



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

**“ESTIMACIÓN DEL VALOR DE CLIENTES FIDELIZADOS EN UNA CADENA DE
SUPERMERCADOS UTILIZANDO MODELOS JERÁRQUICOS BAYESIANOS”**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

FELIPE IGNACIO REUSE TELLO

**PROFESOR GUÍA:
MARCEL GOIC FIGUEROA**

**MIEMBROS DE LA COMISIÓN:
ALEJANDRA PUENTE CHANDÍA
RICHARD VEGA ALABARCÉ**

**SANTIAGO DE CHILE
OCTUBRE 2011**

ESTIMACIÓN DEL VALOR DE CLIENTES FIDELIZADOS EN UNA CADENA DE SUPERMERCADOS UTILIZANDO MODELOS JERÁRQUICOS BAYESIANOS

El presente trabajo nace del interés por conocer qué tan rentables son los clientes de un supermercado, considerando dinámicas particulares en su comportamiento de compra. Tiene por objetivo principal estimar el valor de los consumidores utilizando la métrica Customer Lifetime Value (CLV), herramienta importante que permite tomar decisiones de marketing más focalizadas.

Hasta el momento se han desarrollado algunos trabajos que utilizan modelos econométricos para estimar los montos futuros a gastar por el cliente, pero siempre asumiendo cierta homogeneidad. El principal desafío de esta memoria está en considerar la heterogeneidad que existe en el comportamiento de éstos, para así obtener un CLV a nivel individual. Esto se logra mediante una Regresión Jerárquica Bayesiana, la cual permite obtener parámetros específicos para cada uno de ellos, utilizando tanto la información particular, como también la global de toda la población. De esta manera, se logra captar dinámicas y tendencias particulares del comportamiento de compra individual.

La Regresión Jerárquica Bayesiana logra reducir en un 50% el error de predicción incurrido en la Regresión Lineal Grupal, donde este último modelo asume homogeneidad entre los clientes. El ajuste promedio es de buena calidad, ya que explica el 65% de la varianza. El nuevo modelo utilizado permite que las variables para estimar los montos futuros sean significativas, a diferencia de una Regresión Lineal Individual. En este último caso, la baja cantidad de observaciones por cliente impide obtener resultados significativos.

Una vez calculado el CLV para el año de pronóstico, se forman Quintiles de clientes, donde los superiores tienen un valor promedio de \$911.375, aportando el 45% de la contribución total del supermercado. Por su parte, los menos valiosos tienen un CLV promedio de \$75.456, aportando solo el 4% de las ganancias. La principal característica de los clientes más valiosos, es que son bastante inteligentes para comprar, ya que captan fuertemente las promociones ofrecidas.

Las acciones de marketing propuestas fueron realizadas para 3 de los grupos formados. El primer Quintil corresponde al grupo constituido mayormente por jóvenes, y el objetivo de las acciones es aumentar su monto y frecuencia de compra. Los Quintiles 4 y 5 corresponden a los grupos con mayor valor, y por lo mismo los esfuerzos se enfocan en potenciar la relación futura con la compañía.

Como trabajo futuro, se propone desarrollar modelos más sofisticados para la proyección de las variables predictivas en el periodo de estimación, estudio que probablemente mejoraría los resultados del CLV a nivel individual.

AGRADECIMIENTOS

La Universidad ha sido una de las experiencias más reconfortantes que me ha tocado vivir en la vida, y es la llave para una vida futura llena de éxitos y logros. Ha sido un periodo bastante intenso, con ciertos momentos de alegría y otros de profundo estrés, pero que finalmente llegan al tan esperado cierre de la memoria.

Debo agradecer en primer lugar a Dios, por permitirme desarrollar una vida Universitaria exitosa, que a pesar de tener momentos de flaqueza donde pensé en caer, me levantaba inexplicablemente, y me empujaba a seguir adelante, mencionando incansablemente “Sigue adelante, no te la debe ganar”.

En segundo lugar, debo agradecer a mi familia por su apoyo incondicional, a pesar de la distancia que me permitía verlos en pocas ocasiones, pero que gracias a su apoyo tanto emocional como económico, me permiten hoy cerrar este ciclo de la vida tan importante como es la Universidad. En especial agradezco a mi madre Nora, y a mi Padre Gustavo, por estos 6 años de apoyo permanente que no cesó jamás. También mencionar gratitud hacia mi hermana Solange, quién me entregó compañía durante los últimos 2 años de carrera, y quien ha sido pilar fundamental en todo momento. Gracias también a mis hermanos Gustavo y Juan Pablo, a quienes también dedico el éxito logrado en esta etapa.

También me gustaría agradecer a mis abuelos Nora y Aureliano por haberme recibido durante cuatro años en su casa, sin lugar a dudas no habría sido lo mismo sin ustedes.

Debo mencionar a mis amigos más cercanos con quienes compartí estos 6 años, tanto compañeros de la Universidad como aquellos de la vida. Gracias por haberme brindado momentos de alegría, como también el apoyo en momentos difíciles. Y gracias por haberme dado el ánimo cuando pensaba que sería imposible sacar adelante esta Memoria.

Me gustaría agradecer a la empresa que me dio la oportunidad de desarrollar mi trabajo de Memoria, en especial a mis compañeros de trabajo que me aportaron grandes conocimientos tanto de la materia como del negocio. Mi especial gratitud a Derek Farren por confiar en mi trabajo de título, y también a Richard Vega por su aporte en el ámbito docente y profesional.

Finalmente debo agradecer a mis profesores, en especial al guía Marcel Goic quién me brindó las herramientas necesarias para sacar adelante un trabajo de calidad. Nada de esto habría sido posible de no ser por su buena voluntad y profesionalismo. Además agradecer a Alejandra Puente y Luis Aburto, quienes también me dieron apoyo durante el verano, y a todo el cuerpo docente que hicieron posible mejorar mi trabajo durante todo el semestre.

INDICE GENERAL

1.	INTRODUCCIÓN.....	9
1.1	Antecedentes Generales.....	9
1.2	La Empresa y su Entorno Económico.....	10
1.3	Gerencia de CRM.....	11
1.4	Planteamiento del Problema y Justificación.....	11
1.5	Objetivos.....	13
1.5.1	Objetivo General.....	13
1.5.2	Objetivos Específicos.....	13
1.6	Restricciones y Alcances.....	13
1.7	Resultados Esperados.....	13
2.	MARCO CONCEPTUAL.....	14
2.1	Gestión de Clientes.....	14
2.2	Customer Lifetime Value (CLV).....	15
2.3	Modelos para Predicción de Montos Futuros.....	16
2.3.1	Modelo de Regresión Lineal.....	16
2.3.2	Modelo Jerárquico Bayesiano.....	16
2.4	Metodología KDD.....	18
3.	METODOLOGÍA.....	19
3.1	Análisis y Selección de Datos.....	20
3.2	Tratamientos de Outliers y Missing Values.....	20
3.3	Análisis Descriptivo de Variables.....	20
3.4	Análisis de Tendencias.....	21
3.5	Creación de Variables Dummies.....	21
3.6	Predicción de Montos Mediante Método Simple.....	21
3.7	Predicción de Montos a Nivel Individual.....	21
3.8	Predicción de Montos a Nivel Agregado.....	21
3.9	Predicción Montos con Modelo Jerárquico Bayesiano.....	22
3.10	Validación de Resultados (Ajuste y Error).....	22
3.11	Cálculo del Customer Lifetime Value (CLV).....	22
3.12	Análisis de Sensibilidad Tasa de Retención y Tasa de Descuento.....	23
3.13	Ranking de Clientes.....	23
3.14	Propuesta de Acciones de Marketing.....	23
4.	DESARROLLO METODOLOGÍA.....	24
4.1	Selección de Datos.....	24
4.2	Tratamiento de Outliers y Missing Values.....	25
4.3	Análisis Descriptivo de Variables.....	26
4.4	Análisis de Tendencias.....	29
4.5	Creación de Variables Dummies.....	31
4.6	Predicción de Monto Mediante Método Simple.....	32
4.7	Predicción Montos a Nivel Individual.....	33
4.8	Predicción Montos a Nivel Agregado.....	35
4.9	Predicción Montos Mediante Modelo Jerárquico Bayesiano.....	41
4.10	Validación de Resultados.....	44
4.11	Estimación del Customer Lifetime Value (CLV).....	46
4.11.1	Proyección de Variables Predictivas.....	46
4.11.2	Cálculo del Customer Lifetime Value.....	47
4.11.3	Caracterización de Quintiles.....	50
4.12	Análisis de Sensibilidad CLV.....	53
4.12.1	Tasa de Retención.....	53
4.12.2	Tasa de Descuento.....	55
4.13	Ranking de Clientes.....	56

4.14	Propuesta Acciones de Marketing.....	58
4.14.1	Acciones para Segmento “Joven”	58
4.14.2	Acciones para Segmento “Potencial”	59
4.14.3	Acciones para Segmento “Gold Inteligente”	60
5.	CONCLUSIONES.....	62
5.1	Conclusiones del Trabajo.....	62
5.2	Trabajos Futuros	64
5.3	BRIBLIOGRAFÍA.....	65
5.4	ANEXOS.....	66

INDICE DE GRÁFICOS

Gráfico 1: Participación de Mercado Supermercados en Chile	10
Gráfico 3: Gráfico de Dispersión Monto v/s Cantidad Promociones	27
Gráfico 2: Gráfico de Dispersión Monto v/s Frecuencia	27
Gráfico 4: Distribución de Clientes por Sexo	28
Gráfico 5: Distribución de Clientes por Segmentación “Ciclo de Vida”	29
Gráfico 6: Serie de Montos Promedios a Nivel Agregado	30
Gráfico 7: Montos Gastados Cliente 1 v/s Cliente 2	30
Gráfico 8: Montos Gastados Mensualmente a Nivel Agregado	31
Gráfico 9: Monto Promedio por Segmentación “Ciclo de Vida”	32
Gráfico 10: Distribución R2 Regresión Lineal Individual	35
Gráfico 11: Variación de Montos Gastados Años 2009-2010	37
Gráfico 12: Error MAPE por decil de Clientes (periodo de validación)	38
Gráfico 13: MAPE por quintil, Monto Real v/s Predicción (periodo validación)	39
Gráfico 14: Distribución Coeficiente No Estandarizado Variable Periodo	41
Gráfico 15: Análisis de Significancia Variable Periodo, R. J. Bayesiana	42
Gráfico 16: Pendiente Coeficientes No Est., R. L. Individual v/s R. J. Bayesiana	43
Gráfico 17: Estimación de Monto v/s Monto Real (Periodo validación)	45
Gráfico 18: Variabilidad de Frecuencia Agregada, Año 1 – Año 2	46
Gráfico 19: Gráfico de Caja CLV Clientes	48
Gráfico 20: CLV Promedio por decil de Clientes	48
Gráfico 21: Frecuencia Promedio Mensual por Quintil de Clientes	50
Gráfico 22: Margen Promedio Mensual por Quintil de Clientes	50
Gráfico 23: Presencia en Ticket por Categoría, a Nivel Mensual	51
Gráfico 24: Distribución de Clientes por Segmentación de Valor	66
Gráfico 25: MAPE por quintil, Monto Real v/s Predicción (periodo calibración)	67
Gráfico 26: MAPE _{PRED} por quintil, Monto Real v/s Predicción (periodo calibración)	68
Gráfico 27: MAPE _{PRED} por quintil, Monto Real v/s Predicción (periodo validación)	68
Gráfico 28: Error MAPE por decil de Clientes (periodo de calibración)	69
Gráfico 29: Error MAPE por decil de Clientes (Periodo de calibración)	69
Gráfico 30: Error MAPE por decil de Clientes (Periodo de validación)	70
Gráfico 31: Gráfico de Caja para Porcentaje Margen	71
Gráfico 32: Gráfico de Caja para Monto	71
Gráfico 33: Gráfico de Caja para Frecuencia	72
Gráfico 34: Variabilidad Margen Agregado, Año 1 – Año 2	72
Gráfico 35: Variabilidad Promociones Agregado, Año 1 – Año 2	73
Gráfico 36: Presencia en Ticket por Categoría, a Nivel Mensual (completo)	73
Gráfico 37: Significancia Variable Cantidad Promociones, R. J. Bayesiana	74
Gráfico 38: Significancia Variable Frecuencia, R. J. Bayesiana	74
Gráfico 39: Significancia Variable Margen, R. J. Bayesiana	74
Gráfico 40: Significancia Variable Edad, R. J. Bayesiana	74
Gráfico 41: Significancia Variable Binaria Cheque, R. J. Bayesiana	75
Gráfico 42: Significancia Variable Dummy Vacaciones, R. J. Bayesiana	75
Gráfico 43: Significancia Variable Dummy No_Verano, R. J. Bayesiana	75
Gráfico 44: Significancia Variable Dummy Mto_No_Niños, R. J. Bayesiana	75
Gráfico 45: Distribución Segmento “Valor” por Quintil	77

INDICE DE TABLAS

Tabla 1: Criterio para Outliers.....	26
Tabla 2: Análisis Descriptivo de Variables Transaccionales.....	26
Tabla 3: MAPE Predicción Montos Método Estático.....	33
Tabla 4: Resultados Regresión Lineal Individual.....	34
Tabla 5: Error MAPE Regresión Lineal Individual.....	35
Tabla 6: Resultados Regresión Lineal Grupal.....	36
Tabla 7: Error MAPE Regresión Lineal a Nivel Agregado.....	37
Tabla 8: Ajuste y error MAPE para distintas Regresiones.....	38
Tabla 9: Estimación Montos, según Quintiles por $MAPE_{REAL}$	40
Tabla 10: Coeficiente Estandarizado y Significancia Regresión Lineal Grupal.....	40
Tabla 11: Significancia Variables R.J Bayesiana v/s R. L. Individual.....	43
Tabla 12: Resumen Error MAPE para Métodos Predictivos.....	44
Tabla 13: Estadísticos CLV Clientes.....	48
Tabla 14: Estadísticos CLV Clientes por Quintil.....	49
Tabla 15: Peso de Quintiles según Porcentaje Margen.....	49
Tabla 16: Estadísticos CLV Clientes por Quintil.....	51
Tabla 17: Distribución Segmento de Valor por Quintil.....	53
Tabla 18: Valor CLV por Variación de Tasa de Retención.....	53
Tabla 19: Tasas de Retención por Segmento “Valor”.....	54
Tabla 20: Sensibilidad CLV por variación Tasa Retención por Segmento “Valor”.....	54
Tabla 21: Sensibilidad CLV por Variación Inferior Tasa de Descuento.....	55
Tabla 22: Sensibilidad CLV por Variación Superior Tasa de Descuento.....	55
Tabla 23: Ranking Clientes TOP 20.....	57
Tabla 24: Condiciones de Segmentación de Valor según Monto.....	66
Tabla 25: Porcentaje Clientes que Migran desde cierto Segmento “Valor”.....	67
Tabla 26: Resultados Regresión Lineal Individual.....	76
Tabla 27: Sensibilidad CLV por Variación de Tasa de Retención.....	76
Tabla 28: Disminución Promedio CLV por Sensibilidad en Tasa de Retención.....	77

INDICE DE FIGURAS

Figura 1: Distribución Posteriori de Modelo Jerárquico Bayesiano.....	17
Figura 2: Metodología KDD.....	18
Figura 3: Plan Metodológico	19
Figura 4: Proceso de Limpieza Base de Datos	26

1. INTRODUCCIÓN

1.1 Antecedentes Generales

En la actualidad, en el rubro del 'Retail'¹ ha crecido el interés por implementar estrategias de marketing más focalizadas. Antes se consideraban a todos los clientes por igual, a diferencia de la actualidad donde cada consumidor tiene nombre y apellido, y representan un valor diferente para la compañía. La competencia en la industria es cada vez mayor, por lo que surge la necesidad de buscar estrategias para crear valor al cliente ^[4].

