplitud (o lo que es equivalente a tener amplitud  $\geq 0$ ) un 63 % de los clientes compran los productos estudiados en intervalos de menos de 3 meses entre compra. Esta proporción disminuye a un 51 % de los clientes si se exige una amplitud entre la primera y la última compra del producto de al menos 6 meses. Como es de esperar, esta diferencia se equilibra en los segmentos restantes, donde un 26 % de los clientes compran con un período promedio de más de 6 meses considerando aquellos con amplitud de 6 meses, en comparación a los 14 % que lo hacen sin restringir la amplitud.

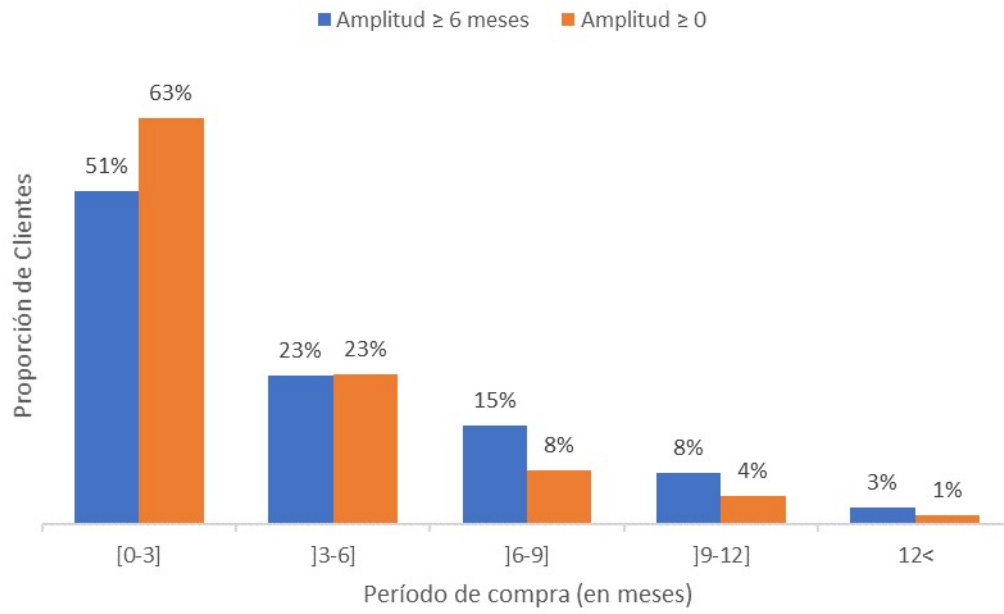


Figura 5.10: Distribución de clientes según períodos de compra para distintas amplitudes de compra

# Capítulo 6

## Análisis de Resultados

Ya definida la estructura sobre la cual se estimarán los quiebres, solo basta por definir la estructura de parámetros observables que se agregan al modelo, recordando que el modelo a testear consiste de un Estado 0 (asociado al quiebre de stock) y dos estados irrestrictos (correspondientes a los niveles de demanda).

Como se estudió en la Sección 3.2.1, existe un vector  $z_{ijt}$  que recolecta las variables observables. Para el estudio se utilizan 36, estas son:

- $Z_{t1} - Z_{t6}$ : variables dicotómicas que representan el día de la semana correspondiente a la observación. Están ordenadas de Lunes a Sábado, dejando el Domingo como nivel base para evitar problemas de colinealidad.
- $Z_{t7}$ : variable semanal que representa la tendencia intrínseca de las ventas del producto en el local. De manera similar a como se hizo con el precio, la magnitud de la tendencia fue reducida para que los parámetros fueran comparables entre sí. Los datos disponibles contemplan 67 semanas de información.
- $Z_{t8} - Z_{t18}$ : variables dicotómicas que indican el mes en el que se encuentra la observación. Están ordenadas de Enero a Noviembre, dejando el mes de Diciembre como nivel base para evitar problemas de colinealidad, al igual que como se hizo con las variables correspondientes a los días de la semana.
- $Z_{t19} - Z_{t21}$ : variables dicotómicas que representan los días anteriores a Navidad, Navidad, y los días posteriores a Navidad respectivamente.
- $Z_{t22} - Z_{t24}$ : variables dicotómicas que representan los días anteriores a Halloween, Halloween, y los días posteriores a Halloween respectivamente.
- $Z_{t25} - Z_{t27}$ : variables dicotómicas que representan los días anteriores a Fiestas Patrias, Fiestas Patrias, y los días posteriores a Fiestas Patrias respectivamente.
- $Z_{t28} - Z_{t30}$ : variables dicotómicas que representan los días anteriores a Año Nuevo, Año Nuevo, y los días posteriores a Año Nuevo respectivamente.
- $Z_{t31} - Z_{t33}$ : variables dicotómicas que representan los días anteriores a Semana Santa, Semana Santa, y los días posteriores a Semana Santa respectivamente.
- $Z_{t34} - Z_{t36}$ : variables dicotómicas que representan los días anteriores a un feriado, el día feriado, y los días posteriores a un feriado respectivamente.

La base de datos con la que se trabaja contiene información transaccional desagregada a nivel de producto por boleta. Sin embargo, el nivel de agregación necesario para realizar los análisis mediante los Modelos de Markov Oculto es a nivel diario, comparando el total de incidencias de un producto con el total de incidencias diarias en cada local. Esta transformación genera una base con las siguientes columnas:

- Fecha.
- Código del Local.
- Código del Producto.
- Cantidad de Boletas del Producto diarias.
- Cantidad de Boletas Totales diarias.

## 6.1. Resultados Iniciales

El primer paso del análisis consiste en estimar los modelos para cada producto utilizando la información completa, es decir, sin realizar segmentación sobre los clientes. Se cuenta con información de 58 productos (descritos en el Capítulo 5), para los 10 locales y los 469 días en estudio, y se estima un modelo para cada uno de ellos. En resumen, los resultados son:

Modelo Completo	Productos	Quiebres	Alertas	Error Tipo I	Falsas Alarmas	Poder de Detección
Total	58	590	744	6.37 %	55.11 %	56.61 %
Incidencia Alta	14	25	14	0.11 %	14.29 %	48.00 %
Incidencia Media	16	226	175	0.34 %	3.43 %	74.78 %
Incidencia Baja	28	339	555	13.77 %	72.43 %	45.13 %

Tabla 6.1: Desempeño Modelo Completo para 58 productos

Se observa que el Poder de Detección promedio es de un 56.61 %, sin embargo, este poder es mayor para los productos que poseen un nivel de incidencia media, con un 74.78 % de poder predictivo (ver Apéndice B para resultados por producto). Por otra parte, existe un gran porcentaje de Falsas Alarmas, principalmente para los productos de incidencia baja, con un 72.43 %. Esto indica que el modelo está sobreestimando los quiebres para este grupo de productos.

### 6.1.1. Definición de Modelos a Comparar

En base a un análisis preliminar (ver Apéndice C), se definen los siguientes grupos segmentados de clientes para comparar sus desempeños:

- **Modelo con ID:** como filtro inicial, se analiza la información de las transacciones que cuentan con un ID asociado. Esta información corresponde a cerca del 75 % de todas las transacciones realizadas.

- **Modelo al 0 % de Lealtad:** Para cada producto, se analiza la información transaccional segmentada a los clientes que hayan realizado *a lo menos una compra del producto* (es decir, a aquellos que tengan un nivel de lealtad mayor al 0 %). Esta segmentación excluye del análisis a los clientes que, a lo largo del horizonte de medición, no realizaron ninguna compra del producto estudiado.

Para lealtades mayores al 0 %, la segmentación se realiza como se explica en el Capítulo 4. Se estiman modelos para los productos en los distintos escenarios:

Modelo con ID	Productos	Quiebres	Alertas	Error Tipo I	Falsas Alarmas	Poder de Detección
Total	58	590	691	6.11 %	57.60 %	49.66 %
Incidencia Alta	14	25	12	0.00 %	0.00 %	48.00 %
Incidencia Media	16	226	174	0.34 %	3.45 %	74.34 %
Incidencia Baja	28	339	505	13.35 %	77.62 %	33.33 %

Tabla 6.2: Desempeño Modelo con ID para 58 productos

Modelo al 0 %	Productos	Quiebres	Alertas	Error Tipo I	Falsas Alarmas	Poder de Detección
Total	58	590	731	6.64 %	59.10 %	50.68 %
Incidencia Alta	14	25	11	0.00 %	0.00 %	44.00 %
Incidencia Media	16	226	172	0.34 %	3.49 %	73.45 %
Incidencia Baja	28	339	548	14.51 %	77.74 %	35.99 %

Tabla 6.3: Desempeño Modelo al 0 % de Lealtad para 58 productos

Como se observa en las Tablas 6.1 y 6.2, el considerar a los clientes con ID no genera mejoras en el Poder de Predicción, tampoco el segmentar sobre los clientes que hayan realizado al menos una compra de cada producto, ya que sus niveles no alcanzan los obtenidos por el Modelo Completo. Los niveles de error también aumentan en ambos modelos, sin embargo, se observa que el Modelo con ID tiene un menor porcentaje de Error Tipo I en comparación al Modelo Completo.

### 6.1.2. Análisis de Sensibilidad

Como se discute en la Sección 3.1.2, la regla que asigna el modelo para decidir si un día  $t$  corresponde o no al Estado 0 o de Quiebre se basa en escoger  $s^* \in \{1, 2, 3\}$  tal que  $s^* = \operatorname{argmax}_{s \in \mathcal{S}} \{P_{ts}\}$ . Es decir, para que se escoja el Estado de Quiebre, la probabilidad de pertenencia al estado debe ser mayor a la probabilidad de estar en el Estado de Demanda Alta o Demanda Media. Si bien esto supone cierta flexibilidad (ya que, en teoría, pueden existir casos donde con  $\frac{1}{3} + \varepsilon$  de probabilidades basta para escoger el estado), en la práctica esta probabilidad es bastante mayor.

Tomando como referencia el Modelo con ID, en el promedio de los días en que el Estado 0 es escogido, su probabilidad de pertenencia es de un 95.36 %, y en el menor valor observado donde se escoge el Estado 0 su probabilidad es de 48.34 %. Por esta razón, cabe preguntarse si al cambiar la regla de decisión pueden cambiar los resultados de forma significativa. El

cambio propuesto consiste en flexibilizar la regla de decisión de pertenencia al Estado 0. En lugar de escogerlo en los casos en que su probabilidad de pertenencia sea mayor al resto de los estados, se establece un umbral definido y estático, estableciendo que si la probabilidad de pertenencia al estado está sobre el umbral (independiente de que sea la mayor probabilidad entre el conjunto de estados posible) se escoge el Estado 0.

En la Figura 6.1 se observa cómo varía el desempeño del modelo al variar el porcentaje de corte donde se genera el umbral de aceptación/rechazo del Estado 0. Es directo notar que, tanto el poder predictivo como las medidas de error son bastante estables a lo largo del rango de probabilidades. También se aprecia que los cambios más significativos en el desempeño ocurren en probabilidades de corte entre 0% y 1%, lo cual es consistente con los resultados del modelo, considerando que en el 91% del total de los días la probabilidad de encontrarse en el Estado 0 es menor a un 1%.

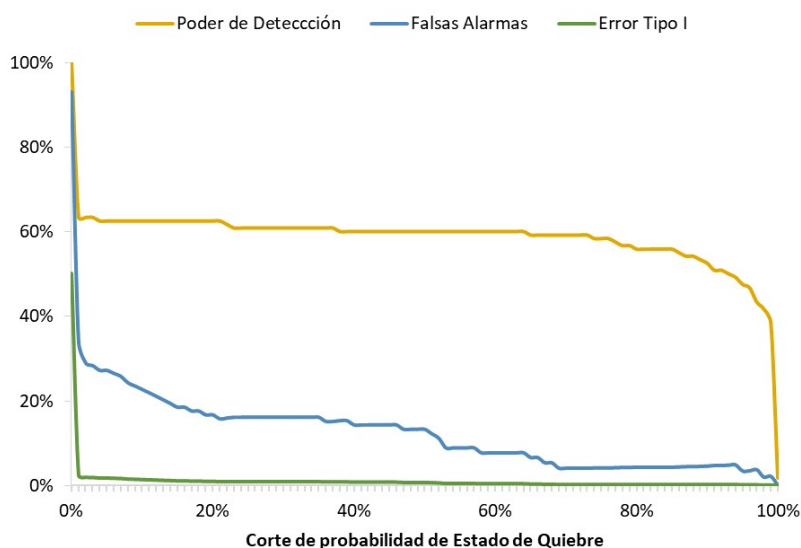


Figura 6.1: Cambio en el desempeño del modelo para distintos cortes en la regla de elección del Estado 0

Cabe mencionar que se encuentran algunas mejoras ajustando el umbral para ciertos productos. Para el Modelo con ID, fijando la probabilidad de corte a un 37%, las Papas Duquesas aumentan su Poder de Detección desde un 0% a un 100%, pero el Error Tipo I aumenta desde un 0.78% a un 5.93%. Por otra parte, tanto la Lasaña como el Chocolate predicen correctamente 1 quiebre más que en el Modelo con ID, fijando la probabilidad de corte en un 22% y 21%, respectivamente. Para los otros modelos, se observan mejoras relativas al modelo original, pero nunca superan los resultados del Modelo con ID.

A pesar de encontrarse mejoras en algunos productos variando el nivel de corte, los resultados generales indican que el modelo es robusto sobre variaciones en el umbral escogido, por lo que se descartan análisis posteriores bajo este enfoque.



## 6.2. Segmentación en base a Nivel de Compra Total

Como se estudia en la Sección 5.2.2, cerca del 64 % de los clientes registrados presentan menos de 10 compras a lo largo del período de estudio. Sin embargo, se cree que la información más útil a la hora de analizar lealtad se obtiene al estudiar el 36 % restante, ya que son estos clientes los que pueden establecer patrones de compra estables en el tiempo. Por esta razón, se estima nuevamente el Modelo con ID y el Modelo a un nivel de Lealtad del 5 %. Sin embargo, esta vez sobre los clientes que hayan realizado al menos 10 compras a lo largo del período en estudio. Los resultados se resumen en las Figuras 6.2 y 6.3:

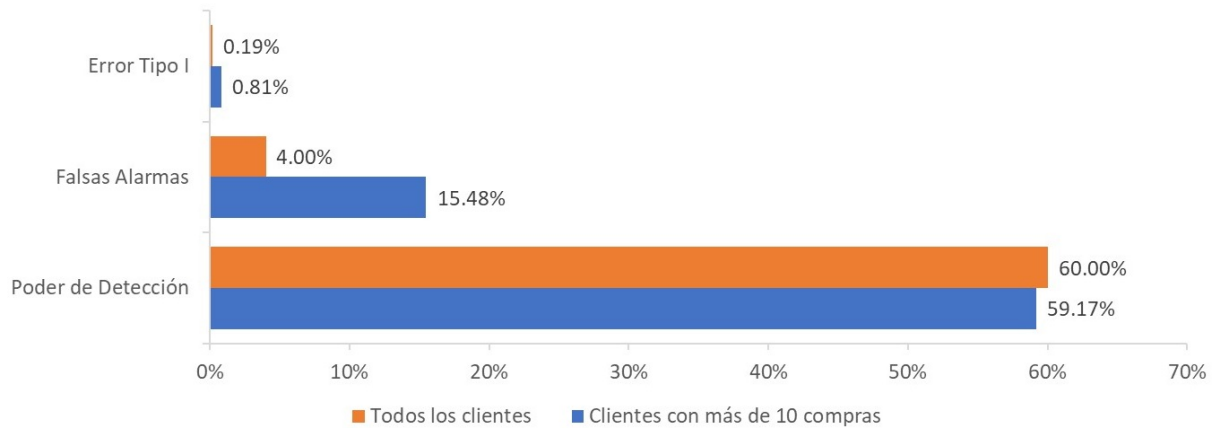


Figura 6.2: Comparación de desempeños del Modelo con ID para distintos segmentos de clientes

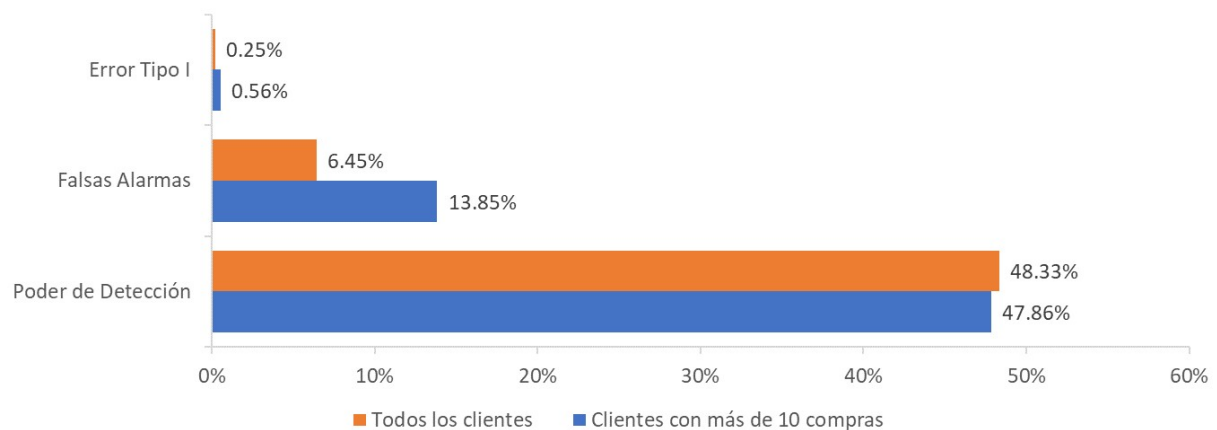


Figura 6.3: Comparación de desempeños del Modelo al 5 % de Lealtad para distintos segmentos de clientes

Se observa que, independiente del modelo, los resultados son mejores analizando todos los clientes. Esto se puede explicar debido a la disminución abrupta en la cantidad de clientes analizados, pasando de un promedio de más de 830.000 a menos de 15.000. En ambos modelos el Poder de Detección es mayor analizando a todos los clientes, y asimismo, los niveles de

error son menores en comparación al análisis sobre los clientes con más de 10 compras. La mayor pérdida en términos del desempeño de los modelos se observa en el aumento de las Falsas Alarmas, que se duplica en el caso del Modelo al 5 % de Lealtad, y casi se cuadriplica en el caso del Modelo con ID.

### 6.3. Enfoque Híbrido para Estimación de Modelos

Los resultados expuestos hasta ahora muestran que estas segmentaciones parecen no superar los poderes predictivos que logra el Modelo Completo. Sumado a esto, un análisis preliminar (ver Apéndice C) parece revelar una tendencia a la subestimación de quiebres al aumentar el nivel de lealtad exigido a los clientes. Una posible explicación a este fenómeno puede deberse a que, en la medida que disminuye la cantidad de clientes, se pierde información valiosa sobre el comportamiento promedio de compra (tendencias de compra, cambios no observables generalizados, etc) de un local en particular. Una forma de incorporar este efecto promedio es alterando la manera en que se calcula la probabilidad de ocurrencia de una observación (ver Sección 3.2.1). La nueva forma de la distribución binomial viene dada por:

$$P_{ijt}(n_{ijt}^\lambda | X_{ijt} = s, z_{ijt}) = \binom{N_{jt}}{n_{ijt}^\lambda} p_{ijts}^{n_{ijt}^\lambda} (1 - p_{ijts})^{N_{jt} - n_{ijt}^\lambda}$$

Donde  $N_{jt}$  corresponde a las boletas totales observadas en el local  $j$  durante el período  $t$  (considerando todo el universo de clientes), y  $n_{ijt}^\lambda$  corresponde a las boletas observadas para el producto  $i$  en el local  $j$  y durante el período  $t$ , tomando en cuenta el comportamiento de clientes con un nivel de lealtad mayor o igual a  $\lambda$ .

Se estiman modelos para cada producto bajo este nuevo enfoque (que en adelante se referirá como *enfoque híbrido*) para el nivel de lealtad de 1 % (ver Apéndice D para detalles de elección del modelo). No se agrega la segmentación al 0 % de lealtad, pues esta coincide a realizar el filtro del Modelo con ID. Los resultados se resumen en la Tabla 6.4:

Modelo Híbrido al 1 %	Productos	Quiebres	Alertas	Error Tipo I	Falsas Alarmas	Poder de Detección
Total	58	590	682	6.11 %	58.36 %	48.14 %
Incidencia Alta	14	25	12	0.00 %	0.00 %	48.00 %
Incidencia Media	16	226	173	0.22 %	2.31 %	74.78 %
Incidencia Baja	28	339	497	13.42 %	79.28 %	30.38 %

Tabla 6.4: Desempeño Modelo Híbrido al 1 % de Lealtad para 58 productos

Se observa que, para el total de los productos, el Modelo Completo presenta mejores resultados que el Modelo Híbrido al 1 % de Lealtad, tanto en su Poder de Detección como en su nivel de Falsas Alarmas (ver Sección 6.1). Sin embargo, y al igual que el Modelo con ID, este presenta un menor nivel de Error Tipo I. Analizando por niveles de incidencia, ambos modelos presentan los mismos poderes de predicción en los productos de incidencia alta e incidencia media. En particular, el Modelo Híbrido no presenta Falsas Alarmas para los

productos de alta incidencia, en comparación con el 14.29 % que levanta el Modelo Completo. Sin embargo, Para los productos de Baja Indicencia este último tiene un mejor desempeño tanto en Poder de Detección como en las Falsas Alarmas.

## 6.4. Enfoque de Lealtad en base a Frecuencia y Amplitud

El análisis y la discusión entablados en la Sección 5.2.4 sugieren incorporar los conceptos de Período y Amplitud de Compras a la hora de segmentar a los clientes. El primero se relaciona con el tiempo promedio que transcurre entre dos compras sucesivas de un producto para un cliente, y el segundo con el tiempo que transcurre entre la primera y la última compra de un producto por parte de un cliente. Estos conceptos se definen en la búsqueda de caracterizar a los clientes y la lealtad de una forma más fiel a la intuición, correspondiendo a un comportamiento constante y duradero de compra.

Para poder concluir sobre la relevancia de incorporar dichas variables, se procede a estimar dos modelos. El primero consiste en segmentar a todos los clientes que tengan un período promedio de compra de no más de 180 días para cada producto (es decir, que el tiempo promedio entre dos compras sucesivas del producto no pase los 6 meses, aproximadamente). El segundo modelo consiste en agregar una restricción adicional al primer modelo, exigiendo que la amplitud de compra sea mayor a 180 días. Cabe destacar que ambos modelos se estiman considerando el enfoque híbrido descrito en la sección anterior. Los resultados se resumen a continuación:

Modelo	Error Tipo I	Falsas Alarmas	Poder de Detección
Modelo 1	6.19 %	54.62 %	51.67 %
Modelo 2	8.56 %	58.85 %	46.67 %

Tabla 6.5: Desempeño de los modelos propuestos

Se observa que ambos modelos presentan peores resultados en comparación al Modelo Completo, tanto en términos del Poder de Predicción como en las medidas de error. Sin embargo, el modelo segmentado sobre el período y no sobre la amplitud presenta leves mejoras en sus medidas de error.

## 6.5. Elección Final de Productos

Observando el desempeño de los modelos por producto (ver Apéndice B), existen 3 productos que tienen un comportamiento anómalo, ya que poseen niveles de quiebre que oscilan entre el nivel medio y bajo y que, sin embargo, generan una gran cantidad de alarmas. Estos corresponden al Aceite de Soya, al Chocapic y a la Mayonesa de Marca Propia. El primero tiene solo un quiebre medido, pero despierta 99 alarmas. Para los otros dos productos el comportamiento es similar, con una razón quiebre/alerta de 1/119 y 15/95, respectivamente.

También se observa que para otros 8 productos medidos, el porcentaje promedio de días en que no se identifican compras del producto alcanza niveles mayores a un 30 %. Esto es, que en un 30 % o más de todo el horizonte de medición no se realizaron compras del producto. Estos productos corresponden al Aceite de Oliva, Cerveza Stella Artois, Champagne, Detergente Líquido, Nidal 1, Pañal, Pañal Activsec y Queso Fresco.