El Marketing Relacional es el proceso de identificar, captar, satisfacer, retener y potenciar las relaciones rentables con los mejores clientes, de manera que se logren los objetivos de las partes involucradas. Tiene por finalidad conservar dichas relaciones de corto, mediano y largo plazo, para así maximizar la calidad de las transacciones en monto y frecuencia. Se deben considerar los siguientes puntos ^[6]:

- El concepto de CRM (Customer Relationship Management) facilita esta forma de marketing orientada a recopilar información y posterior análisis, con la finalidad de crear estrategias comerciales más enfocadas.
- La empresa tiene que estar dispuesta a realizar discriminación positiva de Clientes, es decir, tratar de manera distinta a aquellos más valiosos.
- Clasificar a los clientes según el valor que representan para la empresa, con el objetivo de crear planes de acción diferenciados.

Una de las herramientas más importantes utilizadas en el Marketing Relacional es el denominado CLV (Customer LifeTime Value) por sus siglas en inglés, que cuantifica monetariamente el valor que el cliente representa para la empresa proyectada hacia el futuro y traída al valor presente. Con el conocimiento de esta métrica, la empresa puede tomar mejores decisiones en inversión de publicidad y promoción, además de maximizar la fidelización de los clientes más valiosos. Sin embargo, la empresa considera el gasto de los últimos meses para poder valorar a sus clientes de manera individual. Este proyecto agrega valor al incorporar dinámicas de comportamiento para predecir los montos futuros.

Son muchos los beneficios al estimar el Valor del Cliente, ya que se pueden realizar acciones de marketing más focalizadas y con mayor efecto en el consumidor. Además, se podría reducir los costos considerablemente, al estar focalizando de mejor manera los esfuerzos de marketing. Dentro de los beneficios que se obtienen al conocer el valor del cliente, se tienen: ^[4]

- Aumentar lealtad de cliente
- Identificar oportunidades de venta cruzada
- Mayor efectividad en adquirir clientes valiosos
- Layout y localización consistente con los clientes valiosos
- Optimizar distribución de recursos destinados a publicidad

¹ En español significa "venta al detalle", y se le denomina así a la industria de consumo masivo de productos.

Por lo tanto, el presente trabajo consiste en estimar el valor de los clientes fidelizados en una cadena de supermercados. El estudio tendrá un primer modelo de regresión lineal a nivel agregado para estimar los montos que gastará el cliente en los periodos futuros, y posteriormente se realizará un nuevo modelo de regresión que muestra resultados mejores, debido a que trabaja a nivel individual con cada cliente.

1.2 La Empresa y su Entorno Económico

Las cadenas de retail han logrado tener una presencia cada vez más importante dentro del comercio. En países más desarrollados, su relevancia es significativa y cada vez más masiva. En Chile también ocurre así, y se evidencia con el aumento en la construcción de salas de supermercados, donde se espera para el año 2011 la apertura de 169 locales a nivel nacional².

El supermercadismo en Chile ha tenido un desarrollo superior con respecto al resto de América Latina, pero inferior al de Europa y Norteamérica, por lo que aún hay bastante espacio para seguir creciendo ^[2]. La empresa en estudio tiene una alta participación de mercado dentro del país, y busca seguir ocupando el mismo lugar frente a sus competidores. El siguiente gráfico representa la participación de mercado de las empresas que se desenvuelven dentro de la industria de los supermercados:

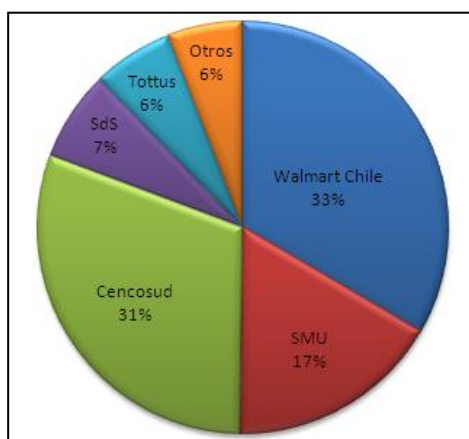


Gráfico 1: Participación de Mercado Supermercados en Chile

Fuente: El Mercurio 2010

La industria ha desarrollado un alto nivel de competencia, lo que puede afectar negativamente los márgenes de venta. Evidentemente, existen cadenas más consolidadas como lo son Walmart Chile, Cencosud, Falabella (Tottus), y ahora último SMU con una participación cercana al 20% ^[3].

La empresa en estudio tiene como negocio principal la distribución de alimentos, además de otras líneas de productos que se ofrecen en menor medida.

• ² Fuente: "Supermercados abrirán, en promedio, un local cada 2 días en 2011", El Mercurio, 16 de Abril del 2010

Posee distintos formatos de supermercados que ha desarrollado en el tiempo según las necesidades de sus clientes, y posee locales ubicados a lo largo de todo el país.

Con el objetivo de aumentar la lealtad de sus clientes, en el año 2006 nace el Club de Fidelización mediante el cual entregan beneficios a sus clientes, a cambio de la información transaccional recolectada en cada compra.

1.3 Gerencia de CRM

El trabajo se realizó en conjunto con el área de Inteligencia de Clientes (Gerencia de CRM) que actualmente posee la empresa en estudio. El equipo que conforma el área de CRM le entregó al alumno la información transaccional y demográfica necesaria para el modelo, y todas las herramientas requeridas durante el periodo de trabajo.

La principal función que tienen los integrantes de esta área es aportar en la toma de decisiones con conocimiento analítico. Se enfoca en mejorar los procesos estratégicos y la gestión de clientes mediante el uso de inteligencia de negocios, lo que se traduce en hacer más rentable a la empresa. La Gerencia CRM tiene a cargo todos los estudios que se originan en la información transaccional recolectada mediante el club de fidelización al cual pertenece la mayoría de los clientes del supermercado.

1.4 Planteamiento del Problema y Justificación

En la actualidad, son varias las empresas de Retail que tienen un Club de Lealtad para fidelizar a sus clientes, y gracias a los datos recolectados cada vez que un cliente realiza una transacción, se puede predecir el comportamiento futuro de éste. Dentro del abanico de estudios posibles, está la estimación del valor del cliente, el cual se logra prediciendo los montos que gastará éste en periodos futuros.

La segmentación de valor que actualmente posee la empresa, divide a los clientes de acuerdo al monto que han gastado durante los últimos 3 meses. De esta manera se define 4 grupos: Platino, Oro, Plata, Cobre. Sin embargo, la principal limitante de este método es que asume que el cliente no variará su comportamiento en el futuro.

Una segmentación de valor estática como la utilizada actualmente, ignora la dinámica de compra de un cliente, y asume que su valor en el futuro será el mismo que tiene hoy. La segmentación de valor que actualmente existe en la empresa, no logra captar estas dinámicas de compra, reflejadas en ciertas tendencias explicadas principalmente por variables demográficas. Por ejemplo, analizando los datos que se disponen para el estudio, se descubrió que el cliente que comienza a tener hijos, llega a triplicar el consumo, y luego vuelve a disminuirlo una vez que los hijos se convierten en adolescentes y posteriormente dejan el hogar. Otro caso identificado son las estacionalidades mensuales, donde se evidencia un aumento de consumo considerable en el mes de Diciembre, además de una disminución importante del monto promedio gastado durante los meses de verano Enero y Febrero. Todas estas tendencias se analizarán en detalle durante el desarrollo de la metodología.

Por lo tanto, el proyecto consiste en analizar la información que se tiene sobre el comportamiento de compra del cliente a lo largo del tiempo, con el fin de estimar el valor que representa para la empresa considerando la existencia de dinámicas de

compra. El modelo que se desarrolla tiene la ventaja de poder capturar este tipo de dinámicas, lo que permite generar estimaciones más certeras.

Las empresas de retail deben ser cuidadosos con la fuga de los clientes que representan mayor rentabilidad en el futuro, los que podrían terminar prefiriendo a la competencia si no se les fideliza como corresponde. Pero para poder lograr el mayor efecto posible con las acciones de marketing, se debe conocer bien a cada cliente en particular, y así evitar que se fuguen los valiosos.

Según un estudio realizado por la compañía Daemon Quest, especializada en estrategias de clientes, marketing y ventas, una empresa tiene 80% de probabilidad de venta sobre un cliente en cartera mientras que tan sólo posee un 25% de probabilidad de éxito sobre un cliente potencial. Otra conclusión que se obtuvo fue que la retención del 25% de clientes de alto valor puede incrementar los beneficios hasta en un 15%, por lo que la fidelización debe convertirse en una herramienta estratégica de crecimiento empresarial ^[4]. Es por esto que toma gran importancia tratar como únicos a cada tipo de cliente, especialmente a los más valiosos y rentables hacia el futuro. Las compañías que utilizan inteligencia de negocio aumentan sus ingresos brutos, además de sus tasas de adquisición y retención de clientes ^[13].

Una manera de valorar a los clientes es asumiendo que éstos seguirán teniendo un comportamiento similar al actual, y en base a este supuesto, aquellos que gastan mayores montos corresponden a los clientes más valiosos. Pero la principal desventaja de este método es que no considera la dinámica de compra del cliente, tal como se explicó anteriormente. Sin embargo, el avance estadístico ha permitido desarrollar modelos econométricos que logran predecir montos cercanos a los que el consumidor gastará en periodos futuros.

Un método para predecir los montos que el cliente gastará en el futuro es mediante una regresión lineal simple, que considere todas las variables que sean significativas para la predicción. La principal ventaja es que al ser una gran cantidad de datos a nivel agregado, los resultados son mucho más robustos, pero no está correcto asumir homogeneidad entre los clientes. Cada consumidor es distinto y tiene su propio comportamiento de compra.

En trabajos anteriores se han realizado distintas metodologías para estimar el valor del cliente, pero la finalidad de este estudio es seguir mejorando la precisión y confiabilidad de los resultados. En uno de los estudios realizados en semestres anteriores, se desarrolló una metodología que consistía en segmentar a priori a los clientes, y una vez identificados los grupos se realizó una regresión lineal para cada uno de ellos. De esta manera se lograba agrupar a clientes con comportamiento parecido (homogeneidad), pero sin embargo podría perderse robustez en los resultados, ya que la cantidad de información en cada regresión es menor.

Una nueva manera de estimar el valor del cliente es prediciendo los montos futuros utilizando Modelos Jerárquicos Bayesianos, donde se puede utilizar información del comportamiento de compra del cliente para lograr el objetivo. La principal ventaja de este tipo de modelos es que capta la heterogeneidad que existe entre los clientes, ya que permite obtener resultados a nivel individual a pesar de tener poca información sobre cada cliente en particular. Sin embargo, utiliza información global de la población que permite obtener resultados significativos a nivel individual.

1.5 Objetivos

1.5.1 Objetivo General

Estimar el valor de los clientes de un supermercado considerando dinámicas de comportamiento, para mejorar la gestión de los esfuerzos en marketing relacional.

1.5.2 Objetivos Específicos

- Identificar las variables transaccionales y demográficas más relevantes en la estimación del valor del cliente.
- Desarrollar los modelos y comparar la calidad en la predicción de montos futuros.
- Establecer un ranking de clientes que permita diferenciar y ordenar a los clientes según su valor.
- Identificar el grupo de clientes más valiosos y rentable para la empresa.
- Establecer una priorización de clientes que permita dirigir acciones de marketing propuestas.

1.6 Restricciones y Alcances

- Se estimará el valor de los clientes que pertenecen actualmente al Club de Fidelización, y se hará a nivel individual.
- No se considerará a los clientes que tengan datos faltantes en alguna de las variables independientes.
- Se entregará un ranking de clientes en base al CLV, que permita identificar a los clientes más valiosos.
- No se propondrá una distribución de los recursos de marketing en base al valor de los clientes.
- No se propondrán campañas de marketing en base al CLV.
- No se incluirá un análisis de fuga de clientes.

1.7 Resultados Esperados

Los resultados que se espera obtener una vez finalizado el estudio son:

- El modelo que mejor predice el valor del cliente y su dinámica de compra.
- El valor del cliente (CLV) a nivel individual, y que permita identificar a los clientes más valiosos.
- Una valuación de la capacidad de predicción del Modelo Jerárquico Bayesiano en relación a otros modelos propuestos por la literatura.
- Un ranking de clientes en base al CLV, que apoye las decisiones internas de la empresa.

2. MARCO CONCEPTUAL

En los últimos años, las empresas que involucran el consumo masivo han destinado muchos recursos en potenciar la relación con sus clientes más valiosos, con el objetivo de mantenerlos atraídos a la compañía. Este tipo de prácticas conforman la Gestión de Clientes, donde una de las herramientas utilizadas para valorar a los consumidores es el Customer Lifetime Value. Para poder estimar este valor considerando el largo plazo, es necesario predecir los montos que gastará el cliente en el futuro, utilizando la información histórica de compra que se tiene tanto a nivel individual como grupal. La principal novedad de este trabajo de título radica en que para estimar los montos futuros, se utiliza un modelo de Regresión Jerárquica Bayesiana, que busca mejorar los resultados de predicción a nivel individual.

2.1 Gestión de Clientes

El éxito de una empresa que vende productos de consumo masivo, depende directamente de cuánto conoce a los compradores y sus necesidades. Para lograr este objetivo, existen muchas herramientas que logran gestionar a los clientes de manera óptima, lo que finalmente se traduce en mayores contribuciones monetarias hacia la empresa. Esta gestión comienza con una buena fidelización del cliente, y uno de los métodos utilizados hoy en día son los Programas de Lealtad. Estos programas están hechos para que el cliente reciba beneficios por preferir a la compañía antes que a la competencia, haciendo su relación con la empresa más fiel y duradera. Por la misma razón, tiendas de retail como los supermercados han creado los “Club de Fidelización”, donde los cliente inscritos acumulan puntos cada vez que realizan una compra, y que en algún momento futuro los hacen tangibles con el canje en la tienda ^[1].

Cuando se fideliza a los clientes mediante un club, todas las transacciones que este realiza quedan ligadas a su rut, lo que permite conocer a cada comprador de forma personalizada. Este proceso de recolección de datos y análisis de la información es conocido como Inteligencia de Clientes, y permite diseñar campañas de marketing más efectivas y localizadas. Se puede identificar a grupos de clientes más homogéneos y parecidos en su comportamiento de compra, además de identificar necesidades específicas para cada tipo de cliente. El principal beneficio para la empresa es maximizar su relación con el cliente, lo que se ve reflejado en mayores rentabilidades hacia el futuro ^[6].

A medida que ha ido evolucionando la Inteligencia de Clientes, han aparecido estrategias como CRM (Customer Relationship Management), que se enfoca en seleccionar a los clientes más valiosos y potenciar una buena hacia el futuro que maximice su CLV.

Dentro de los potenciales beneficios que logra una estrategia CRM, se encuentra ^[5]:

- Mayor conocimiento del cliente.
- Aumento de la satisfacción y lealtad del cliente.
- Aumento de ventas
- Reducción de costos

Una de las herramientas utilizadas para poder identificar a los clientes más valiosos es la métrica CLV, que se puede estimar tanto a nivel de segmento como a nivel individual.

2.2 Customer Lifetime Value (CLV)

El Valor de Vida del Cliente (CLV) generalmente está definido como el valor presente de las ganancias futuras que el cliente generará para la empresa en un periodo de tiempo. Se estima tanto a nivel individual como de segmento, lo que permite identificar a aquellos clientes más valiosos. Este indicador es fundamental para las empresas y permite desarrollar una mejor estrategia de CRM, dado que al conocer el valor de los clientes se puede determinar cuántos recursos utilizar en ellos, cuántos esfuerzos de marketing brindarles y cuánto tiempo dedicarles. ^[10]

Matemáticamente, el CLV de un cliente es (Gupta et al. 2004; Reinartz and Kumar 2003)^[10]:

$$CLV = \sum_{t=0}^T \frac{(p_t - c_t)r_t}{(1+i)^t} - AC$$

Donde:

- p_t : Monto gastado por cliente en periodo t.
- c_t : Costo asociado a los servicios del cliente en el periodo t.
- i : Tasa de descuento.
- r_t : Tasa retención del periodo t.
- AC : Costo de adquisición del cliente.
- T : Tiempo horizonte para la estimación del CLV.

Conocer el CLV de un cliente permite discriminar entre los que son más rentables y aquellos que no lo son en el largo plazo. Además, ayuda a identificar oportunidades de venta cruzada, lo que se traduce en mayores montos de compra^[4].

Los investigadores de la materia han aplicado ciertas variaciones en modelos que estiman el CLV, como por ejemplo considerar distintos horizontes de tiempo. Algunos han utilizado un tiempo de horizonte fijo (Reinartz and Kumar 2000; Thomas 2001) mientras que otros expertos han preferido un horizonte infinito (Fader, Hardie and Lee 2005; Gupta, Lehmann and Stuart 2004).

Una manera de calcular el CLV es asumiendo que el cliente no cambiará su comportamiento en el futuro. Por lo tanto, el monto que gastará en los próximos periodos será el mismo gastado en el pasado.

Otra manera de calcular la métrica es prediciendo el monto que el cliente gastará en los periodos futuros (p_t). Existen modelos sofisticados como los probabilísticos, econométricos, de data mining, entre otros ^[12]. Para este estudio en particular se utilizarán modelos econométricos, específicamente Regresión Lineal Simple y Modelos Jerárquicos Bayesianos.

Para este caso en particular no se considerarán los costos asociados a cada cliente, ya que la empresa asume que son los mismos para todos los clientes, y afectarían en la misma medida al CLV individual.

La tasa de retención r_t podría variar entre un cliente y otro, pero actualmente no se dispone de dicha información a nivel individual. En base a discusiones que se tuvo con expertos en la empresa, se decide utilizar una tasa de retención del 100%.

2.3 Modelos para Predicción de Montos Futuros

Los parámetros corresponden a la incidencia que cada variable tiene para explicar el comportamiento de los datos disponibles en una base de datos, y existen varios modelos econométricos que cumplen este objetivo. A continuación se explican dos de las opciones disponibles.

2.3.1 Modelo de Regresión Lineal

Es un método matemático que representa la relación entre una variable dependiente de otras variables que explican su comportamiento. Esta relación queda representada por la siguiente ecuación lineal^[4]:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_N x_{Ni} + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

Los X_{i} corresponde al valor que toma la variable explicativa i -ésima, y los a_i son los parámetros que representan la incidencia que cada variable i tiene para explicar el valor de la dependiente Y_i . La sumatoria del producto entre la variable independiente con su respectivo parámetro, más un intercepto de estimación a_0 (constante), logra predecir la variable de interés Y_i .

La técnica utilizada para estimar los parámetros es Mínimos Cuadrados, que busca minimizar el error cuadrático medio. Se debe considerar que al ocupar este método para ajustar la regresión, el valor del R^2 aumenta mientras mayor es el número de variables independientes. Sin embargo, se debe medir la bondad del ajuste, corroborando que el coeficiente de determinación también aumente al agregar más variables X_i .

En este trabajo se desarrollarán dos modelos utilizando la regresión lineal como técnica de predicción, donde el primero corresponde a una regresión a nivel individual, y el segundo será una regresión a nivel global de clientes sin asumir heterogeneidad.

2.3.2 Modelo Jerárquico Bayesiano

Este modelo difiere en gran medida con respecto al anterior, ya que los resultados son entregados, en este caso, a nivel de cliente. El objetivo es predecir los montos futuros que gastará éste, considerando la heterogeneidad existente en la población. El modelo de regresión lineal entrega una función que predice los montos de manera agregada, sin embargo podría existir diferencias importantes entre los consumidores que resulta interesante considerar. Este tipo de modelos se ha estado utilizando bastante últimamente, sobre todo en negocios relacionados con el retail.

Permite explicar de manera más completa el comportamiento de los datos, en comparación con los modelos simples.

Lo primero que se debe tener claro es que un modelo de Regresión Jerárquica Bayesiana es una mezcla de 3 conceptos estadísticos:

- Modelo de Regresión
- Modelo Jerárquico
- Modelo Bayesiano

El primer concepto hace referencia a un modelo que explica el comportamiento de una variable independiente en función de otras variables observables, tal como se explicó en el punto anterior del Marco Conceptual. Esta relación queda definida con la siguiente ecuación:

$$Y_i = f(\bar{X}_i, \bar{\theta}) + \varepsilon_i$$

El segundo factor se refiere a un modelo que contempla Jerarquía en dos niveles, una poblacional y otra particular a nivel i . Quiere decir que los parámetros $\bar{\theta}$ ya no son completamente independientes entre ellos, sino que existe una heterogeneidad importante de considerar. Por lo tanto, la regresión ya no es en función de \bar{X}_i y $\bar{\theta}$, sino que habrá un parámetro $\bar{\theta}_i$ particular para cada individuo i . En otras palabras, existe una heterogeneidad ligada a la información disponible a nivel de grupo, la que se considera mediante el parámetro w . Matemáticamente hablando, la Jerarquía queda representada por la siguiente ecuación:

$$\theta_i = g(Y_i, w) + \zeta$$

Por último, la relación que existe entre estos parámetros $\bar{\theta}_i$ y w se capta mediante un Modelo Bayesiano, de tal manera que los w de la población corresponden a los priori³ de los parámetros individuales [8]. De alguna manera, el comportamiento particular de cada cliente debiera parecerse al que tiene el grupo poblacional. El resultado posteriori resulta ser una multiplicación entre la verosimilitud y el priori, debiendo cumplirse ambos criterios. En la siguiente figura, la curva color verde representa la distribución final resultante:

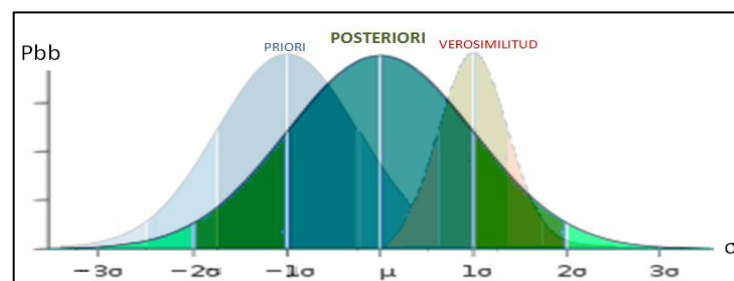


Figura 1: Distribución Posteriori de Modelo Jerárquico Bayesiano

Fuente: Elaboración Propia

³ Conocimiento previo que se dispone de la población, que ayuda a entender el comportamiento individual.

2.4 Metodología KDD

Sus siglas en inglés significan Knowledge Discovery in Databases, y corresponde a una metodología que utiliza bases de datos para extraer información valiosa. Mediante la utilización de algoritmos de Minería de Datos, se puede estimar parámetros que puedan explicar el comportamiento de los datos.

EL siguiente diagrama resume los principales pasos que involucra el proceso KDD ^[9]:

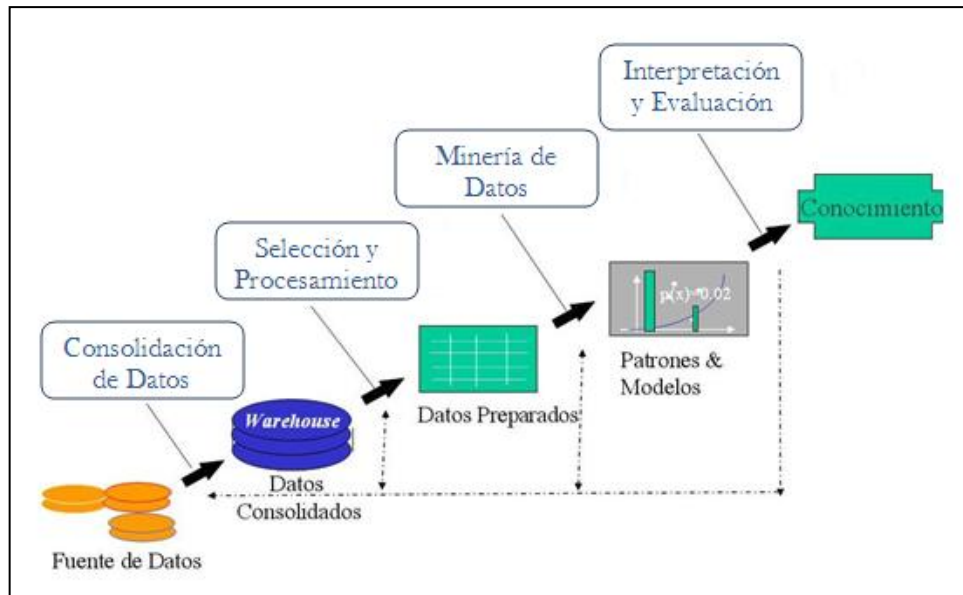


Figura 2: Metodología KDD

Fuente: Menasalvas, E.

1. Selección de la fuente de información acorde al objetivo del estudio.
2. Diseñar un esquema que reúna de manera óptima toda la información, e implementar el almacén de datos para que pueda ser navegado.
3. Selección de las variables relevantes para el estudio y procesamiento (limpieza y transformación de variables).
4. Minería de datos, que corresponde al desarrollo del modelo econométrico mismo (regresión lineal, árbol de decisión, etc.).
5. Interpretación de los resultados y evaluación (calidad).

En este trabajo se deberán realizar ciertas adaptaciones a la metodología KDD, para que se ajuste correctamente. En el siguiente punto se explica en detalle cada uno de los pasos a realizar.

3. METODOLOGÍA

La metodología del presente trabajo comienza con una selección y procesamiento de los datos, posteriormente el desarrollo de los modelos econométricos que permiten predecir los montos futuros, para finalmente calcular el valor de los clientes e interpretar los resultados.

El siguiente diagrama resume las etapas de la metodología, representada como una adaptación del método KDD que se ajusta a este estudio en particular.

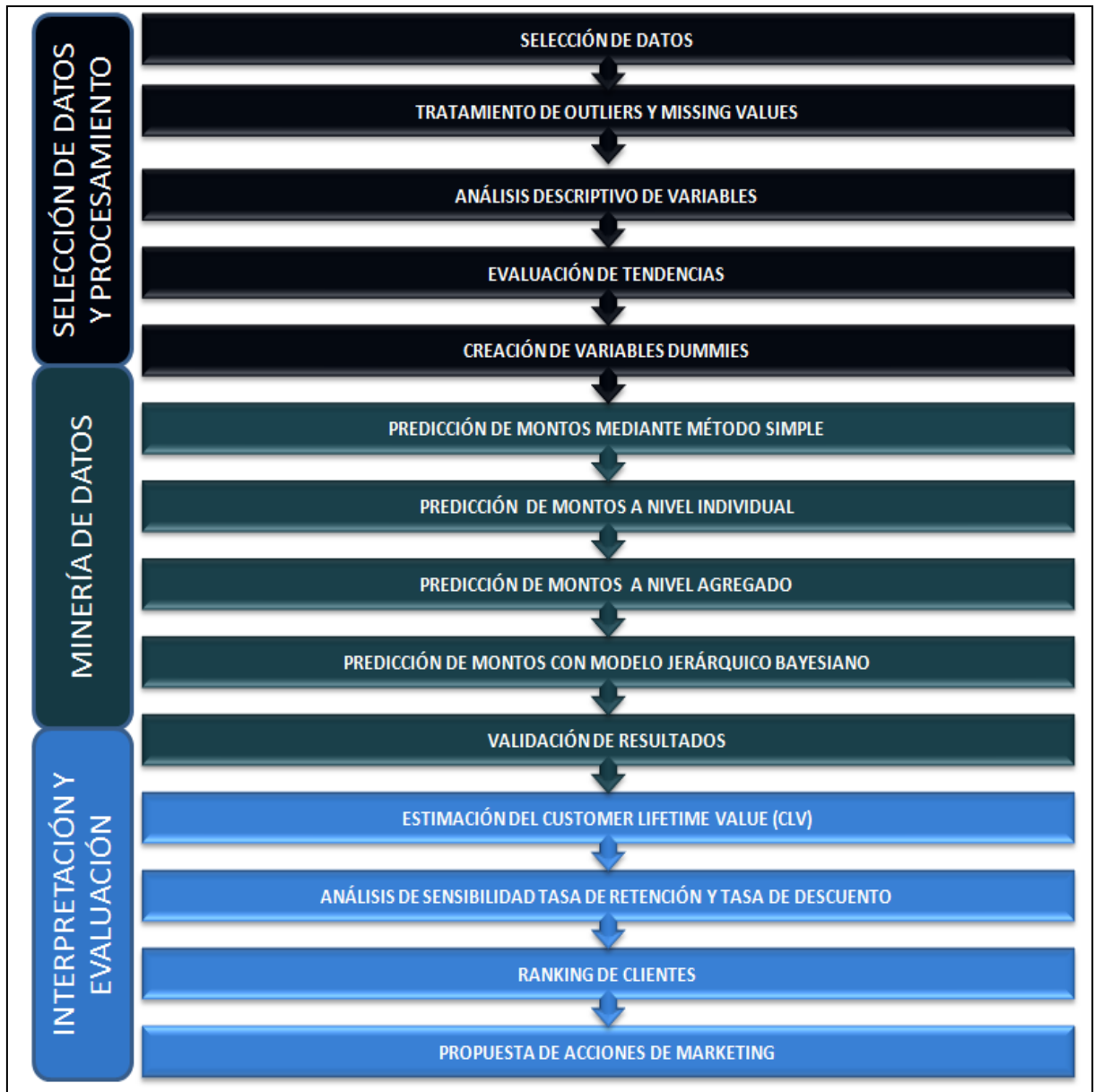


Figura 3: Plan Metodológico

Fuente: Elaboración Propia

3.1 Análisis y Selección de Datos

La cadena de supermercados posee una gran cantidad de datos tanto transaccionales como demográficos. Esto se debe a que el Club de Fidelización comenzó a operar en el año 2006, y a la fecha ha recolectado información valiosa sobre el comportamiento de compra de cada cliente.

Se analizará la base de datos entregada por la empresa, que incluirá todas las variables disponibles para el estudio. Previo al análisis de los datos disponibles, se realizarán reuniones con expertos que trabajan en el área de Inteligencia de Negocios y Estudios de Marketing, con el objetivo de satisfacer la verdadera necesidad del cliente.

Mediante una exploración exhaustiva de la base de datos disponibles, se analizarán todas las variables que la conforman, con la finalidad de seleccionar aquellas más relevantes para este estudio en particular. En una primera instancia se calculará la correlación entre ellas, para asegurarse de no considerar variables que tengan intrínseca la misma información, y así evitar redundancia en el modelo.

3.2 Tratamientos de Outliers y Missing Values

Una vez seleccionadas las variables que se utilizarán en el modelo, se procederá con la eliminación de datos que puedan aumentar el ruido en las predicciones. El objetivo de este procedimiento es predecir valores más certeros, además de mejorar la calidad del modelo. Para este estudio en específico, se considerarán dos tipos de datos que requieren tratamiento:

- **Outliers:** Son aquellos elementos de los datos que difieren en gran medida del resto, lo que puede ser producto de algún cliente con un comportamiento extremo distinto al normal, o simplemente porque los datos fueron mal registrados. Al ser muchos los clientes que pertenecen al Club de Fidelización, posiblemente existan muchos valores extremos, los que serán eliminados en su totalidad.
- **Missing Values:** Se eliminarán aquellos registros que no contengan valor en alguna de las variables a considerar en el estudio. Debido a que se cuenta con un gran número de datos, se eliminará el 100% de los registros que cumplan con esta condición.