Se calculan los desempeños de los distintos modelos dejando fuera los productos conflictivos. Se escoge el Modelo Completo y el Modelo Híbrido al 1 % de Lealtad, por tener los mejores desempeños. Se estudia también el Modelo con ID y el Modelo al 0 % de Lealtad. El primero se agrega pues es utilizado como modelo de referencia en distintos análisis, y el segundo ya que sirve como referencia para establecer un nivel base de lealtad. Los resultados se muestran a continuación:

Modelo Completo	Productos	Quiebres	Alertas	Error Tipo I	Falsas Alarmas	Poder de Detección
Total	47	373	260	0.30 %	6.15 %	65.42 %
Incidencia Alta	14	25	14	0.11 %	14.29 %	48.00 %
Incidencia Media	16	226	175	0.34 %	3.43 %	74.78 %
Incidencia Baja	17	122	71	0.42 %	11.27 %	51.64 %

Tabla 6.6: Desempeño Modelo Completo para Productos Finales

Modelo con ID	Productos	Quiebres	Alertas	Error Tipo I	Falsas Alarmas	Poder de Detección
Total	47	373	259	0.29 %	6.18 %	65.15 %
Incidencia Alta	14	25	12	0.00 %	0.00 %	48.00 %
Incidencia Media	16	226	178	0.39 %	3.93 %	75.66 %
Incidencia Baja	17	122	69	0.47 %	13.04 %	49.18 %

Tabla 6.7: Desempeño Modelo con ID para Productos Finales

Modelo Híbrido al 1 %	Productos	Quiebres	Alertas	Error Tipo I	Falsas Alarmas	Poder de Detección
Total	47	373	260	0.29 %	6.15 %	65.42 %
Incidencia Alta	14	25	12	0.00 %	0.00 %	48.00 %
Incidencia Media	16	226	175	0.22 %	2.29 %	75.66 %
Incidencia Baja	17	122	73	0.63 %	16.44 %	50.00 %

Tabla 6.8: Desempeño Modelo Híbrido al 1 % de Lealtad para Productos Finales

Modelo al 0 %	Productos	Quiebres	Alertas	Error Tipo I	Falsas Alarmas	Poder de Detección
Total	47	373	251	0.31 %	6.77 %	62.73 %
Incidencia Alta	14	25	11	0.00 %	0.00 %	44.00 %
Incidencia Media	16	226	172	0.34 %	3.49 %	73.45 %
Incidencia Baja	17	122	68	0.57 %	16.18 %	46.72 %

Tabla 6.9: Desempeño Modelo al 0 % de Lealtad para Productos Finales

En términos del poder predictivo, los modelos que mejor resultados entregan son (al igual que para el análisis inicial) el Modelo Completo y el Modelo Híbrido al 1 %, con un Poder de

Detección de quiebres del 65.42%. Observando los niveles de Falsas Alarmas, se mantiene la predominancia de ambos modelos, los cuales registran un nivel de 6.15%. Sin embargo, los niveles más bajos del Error Tipo I se observan tanto en el Modelo con ID como en el Modelo Híbrido al 1%, con un 0.29%.

Analizando los dos modelos con mejor desempeño, se observan ciertas diferencias entre los grupos de productos. En primer lugar, ambos tienen el mismo poder predictivo en el grupo de alta incidencia, pero el Modelo al 1% de Lealtad no posee errores. Estas diferencias ocurren en los productos Lavalozas y Toallas de Papel, donde el Modelo Completo genera una falsa alarma por cada uno. En el grupo de incidencia media, en cambio, el Modelo al 1% de Lealtad logra un mejor desempeño tanto en su Poder de Detección (con un 75.66% versus un 74.78% del Modelo Completo) como en sus medidas de error. Esta mejora en la detección se observa en cuatro productos: Atún, Chocolate, Papas Stax y Queso Gauda (con una detección más en cada caso, ver Apéndice E). Sin embargo, el Modelo Completo presenta mejores resultados para productos de baja incidencia, tanto en el poder predictivo como en las medidas de error.

## 6.6. Robustez de los resultados

Los resultados obtenidos hasta el momento muestran que, para productos de incidencia alta y media, el mejor modelo resulta ser el Modelo Híbrido al 1% de Lealtad, pero *¿son los resultados obtenidos suficientes para ser extrapolados?*. Esta pregunta es de gran importancia, y hasta el momento no se ha abordado dentro del desarrollo del problema. Al diseñar modelos estadísticos predictivos sobre muestras acotadas, hay factores que se deben considerar a la hora de tomar conclusiones sobre la eficiencia y representatividad de sus resultados. El más importante corresponde a la componente estocástica propia de esta clase de modelos, y se intenta solucionar buscando parámetros y ajustes que mejor describan la realidad y el comportamiento de sus variables más relevantes.

Para el caso de la estimación Bayesiana mediante métodos de Cadenas de Markov Monte Carlo (MCMC), la estimación entrega una distribución para cada uno de los parámetros involucrados. Esta distribución contiene la incertidumbre propia del valor del parámetro, y observando su dispersión y sus intervalos de confianza se puede concluir acerca del nivel de incertidumbre en la estimación.

Para entregar una mayor robustez en la estimación de los resultados, se calculan las alertas de quiebres generadas por el modelo durante calibración de parámetros y cada 200 iteraciones (ver Sección 3.3 para más detalles), generando un total de 15.000 estimaciones de alertas por modelo. Con esto, se calculan las métricas de error y el Poder de Predicción en cada una y se obtienen los valores promedio junto con su medida de error. La Tabla 6.10 resume los resultados obtenidos para el Modelo Completo y para el Modelo Híbrido al 1% de Lealtad:

Modelo		Error Tipo I	Falsas Alarmas	Poder de Detección
<b>Completo</b>	Promedio	1.20 %	10.12 %	63.22 %
	Desv. Estándar	0.81 %	4.50 %	6.42 %
<b>Híbrido al 1 % de Lealtad</b>	Promedio	1.22 %	8.91 %	63.39 %
	Desv. Estándar	0.91 %	3.55 %	5.89 %

Tabla 6.10: Resultados de los submodelos para estimación de Modelo Completo

Se observa que ambos modelos disminuyen su Poder de Detección al compararse con los resultados anteriores de cada uno. Asimismo, las medidas de error aumentan al promediar los resultados de los 15.000 modelos estimados. Al compararse entre ellos, el Modelo Híbrido al 1 % de Lealtad posee mejor desempeño en su Poder de Detección (con diferencias menores a un 0.2 %) y en sus niveles de Falsas Alarmas.

En base a los resultados obtenidos, se generan intervalos de confianza a un 95 % en ambos modelos y para cada una de las métricas de desempeño:

- Modelo Completo: Los rangos obtenidos son:

$$\begin{aligned}
 \text{Poder de Detección} &\in [63,12 \% - 63,32 \%] \\
 \text{Falsas Alarmas} &\in [10,05 \% - 10,19 \%] \\
 \text{Error Tipo I} &\in [1,19 \% - 1,21 \%]
 \end{aligned} \tag{6.1}$$

- Modelo Híbrido al 1 % de Lealtad: Los rangos obtenidos son:

$$\begin{aligned}
 \text{Poder de Detección} &\in [63,30 \% - 63,48 \%] \\
 \text{Falsas Alarmas} &\in [8,85 \% - 8,97 \%] \\
 \text{Error Tipo I} &\in [1,21 \% - 1,23 \%]
 \end{aligned} \tag{6.2}$$

Se puede concluir que, en términos del Poder de Detección, si bien el Modelo Híbrido tiene un mayor valor promedio, los intervalos de confianza generados no son disjuntos, por lo que no se puede concluir si este posee un mayor Poder de Detección que el Modelo Completo. Por otra parte, el Modelo Híbrido tiene un menor nivel de Falsas Alarmas en comparación al Modelo Completo, y esta diferencia es significativa. A pesar de esto, el Modelo Completo presenta menores niveles de Error Tipo I, también de forma significativa.

## 6.7. Discusión

En base a los resultados obtenidos, y en conjunto con los análisis realizados a lo largo del presente Trabajo, en esta sección se discute acerca de las consecuencias y el alcance que tienen estos resultados .

Como se describe en la Sección 4.2, inicialmente se postula que al segmentar los clientes en base a sus niveles de lealtad hacia los productos, y solo observando su comportamiento, las

estimaciones podrían generar una menor cantidad de quiebres en comparación con el modelo sin segmentación. Esto trae como consecuencia una disminución en las medidas de error (y también una disminución en su Poder de Detección). Como se observa en las Figuras 6.4 y 6.5 se cumple que, para el Modelo con ID y para los Modelos con Lealtad del 1% y 2%, las estimaciones de error son menores que para el Modelo Completo.

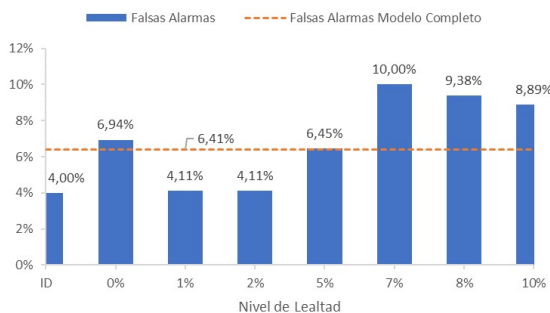


Figura 6.4: Comparación de Falsas Alarmas entre modelos para 14 productos

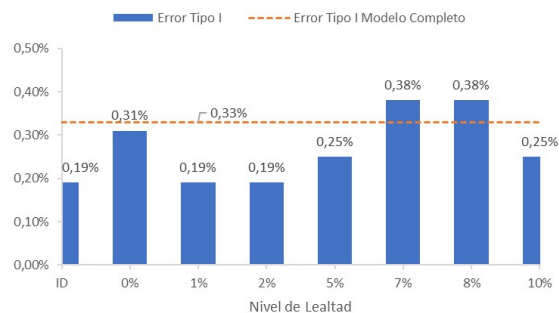


Figura 6.5: Comparación de Error Tipo I entre modelos para 14 productos

Sin embargo, este comportamiento no se replica para el Modelo al 0% de Lealtad, donde las Falsas Alarmas son levemente mayores al Modelo Completo. También se observa que, a medida que aumenta el nivel de lealtad, los estimadores de error alcanzan valores más altos que los entregados por el Modelo Completo. Se realiza el análisis sobre los 14 productos estudiados en el Apéndice C, ya que sobre ellos se estiman modelos con más variedad de segmentaciones.

Estos resultados junto con los descritos en el Apéndice C.1, que muestran una disminución progresiva en el Poder de Detección de los modelos a medida que se aumenta el nivel de lealtad sobre el cual se segmentan los clientes, sugieren que realizar este tipo de segmentación no mejora los rendimientos a nivel general sino que se limita a generar una potencial ganancia en términos de medidas de error a nivel local.

Es importante señalar que no se estiman posteriores modelos con lealtades mayores al 15% (que obtiene un Poder de Predicción de tan solo un 18.33%, ver Apéndice F) ya que se descarta la posibilidad de que sus modelos alcancen buenos indicadores de desempeño. Esto se argumenta por dos razones: la primera corresponde a la disminución progresiva del desempeño de los modelos al aumentar el nivel de lealtad (como ya se explicó), y la segunda debido a la pérdida de información producto de la gran disminución de clientes al exigir mayores niveles de lealtad a los productos (detallado en la Sección 5.8).

### 6.7.1. Desempeño de los Modelos

Los distintos análisis realizados en el presente capítulo dan cuenta de cómo varía el desempeño de los modelos estimados segmentando bajo distintos criterios. En primer lugar, al anali-

zar el modelo utilizando la segmentación en base a la definición usual de lealtad se encuentran variados resultados, ya descritos con anterioridad.

Al testear modelos añadiendo las nociones de frecuencia y amplitud descritas en al Sección 5.2.4, se encuentran resultados negativos en términos del desempeño de los modelos. Esto se puede explicar porque, al exigir a los clientes un nivel de frecuencia de compra, esto lleva implícita la exigencia de que hayan comprado el producto al menos dos veces (para establecer una medida de distancia de compra). Con esta exigencia (que resulta ser la mínima al hablar de frecuencia), el universo promedio de clientes por producto se reduce a 15.861.

Sin embargo, al añadir la variación híbrida de los modelos estos mejoran en su Poder de Detección. El mejor resultado se aprecia en el Modelo Híbrido al 1 % de Lealtad, que mejora su desempeño tanto en la predicción de quiebres como en sus niveles de error. Este resultado encontrado para los 58 productos iniciales se logra mantener en la elección final de productos. En particular, destaca su desempeño en términos de los niveles de error para productos de Alta Incidencia, y su desempeño general para los productos de Incidencia Media, donde resulta ser el modelo con mejores resultados dentro de los estimados.



# Capítulo 7

## Conclusiones

### 7.1. Conclusiones

El desarrollo del presente Trabajo de Título se centró en la búsqueda y generación de modelos que permitieran aumentar el desempeño de los sistemas de reposición, a la hora de identificar quiebres de stock en góndola. Los problemas relacionados con quiebres de stock tienen implicancias económicas directas e indirectas para los supermercados, por lo que refinarlo representa un desafío de suma importancia para la industria. Se propone estudiar el problema otorgando una especial importancia al comportamiento de compra de los clientes, en especial a aquellos que poseen lo que se denomina como *lealtad* a un producto. Esta definición se adopta en cuanto clasifica a los clientes según sus niveles de incidencia sobre los productos, y en base a ellos, se estudian distintos segmentos particulares de clientes.