Luego de haber eliminado todos los datos extremos, y con el objetivo de simplificar el trabajo estadístico, se extraerá una muestra de clientes. Debido a la gran cantidad de datos que se dispone, el tamaño de la muestra será de 50.000 clientes. Posteriormente se seleccionará un periodo de calibración y otro de testeo, los que serán utilizados para calcular el error en las predicciones y validar los resultados.

3.3 Análisis Descriptivo de Variables

Una vez teniendo todas las variables seleccionadas para el estudio, se hará un análisis descriptivo, para así entender cómo se comporta cada una de ellas. Esta etapa corresponde a una familiarización entre el investigador y los datos disponibles,

y así tener un conocimiento preliminar de los clientes antes de comenzar a desarrollar el modelo.

3.4 Análisis de Tendencias

La variable Periodo es aquella que representa el tiempo con unidad mes, y podría entregar información importante a nivel individual. Mientras algunos clientes tienden a comprar mayores montos hacia el futuro, otros pueden tender a disminuir sus montos de compra.

En esta etapa se analizan diferentes tendencias que se identifiquen a nivel individual, las que podrían ser captadas por el modelo de Regresión Jerárquica Bayesiana, no así mediante una Regresión Lineal a nivel Global.

3.5 Creación de Variables Dummies

Existen dinámicas en el comportamiento de compra que no se pueden captar considerando sólo variables continuas como las transaccionales, y para descubrir cuáles son éstas, se analizarán variables categóricas disponibles tales como el Periodo (mes) y Segmento Ciclo Vida.

El objetivo de crear dummies es poder considerar fenómenos como estacionalidades, o también momento del ciclo de vida que puedan afectar de manera especial en el monto gastado.

3.6 Predicción de Montos Mediante Método Simple

Para tener un primer pronóstico de los montos futuros, se trabajará con un método estático que no involucra modelos econométricos. La manera más común utilizada para minimizar el error de predicción corresponde al promedio natural de los meses históricos disponibles para cada cliente. Sin embargo, considerando la opinión de expertos en la empresa, se decide realizar la estimación asumiendo que el cliente gastará en el periodo t el monto gastado durante ese mismo mes del año anterior.

El objetivo de este método es obtener un resultado base acorde con la segmentación de valor que actualmente utiliza la empresa, para así distinguir a sus clientes y evaluar la mejoría de la predicción con los modelos econométricos.

3.7 Predicción de Montos a Nivel Individual

El método econométrico a utilizar será una regresión lineal simple, y para una selección aleatoria de clientes, se estimarán los parámetros a nivel individual. Los resultados obtenidos serán comparados con la regresión Bayesiana a realizarse en la etapa subsiguiente, la cual también trabaja a nivel individual.

3.8 Predicción de Montos a Nivel Agregado

En esta etapa de la metodología, también se utilizará una regresión lineal, pero esta vez se estimarán los parámetros a nivel global de clientes. El objetivo es predecir los montos futuros que gastará cada uno, pero asumiendo homogeneidad entre ellos.

3.9 Predicción Montos con Modelo Jerárquico Bayesiano

A diferencia de la regresión lineal individual, este modelo busca mejorar los resultados utilizando tanto la información global como la perteneciente a cada cliente en particular. El objetivo es considerar heterogeneidad entre los clientes, debido a que debieran existir diferencias importantes en sus comportamientos de compra, lo que podría afectar en la predicción del monto futuro.

En una primera instancia, se obtiene el parámetro a priori que representa la distribución poblacional de cada una de las variables predictivas. Luego, se deben obtener los parámetros específicos de cada cliente, utilizando como variable el parámetro común para toda la población.

El lenguaje que se utiliza para este tipo de regresión corresponde a R, específicamente se realiza mediante el paquete bayesm. La función `rhierLinearModel` incorporada en este paquete realiza las Regresiones Jerárquicas Bayesianas.

3.10 Validación de Resultados (Ajuste y Error)

Para poder validar los datos obtenidos en los modelos predictivos, es necesario realizar un análisis tanto del error como del ajuste. Existen algunos indicadores que representan esta información, los que ayudarán a concluir qué modelo entregó los mejores resultados.

Para analizar el ajuste del modelo, se observará el R^2 de cada regresión, indicador que debe ser cercano a 1 para explicar un mayor porcentaje de la varianza en la predicción. En el caso del error se considerarán el MAPE, donde la calidad del modelo será mejor mientras más bajo sea su valor.

3.11 Cálculo del Customer Lifetime Value (CLV)

Luego de haber estimado los montos futuros, y de seleccionar el modelo que haya tenido mejor calidad en la predicción, se procede a calcular el valor del cliente.

La tasa de descuento a utilizar será decidida en base a juicio de expertos, según las características particulares del negocio. Por otra parte, la tasa de retención considerada será 100%, ya que todos los clientes analizados son parte del Club de Fidelización, y la probabilidad de que dejen de comprar es muy baja.

A pesar de que en la ecuación del CLV se considera la contribución que el cliente aporta en el periodo t , en este caso se considerará sólo el monto, ya que actualmente es la variable monetaria que le interesa a la empresa para valorar a sus clientes.

El costo de adquisición no se considerará en el cálculo del CLV, dado que los resultados son para fines comparativos entre los clientes, y este valor es el mismo para todos ellos.

3.12 Análisis de Sensibilidad Tasa de Retención y Tasa de Descuento

El análisis de sensibilidad tiene por objetivo dimensionar qué tanto varía el valor del cliente utilizando distintos valores para la tasas de Retención y de Descuento, en base a los utilizados originalmente.

En el primer análisis de sensibilidad, se realizan 2 variaciones para la tasa de retención. Primero se varía constantemente en un 5%, tanto inferior como superiormente. Luego se analiza el cambio del CLV considerando tasas de retención específicas para cada Segmento de Valor que actualmente se maneja en la empresa.

En el segundo análisis se estudia la variación del CLV con una variación constante del 1% en la tasa de descuento, tanto superior como inferiormente. El objetivo es dimensionar cuánto afecta el valor que se le asigna a este parámetro, y cómo podrían cambiar las interpretaciones del estudio.

3.13 Ranking de Clientes

Una vez que se tenga el valor de cada cliente, se ordenarán los resultados obtenidos, con el fin de crear un ranking por CLV que permita diferenciar a los clientes más valiosos de los menos rentables.

El primer paso será analizar a los clientes utilizando el CLV como la variable más relevante, y se identificarán segmentos más y menos valiosos. Utilizando el resto de las variables disponibles, se podrá caracterizar a cada grupo de clientes. El objetivo es presentar la información de manera más ordenada y concisa, para que el área de marketing pueda mejorar la toma de decisiones en campañas focalizadas.

3.14 Propuesta de Acciones de Marketing

Finalmente, se propondrán algunas acciones de marketing que busquen potenciar la relación con los clientes más valiosos. El objetivo es monitorear de alguna manera a los clientes que representan mayor valor para la compañía, para así maximizar la rentabilidad que representan para la empresa.

4. DESARROLLO METODOLOGÍA

4.1 Selección de Datos

El club de fidelización que actualmente funciona en la empresa, nació en el año 2006, y hoy lo conforman en total 5.358.343 clientes. Por lo tanto, la base de datos que se dispone para realizar el estudio, es muy nutrida en información transaccional y demográfica. Es necesario destacar que el club sólo funciona para los formatos Hipermercado y Supermercado, dejando a los otros fuera del estudio.

El periodo de estudio corresponde a 24 meses desde 01 de Diciembre del 2008 hasta el 30 de Noviembre del 2010. Se consideraron todos los clientes que hubieran comprado al menos 1 vez durante ese periodo, los que forman un grupo total de 2.598.922 compradores fidelizados. Además es importante mencionar que la información se encuentra a nivel mensual. Dadas las características del problema que se busca resolver, se seleccionaron las siguientes variables transaccionales:

- **Periodo:** Variable temporal continua (mes t).
- **Recencia:** N° de meses desde la última compra a periodo t.
- **Frecuencia:** Número de transacciones realizadas en periodo t.
- **Monto:** Cantidad de dinero gastado en periodo t.
- **Utilidad:** Monto descontando costo de productos y promociones en periodo t.
- **Margen:** Porcentaje de Utilidad sobre Monto en periodo t.
- **Monto Cheque:** Valor monetario de puntos acumulados en periodo t.
- **Binaria Cheque:** Con valor 1 si cliente cobró cheque en periodo t-1.
- **Cantidad Promociones:** N° descuentos captados en periodo t.
- **Binaria Promociones:** Con valor 1 si captó alguna promoción en periodo t-1.

Paralelamente, se consideraron todas las variables demográficas que pudieran entregar información sobre dinámicas del comportamiento del cliente:

- **Edad**
- **Sexo**

Finalmente, se incluyeron algunas segmentaciones que actualmente existen en la empresa, las que podrían captar ciertas tendencias en la experiencia de compra del cliente:

- **Segmento Habitualidad:** Corresponde a segmentación que diferencia a los clientes por su habitualidad de visita (utiliza frecuencia).
- **Segmento Valor:** Clasificación del cliente según monto gastado en los últimos 3 meses.
- **Segmento Ciclo de Vida:** Define el ciclo de vida del cliente en periodo t.

Existe una alta correlación entre las variables Monto y Utilidad, sin embargo es necesario captar de alguna manera la rentabilidad que cada cliente representa para la empresa. Dado que tiene una baja correlación con el Monto, se ha seleccionado la variable Margen para dicha finalidad.

Con respecto a la variable Monto Cheque, esta se encuentra altamente correlacionada con la variable Binaria Cheque, pero se decidió utilizar esta última para la predicción del Monto. Por su parte, la variable Cantidad Promociones se encuentra fuertemente correlacionada con Binaria Promociones, pero para poder

captar de mejor manera el efecto de las promociones en la compra del cliente, se seleccionó la variable Cantidad Promociones.

Finalmente, se eliminaron de la base las variables Segmento Habitualidad por estar altamente correlacionada con la variable Frecuencia, y también se descartó Segmento Valor por tener fuerte correlación con Monto.

En resumen, las variables que se utilizarán en la predicción de los montos futuros serán:

- Periodo
- Recencia
- Frecuencia
- Margen
- Binaria Cheque
- Cantidad Promociones
- Edad
- Sexo
- Región
- Segmento Ciclo de Vida

Del conjunto de variables, sólo Sexo y Región permanecen constantes. Edad varía de un año a otro, y la variable Segmento Ciclo de Vida se actualiza cada 6 meses. Las variables restantes van cambiando mes a mes, de acuerdo a las transacciones de cada cliente en particular.

4.2 Tratamiento de Outliers y Missing Values

Existen varios registros de los clientes que poseen información faltante, o valores que se escapan considerablemente de lo normal. Para realizar un estudio más confiable y con datos que no produzcan ruido en los resultados, es importante eliminar todos estos valores erróneos.

Primero se identifican a los clientes que contengan alguno de sus registros por mes con información faltante, y se elimina totalmente del estudio. El total de clientes eliminados por esta razón son 388.289.

Posteriormente se analiza los valores extremos mediante gráficos de caja, los que tienen la ventaja de poder agrupar los valores que se encuentran en un rango normal, y reflejar aquellos valores que están muy alejados del promedio. Algunos de estos valores corresponden a clientes que han realizado devoluciones en sus transacciones, quedando registrado un monto muy bajo o con número 1 (por política de la empresa, para temas contables). La siguiente tabla resume los criterios de outliers para cada variable⁴:

⁴ Para detalle de los Gráficos de Caja, revisar Anexo C.

CRITERIO				
0	≤	Monto	≤	600.000
-48%	≤	Margen	≤	83%
1	≤	Frecuencia	≤	20
0	≤	Cantidad Promociones	≤	40

Tabla 1: Criterio para Outliers

Fuente: Elaboración propia

De esta manera se identifican a 102.350 clientes que poseen valores extremos en alguno de sus registros.

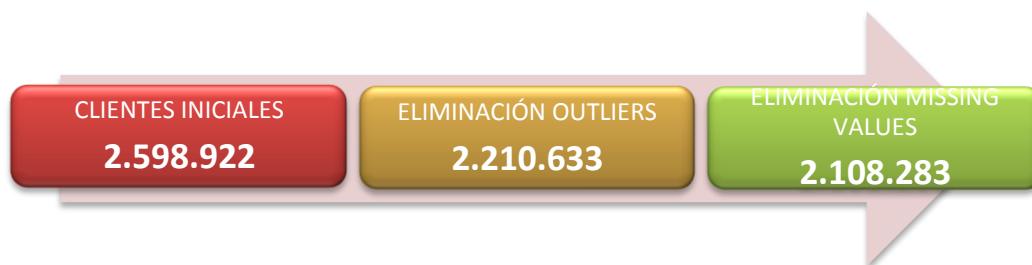


Figura 4: Proceso de Limpieza Base de Datos

Fuente: Elaboración propia

Finalmente se obtiene una base de datos con información sobre 2.108.283 clientes, que corresponde al 81% de los iniciales. Por simplicidad, en adelante se trabajará con una muestra aleatoria de 50.000 clientes, ya que la extensión del modelo para el resto es directa.

4.3 Análisis Descriptivo de Variables

Posterior a la selección de la base de datos que se utilizará para el estudio, es importante realizar un análisis descriptivo de las principales variables, con el fin de conocerlas y entenderlas. La siguiente tabla resume los principales estadísticos que describen a las variables transaccionales:

VARIABLE	MÍNIMO	MÁXIMO	PROMEDIO	D. ESTÁNDAR
Monto	0	\$ 599.996	\$ 46.622	\$ 69.808
Frecuencia	0	20	2,8	3,4
Cantidad Promociones	0	40	1,0	2,1
Recencia	0	21	1,3	2,8
Margen	-50%	83%	11%	11%

Tabla 2: Análisis Descriptivo de Variables Transaccionales

Fuente: Elaboración propia

El Monto que gasta cada cliente depende de otras variables como la frecuencia con que visita el supermercado, o la cantidad de promociones que aparecen en su ticket.

El Gráficos N°2 muestra la variación del monto gastado mensualmente según la frecuencia de compra, pero no es clara la presencia de montos superiores frente a una mayor cantidad de visitas al supermercado. Esto se debe a que hay clientes que a pesar de ir una o dos veces al mes, gastan tanto como otros clientes que van una o más veces a la semana.

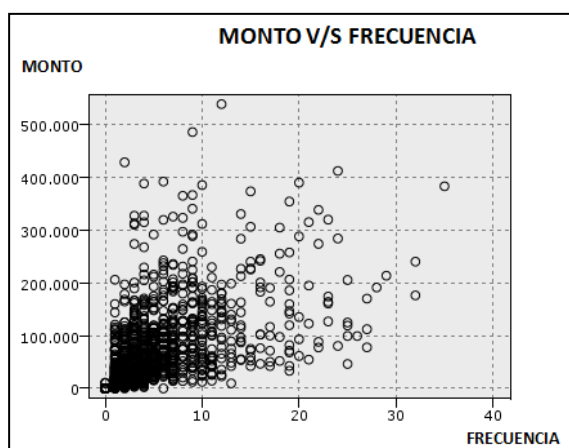


Gráfico 2: Gráfico de Dispersión Monto v/s Frecuencia

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Por otra parte, no es claro el comportamiento del monto según la cantidad de promociones que capta el cliente mensualmente. En el gráfico N°3 se logra distinguir una mínima tendencia positiva del monto, a medida que aumentan la cantidad de promociones captadas. Es evidente que mientras mayor es el monto de compra, probablemente es mayor la cantidad de productos comprados, y por ende aumenta la probabilidad de que un cliente haya captado más promociones.

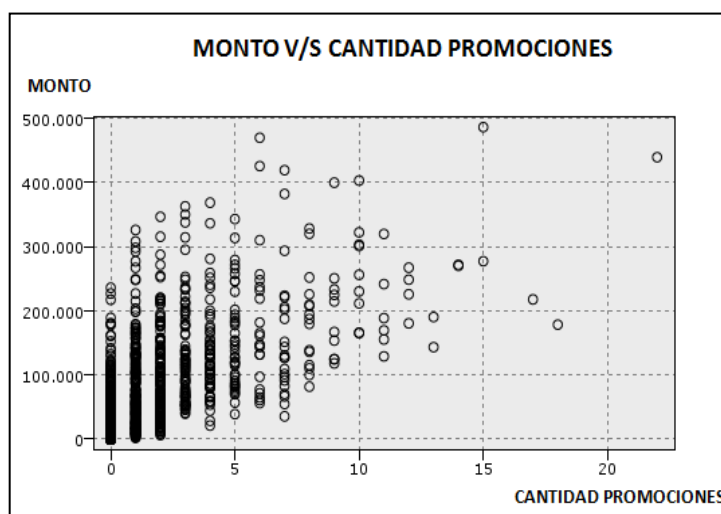


Gráfico 3: Dispersión Monto v/s Cantidad Promociones

Fuente: Elaboración propia en base a información proporcionada por la Empresa

La distribución de clientes por sexo es bastante marcada, ya que en su mayoría corresponde a mujeres. Sin embargo, hoy en día existen muchas mujeres profesionales activas, lo que provoca un aumento en la cantidad de hombres con la responsabilidad de comprar en el supermercado.

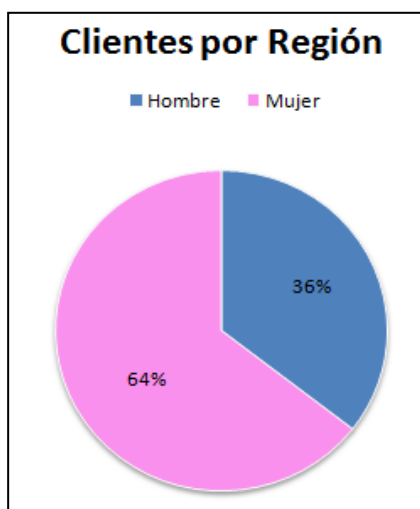


Gráfico 4: Distribución de Clientes por Sexo

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Al observar el monto promedio por sexo, se deduce que no hay mayores diferencias entre el hombre y la mujer (\$54.345 y \$48.737 respectivamente). Además, los integrantes de una familia suelen utilizar el RUT de la madre al realizar una compra, por lo que esta variable podría entregar información poco para la predicción del monto futuro. En base a este análisis, se decide no utilizar la variable Sexo en la predicción.

La evolución de la Edad si podría presentar dinámicas en el comportamiento de compra del cliente, ya que van cambiando tanto las necesidades del cliente, como también el presupuesto disponible para gastar.

La variable Segmento Ciclo de Vida posee bastante información sobre la evolución de un cliente con el pasar del tiempo, ya que considera aspectos importantes que podrían influir en el monto que gasta en cada periodo. Como se explicó en la selección de variables, se refiere al periodo de vida que está viviendo el cliente dentro del ciclo natural, y se distinguen 7 grupos:

1. Adulto joven hombre.
2. Adulto joven mujer.
3. Pareja sin hijos.
4. Familia con hijos igual o menor a 3.
5. Familia con hijos entre 4 y 12.
6. Familia con hijos entre 13 y 18.
7. Familia con hijos de 19 o más.
8. Adulto mayor.

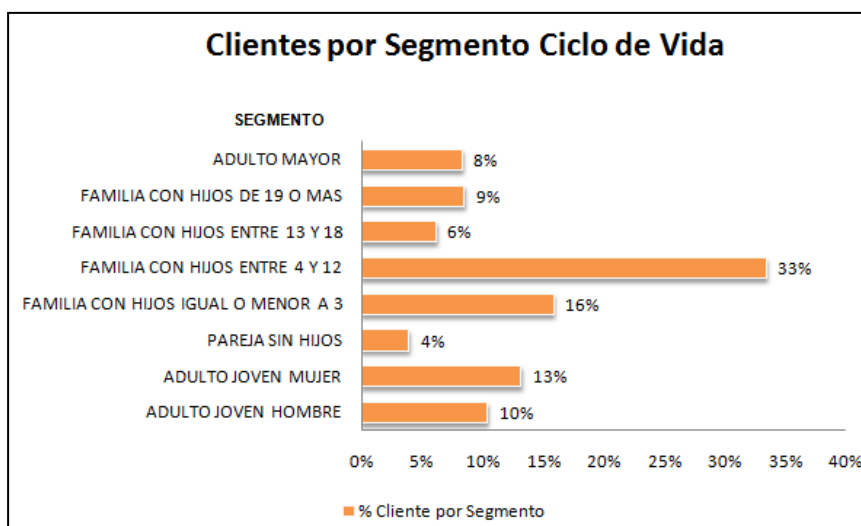


Gráfico 5: Distribución de Clientes por Segmentación “Ciclo de Vida”

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Según la distribución de clientes por segmento que se muestra en el Gráfico N°5, es evidente que en su mayoría son familias con niños chicos menores de 12 años, sin embargo no podemos sacar ninguna conclusión a priori de cómo afecta el ciclo de vida en el valor del cliente.

4.4 Análisis de Tendencias

Uno de los factores que no considera el método simple es la tendencia en el monto gastado por el cliente a través del tiempo, a medida que pasan los meses. La regresión Lineal Grupal indica que los clientes tienen una mínima tendencia positiva, pero que prácticamente es nula⁵. Esto quiere decir que los clientes tienden a gastar un monto constante en el tiempo, y que podría disminuir un poco al largo plazo. El siguiente gráfico corrobora lo anterior, al mostrar el monto promedio gastado en el periodo de data histórica, donde no existen mayores diferencias entre el año 1 (2009) y año 2 (2010):

⁵ Para ver los resultados de la Regresión Lineal global, ver desarrollo metodología punto 4.8 “Predicción de Montos a Nivel Agregado”

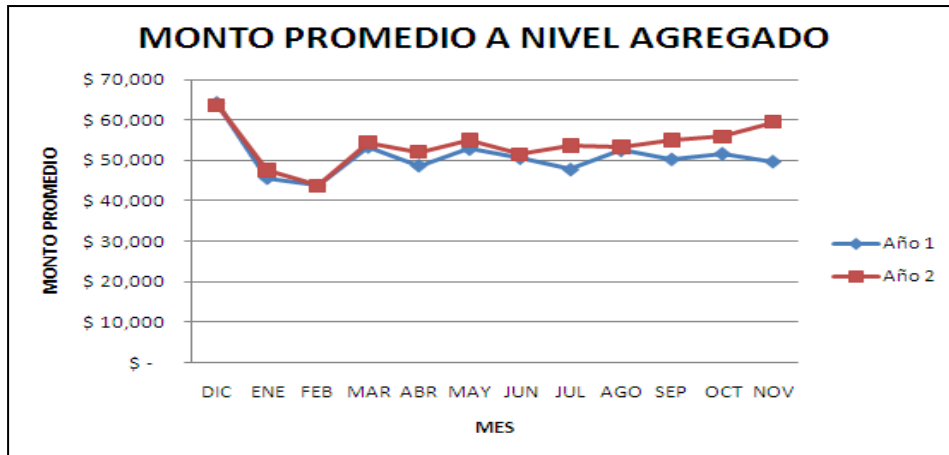


Gráfico 6: Serie de Montos Promedios a Nivel Agregado

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Sin embargo, al considerar el monto promedio agregado a nivel de clientes, no se está considerando la heterogeneidad que pudiese existir en su comportamiento. Para demostrar las diferencias de tendencias en el monto gastado, se identificaron a 2 clientes del grupo total, los pudieran representar distintos comportamientos hacia el futuro. El Gráfico N°7 muestra la serie de montos gastados por cada uno en particular, donde claramente uno tiende a comprar más hacia el futuro, mientras que el otro presenta una caída en sus montos gastados.

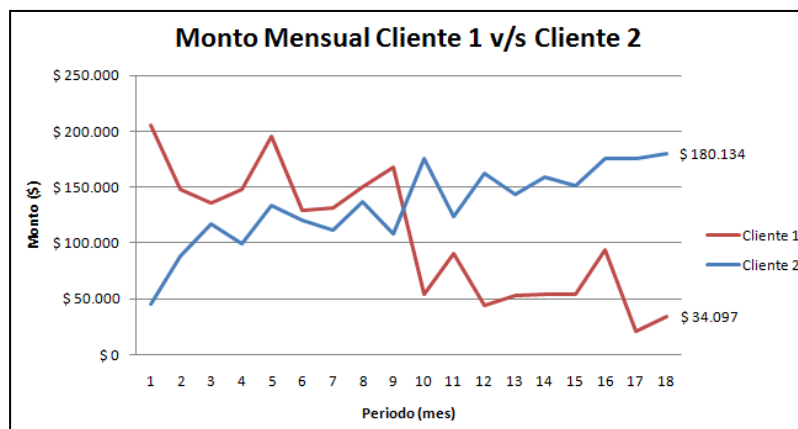


Gráfico 7: Montos Gastados Cliente 1 v/s Cliente 2

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Tanto la Regresión Lineal a nivel individual como también el Modelo Jerárquico Bayesiano, buscan captar este tipo de tendencias particulares de cada cliente, factor que no logra abarcar un Modelo Simple o Regresión Lineal Grupal.

4.5 Creación de Variables Dummies

La creación de variables dummies tiene la finalidad de captar ciertas dinámicas en la estimación de los montos futuros. Para identificarlas, fue necesario analizar gráficamente cruces entre las variables disponibles, principalmente las categóricas.

Una de las dinámicas identificadas al observar los montos promedio por periodo fueron las estacionalidades. Principalmente se observa un aumento considerable del monto durante el mes de Diciembre, y una baja durante los meses de verano Enero y Febrero. El resto del año se comporta de manera pareja, incluso en septiembre donde generalmente existe estacionalidad. Es importante considerar este factor en la estimación de los montos, ya que podría afectar positiva o negativamente, dependiendo del mes que se está pronosticando.

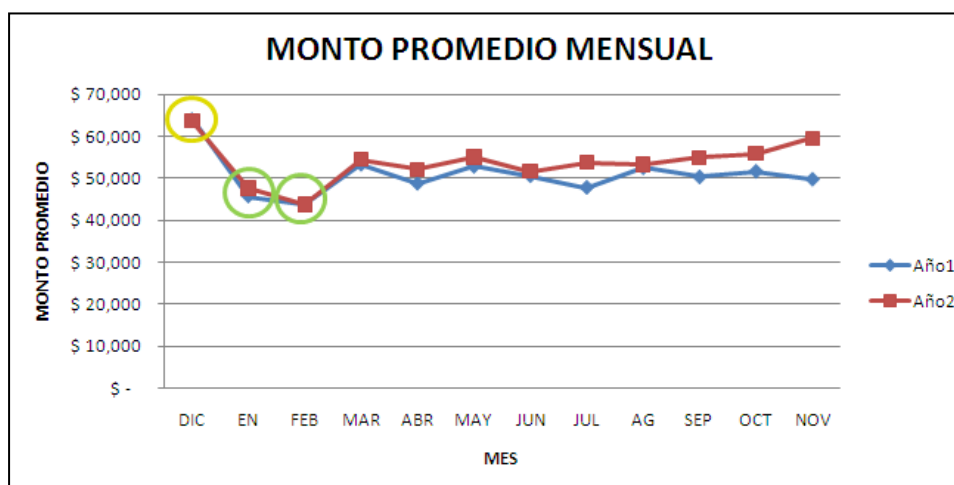


Gráfico 8: Montos Gastados Mensualmente a Nivel Agregado

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Para considerar la estacionalidad, se crearán 3 variables dummies que quedará valorada por 1 en caso de que se cumpla la condición, y 0 en caso contrario. Las nuevas variables son:

- **DICIEMBRE** (mes de Diciembre)
- **VACACIONES** (meses de Enero y Febrero)
- **NO_VERANO** (meses de Marzo a Noviembre)

Otra dinámica identificada corresponde a la variación del monto promedio gastado, en función del ciclo de vida que vive el cliente. Hay un aumento importante del monto de compra cuando se forma una familia y nace el primer hijo. Este monto se mantiene elevado mientras los hijos son niños menores a 13 años, y una vez que los hijos pasan a la adolescencia, el monto promedio del cliente vuelve a disminuir a un nivel semejante al que gastaba antes de tener hijos, manteniéndolo hasta la adultez mayor.

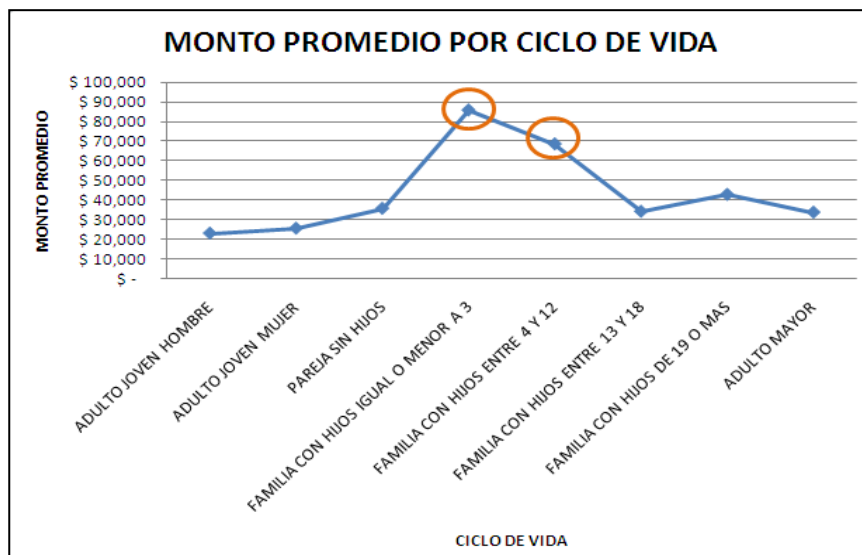


Gráfico 9: Monto Promedio por Segmentación "Ciclo de Vida"

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Por lo tanto, las nuevas variables dummies que tomarán valor 1 si cumplen la condición, y 0 en caso contrario, son:

- **MTO_SI_NIÑOS** (CICLO_VIDA "Familia con hijos igual o menor a 3" ó "Familia con hijos entre 4 y 12")
- **MTO_NO_NIÑOS** (CICLO_VIDA "Adulto joven hombre" ó "Adulto joven mujer" ó "Pareja sin hijos" ó "Familia con hijos entre 13 y 18" ó "Familia con hijos de 19 o más" ó "Adulto mayor")

4.6 Predicción de Monto Mediante Método Simple

Como ejercicio base y previo a las regresiones, se consideraron los montos futuros iguales a los gastados en el mismo mes del año anterior. Este criterio se decide luego de observar la estacionalidad en los montos gastados, ya que podría disminuir el error de predicción en comparación con otros criterios.

El Segmento de Valor que se le asigna al cliente en la actualidad se basa precisamente en el monto gastado durante el periodo anterior, y se pretende dimensionar el error cometido al utilizar ese método.

El método considera los montos gastado por los clientes en los periodos t del año $m-1$, el cual contempla los meses desde Diciembre 2008 hasta Noviembre 2009. De esta manera se estima que el monto a consumir en el mes t del año m serán igual al valor gastado en el mes t del año $m-1$.