El modelo propuesto consistió en un Modelo de Markov Oculto, el cual se evaluó sobre un conjunto de 47 productos, obteniendo diversos resultados. El mejor de ellos se obtuvo analizando a aquellos clientes que hubieran presentado los productos en estudio en al menos el 1% de sus boletas (o clientes con Nivel de Lealtad de 1%), en su variación híbrida. Los resultados arrojaron un Poder de Predicción de quiebres de 65.42%, con un 6.15% de Falsas Alarmas y un 0.29% de Error Tipo I, igualando en poder predictivo al Modelo con información completa, pero mejorando los niveles de error.

Los buenos resultados obtenidos abren la pregunta de la capacidad de extrapolación que estos tienen para casos más generales. Desde el punto de vista estadístico, se discute sobre las limitaciones de los resultados y se generan estimadores para entender los errores propios de la muestra. Los resultados obtenidos no permiten concluir a favor o en contra del modelo en términos de si resulta mejor que el Modelo con información completa, pero confirma que sus buenos resultados pueden ser extrapolables.

## 7.2. Trabajo Futuro

En cuanto al modelo y su calibración, una limitación importante se encuentra en la acotada información del real estado de góndola con la que se mide el desempeño de los modelos (que correspondió a menos del 3% del total de días analizados). A futuro, podrían considerarse tomar mediciones en distintos meses, de manera de contar con información a lo largo del período en esto. Con esto, se podrían utilizar técnicas de *cross-validation* para disminuir el nivel de sobreajuste o de aleatoriedad de la muestra (Refaeilzadeh et al., 2009).

Desde el punto de vista comercial, el presente trabajo abre puertas a futuras estimaciones del impacto que podría tener analizar a ciertos segmentos de los clientes (que, como se vio, explican el comportamiento de los productos con similar precisión en comparación a analizar a todos los clientes) sobre las estrategias que supermercados y marcas desarrollan. A modo de ejemplo, se podría pensar en el desarrollo de una campaña de marketing focalizada al menor segmento de clientes que presente un desempeño mayor a cierta cota en sus modelos de predicción de quiebres.

Finalmente, en el desarrollo del presente trabajo se considera una misma cantidad de estados posibles para todos los productos. Sin embargo, existe la posibilidad de que para ciertos productos (por ejemplo, de baja incidencia) el mejor ajuste sea con un número menor de estados, asumiendo una menor movilidad en los niveles de demanda. El mismo criterio se podría considerar al aumentar los niveles de lealtad de los clientes, ya que a medida que este aumenta, la cantidad de ventas observadas disminuye.

# Bibliografía

- C. Lamb, Hair Jr. J., and C. McDaniel. *Essential of Marketing*. Neil Marquardt, OH, USA, 2009.
- Diario Online Estrategia. Resultados 2017: Tres grandes del retail concentraron 94% de las ganancias y 95% de los ingresos del sector [en línea]. <<http://www.estrategia.cl/texto-diario/mostrar/1045297>>, 2018. [Consulta: 2018-11-10].
- J. Von Hausen. *Efecto de disponibilidad de variedad de productos en góndola en el comportamiento de clientes*. Tesis para optar al grado de Magíster en Gestión de Operaciones, Universidad de Chile, 2014.
- América Retail. Ventas en tiendas de conveniencia duplican en crecimiento a los supermercados [en línea]. <<http://xurl.es/9vkqm>>, 2018. [Consulta: 2018-11-10].
- Online Feller Rate. Informe de clasificación, walmart chile [en línea]. <<http://www.feller-rate.cl/general2/corporaciones/walmart1610.pdf>>, 2016. [Consulta: 2018-09-08].
- Cencosud. Memoria anual 2016 [en línea]. <[http://s2.q4cdn.com/740885614/files/doc\\_financials/2016/AR/Memoria-Anual-2016.pdf](http://s2.q4cdn.com/740885614/files/doc_financials/2016/AR/Memoria-Anual-2016.pdf)>, 2017. [Consulta: 2018-09-08].
- SMU. Memoria anual 2016 [en línea]. <[https://www.smu.cl/wp-content/files\\_mf/1491566288MemoriaAnualSMU2016.pdf](https://www.smu.cl/wp-content/files_mf/1491566288MemoriaAnualSMU2016.pdf)>, 2017. [Consulta: 2018-09-08].
- Tottus. Reporte de sostenibilidad 2016 [en línea]. <[http://www.tottus.cl/static/pdf/reporte/reporte\\_sostenibilidad\\_2016.pdf](http://www.tottus.cl/static/pdf/reporte/reporte_sostenibilidad_2016.pdf)>, 2017. [Consulta: 2018-11-14].
- T. Gruen and D. Corsten. *A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry*. Grocery Manufacturers Association, 2007.
- T. Gruen, D. Corsten, and S. Dharadwaj. *Retail out-of-stocks: A world-wide examination of extent, causes and consumer responses*. Grocery Manufacturers of America 11-19, 2002.
- M. Bosch, R. H. Hilger, and A. Shilkrut. *La medición de faltantes en góndola*. Working Paper, Universidad de Chile, 2005.
- L. Gatica. *Estudio del comportamiento de los clientes frente a un quiebre de stock y su impacto económico en un supermercado*. Memoria para optar al título de Ingeniera Civil Industrial, Universidad de Chile, 2014.

- M. Olivares and A. Schilkrut. Smart stores: Aplicaciones de iot e inteligencia artificial para mejorar la ejecucion operacional en retail [en línea]. <<http://www.dii.uchile.cl/wp-content/uploads/2018/08/charla-BAFI-Julio2018-con-Zippedi.pdf>>, 2018. [Consulta: 2018-11-10].
- B. Hardgrave, M. Waller, and R. Miller. *RFID's Impact on Out of Stocks: A Sales Velocity Analysis*. Sam M. Walton College of Business, 2006.
- CERET. *Sistema osca 2.0 - control estadístico de procesos*. Documento de Trabajo, Departamento de Ingeniería Civil Industria, Universidad de Chile, 2014.
- T. Figueroa. *Modelo predictivo de quiebres de stock en un supermercado*. Memoria para optar al título de Ingeniera Civil Industrial, Universidad de Chile, 2009.
- L. Chen. *Fixing phantom stockouts: Optimal data-driven shelf inspection policies*. Working Paper, Fuqua School of Business, Duke University, Durham, 2014.
- M. Fisher and A. Raman. *The New Science of Retailing: How Analytics are Transforming Supply Chains and Improving Performance*. Harvard Business Press, 2010.
- Y. Kang and S. Gershwin. *Information inaccuracy in inventory systems: Stock loss and stockout*. IIE Trans. 37 843-859., 2005.
- N. Bhan. *Inventory estimation from transactions via hidden markov models*. Ph.D. thesis, Massachusetts Institute of Technology, 2015.
- R. Arriagada. *Metodología para detección de quiebres de stock en el retail*. Tesis para optar al grado de Magíster en Gestión de Operaciones, Universidad de Chile, 2018.
- R. Montoya and C. González. *A Hidden Markov Model to Detect On-Shelf Out-of-Stocks Using Point of Sales Data*. Manufacturing and Service Operations Management, 2018. Forthcoming.
- W. Zucchini and I. MacDonald. *Hidden Markov models for time series: an introduction using R*. CRC press, 2009.
- C. González. *Identificación de quiebres de stock en góndola utilizando cadenas de markov ocultas*. Memoria para optar al título de Ingeniera Civil Industrial, Universidad de Chile, 2016.
- I. Yildirim. *Bayesian Inference: Metropolis-Hastings Sampling*. Dept. of Brain and Cognitive Sciences, Univ. of Rochester, Rochester, NY, 2012.
- P. Refaeilzadeh, L. Tang, and H. Liu. *Cross-Validation*. Encyclopedia of Database Systems, Springer, 2009.

# Apéndice A

## Grupos de Incidencia por Producto

Producto	Incidencia	Rango*	Grupo de Incidencia
Leche Entera	1.81 %	2.4	Alta
Gaseosa Cola	1.63 %	2.4	Alta
Toalla de Papel	1.32 %	3	Alta
Crackeler	1.17 %	2.9	Alta
Lavalozas	1.04 %	3.9	Alta
Fósforos	0.96 %	4.4	Alta
Té	0.93 %	2.7	Alta
Leche Chocolate	0.81 %	2.4	Alta
Nescafé	0.79 %	2.6	Alta
Fideo	0.79 %	4	Alta
Néctar	0.79 %	2.2	Alta
Avena	0.75 %	2	Alta
Palmitos	0.72 %	4.3	Alta
Vienesas de Ave	0.66 %	3.4	Alta
Margarina	0.59 %	2.8	Media
Cerveza Escudo	0.56 %	6.8	Media
Mayonesa Hellmanns	0.56 %	3	Media
Puré de Papas	0.55 %	5	Media
Jugo en Polvo	0.51 %	5.7	Media
Queso Gauda	0.51 %	3.9	Media
Promedio	0.42 %	2.3	Media
Gaseosa Naranja	0.41 %	3.4	Media
Quesillo	0.39 %	3	Media
Vienesas	0.38 %	2.9	Media
Azúcar	0.37 %	4.2	Media
Leche sin Lactosa	0.37 %	2.5	Media
Camarón	0.37 %	2.3	Media
Chorizo	0.36 %	6.2	Media
Atún	0.35 %	2.2	Media
Chocolate Trencito	0.33 %	3.7	Media

Producto	Incidencia	Rango	Grupo de Incidencia
Papas Stax	0.31 %	1.9	Media
Jurel	0.29 %	3.8	Baja
Lasaña	0.22 %	3.2	Baja
Papas Duquesas	0.21 %	3.1	Baja
Chocapic	0.20 %	5.3	Baja
Salmón	0.19 %	3.9	Baja
Aceite Soya	0.17 %	3.1	Baja
Gaseosa Cola Light	0.16 %	12	Baja
Nido Polvo	0.15 %	3.5	Baja
Caluga Suny	0.15 %	1.8	Baja
Pisco	0.15 %	3.8	Baja
Papel Higiénico	0.14 %	7.8	Baja
Alimento para Perro	0.14 %	11.1	Baja
Queso Fresco	0.13 %	8	Baja
Maizena	0.13 %	2.6	Baja
Tortilla	0.13 %	9.4	Baja
Mayonesa Marca Propia	0.11 %	2.5	Baja
Shampoo	0.11 %	3.2	Baja
Café	0.11 %	3	Baja
Cerveza S/Alcohol	0.11 %	4.7	Baja
Champagne	0.10 %	3	Baja
Salchichas	0.09 %	4.4	Baja
Cerveza Stella Artois	0.08 %	3.4	Baja
Leche en Polvo	0.07 %	4.1	Baja
Detergente Líquido	0.05 %	7.4	Baja
Aceite Oliva	0.03 %	4.2	Baja
Nidal 1	0.03 %	3.4	Baja
Pañal	0.02 %	3	Baja
Pañal Activsec	0.02 %	2.4	Baja

\*Rango=Max(incidencia)/Min(incidencia) entre los locales en estudio.