Para observar la precisión de este criterio de estimación, se calcula el error MAPE sobre el monto real y sobre el monto estimado en los meses de validación, tal como lo especifican las siguientes fórmulas:

$$MAPE_{REAL} = \frac{1}{N} \sum_{i=1}^N abs \left[\frac{MG_{REAL} - MG_{ESTIMADO}}{MG_{REAL}} \right]$$

$$MAPE_{ESTIMADO} = \frac{1}{N} \sum_{i=1}^N abs \left[\frac{MG_{REAL} - MG_{ESTIMADO}}{MG_{ESTIMADO}} \right]$$

Donde $MG_{ESTIMADO}$ corresponde al monto estimado para el mes t , mientras que el MG_{REAL} es el monto gastado realmente durante el periodo t . De esta manera se logra calcular la diferencia porcentual entre la predicción y el valor real.

El error cometido utilizando este método es cercano al 200%, siendo levemente peor el MAPE calculado sobre el monto estimado. La siguiente tabla resume los errores calculados para el periodo de validación:

$MAPE_{REAL}$	$MAPE_{ESTIM}$
196%	208%

Tabla 3: MAPE Predicción Montos Método Estático

Fuente: Elaboración propia

4.7 Predicción Montos a Nivel Individual

Tal como se explica en el Marco Conceptual, una de las ventajas de utilizar regresiones Bayesianas es que logra captar la heterogeneidad de los clientes, es decir, obtener una función lineal única para cada cliente que logre explicar en función de las variables independientes, el monto que gastará en el futuro. Pero no se sabe si una regresión lineal típica a nivel individual basta para obtener resultados significativos y aceptables.

Se seleccionaron aleatoriamente a 2000 clientes, y se realizó una regresión para cada uno utilizando su información individual. El periodo de calibración corresponde a los datos entre el mes Diciembre del 2008 hasta Mayo del 2010, información que se utilizó para obtener los coeficientes de la función lineal que predice los montos.

Para validar los resultados, se utiliza el periodo de 6 meses desde Junio del 2010 a Noviembre del 2010, que corresponde al 75% de los datos disponibles.

Las variables independientes utilizadas son:

- Periodo
- Cantidad Promociones
- Frecuencia
- Recencia
- Margen
- Edad

- Binaria Cheque
- Dummy DICIEMBRE
- Dummy VACACIONES
- Dummy NO_VERANO
- Dummy MTO_SI_NIÑOS
- Dummy MTO_NO_NIÑOS

La Tabla N°4 resume la significancia de las variables considerando un 90% como nivel de confianza. El criterio de decisión fue que una variable es significativa si en más del 50% de las regresiones individuales lo fue. Por lo tanto, la única variable independiente que resulta ser significativa en la regresión es la Frecuencia.

VARIABLE	% Clientes Significativa	% Clientes No Significativa	Conclusión
Constante	29%	71%	NO SIG
Periodo	1%	99%	NO SIG
Cantidad Promociones	23%	77%	NO SIG
Frecuencia	68%	32%	SIG
Recencia	26%	74%	NO SIG
Margen	24%	76%	NO SIG
Edad	10%	90%	NO SIG
Binaria Cheque	16%	84%	NO SIG
Dummy NO VERANO	25%	75%	NO SIG
Dummy VACACIONES	34%	66%	NO SIG
Ajuste R² Promedio 0,784			

Tabla 4: Resultados Regresión Lineal Individual

Fuente: Elaboración propia

Por su parte, el R² promedio es del rango 0.7, y más del 70% de las regresiones tienen un Ajuste superior al 0.8. Sin embargo, al revisar la distribución de los R² se identifican algunos muy bajos y cercanos a cero. En el Gráfico N°10 se observa que los ajustes se concentran entre los valores 0.8 y 0.9, valores muy cercanos a uno, y que podrían significar sobre ajustes.

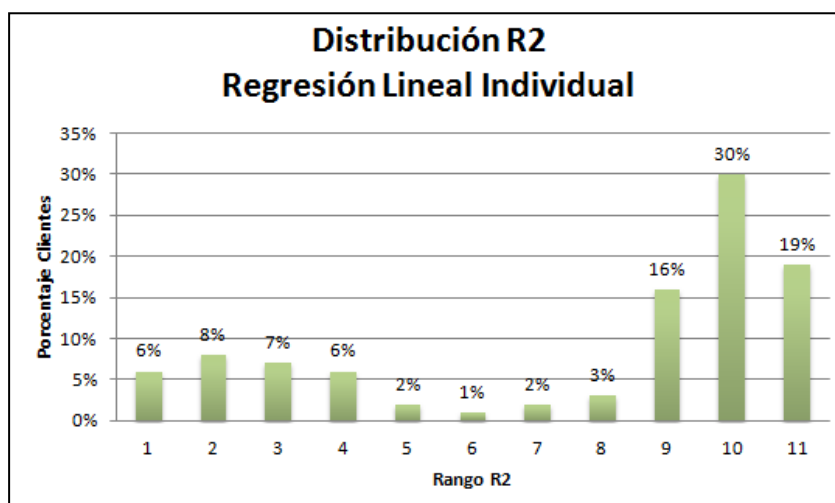


Gráfico 10: Distribución R2 Regresión Lineal Individual

Fuente: Elaboración propia

La cantidad de datos utilizados para la calibración de la Regresión Lineal Individual es muy reducida, por lo que el modelo responde sensiblemente a pequeñas diferencias en la información de entrenamiento, y la curva de predicción se ajusta estrictamente al comportamiento de dichos datos. Otro fenómeno que ocurre es el sub ajuste, el cual tiene el efecto opuesto donde el modelo no es sensible a las diferencias en los datos de calibración, prediciendo una curva que se aleja bastante de los datos reales. Estos fenómenos de sobre y sub ajuste, impiden que el modelo se pueda generalizar posteriormente a otro conjunto de datos.

Para ver la calidad en la predicción, se calcula el error MAPE tanto para el periodo de calibración y de validación. La siguiente tabla resume los valores obtenidos:

	MAPE _{REAL}	MAPE _{ESTIM}
Calibración	42%	45%
Validación	108%	124%

Tabla 5: Error MAPE Regresión Lineal Individual

Fuente: Elaboración propia

A pesar de que disminuye en comparación con el error de predicción obtenido en el método simple, los siguientes métodos buscan mejorar la calidad de la predicción.

4.8 Predicción Montos a Nivel Agregado

Como acción previa a la estimación de montos mediante las regresiones Bayesianas, se necesita tener un parámetro de comparación que permita analizar la calidad de los resultados obtenidos, desde una perspectiva agregada a otra

individual. Por lo mismo, se predicen los montos considerando homogeneidad entre los clientes, mediante una regresión lineal a nivel agregado⁶.

Bajo el mismo criterio que la regresión anterior respecto a los periodos de calibración y validación, se procede a estimar los montos futuros en base a las mismas variables independientes. Los resultados se resumen en la Tabla N°6.

VARIABLE	Coef. Estand.	Sig.
Constante		0.000
Periodo	0.004	0.000
Cantidad Promociones	0.483	0.000
Recencia	-0.030	0.000
Frecuencia	0.351	0.000
Margen	0.035	0.000
Edad	0.036	0.000
Binaria Cheque	0.078	0.000
Dummy VACACIONES	-0.011	0.000
Dummy NO_VERANO	-0.003	0.000
Dummy MTO_NO_NIÑOS	-0.058	0.000
Ajuste R² 0,605		

Tabla 6: Resultados Regresión Lineal Grupal

Fuente: Elaboración propia

A diferencia de la regresión individual, las variables resultan ser todas significativas en la predicción del Monto. Las variables transaccionales más importantes corresponden a la Cantidad de Promociones y Frecuencia, afectando en mayor medida la predicción de los montos. Los signos en los coeficientes estandarizados tienen sentido, como por ejemplo la Recencia, donde una disminución en el valor de esta variable afectaría negativamente la probabilidad de compra, y por ende menor monto posible a gastar. Por su parte, las variables binarias eliminadas según los grados de libertad necesarios, corresponde a las circunstancias en que el monto es mayor, y por lo mismo el signo de los coeficientes para las variables binarias consideradas son negativos.

El error de la predicción para el periodo de calibración y validación se resumen en la siguiente tabla, donde se presenta el MAPE sobre el valor real y sobre la predicción misma:

⁶ Para los análisis estadísticos hasta las regresiones lineales, se utilizó el software Clementine de SPSS.

	MAPE _{REAL}	MAPE _{ESTIM}
Calibración	232%	72%
Validación	153%	72%

Tabla 7: Error MAPE Regresión Lineal a Nivel Agregado

Fuente: Elaboración propia

Sin embargo, el error MAPE sobre el valor real sigue siendo elevado, por lo que se analizó la variabilidad de los montos desde el año 2009 al año 2010. El Gráfico N°11 representa esta variabilidad según la frecuencia de compra durante los 2 años de historia, y se puede ver importantes diferencias para los clientes que sólo han comprado una vez durante el periodo.

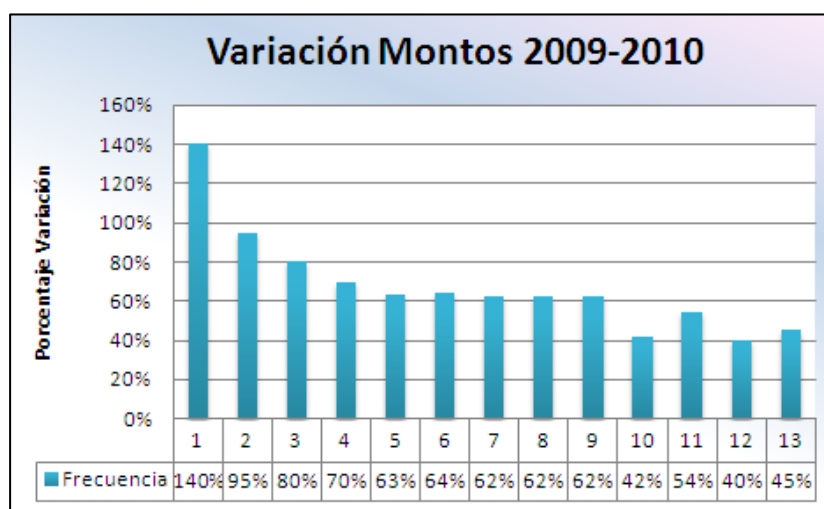


Gráfico 11: Variación de Montos Gastados Años 2009-2010

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Por otra parte, la variable Recencia es bastante difícil de proyectar para los meses del periodo de predicción, y dada su baja importancia relativa en el modelo de regresión, se eliminará del estudio.

Por lo tanto, para mejorar la calidad de la predicción, se realizarán las siguientes acciones en 4 nuevas regresiones:

- Se eliminarán clientes que hayan ido sólo 1 vez durante el periodo de data histórica.
- Se eliminará la variable Recencia, debido a que proyectarla en el periodo de predicción resulta muy complejo⁷ y no tiene mucha importancia según su coeficiente estandarizado (-0,03).

⁷ Proyección de variables se explica en sección del cálculo del CLV.

La Tabla N°8 representa la calidad de las regresiones realizadas en los distintos casos.

	CALIBRACIÓN		VALIDACIÓN		
	R ²	MAPE _{REAL}	MAPE _{ESTIM}	MAPE _{REAL}	MAPE _{ESTIM}
C/Rec, Visitas ≥1	0.630	195%	72%	190%	72%
C/Rec, Visitas ≥2	0.615	170%	73%	162%	73%
S/Rec, Visitas ≥1	0.630	192%	72%	185%	72%
S/Rec, Visitas ≥2	0.615	169%	69%	161%	74%

Tabla 8: Ajuste y error MAPE para distintas Regresiones

Fuente: Elaboración propia

A pesar de que el mejor ajuste lo tienen las regresiones que consideran a clientes con mínimo 1 visita durante el periodo de historia (0,630), la diferencia es muy baja. Además, el error de predicción disminuye notoriamente al sacar los clientes que presentan esa situación.

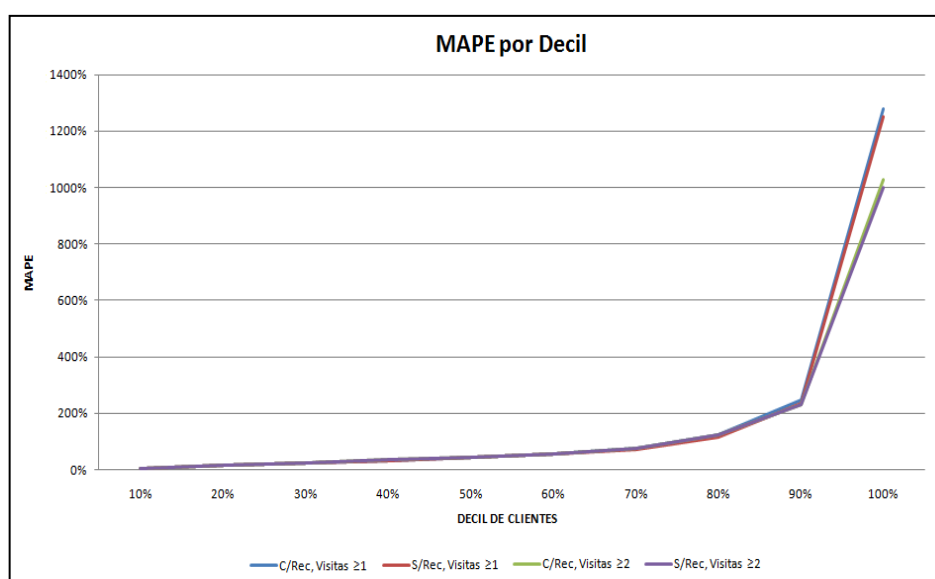


Gráfico 12: Error MAPE por decil de Clientes (periodo de validación)

Fuente: Elaboración propia

Se observa que en el modelo de regresión realizado sin considerar Recencia, y eliminando a todos los clientes con frecuencia 1, el 80% de los clientes explica un error MAPE cercano al 100%. Por otro lado, los modelos donde se considera a clientes con frecuencia 1, presentan un aumento considerable en el error, ya que el 80% de los clientes explican un error superior al 200%.⁸

⁸ Para ver los mismo gráficos de error MAPE sobre el valor Monto Predicción, revisar Anexo B.

El gráfico del error sobre el periodo de validación, a pesar de presentar menores diferencias, logra confirmar que la predicción mejora al eliminar clientes con frecuencia igual 1. Por otro lado, los efectos de eliminar Recencia como variable son despreciables.

Finalmente, la regresión que mejores resultados arroja luego de revisar tanto el ajuste como el error $MAPE_{real}$ cometido en las predicciones, corresponde a aquella donde no se consideran los clientes con visitas menores a 2, y en la cual se elimina Recencia como variable predictiva. Además, la empresa está particularmente interesada en los clientes de alto valor (quienes compran mayor monto), por lo que la variable Recencia no aporta mucha información.

Para poder analizar cuándo funciona mejor el modelo, se formaron quintiles de clientes según el $MAPE_{real}$ en la predicción de sus montos futuros. El QUINTIL 1 (puntos celestes) corresponde a aquellas predicciones con mayor error MAPE, mientras que el QUINTIL 5 (puntos amarillos) son aquellos con error más bajo. Al observar en detalle el Gráfico N°13, se concluye que el modelo funciona mejor con la predicción de montos pequeños, tal como muestra la concentración del color amarillo. Las mejores predicciones debieran encontrarse en la diagonal, dado que son aquellos valores predichos más cercanos a los reales de validación⁹.

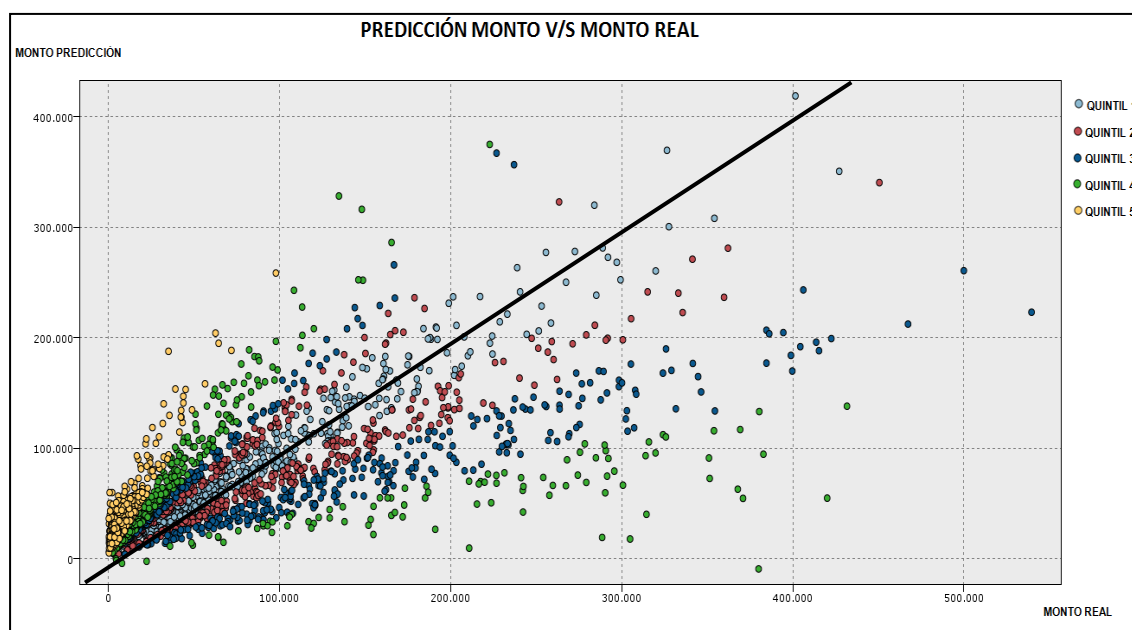


Gráfico 13: MAPE por quintil, Monto Real v/s Predicción (periodo validación)

Fuente: Elaboración propia

En el siguiente gráfico se resume la información obtenida en el análisis anterior, es decir, el Monto promedio real y el estimado para cada Quintil de clientes, además de la frecuencia de compra de cada grupo.

⁹ Para observar los mismos gráficos de error MAPE sobre el valor Monto Predicción, revisar Anexo B.

	Frecuencia	Monto Real	Monto Estimado	MAPE _{REAL}
Quintil 1	6,7	\$101.071	\$98.926	9%
Quintil 2	6,6	\$111.934	\$96.270	29%
Quintil 3	6,1	\$129.635	\$87.960	50%
Quintil 4	5,5	\$87.584	\$74.850	93%
Quintil 5	3,2	\$11.516	\$42.886	588%

Tabla 9: Estimación Montos, según Quintiles por MAPE_{REAL}

Fuente: Elaboración propia

Se puede observar que el modelo funciona mejor para predecir montos de clientes con mayor frecuencia de visita y con un gasto promedio superior a \$100.000, lo que se ve reflejado en un MAPE_{REAL} promedio de 9,4%. En cambio, en el caso de aquellos clientes con baja frecuencia y monto de gasto, el modelo no es tan eficiente, ya que el MAPE_{REAL} promedio asciende a 588%.

Por lo tanto, el Modelo de Regresión Lineal Grupal definitivo se muestra en la siguiente tabla:

VARIABLE	Coef. Estand.	Sig.
Constante		0.000
Periodo	0.006	0.000
Cantidad Promociones	0.467	0.000
Frecuencia	0.380	0.000
Margen	0.059	0.000
Edad	0.033	0.000
Binaria Cheque	0.080	0.000
Dummy VACACIONES	-0.011	0.000
Dummy NO_VERANO	-0.005	0.000
Dummy MTO_NO_NIÑOS	-0.056	0.000
Ajuste R² 0,615		

Tabla 10: Coeficiente Estandarizado y Significancia Regresión Lineal Grupal

Fuente: Elaboración propia

El ajuste sigue siendo aceptable, y todas las variables independientes resultaron ser significativas. Esto se debe a que existe una gran cantidad de datos utilizada en la Regresión Lineal grupal (todos los clientes) a diferencia de la individual donde sólo existen 18 observaciones por cada regresión (es decir, cada cliente). Al comparar los resultados con Tabla N^o6, no ven mayores diferencias en

los coeficientes estandarizados, pero el ajuste R^2 mejora levemente desde 0,605 a 0,615.

Una de las variables más interesantes de analizar es el Periodo, donde se distingue una mínima tendencia positiva, lo que se traduce en un leve aumento de los montos gastados a medida que avanzan los meses. Sin embargo, en el punto 4.4 del desarrollo de la metodología se distinguió tendencias tanto positivas como negativas en la totalidad de los clientes, heterogeneidad que no logra captar esta regresión grupal.

4.9 Predicción Montos Mediante Modelo Jerárquico Bayesiano

La mayor desventaja de utilizar una regresión lineal para predecir los montos futuros es que asume homogeneidad entre los clientes. Sin embargo, existen grandes diferencias entre ellos, lo que se ha podido observar en la variabilidad que tienen los datos históricos de los clientes. La motivación principal por utilizar modelos de regresión Bayesiana es precisamente el hecho de que considera heterogeneidad entre los clientes, logrando estimar parámetros individuales según su comportamiento particular.

Los datos con los que se trabaja fueron los mismos utilizados en la Regresión Lineal, considerando las observaciones a nivel de clientes. El programa estadístico que se utiliza para realizar la Regresión Bayesiana es R, específicamente la función `rhierLinearModel` que contempla el paquete `bayesm`. Para la obtención de los parámetros que estimarán el monto futuro, el modelo se corre con 2.000 iteraciones.

Los coeficientes obtenidos son a nivel de clientes, cuya distribución se puede observar en el siguiente gráfico, donde se ejemplifica el caso de la variable Periodo.

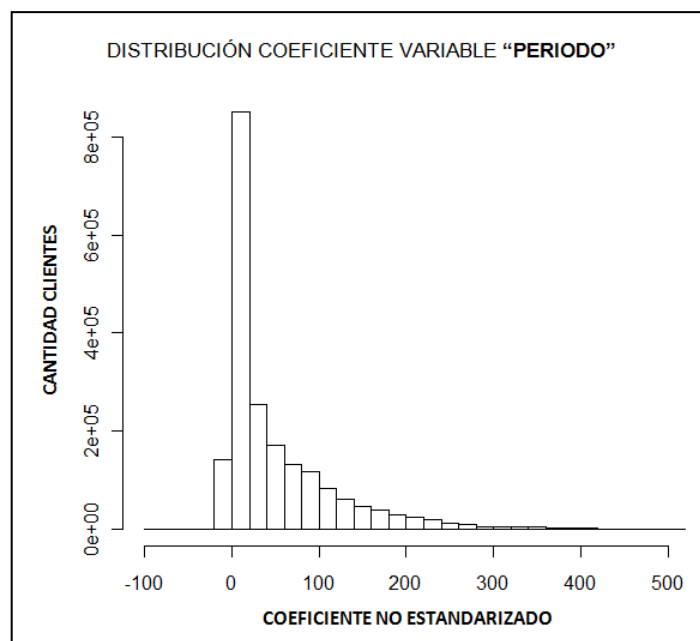


Gráfico 14: Distribución Coeficiente No Estandarizado Variable Periodo

Fuente: Elaboración propia

En este caso se puede ver que algunos de los coeficientes obtenidos son negativos, lo que explica una disminución de los montos a gastar en el futuro a

medida que pasan los meses. Sin embargo, una buena mayoría corresponden a coeficientes positivos, lo que evidencia la existencia de clientes que aumentan sus montos de gasto con el pasar del tiempo. Esto justifica lo observado en el punto 4.4 de la metodología, y demuestra la existencia de heterogeneidad entre los clientes, característica que no se logra ver en la estimación mediante Regresión Lineal Grupal.

Resulta importante analizar la significancia de las variables predictivas, para así compararlas con las obtenidas en las Regresiones Lineales Individuales. Este ejercicio se realizó para nivel de significancia del 90% y el criterio utilizado fue eliminar el 5% de los coeficientes más bajos y el 5% más alto. Se asume que las variables significativas son aquellas que no incluyen el cero dentro de los valores, luego de realizar la acción anteriormente descrita. El Gráfico N°15 ejemplifica el análisis de la significancia para la variable Periodo, proceso que se realiza para todas las variables utilizadas en la predicción¹⁰. Los valores en rojo corresponden a los eliminados, y los destacados en verde son los que se consideran para evaluar la significancia. En este ejemplo de la variable Periodo, el cero sigue estando incluido en la distribución de los coeficientes no estandarizados, por lo que esta variable se considera no significativa.

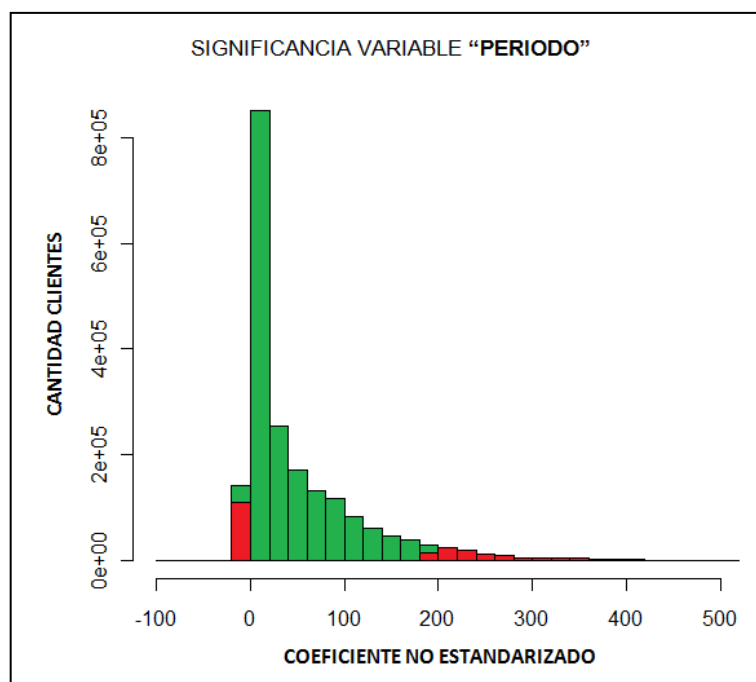


Gráfico 15: Análisis de Significancia Variable Periodo, R. J. Bayesiana

Fuente: Elaboración propia

La siguiente tabla comparativa resume la significancia de todas las variables utilizadas para predecir los montos, tanto con la Regresión Jerárquica Bayesiana (RJB) como también con la Regresión Lineal Individual (RLI):

¹⁰ Para ver análisis de significancia para todas las variables, revisar Anexo F.

VARIABLE	Int. Confianza RJB	Significancia RJB	Significancia RLI
Periodo	0,9	NO SIG	NO SIG
Cantidad Promociones	0,9	NO SIG	NO SIG
Frecuencia	0,9	SIG	SIG
Margen	0,9	SIG	NO SIG
Edad	0,9	SIG	NO SIG
Binaria Cheque	0,9	SIG	NO SIG
Dummy VACACIONES	0,9	SIG	NO SIG
Dummy NO_VERANO	0,9	NO SIG	NO SIG
Dummy MTO_NO_NIÑOS	0,9	NO SIG	NO SIG

Tabla 11: Significancia Variables R.J Bayesiana v/s R. L. Individual

Fuente: Elaboración propia

En la Regresión Jerárquica Bayesiana hay mayor cantidad de variables significativas, a diferencia de la Regresión Lineal Individual donde sólo una variable cumple con este criterio. Esto se debe a que en el último caso, los datos utilizados son las 18 observaciones para cada cliente, y dada la baja cantidad de registros, resulta muy difícil que una variable sea significativa. En cambio, en la Regresión Bayesiana, además de utilizar la información específica de cada cliente, también utiliza la información de la población total, lo que permite una estimación menos sensible frente a pequeñas variaciones en los datos utilizados. El tabla N°16 muestra la comparación de las pendientes de los coeficientes para ambas regresiones, en el caso de 6 clientes seleccionados al azar. En este caso, se utiliza nuevamente como ejemplo la variable Periodo.

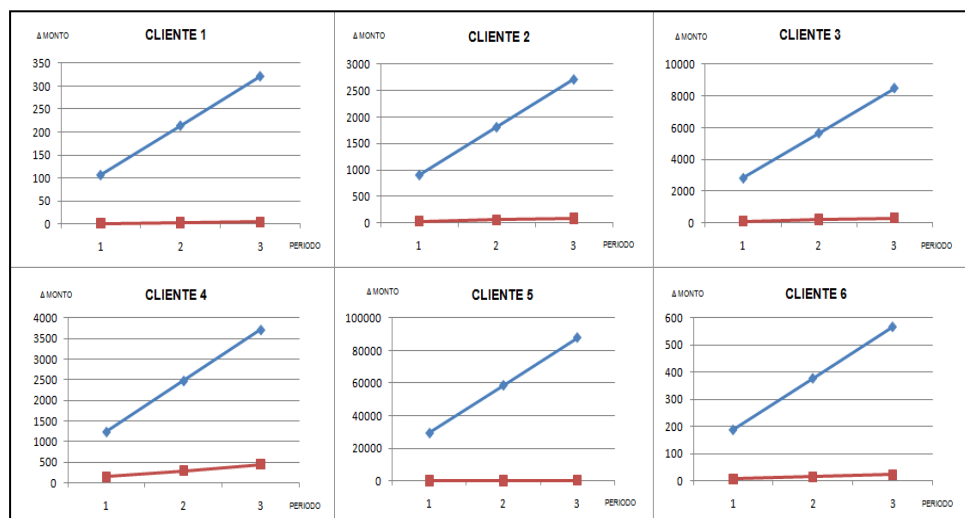


Gráfico 16: Pendiente Coeficientes No Est., R. L. Individual v/s R. J. Bayesiana

Fuente: Elaboración propia

Como se observa en los gráficos, la pendiente de los coeficientes obtenidos mediante la Regresión Bayesiana (serie color rojo) tiende a suavizarse en comparación con la Regresión Lineal Individual, por lo que se espera obtener un error menor en la predicción de los montos. Esto se debe a que suele haber sub o sobre ajuste en el modelo de Regresión lineal Individual, pero en el caso de la Jerárquica Bayesiana, la curva es menos sensible con pequeños cambios en los inputs, y por ende el modelo se puede generalizar a otro conjunto de datos. Lo mismo ocurre con el intercepto, que no absorbe tanta variabilidad en la estimación de los montos, y es precisamente una de las ventajas del modelo seleccionado.

4.10 Validación de Resultados

La validación de los resultados obtenidos en la estimación mediante el modelo de Regresión Jerárquica Bayesiana se realiza considerando dos criterios: El ajuste R^2 y el Error MAPE.

Dado que la estimación Bayesiana es a nivel individual, se debe calcular el promedio de todos los clientes, obteniéndose un R^2 ajustado de 0,652. Este valor es bastante similar al obtenido en la Regresión Lineal Grupal (0,615), y dado que explica el 65% de la varianza, sigue siendo un valor bastante aceptable.

El error se calcula como la variación porcentual entre el valor del monto predicho para el periodo de validación y el real gastado. El MAPE promedio del periodo de validación con respecto al valor real es de 82%, mientras que el calculado con respecto al valor predicho resulta ser 68%. La siguiente tabla compara los errores de los métodos utilizados:

	MAPE _{REAL}	MAPE _{ESTIM}
Método Simple	196%	208%
RL Individual	108%	124%
RL Grupal	161%	74%
RJ Bayesiana	82%	68%

Tabla 12: Resumen Error MAPE para Métodos Predictivos

Fuente: Elaboración propia

Se observa claramente que el pronóstico tiene menor error en la estimación mediante regresión Bayesiana, por lo que corresponde a un buen modelo para predecir los montos futuros. El siguiente gráfico muestra la evolución del MAPE para un cliente en particular, donde se observa el monto real del periodo de validación v/s los estimados (mes 19 al 24)¹¹:

¹¹ Para mostrar de manera más clara la calidad de predicción, se escoge un cliente con comportamiento normal, y que realiza al menos 1 visita mensual.