Tabla A.1: Segmentación de productos según nivel de incidencia

## Apéndice B

### Modelos por Producto para todos los productos

	Quiebres	Alertas	TP	FP	FN	TN	Total	Error Tipo I	Falsas Alarmas	Detección	% de días sin boletas
Atún (0.35%)	15	14	12	2	3	98	115	2.00%	14.29%	80.00%	4.85%
Azúcar (0.37%)	7	6	6	-	1	118	125	0%	0%	85.71%	4.90%
Camarón (0.37%)	5	5	5	-	-	108	113	0%	0%	100.00%	6.25%
Chocolate (0.33%)	41	18	18	-	23	44	85	0%	0%	43.90%	18.89%
Gaseosa naranja (0.41%)	1	1	-	1	1	128	130	0.78%	100.00%	0%	4.17%
Jugo polvo (0.51%)	7	3	3	-	4	98	105	0%	0%	42.86%	11.13%
Jurel (0.29%)	12	11	11	-	1	116	128	0%	0%	91.67%	8.17%
Lasaña (0.22%)	12	2	2	-	10	102	114	0%	0%	16.67%	15.85%
Lavaloza (1.04%)	6	7	6	1	-	118	125	0.84%	14.29%	100.00%	4.33%
Leche Chocolate (0.81%)	5	3	3	-	2	121	126	0%	0%	60.00%	1.81%
Toallas de papel (1.32%)	1	2	1	1	-	118	120	0.84%	50.00%	100.00%	1.44%
Papas duquesas (0.21%)	1	0	-	-	1	121	122	0%	0%	0%	13.90%
Quesillo (0.39%)	3	2	2	-	1	120	123	0%	0%	66.67%	15.11%
Té (0.93%)	4	1	1	-	3	113	117	0%	0%	25.00%	4.08%
<b>Aceite Maravilla y Soya (0.17%)</b>	1	99	1	98	-	21	120	82.35%	98.99%	100.00%	55.59%
Aceite Oliva Extra Virgen (0.03%)	14	5	-	5	14	89	108	5.32%	100.00%	0%	43.46%
Alimento para Perro (0.14%)	0	0	-	-	-	129	129	0%	0%	0%	12.25%
Avena Instantánea (0.75%)	0	0	-	-	-	129	129	0%	0%	0%	1.01%
Café Instantáneo (0.11%)	6	4	4	-	2	128	134	0%	0%	66.67%	15.14%
Caluga Suny (0.15%)	12	4	3	1	9	105	118	0.94%	25.00%	25.00%	11.31%
<b>Cereal Chocapic (0.20%)</b>	1	119	1	118	-	10	129	92.19%	99.16%	100.00%	35.33%
Cerveza Escudo (0.56%)	0	0	-	-	-	134	134	0%	0%	0%	4.28%
Cerveza Cristal S/Alcohol (0.11%)	3	0	-	-	3	109	112	0%	0%	0%	27.57%
Cerveza Stella Artois (0.08%)	13	11	9	2	4	106	121	1.85%	18.18%	69.23%	33.68%
Champagne Demisec (0.10%)	4	1	-	1	4	117	122	0.85%	100.00%	0%	43.32%
Chorizo Parrillero (0.36%)	1	1	-	1	1	135	137	0.74%	100.00%	0%	3.73%
Detergente Líquido (0.05%)	39	13	13	-	26	76	115	0%	0%	33.33%	40.99%
Fideo Espiral (0.79%)	0	0	-	-	-	126	126	0%	0%	0%	1.40%
Fósforos Pequeños (0.96%)	0	0	-	-	-	137	137	0%	0%	0%	1.14%
Galleta Crackeler (1.17%)	1	0	-	-	1	135	136	0%	0%	0%	3.71%
Gaseosa Cola (1.63%)	0	0	-	-	-	136	136	0%	0%	0%	0.92%
Gaseosa Cola Light (0.16%)	10	3	3	-	7	101	111	0%	0%	30.00%	20.93%
Leche en Polvo Entera (0.07%)	5	1	-	1	5	111	117	0.89%	100.00%	0%	20.65%
Leche Entera (1.81%)	1	0	-	-	1	134	135	0%	0%	0%	1.81%
Nidal 1 Fórmula (0.03%)	21	0	-	-	21	42	63	0%	0%	0%	38.58%
Leche sin Lactosa (0.37%)	0	0	-	-	-	133	133	0%	0%	0%	7.21%
Maizena (0.13%)	22	15	15	-	7	84	106	0%	0%	68.18%	16.35%
Margarina Qualy (0.59%)	6	1	1	-	5	112	118	0%	0%	16.67%	6.85%
<b>Mayonesa Marca Propia (0.11%)</b>	15	95	13	82	2	6	103	93.18%	86.32%	86.67%	57.65%
Mayonesa Hellmanns (0.56%)	0	0	-	-	-	132	132	0%	0%	0%	3.41%
Néctar Piña (0.79%)	0	0	-	-	-	136	136	0%	0%	0%	1.67%
Nescafé Trad. Tarro (0.79%)	0	0	-	-	-	136	136	0%	0%	0%	3.73%
Nido Polvo (0.15%)	11	9	9	-	2	99	110	0%	0%	81.82%	11.17%
Palmitos Rodaja (0.72%)	5	1	1	-	4	120	125	0%	0%	20.00%	11.82%
Pañal Recién Nacido (0.02%)	1	0	-	-	1	111	112	0%	0%	0%	57.64%
Pañal Activsec (0.02%)	81	31	29	2	52	39	122	4.88%	6.45%	35.80%	55.76%
Papa Stax Original (0.31%)	135	123	121	2	14	-	137	100.00%	1.63%	89.63%	14.91%
Papel Higiénico (0.14%)	2	0	-	-	2	128	130	0%	0%	0%	18.20%
Pisco 35 (0.15%)	3	0	-	-	3	120	123	0%	0%	0%	15.66%
Puré de Papas (0.55%)	0	0	-	-	-	129	129	0%	0%	0%	5.84%
Queso Fresco (0.13%)	27	110	24	86	3	7	120	92.47%	78.18%	88.89%	38.24%
Queso Gauda Laminado (0.51%)	5	1	1	-	4	112	117	0%	0%	20.00%	6.21%
Salchichas (0.09%)	9	9	8	1	1	114	124	0.87%	11.11%	88.89%	18.23%
Salmón (0.19%)	3	6	2	4	1	123	130	3.15%	66.67%	66.67%	14.88%
Shampoo Colorvive (0.11%)	6	6	5	1	1	114	121	0.87%	16.67%	83.33%	11.88%
Tortilla Burrito (0.13%)	5	1	1	-	4	89	94	0%	0%	20.00%	17.43%
Vienesas JK (0.38%)	0	0	-	-	-	137	137	0%	0%	0%	1.51%
Vienesas de Ave (0.66%)	2	0	-	-	2	114	116	0%	0%	0%	13.17%
<b>Total</b>	<b>590</b>	<b>744</b>	<b>334</b>	<b>410</b>	<b>256</b>	<b>6,028</b>	<b>7,028</b>			<b>Promedio</b>	<b>16.23%</b>

Tabla B.1: Resultado del Modelo Completo para los 58 productos

	Quiebres	Alertas	TP	FP	FN	TN	Total	Error Tipo I	Falsas Alarmas	Detección	% de días sin boletas
Atún (0.35%)	15	16	13	3	2	104	122	2.80%	18.75%	86.67%	11.41%
Azúcar (0.37%)	7	6	6	-	1	123	130	0%	0%	85.71%	11.45%
Camarón (0.37%)	5	5	5	-	-	114	119	0%	0%	100.00%	12.71%
Chocolate Trencito (0.33%)	41	18	18	-	23	44	85	0%	0%	43.90%	24.48%
Gaseosa Naranja (0.41%)	1	1	-	1	1	134	136	0.74%	100.00%	0%	10.77%
Jugo en Polvo (0.51%)	7	3	3	-	4	103	110	0%	0%	42.86%	17.25%
Jurel (0.29%)	12	11	11	-	1	122	134	0%	0%	91.67%	14.50%
Lasaña (0.22%)	12	4	3	1	9	105	118	0.94%	25.00%	25.00%	21.64%
Lavalozas (1.04%)	6	6	6	-	-	125	131	0%	0%	100.00%	10.92%
Leche Chocolate (0.81%)	5	3	3	-	2	126	131	0%	0%	60.00%	8.57%
Toalla de Papel (1.32%)	1	1	1	-	-	122	123	0%	0%	100.00%	8.23%
Papas Duquesas (0.21%)	1	1	-	1	1	127	129	0.78%	100.00%	0%	19.83%
Quesillo (0.39%)	3	2	2	-	1	125	128	0%	0%	66.67%	20.96%
Té (0.93%)	4	1	1	-	3	119	123	0%	0%	25.00%	10.68%
<b>Aceite Soya (0.17%)</b>	1	99	1	98	-	21	120	82.35%	98.99%	100.00%	55.59%
Aceite Oliva (0.03%)	14	0	-	-	14	94	108	0%	0%	0%	47.36%
Alimento para Perro (0.14%)	0	0	-	-	-	129	129	0%	0%	0%	18.29%
Avena (0.75%)	0	0	-	-	-	129	129	0%	0%	0%	7.83%
Café (0.11%)	6	4	4	-	2	128	134	0%	0%	66.67%	20.98%
Caluga Suny (0.15%)	12	1	1	-	11	106	118	0%	0%	8.33%	17.42%
<b>Chocapic (0.20%)</b>	1	119	1	118	-	10	129	92.19%	99.16%	100.00%	35.33%
Cerveza Escudo (0.56%)	0	0	-	-	-	134	134	0%	0%	0%	10.87%
Cerveza S/Alcohol (0.11%)	3	0	-	-	3	109	112	0%	0%	0%	32.56%
Cerveza Stella Artois (0.08%)	13	2	-	2	13	106	121	1.85%	100.00%	0%	38.25%
Champagne (0.10%)	4	0	-	-	4	118	122	0%	0%	0%	47.23%
Chorizo (0.36%)	1	0	-	-	1	136	137	0%	0%	0%	10.36%
Detergente Líquido (0.05%)	39	11	11	-	28	76	115	0%	0%	28.21%	45.05%
Fideo (0.79%)	0	0	-	-	-	126	126	0%	0%	0%	8.19%
Fósforos (0.96%)	0	0	-	-	-	137	137	0%	0%	0%	7.95%
Crackeler (1.17%)	1	0	-	-	1	135	136	0%	0%	0%	10.34%
Gaseosa Cola (1.63%)	0	0	-	-	-	136	136	0%	0%	0%	7.74%
Gaseosa Cola Light (0.16%)	10	4	4	-	6	101	111	0%	0%	40.00%	26.38%
Leche en Polvo (0.07%)	5	3	1	2	4	110	117	1.79%	66.67%	20.00%	26.12%
Leche Entera (1.81%)	1	0	-	-	1	134	135	0%	0%	0%	8.57%
Nidal 1 (0.03%)	21	0	-	-	21	42	63	0%	0%	0%	42.81%
Leche sin Lactosa (0.37%)	0	0	-	-	-	133	133	0%	0%	0%	13.60%
Maizena (0.13%)	22	11	11	-	11	84	106	0%	0%	50.00%	22.11%
Margarina (0.59%)	6	1	1	-	5	112	118	0%	0%	16.67%	13.26%
<b>Mayonesa Marca Propia (0.11%)</b>	15	95	13	82	2	6	103	93.18%	86.32%	86.67%	57.65%
Mayonesa Hellmanns (0.56%)	0	0	-	-	-	132	132	0%	0%	0%	10.06%
Néctar (0.79%)	0	0	-	-	-	136	136	0%	0%	0%	8.44%
Nescafé (0.79%)	0	0	-	-	-	136	136	0%	0%	0%	10.36%
Nido Polvo (0.15%)	11	7	7	-	4	99	110	0%	0%	63.64%	17.29%
Palmitos (0.72%)	5	1	1	-	4	120	125	0%	0%	20.00%	17.89%
Pañal (0.02%)	1	0	-	-	1	111	112	0%	0%	0%	60.55%
Pañal Activsec (0.02%)	81	13	13	-	68	41	122	0%	0%	16.05%	58.81%
Papas Stax (0.31%)	135	122	120	2	15	-	137	100.00%	1.64%	88.89%	20.77%
Papel Higiénico (0.14%)	2	0	-	-	2	128	130	0%	0%	0%	23.84%
Pisco (0.15%)	3	2	1	1	2	119	123	0.83%	50.00%	33.33%	21.47%
Puré de Papas (0.55%)	0	0	-	-	-	129	129	0%	0%	0%	12.32%
Queso Fresco (0.13%)	27	101	21	80	6	13	120	86.02%	79.21%	77.78%	42.49%
Queso Gauda (0.51%)	5	0	-	-	5	112	117	0%	0%	0%	12.67%
Salchichas (0.09%)	9	5	4	1	5	114	124	0.87%	20.00%	44.44%	23.86%
Salmón (0.19%)	3	5	2	3	1	124	130	2.36%	60.00%	66.67%	20.75%
Shampoo (0.11%)	6	5	4	1	2	114	121	0.87%	20.00%	66.67%	17.95%
Tortilla (0.13%)	5	2	-	2	5	87	94	2.25%	100.00%	0%	23.11%
Vienesas (0.38%)	0	0	-	-	-	137	137	0%	0%	0%	8.29%
Vienesas de Ave (0.66%)	2	0	-	-	2	114	116	0%	0%	0%	19.15%
<b>Total</b>	590	691	293	398	297	6,111	7,099			<b>Promedio</b>	21.86%

Tabla B.2: Resultado del Modelo con ID para los 58 productos



	Quiebres	Alertas	TP	FP	FN	TN	Total	Error Tipo I	Falsas Alarmas	Detección	% de días sin boletas
Atún (0.35%)	15	15	13	2	2	105	122	1.87%	13.33%	86.67%	11.41%
Azúcar (0.37%)	7	6	6	-	1	123	130	0%	0%	85.71%	11.45%
Camarón (0.37%)	5	5	5	-	-	114	119	0%	0%	100.00%	12.71%
Chocolate Trencito (0.33%)	41	19	19	-	22	44	85	0%	0%	46.34%	24.48%
Gaseosa Naranja (0.41%)	1	0	-	-	1	135	136	0%	0%	0%	10.77%
Jugo en Polvo (0.51%)	7	3	3	-	4	103	110	0%	0%	42.86%	17.25%
Jurel (0.29%)	12	11	11	-	1	122	134	0%	0%	91.67%	14.50%
Lasaña (0.22%)	12	4	3	1	9	105	118	0.94%	25.00%	25.00%	21.64%
Lavalozas (1.04%)	6	6	6	-	-	125	131	0%	0%	100.00%	10.92%
Leche Chocolate (0.81%)	5	3	3	-	2	126	131	0%	0%	60.00%	8.57%
Toalla de Papel (1.32%)	1	1	1	-	-	122	123	0%	0%	100.00%	8.23%
Papas Duquesas (0.21%)	1	0	-	-	1	128	129	0%	0%	0%	19.83%
Quesillo (0.39%)	3	2	2	-	1	125	128	0%	0%	66.67%	20.96%
Té (0.93%)	4	1	1	-	3	119	123	0%	0%	25.00%	10.68%
<b>Aceite Soya (0.17%)</b>	1	99	1	98	-	21	120	82.35%	98.99%	100.00%	55.59%
Aceite Oliva (0.03%)	14	1	-	1	14	93	108	1.06%	100.00%	0%	47.36%
Alimento para Perro (0.14%)	0	0	-	-	-	129	129	0%	0%	0%	18.29%
Avena (0.75%)	0	0	-	-	-	129	129	0%	0%	0%	7.83%
Café (0.11%)	6	4	4	-	2	128	134	0%	0%	66.67%	20.98%
Caluga Suny (0.15%)	12	1	1	-	11	106	118	0%	0%	8.33%	17.42%
<b>Chocapic (0.20%)</b>	1	119	1	118	-	10	129	92.19%	99.16%	100.00%	35.33%
Cerveza Escudo (0.56%)	0	0	-	-	-	134	134	0%	0%	0%	10.87%
Cerveza S/Alcohol (0.11%)	3	0	-	-	3	109	112	0%	0%	0%	32.56%
Cerveza Stella Artois (0.08%)	13	1	-	1	13	107	121	0.93%	100.00%	0%	38.25%
Champagne (0.10%)	4	3	-	3	4	115	122	2.54%	100.00%	0%	47.23%
Chorizo (0.36%)	1	0	-	-	1	136	137	0%	0%	0%	10.36%
Detergente Líquido (0.05%)	39	18	12	6	27	70	115	7.89%	33.33%	30.77%	45.05%
Fideo (0.79%)	0	0	-	-	-	126	126	0%	0%	0%	8.19%
Fósforos (0.96%)	0	0	-	-	-	137	137	0%	0%	0%	7.95%
Crackeler (1.17%)	1	0	-	-	1	135	136	0%	0%	0%	10.34%
Gaseosa Cola (1.63%)	0	0	-	-	-	136	136	0%	0%	0%	7.74%
Gaseosa Cola Light (0.16%)	10	4	4	-	6	101	111	0%	0%	40.00%	26.38%
Leche en Polvo (0.07%)	5	0	-	-	5	112	117	0%	0%	0%	26.12%
Leche Entera (1.81%)	1	0	-	-	1	134	135	0%	0%	0%	8.57%
Nidal 1 (0.03%)	21	0	-	-	21	42	63	0%	0%	0%	42.81%
Leche sin Lactosa (0.37%)	0	0	-	-	-	133	133	0%	0%	0%	13.60%
Maizena (0.13%)	22	15	15	-	7	84	106	0%	0%	68.18%	22.11%
Margarina (0.59%)	6	1	1	-	5	112	118	0%	0%	16.67%	13.26%
<b>Mayonesa Marca Propia (0.11%)</b>	15	95	13	82	2	6	103	93.18%	86.32%	86.67%	57.65%
Mayonesa Hellmanns (0.56%)	0	0	-	-	-	132	132	0%	0%	0%	10.06%
Néctar (0.79%)	0	0	-	-	-	136	136	0%	0%	0%	8.44%
Nescafé (0.79%)	0	0	-	-	-	136	136	0%	0%	0%	10.36%
Nido Polvo (0.15%)	11	8	8	-	3	99	110	0%	0%	72.73%	17.29%
Palmitos (0.72%)	5	1	1	-	4	120	125	0%	0%	20.00%	17.89%
Pañal (0.02%)	1	0	-	-	1	111	112	0%	0%	0%	60.55%
Pañal Activsec (0.02%)	81	0	-	-	81	41	122	0%	0%	0%	58.81%
Papas Stax (0.31%)	135	122	120	2	15	-	137	100.00%	1.64%	88.89%	20.77%
Papel Higiénico (0.14%)	2	0	-	-	2	128	130	0%	0%	0%	23.84%
Pisco (0.15%)	3	1	-	1	3	119	123	0.83%	100.00%	0%	21.47%
Puré de Papas (0.55%)	0	0	-	-	-	129	129	0%	0%	0%	12.32%
Queso Fresco (0.13%)	27	101	21	80	6	13	120	86.02%	79.21%	77.78%	42.49%
Queso Gauda (0.51%)	5	0	-	-	5	112	117	0%	0%	0%	12.67%
Salchichas (0.09%)	9	5	4	1	5	114	124	0.87%	20.00%	44.44%	23.86%
Salmón (0.19%)	3	1	-	1	3	126	130	0.79%	100.00%	0%	20.75%
Shampoo (0.11%)	6	5	4	1	2	114	121	0.87%	20.00%	66.67%	17.95%
Tortilla (0.13%)	5	1	1	-	4	89	94	0%	0%	20.00%	23.11%
Vienesas (0.38%)	0	0	-	-	-	137	137	0%	0%	0%	8.29%
Vienesas de Ave (0.66%)	2	0	-	-	2	114	116	0%	0%	0%	19.15%
<b>Total</b>	590	682	284	398	306	6,111	7,099			<b>Promedio</b>	21.86%

Tabla B.3: Resultados del Modelo Híbrido al 1% de Lealtad para los 58 productos

# Apéndice C

## Análisis de Resultados para Niveles de Lealtad

A lo largo de esta sección se presentan los resultados de distintos modelos de predicción de Quiebres de Stock, y para sus efectos, se trabaja con la información de los 14 productos estudiados por Montoya and González (2018).

Como se analizó en la Sección 5.1.1, casi un 75 % de los datos transaccionales cuentan con identificación (columna que identifica al RUT no vacía). Estas tres cuartas partes son la base desde donde se obtiene información realmente útil sobre los clientes, ya que los datos no identificables no pueden ser asociados a lealtad de ningún tipo (ya que no se conoce quién compró la canasta de productos).

Se procede a correr el modelo considerando este filtro sobre las transacciones (que desde ahora se denominará *Modelo con ID*), comparando los resultados obtenidos con el Modelo Completo (modelo estimado en la sección anterior). La Tabla C.1 resume los resultados:

<b>Modelo</b>	<b>Error Tipo I</b>	<b>Falsas Alarmas</b>	<b>Poder de Detección</b>
Modelo Completo	0.33 %	6.41 %	60.83 %
Modelo con ID	0.19 %	4.00 %	60.00 %

Tabla C.1: Desempeño Modelo con ID en comparación a Modelo Completo

Se observa que el Modelo con ID posee un menor Poder de Detección que el Modelo Inicial. Sin embargo, su rendimiento mejora en cuanto a Falsas Alarmas y Error Tipo I. Analizando el detalle por producto en la Tabla C.2, en términos de Poder de Detección, los resultados son los mismos para 10 de los 14 productos. Además, en casi la totalidad de ellos mejoran los rendimientos relacionados a errores por Falsos Positivos.

Si bien la predicción de este modelo es levemente peor en términos de predicción, se mantendrá preliminarmente como referencia para los resultados futuros, ya que para generar los segmentos de clientes se realizan filtros sobre la base de datos completa de clientes (que justamente corresponde a los datos analizados en el Modelo con ID).

Producto (Incidencia)	Quebres	Alertas	Error Tipo I	Falsas Alarmas	Detección
Atún (0.35 %)	15	16	2.80 %	18.75 %	86.67 %
Azúcar (0.37 %)	7	6	0 %	0 %	85.71 %
Camarón (0.37 %)	5	5	0 %	0 %	100.00 %
Chocolate (0.33 %)	41	18	0 %	0 %	43.90 %
Gaseosa naranja (0.41 %)	1	1	0.74 %	100.00 %	0 %
Jugo polvo (0.51 %)	7	3	0 %	0 %	42.86 %
Jurel (0.29 %)	12	11	0 %	0 %	91.67 %
Lasaña (0.22 %)	12	4	0.94 %	25.00 %	25.00 %
Lavalozza (1.04 %)	6	6	0 %	0 %	100.00 %
Leche Chocolate (0.81 %)	5	3	0 %	0 %	60.00 %
Toallas de papel (1.32 %)	1	1	0 %	0 %	100.00 %
Papas duquesas (0.21 %)	1	1	0.78 %	100.00 %	0 %
Quesillo (0.39 %)	3	2	0 %	0 %	66.67 %
Té (0.93 %)	4	1	0 %	0 %	25.00 %

Tabla C.2: Desempeño del Modelo con ID por producto

## C.1. Predicción de los Modelos

En base a las definiciones de lealtad adoptadas en el Capítulo 4, se generan subconjuntos de clientes para distintos niveles. Como ya se ha mencionado, la noción de lealtad se asocia a tendencias de compra por parte de los clientes, y se define como el porcentaje de veces que un cliente en particular compra un producto sobre todas sus compras, sobre todos los locales y la totalidad del período en estudio. Así, mientras más se aumente el nivel de lealtad, menos clientes serán catalogados como *Cientes Leales*.

Se realizan estimaciones para los siguientes niveles de lealtades: 0 %, 1 %, 2 %, 5 %, 7 %, 8 % y 10 %. La Tabla C.3 resume los resultados en la predicción de los modelos:

Modelo	Error Tipo I	Falsas Alarmas	Poder de Detección
Modelo Completo	0.33 %	6.41 %	60.83 %
Modelo con ID	0.19 %	4 %	60 %
Lealtad 0 %	0.31 %	6.94 %	55.83 %
Lealtad 1 %	0.19 %	4.11 %	58.33 %
Lealtad 2 %	0.19 %	4.11 %	58.33 %
Lealtad 5 %	0.25 %	6.45 %	48.33 %
Lealtad 7 %	0.38 %	10 %	45 %
Lealtad 8 %	0.38 %	9.38 %	48.33 %
Lealtad 10 %	0.25 %	8.89 %	34.17 %

Tabla C.3: Resumen con desempeño de modelos estimados

La tabla está coloreada de tal forma que se destaca con color rojo los más altos errores, y con verde los más altos poderes de predicción. Se aprecia que, a medida que se aumenta el nivel de lealtad, la estimación pierde Poder de Detección alcanzando tan solo un 34.17 % a

una lealtad del 10 %. Un comportamiento distinto se ve para los porcentajes de Error Tipo I y de Falsas Alarmas, los cuales disminuyen para bajos niveles de lealtad (en especial para 1 % y 2 %). Esto se debe, principalmente, a que en estos modelos se generan menos alertas (73 alertas de quiebre en comparación a las 78 alertas en el Modelo con ID y el Modelo Completo), lo que implica una disminución en el error pero, al mismo tiempo, una disminución en la predicción.

# Apéndice D

## Elección de Modelo Híbrido a utilizar

En esta sección se detalla el proceso de elección del Modelo Híbrido a utilizar. Se escogen los modelos con una lealtad del 1% y del 2%, ya que estos modelos entregan los mejores resultados bajo la segmentación usual (ver Apéndice C). Se agrega el nivel de lealtad al 10%, para tener en consideración un caso extremo.

<b>Modelo</b>	<b>Error Tipo I</b>	<b>Falsas Alarmas</b>	<b>Poder de Detección</b>
Híbrido al 1 % de Lealtad	0.19 %	3.95 %	60.83 %
Híbrido al 2 % de Lealtad	0.19 %	4.23 %	56.67 %
Híbrido al 10 % de Lealtad	0.56 %	16.67 %	37.50 %

Tabla D.1: Desempeño de Modelos Híbridos

Se observa que para el Modelo Híbrido al 10% de Lealtad el poder predictivo aumenta en un 3.33% en comparación al resultado obtenido por el modelo con la segmentación usual (ver Apéndice C), sin embargo, sus niveles de error también aumentan a casi el doble. Para el caso del Modelo Híbrido al 2% de Lealtad, tanto el poder predictivo como los errores empeoran levemente. Los mejores resultado se encuentran para el Modelo Híbrido al 1% de Lealtad, ya que mejora tanto en poder predictivo (en un 2.5%) como en Falsas Alarmas, manteniendo constante su Error Tipo I.

Analizando el detalle por producto del modelo que obtuvo mejores resultados en la Tabla D.2 se observa que, en comparación con el Modelo con ID (ver Tabla C.2), este mejora en tres de los catorce productos. En particular, identifica 19 quiebres de forma correcta para el Chocolate, mientras que el Modelo con ID lo hace en 18 ocasiones. Sumado a esto, no genera alertas de quiebre para la Gaseosa de Naranja ni para las Papas duquesas, lo que si bien no es óptimo (ya que el modelo no logra identificar el quiebre que tiene cada producto), no genera falsas alarmas como en el caso del Modelo con ID.

<b>Producto (Incidencia)</b>	<b>Quiebres</b>	<b>Alertas</b>	<b>Error Tipo I</b>	<b>Falsas Alarmas</b>	<b>Detección</b>
Atún (0.35 %)	15	15	1.87 %	13.33 %	86.67 %
Azúcar (0.37 %)	7	6	0 %	0 %	85.71 %
Camarón (0.37 %)	5	5	0 %	0 %	100.00 %
Chocolate (0.33 %)	41	19	0 %	0 %	46.34 %
Gaseosa naranja (0.41 %)	1	0	0 %	0 %	0 %
Jugo polvo (0.51 %)	7	3	0 %	0 %	42.86 %
Jurel (0.29 %)	12	11	0 %	0 %	91.67 %
Lasaña (0.22 %)	12	4	0.94 %	25.00 %	25.00 %
Lavalozza (1.04 %)	6	6	0 %	0 %	100.00 %
Leche Chocolate (0.81 %)	5	3	0 %	0 %	60.00 %
Toallas de papel (1.32 %)	1	1	0 %	0 %	100.00 %
Papas duquesas (0.21 %)	1	0	0 %	0 %	0 %
Quesillo (0.39 %)	3	2	0 %	0 %	66.67 %
Té (0.93 %)	4	1	0 %	0 %	25.00 %

Tabla D.2: Desempeño por Producto para el Modelo Híbrido al 1 % de Lealtad

# Apéndice E

## Modelos por Producto para Productos Finales

	Quiebres	Alertas	TP	FP	FN	TN	Total	Error Tipo I	Falsas Alarmas	Detección	Precio medio
Atún (0.35 %)	15	14	12	2	3	98	115	2.00 %	14.29 %	80.00 %	809
Azúcar (0.37 %)	7	6	6	-	1	118	125	0 %	0 %	85.71 %	2,898
Camarón (0.37 %)	5	5	5	-	-	108	113	0 %	0 %	100.00 %	2,050
Chocolate (0.33 %)	41	18	18	-	23	44	85	0 %	0 %	43.90 %	1,101
Gaseosa naranja (0.41 %)	1	1	-	1	1	128	130	0.78 %	100.00 %	0 %	1,463
Jugo polvo (0.51 %)	7	3	3	-	4	98	105	0 %	0 %	42.86 %	165
Jurel (0.29 %)	12	11	11	-	1	116	128	0 %	0 %	91.67 %	815
Lasaña (0.22 %)	12	2	2	-	10	102	114	0 %	0 %	16.67 %	866
Lavalozza (1.04 %)	6	7	6	1	-	118	125	0.84 %	14.29 %	100.00 %	1,794
Leche Chocolate (0.81 %)	5	3	3	-	2	121	126	0 %	0 %	60.00 %	810
Toallas de papel (1.32 %)	1	2	1	1	-	118	120	0.84 %	50.00 %	100.00 %	1,164
Papas duquesas (0.21 %)	1	0	-	-	1	121	122	0 %	0 %	0 %	2,381
Quesillo (0.39 %)	3	2	2	-	1	120	123	0 %	0 %	66.67 %	1,174
Té (0.93 %)	4	1	1	-	3	113	117	0 %	0 %	25.00 %	1,947
Alimento para Perro (0.14 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	6,309
Avena Instantánea (0.75 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	1,193
Café Instantáneo (0.11 %)	6	4	4	-	2	128	134	0 %	0 %	66.67 %	1,592
Caluga Suny (0.15 %)	12	4	3	1	9	105	118	0.94 %	25.00 %	25.00 %	1,299
Cerveza Escudo (0.56 %)	0	0	-	-	-	134	134	0 %	0 %	0 %	3,830
Cerveza Cristal S/Alcohol (0.11 %)	3	0	-	-	3	109	112	0 %	0 %	0 %	2,822
Chorizo Parrillero (0.36 %)	1	1	-	1	1	135	137	0.74 %	100.00 %	0 %	2,673
Fideo Espiral (0.79 %)	0	0	-	-	-	126	126	0 %	0 %	0 %	511
Fósforos Pequeños (0.96 %)	0	0	-	-	-	137	137	0 %	0 %	0 %	978
Galleta Crackeler (1.17 %)	1	0	-	-	1	135	136	0 %	0 %	0 %	271
Gaseosa Cola (1.63 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	1,504
Gaseosa Cola Light (0.16 %)	10	3	3	-	7	101	111	0 %	0 %	30.00 %	1,204
Leche en Polvo Entera (0.07 %)	5	1	-	1	5	111	117	0.89 %	100.00 %	0 %	6,216
Leche Entera (1.81 %)	1	0	-	-	1	134	135	0 %	0 %	0 %	597
Leche sin Lactosa(0.37 %)	0	0	-	-	-	133	133	0 %	0 %	0 %	776
Maizena (0.13 %)	22	15	15	-	7	84	106	0 %	0 %	68.18 %	881
Margarina Qualy (0.59 %)	6	1	1	-	5	112	118	0 %	0 %	16.67 %	1,028
Mayonesa Hellmanns (0.56 %)	0	0	-	-	-	132	132	0 %	0 %	0 %	1,537
Néctar Piña (0.79 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	858
Nescafé Trad. Tarro (0.79 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	2,888
Nido Polvo (0.15 %)	11	9	9	-	2	99	110	0 %	0 %	81.82 %	4,217
Palmitos Rodaja (0.72 %)	5	1	1	-	4	120	125	0 %	0 %	20.00 %	726
Papa Stax Original (0.31 %)	135	123	121	2	14	-	137	100.00 %	1.63 %	89.63 %	1,136
Papel Higiénico (0.14 %)	2	0	-	-	2	128	130	0 %	0 %	0 %	6,899
Pisco 35 (0.15 %)	3	0	-	-	3	120	123	0 %	0 %	0 %	3,971
Puré de Papas (0.55 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	958
Queso Gauda Laminado (0.51 %)	5	1	1	-	4	112	117	0 %	0 %	20.00 %	2,890
Salchichas (0.09 %)	9	9	8	1	1	114	124	0.87 %	11.11 %	88.89 %	2,923
Salmón (0.19 %)	3	6	2	4	1	123	130	3.15 %	66.67 %	66.67 %	3,424
Shampoo Colorvive (0.11 %)	6	6	5	1	1	114	121	0.87 %	16.67 %	83.33 %	2,459
Tortilla Burrito (0.13 %)	5	1	1	-	4	89	94	0 %	0 %	20.00 %	1,586
Vienesas JK (0.38 %)	0	0	-	-	-	137	137	0 %	0 %	0 %	1,695
Vienesas de Ave (0.66 %)	2	0	-	-	2	114	116	0 %	0 %	0 %	1,915
<b>Total</b>	<b>373</b>	<b>260</b>	<b>244</b>	<b>16</b>	<b>129</b>	<b>5,404</b>	<b>5,793</b>				

Tabla E.1: Resultados del Modelo Completo para 47 productos finales



	Quiebres	Alertas	TP	FP	FN	TN	Total	Error Tipo I	Falsas Alarmas	Detección	Precio medio
Atún (0.35 %)	15	16	13	3	2	104	122	2.80 %	18.75 %	86.67 %	809
Azúcar (0.37 %)	7	6	6	-	1	123	130	0 %	0 %	85.71 %	2,898
Camarón (0.37 %)	5	5	5	-	-	114	119	0 %	0 %	100.00 %	2,050
Chocolate (0.33 %)	41	18	18	-	23	44	85	0 %	0 %	43.90 %	1,101
Gaseosa naranja (0.41 %)	1	1	-	1	1	134	136	0.74 %	100.00 %	0 %	1,463
Jugo polvo (0.51 %)	7	3	3	-	4	103	110	0 %	0 %	42.86 %	165
Jurel (0.29 %)	12	11	11	-	1	122	134	0 %	0 %	91.67 %	815
Lasaña (0.22 %)	12	4	3	1	9	105	118	0.94 %	25.00 %	25.00 %	866
Lavalozza (1.04 %)	6	6	6	-	-	125	131	0 %	0 %	100.00 %	1,794
Leche Chocolate (0.81 %)	5	3	3	-	2	126	131	0 %	0 %	60.00 %	810
Toallas de papel (1.32 %)	1	1	1	-	-	122	123	0 %	0 %	100.00 %	1,164
Papas duquesas (0.21 %)	1	1	-	1	1	127	129	0.78 %	100.00 %	0 %	2,381
Quesillo (0.39 %)	3	2	2	-	1	125	128	0 %	0 %	66.67 %	1,174
Té (0.93 %)	4	1	1	-	3	119	123	0 %	0 %	25.00 %	1,947
Alimento para Perro (0.14 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	6,309
Avena Instantánea (0.75 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	1,193
Café Instantáneo (0.11 %)	6	4	4	-	2	128	134	0 %	0 %	66.67 %	1,592
Caluga Suny (0.15 %)	12	2	2	-	10	106	118	0 %	0 %	16.67 %	1,299
Cerveza Escudo (0.56 %)	0	0	-	-	-	134	134	0 %	0 %	0 %	3,830
Cerveza Cristal S/Alcohol (0.11 %)	3	0	-	-	3	109	112	0 %	0 %	0 %	2,822
Chorizo Parrillero (0.36 %)	1	1	-	1	1	135	137	0.74 %	100.00 %	0 %	2,673
Fideo Espiral (0.79 %)	0	0	-	-	-	126	126	0 %	0 %	0 %	511
Fósforos Pequeños (0.96 %)	0	0	-	-	-	137	137	0 %	0 %	0 %	978
Galleta Crackeler (1.17 %)	1	0	-	-	1	135	136	0 %	0 %	0 %	271
Gaseosa Cola (1.63 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	1,504
Gaseosa Cola Light (0.16 %)	10	2	2	-	8	101	111	0 %	0 %	20.00 %	1,204
Leche en Polvo Entera (0.07 %)	5	2	1	1	4	111	117	0.89 %	50.00 %	20.00 %	6,216
Leche Entera (1.81 %)	1	0	-	-	1	134	135	0 %	0 %	0 %	597
Leche sin Lactosa(0.37 %)	0	0	-	-	-	133	133	0 %	0 %	0 %	776
Maizena (0.13 %)	22	14	14	-	8	84	106	0 %	0 %	63.64 %	881
Margarina Qualy (0.59 %)	6	1	1	-	5	112	118	0 %	0 %	16.67 %	1,028
Mayonesa Hellmanns (0.56 %)	0	0	-	-	-	132	132	0 %	0 %	0 %	1,537
Néctar Piña (0.79 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	858
Nescafé Trad. Tarro (0.79 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	2,888
Nido Polvo (0.15 %)	11	7	7	-	4	99	110	0 %	0 %	63.64 %	4,217
Palmitos Rodaja (0.72 %)	5	1	1	-	4	120	125	0 %	0 %	20.00 %	726
Papa Stax Original (0.31 %)	135	125	123	2	12	-	137	100.00 %	1.60 %	91.11 %	1,136
Papel Higiénico (0.14 %)	2	0	-	-	2	128	130	0 %	0 %	0 %	6,899
Pisco 35 (0.15 %)	3	0	-	-	3	120	123	0 %	0 %	0 %	3,971
Puré de Papas (0.55 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	958
Queso Gauda Laminado (0.51 %)	5	0	-	-	5	112	117	0 %	0 %	0 %	2,890
Salchichas (0.09 %)	9	9	8	1	1	114	124	0.87 %	11.11 %	88.89 %	2,923
Salmón (0.19 %)	3	5	2	3	1	124	130	2.36 %	60.00 %	66.67 %	3,424
Shampoo Colorvive (0.11 %)	6	6	5	1	1	114	121	0.87 %	16.67 %	83.33 %	2,459
Tortilla Burrito (0.13 %)	5	2	1	1	4	88	94	1.12 %	50.00 %	20.00 %	1,586
Vienesas JK (0.38 %)	0	0	-	-	-	137	137	0 %	0 %	0 %	1,695
Vienesas de Ave (0.66 %)	2	0	-	-	2	114	116	0 %	0 %	0 %	1,915
<b>Total</b>	<b>373</b>	<b>259</b>	<b>243</b>	<b>16</b>	<b>130</b>	<b>5,475</b>	<b>5,864</b>				

Tabla E.2: Resultados del Modelo con ID para 47 productos finales

	Quiebres	Alertas	TP	FP	FN	TN	Total	Error Tipo I	Falsas Alarmas	Detección	Precio medio
Atún (0.35 %)	15	15	13	2	2	105	122	1.87 %	13.33 %	86.67 %	809
Azúcar (0.37 %)	7	6	6	-	1	123	130	0 %	0 %	85.71 %	2,898
Camarón (0.37 %)	5	5	5	-	-	114	119	0 %	0 %	100.00 %	2,050
Chocolate (0.33 %)	41	19	19	-	22	44	85	0 %	0 %	46.34 %	1,101
Gaseosa naranja (0.41 %)	1	0	-	-	1	135	136	0 %	0 %	0 %	1,463
Jugo polvo (0.51 %)	7	3	3	-	4	103	110	0 %	0 %	42.86 %	165
Jurel (0.29 %)	12	11	11	-	1	122	134	0 %	0 %	91.67 %	815
Lasaña (0.22 %)	12	4	3	1	9	105	118	0.94 %	25.00 %	25.00 %	866
Lavalozza (1.04 %)	6	6	6	-	-	125	131	0 %	0 %	100.00 %	1,794
Leche Chocolate (0.81 %)	5	3	3	-	2	126	131	0 %	0 %	60.00 %	810
Toallas de papel (1.32 %)	1	1	1	-	-	122	123	0 %	0 %	100.00 %	1,164
Papas duquesas (0.21 %)	1	0	-	-	1	128	129	0 %	0 %	0 %	2,381
Quesillo (0.39 %)	3	2	2	-	1	125	128	0 %	0 %	66.67 %	1,174
Té (0.93 %)	4	1	1	-	3	119	123	0 %	0 %	25.00 %	1,947
Alimento para Perro (0.14 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	6,309
Avena Instantánea (0.75 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	1,193
Café Instantáneo (0.11 %)	6	4	4	-	2	128	134	0 %	0 %	66.67 %	1,592
Caluga Suny (0.15 %)	12	0	-	-	12	106	118	0 %	0 %	0 %	1,299
Cerveza Escudo (0.56 %)	0	0	-	-	-	134	134	0 %	0 %	0 %	3,830
Cerveza Cristal S/Alcohol (0.11 %)	3	1	-	1	3	108	112	0.92 %	100.00 %	0 %	2,822
Chorizo Parrillero (0.36 %)	1	0	-	-	1	136	137	0 %	0 %	0 %	2,673
Fideo Espiral (0.79 %)	0	0	-	-	-	126	126	0 %	0 %	0 %	511
Fósforos Pequeños (0.96 %)	0	0	-	-	-	137	137	0 %	0 %	0 %	978
Galleta Crackeler (1.17 %)	1	0	-	-	1	135	136	0 %	0 %	0 %	271
Gaseosa Cola (1.63 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	1,504
Gaseosa Cola Light (0.16 %)	10	4	3	1	7	100	111	0.99 %	25.00 %	30.00 %	1,204
Leche en Polvo Entera (0.07 %)	5	0	-	-	5	112	117	0 %	0 %	0 %	6,216
Leche Entera (1.81 %)	1	0	-	-	1	134	135	0 %	0 %	0 %	597
Leche sin Lactosa(0.37 %)	0	0	-	-	-	133	133	0 %	0 %	0 %	776
Maizena (0.13 %)	22	14	14	-	8	84	106	0 %	0 %	63.64 %	881
Margarina Qualy (0.59 %)	6	1	1	-	5	112	118	0 %	0 %	16.67 %	1,028
Mayonesa Hellmanns (0.56 %)	0	0	-	-	-	132	132	0 %	0 %	0 %	1,537
Néctar Piña (0.79 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	858
Nescafé Trad. Tarro (0.79 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	2,888
Nido Polvo (0.15 %)	11	9	9	-	2	99	110	0 %	0 %	81.82 %	4,217
Palmitos Rodaja (0.72 %)	5	1	1	-	4	120	125	0 %	0 %	20.00 %	726
Papa Stax Original (0.31 %)	135	124	122	2	13	-	137	100.00 %	1.61 %	90.37 %	1,136
Papel Higiénico (0.14 %)	2	0	-	-	2	128	130	0 %	0 %	0 %	6,899
Pisco 35 (0.15 %)	3	3	2	1	1	119	123	0.83 %	33.33 %	66.67 %	3,971
Puré de Papas (0.55 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	958
Queso Gauda Laminado (0.51 %)	5	0	-	-	5	112	117	0 %	0 %	0 %	2,890
Salchichas (0.09 %)	9	9	8	1	1	114	124	0.87 %	11.11 %	88.89 %	2,923
Salmón (0.19 %)	3	6	2	4	1	123	130	3.15 %	66.67 %	66.67 %	3,424
Shampoo Colorvive (0.11 %)	6	5	4	1	2	114	121	0.87 %	20.00 %	66.67 %	2,459
Tortilla Burrito (0.13 %)	5	3	1	2	4	87	94	2.25 %	66.67 %	20.00 %	1,586
Vienesas JK (0.38 %)	0	0	-	-	-	137	137	0 %	0 %	0 %	1,695
Vienesas de Ave (0.66 %)	2	0	-	-	2	114	116	0 %	0 %	0 %	1,915
<b>Total</b>	<b>373</b>	<b>260</b>	<b>244</b>	<b>16</b>	<b>129</b>	<b>5,475</b>	<b>5,864</b>				

Tabla E.3: Resultados del Modelo Híbrido al 1 % de Lealtad para 47 productos finales

	Quiebres	Alertas	TP	FP	FN	TN	Total	Error Tipo I	Falsas Alarmas	Detección	Precio medio
Atún (0.35 %)	15	14	12	2	3	105	122	1.87 %	14.29 %	80.00 %	809
Azúcar (0.37 %)	7	6	6	-	1	123	130	0 %	0 %	85.71 %	2,898
Camarón (0.37 %)	5	5	5	-	-	114	119	0 %	0 %	100.00 %	2,050
Chocolate (0.33 %)	41	15	15	-	26	44	85	0 %	0 %	36.59 %	1,101
Gaseosa naranja (0.41 %)	1	0	-	-	1	135	136	0 %	0 %	0 %	1,463
Jugo polvo (0.51 %)	7	3	3	-	4	103	110	0 %	0 %	42.86 %	165
Jurel (0.29 %)	12	11	11	-	1	122	134	0 %	0 %	91.67 %	815
Lasaña (0.22 %)	12	4	3	1	9	105	118	0.94 %	25.00 %	25.00 %	866
Lavalozza (1.04 %)	6	6	6	-	-	125	131	0 %	0 %	100.00 %	1,794
Leche Chocolate (0.81 %)	5	3	3	-	2	126	131	0 %	0 %	60.00 %	810
Toallas de papel (1.32 %)	1	0	-	-	1	122	123	0 %	0 %	0 %	1,164
Papas duquesas (0.21 %)	1	2	-	2	1	126	129	1.56 %	100.00 %	0 %	2,381
Quesillo (0.39 %)	3	2	2	-	1	125	128	0 %	0 %	66.67 %	1,174
Té (0.93 %)	4	1	1	-	3	119	123	0 %	0 %	25.00 %	1,947
Alimento para Perro (0.14 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	6,309
Avena Instantánea (0.75 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	1,193
Café Instantáneo (0.11 %)	6	4	4	-	2	128	134	0 %	0 %	66.67 %	1,592
Caluga Suny (0.15 %)	12	5	4	1	8	105	118	0.94 %	20.00 %	33.33 %	1,299
Cerveza Escudo (0.56 %)	0	2	-	2	-	132	134	1.49 %	100.00 %	0 %	3,830
Cerveza Cristal S/Alcohol (0.11 %)	3	0	-	-	3	109	112	0 %	0 %	0 %	2,822
Chorizo Parrillero (0.36 %)	1	0	-	-	1	136	137	0 %	0 %	0 %	2,673
Fideo Espiral (0.79 %)	0	0	-	-	-	126	126	0 %	0 %	0 %	511
Fósforos Pequeños (0.96 %)	0	0	-	-	-	137	137	0 %	0 %	0 %	978
Galleta Crackeler (1.17 %)	1	0	-	-	1	135	136	0 %	0 %	0 %	271
Gaseosa Cola (1.63 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	1,504
Gaseosa Cola Light (0.16 %)	10	2	2	-	8	101	111	0 %	0 %	20.00 %	1,204
Leche en Polvo Entera (0.07 %)	5	1	-	1	5	111	117	0.89 %	100.00 %	0 %	6,216
Leche Entera (1.81 %)	1	0	-	-	1	134	135	0 %	0 %	0 %	597
Leche sin Lactosa(0.37 %)	0	0	-	-	-	133	133	0 %	0 %	0 %	776
Maizena (0.13 %)	22	14	14	-	8	84	106	0 %	0 %	63.64 %	881
Margarina Qualy (0.59 %)	6	1	1	-	5	112	118	0 %	0 %	16.67 %	1,028
Mayonesa Hellmanns (0.56 %)	0	0	-	-	-	132	132	0 %	0 %	0 %	1,537
Néctar Piña (0.79 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	858
Nescafé Trad. Tarro (0.79 %)	0	0	-	-	-	136	136	0 %	0 %	0 %	2,888
Nido Polvo (0.15 %)	11	6	6	-	5	99	110	0 %	0 %	54.55 %	4,217
Palmitos Rodaja (0.72 %)	5	1	1	-	4	120	125	0 %	0 %	20.00 %	726
Papa Stax Original (0.31 %)	135	124	122	2	13	-	137	100.00 %	1.61 %	90.37 %	1,136
Papel Higiénico (0.14 %)	2	0	-	-	2	128	130	0 %	0 %	0 %	6,899
Pisco 35 (0.15 %)	3	0	-	-	3	120	123	0 %	0 %	0 %	3,971
Puré de Papas (0.55 %)	0	0	-	-	-	129	129	0 %	0 %	0 %	958
Queso Gauda Laminado (0.51 %)	5	0	-	-	5	112	117	0 %	0 %	0 %	2,890
Salchichas (0.09 %)	9	6	5	1	4	114	124	0.87 %	16.67 %	55.56 %	2,923
Salmón (0.19 %)	3	6	2	4	1	123	130	3.15 %	66.67 %	66.67 %	3,424
Shampoo Colorvive (0.11 %)	6	6	5	1	1	114	121	0.87 %	16.67 %	83.33 %	2,459
Tortilla Burrito (0.13 %)	5	1	1	-	4	89	94	0 %	0 %	20.00 %	1,586
Vienesas JK (0.38 %)	0	0	-	-	-	137	137	0 %	0 %	0 %	1,695
Vienesas de Ave (0.66 %)	2	0	-	-	2	114	116	0 %	0 %	0 %	1,915
<b>Total</b>	<b>373</b>	<b>251</b>	<b>234</b>	<b>17</b>	<b>139</b>	<b>5,474</b>	<b>5,864</b>				

Tabla E.4: Resultados del Modelo al 0 % de Lealtad para 47 productos finales

# Apéndice F

## Resultados por Producto del Modelo al 15 % de Lealtad

	Quiebres	Alertas	TP	FP	FN	TN	Total	Error Tipo I	Falsas Alarmas	Detección
Atún (0.35 %)	15	0	-	-	15	107	122	0%	0%	0%
Azúcar (0.37 %)	7	4	4	-	3	123	130	0%	0%	57.14%
Camarón (0.37 %)	5	3	1	2	4	112	119	1.75%	66.67%	20.00%
Chocolate (0.33 %)	41	4	4	-	37	44	85	0%	0%	9.76%
Gaseosa naranja (0.41 %)	1	0	-	-	1	135	136	0%	0%	0%
Jugo polvo (0.51 %)	7	3	3	-	4	103	110	0%	0%	42.86%
Jurel (0.29 %)	12	0	-	-	12	122	134	0%	0%	0%
Lasaña (0.22 %)	12	0	-	-	12	106	118	0%	0%	0%
Lavalozza (1.04 %)	6	6	6	-	-	125	131	0%	0%	100.00%
Leche Chocolate (0.81 %)	5	3	3	-	2	126	131	0%	0%	60.00%
Toallas de papel (1.32 %)	1	0	-	-	1	122	123	0%	0%	0%
Papas duquesas (0.21 %)	1	0	-	-	1	128	129	0%	0%	0%
Quesillo (0.39 %)	3	0	-	-	3	125	128	0%	0%	0%
Té (0.93 %)	4	1	1	-	3	119	123	0%	0%	25.00%
<b>Total</b>	120	24	22	2	98	1,597	1,719			
	<b>Error Tipo I</b>	<b>Falsas Alarmas</b>	<b>Poder de Detección</b>							
Resultado total	0.13%	8.33%	18.33%							

Tabla F.1: Resultados del Modelo al 15 % de Lealtad para 14 productos