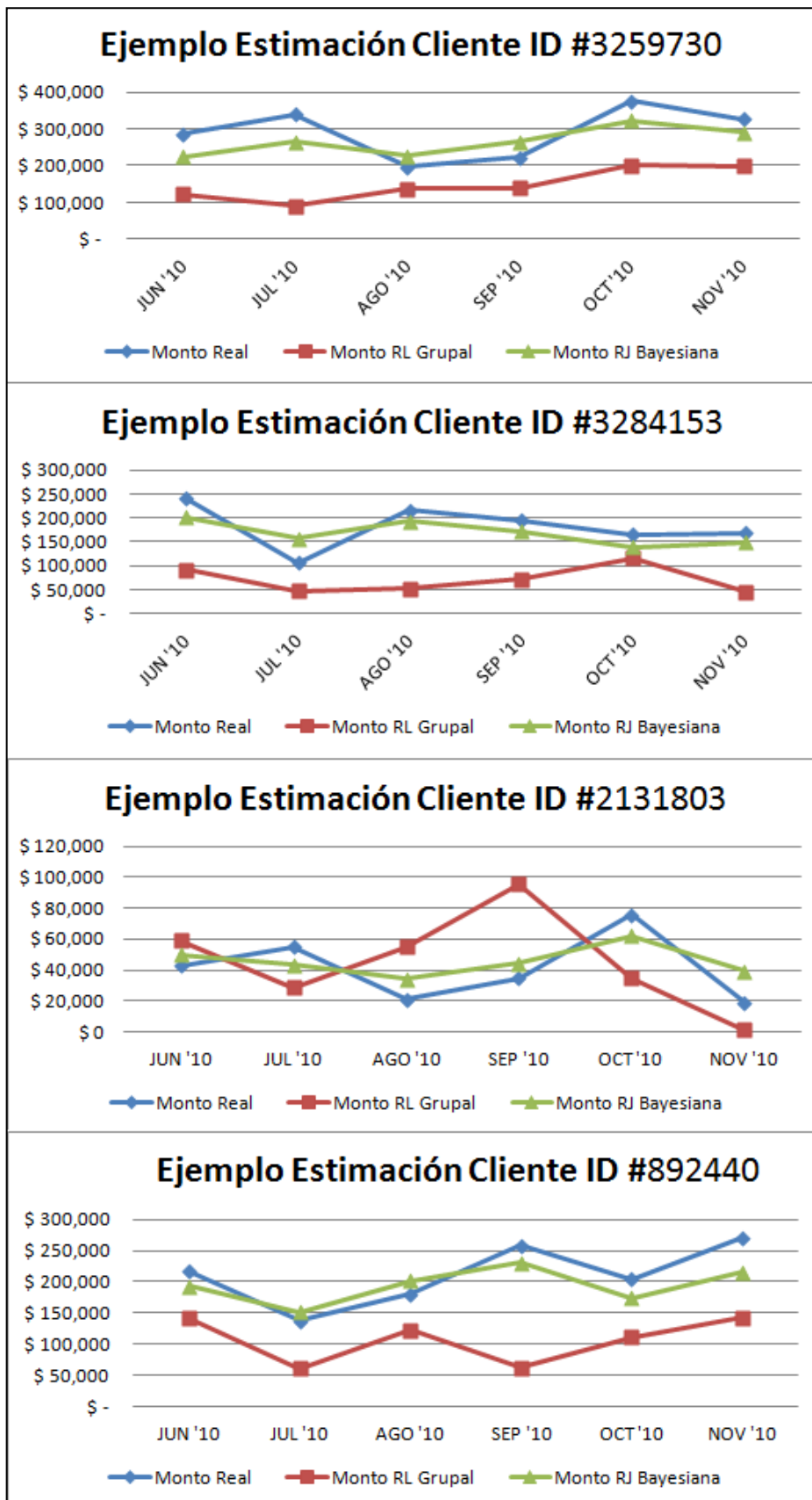


Gráfico 17: Estimación de Monto v/s Monto Real (Periodo validación)

Fuente: Elaboración propia

Los gráficos anteriores demuestran que el modelo de regresión Jerárquica Bayesiana logra predecir los montos de manera más certera que el modelo de Regresión Lineal Grupal. En los 3 casos mostrados se puede ver como la curva de color verde (estimación con Modelo Jerárquico Bayesiano) se ajusta de mejor manera al monto real gastado.

4.11 Estimación del Customer Lifetime Value (CLV)

Luego de haber validado los modelos para predecir los montos futuros, se selecciona la Regresión Jerárquica Bayesiana como el modelo a utilizar. Sin embargo, se debe realizar una proyección de las variables predictivas para poder obtener el monto futuro que gastará el cliente cada mes.

4.11.1 Proyección de Variables Predictivas

Para obtener los valores de las variables independientes correspondientes al periodo de predicción, se pueden realizar proyecciones futuras en base a factores de crecimiento o decrecimiento, o simplemente utilizar el promedio de la data histórica.

La variable Edad corresponde a un predictivo móvil en el tiempo que se va actualizando mes a mes. Por lo tanto, se tendrá la edad que el cliente debiera tener el mes para el cual se está realizando la predicción del monto.

El Ciclo de Vida es una segmentación que posee la empresa, la cual se actualiza cada 6 meses. Dado que es una característica del cliente con baja variabilidad en el corto plazo, se utilizará la última actualización para cada cliente, y se obtendrán las dummies correspondientes.

Las variables Frecuencia, Margen, Cantidad de Promociones y Binaria Cheque son aquellas incontrolables que dependen de cada cliente en particular, y se deben proyectar en forma individual para los meses futuros en los que se quiere predecir el monto. El siguiente gráfico representa la Frecuencia promedio agregada mensual, para los 2 años de data histórica disponible:

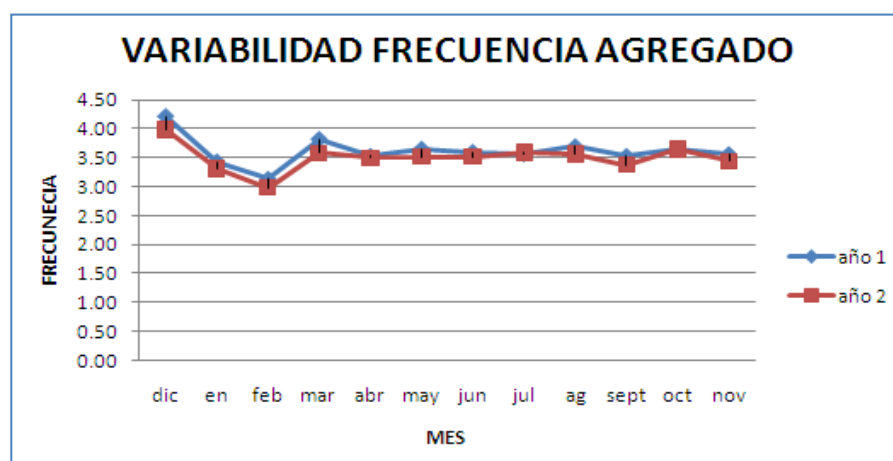


Gráfico 18: Variabilidad de Frecuencia Agregada, Año 1 – Año 2

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Al observar el comportamiento de dicho predictivo, se observa una baja variabilidad entre un año y otro, lo que logra concluir que el comportamiento de cada cliente es relativamente estable. Lo mismo se puede observar para las otras variables transaccionales independientes, por lo que se asume también cierta estabilidad¹². Por dicha razón, se calcula el promedio de las variables Margen, Frecuencia y Cantidad Promociones para cada mes t del año a predecir m , tal como lo muestran las siguientes expresiones:

$$MARGEN_{t,m} = \frac{MARGEN_{t,m-1} + MARGEN_{t,m-2}}{2}$$

$$FRECUENCIA_{t,m} = \frac{FRECUENCIA_{t,m-1} + FRECUENCIA_{t,m-2}}{2}$$

$$PROMOCIONES_{t,m} = \frac{PROMOCIONES_{t,m-1} + PROMOCIONES_{t,m-2}}{2}$$

La variable Binaria Cheque toma valor igual a 1 si el cliente cobró el cheque durante el periodo de cobro del año anterior al cual pertenece el mes en cuestión. Por ejemplo, si se está proyectando la variable Binaria Cheque del mes septiembre 2011 para el cliente i , tomará valor igual a 1 si el cliente i cobró cheque entre los meses Septiembre y Noviembre del 2010.

4.11.2 Cálculo del Customer Lifetime Value

Una vez con las variables predictivas proyectadas, se predice el monto desde el mes Diciembre 2010 al mes Noviembre 2011.

La tasa de descuento a utilizar fue decidida en base a opinión de expertos que trabajan en la empresa, principalmente pertenecientes al área de finanzas, y que corresponde a un 11%.

Por simplicidad, la tasa de retención a considerar será del 100%. Además, los clientes en estudio son parte del Club de Fidelización, y han realizado al menos 2 compras en el último par de años, por lo que se consideran clientes retenidos. Por lo tanto, la ecuación matemática queda definida como sigue:

$$CLV_i = \sum_{t=0}^T \frac{P_t \cdot r_t}{(1+i)^t}$$

El resultado del CLV para cada cliente corresponde entonces al valor presente de los montos que gastará en el futuro, según la predicción obtenida con la Regresión Jerárquica Bayesiana. Los resultados obtenidos son:

¹² Para ver el comportamiento de las variables Margen y Cantidad Promociones, dirigirse a Anexo D.

	PROMEDIO	MIN	MAX	MEDIANA	DESV. EST.
CLV	\$424.330	\$2.269	\$3.257.115	\$352.536	\$314.338

Tabla 13: Estadísticos CLV Clientes

Fuente: Elaboración propia

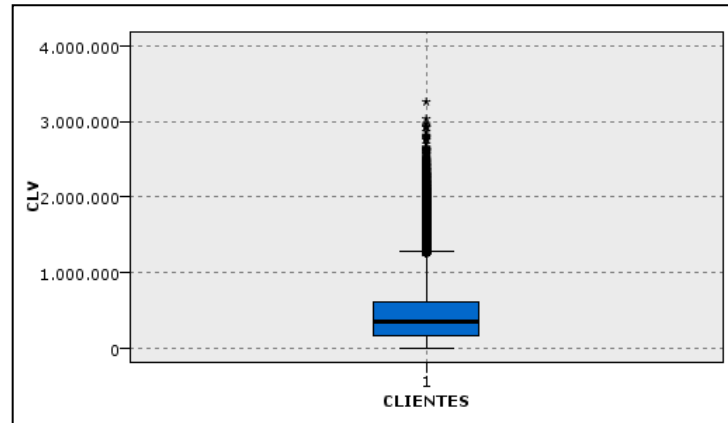


Gráfico 19: Gráfico de Caja CLV Clientes

Fuente: Elaboración propia

El promedio del CLV (\$352.536) es relativamente bueno, considerando que el periodo para el cual se calcula es de sólo un año. Evidentemente hay clientes con muy bajo valor que no vale la pena invertir en ellos. Sin embargo existen muy buenos clientes que se predice comprarán altos valores en el futuro, y a quienes hay que fidelizar en mayor medida. Más adelante se mostrará un ranking de los mejores clientes.

Para ver la distribución del CLV, se observan deciles de clientes, donde el “DECIL 1” corresponde al peor 10% y el “DECIL 10” al mejor 10%, tal como se puede observar en el siguiente gráfico:

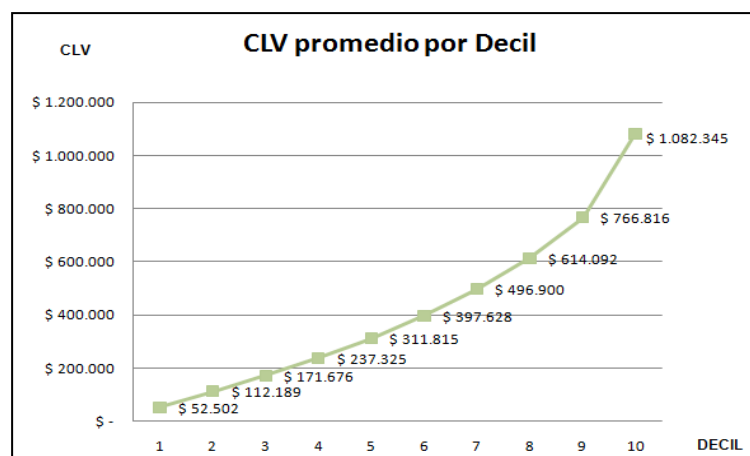


Gráfico 20: CLV Promedio por decil de Clientes

Fuente: Elaboración propia

Se puede ver que el mejor 10% de clientes tiene un CLV mayor a \$1.000.000, lo que representa a un grupo de clientes bastante rentable para la empresa. Además, aportar a la empresa el 25% de la contribución total durante el periodo de predicción, a pesar de ser sólo el 10% de los clientes totales. Por último, se puede distinguir que los primeros deciles se diferencian menos con respecto al CLV promedio de los clientes, pero a medida que se va acercando a los mejores deciles, se observa una diferencia mucho mayor entre los promedios.

Para poder organizar a los clientes por su valor, se decide realizar una segmentación por quintiles, para luego caracterizarlos según su comportamiento. La siguiente tabla resume el análisis descriptivo del CLV para cada grupo.

	PROMEDIO	MIN	MAX	MEDIANA	DESV. EST.
CLV Quintil 1	\$75.456	\$2.269	\$127.361	\$77.016	\$31.243
CLV Quintil 2	\$188.187	\$127.362	\$253.440	\$187.077	\$36.289
CLV Quintil 3	\$334.815	\$253.441	\$424.685	\$332.700	\$49.350
CLV Quintil 4	\$537.293	\$424.686	\$665.678	\$533.385	\$69.041
CLV Quintil 5	\$911.375	\$665.679	\$3.257.115	\$854.081	\$217.739

Tabla 14: Estadísticos CLV Clientes por Quintil

Fuente: Elaboración propia

El quintil 1 tiene un CLV promedio de \$75.456, y corresponde al quintil de clientes con menor valor. Se puede observar que el peor cliente tiene un CLV de \$2.269, el cual es muy bajo para la empresa considerando que el cálculo se realizó para un año de predicción. El peso del Quintil 1 que muestra la Tabla N°14, evidencia la baja importancia de estos clientes con menor CLV, ya que representan tan solo el 4% de la contribución total para el año de estudio.

El quintil con los clientes más valiosos tiene un CLV promedio de \$911.375, pero a pesar de no superar el millón de pesos, representan una rentabilidad bastante significativa para la empresa, ya que corresponde al 20% de los clientes totales. Y la contribución total de éste es prácticamente el 50% del total proyectado. Además, existen muy buenos clientes con un valor superior a \$1.000.000.

	Peso Segmento Según Contribución
Quintil 1	4%
Quintil 2	9%
Quintil 3	16%
Quintil 4	26%
Quintil 5	45%

Tabla 15: Peso de Quintiles según Porcentaje Margen

Fuente: Elaboración propia

4.11.3 Caracterización de Quintiles

Para poder observar en detalle el comportamiento de los clientes que pertenecen a cada Quintil, se analizó la información del último año de data histórica transaccional.

El siguiente gráfico representa la frecuencia promedio mensual de los clientes pertenecientes a cada Quintil, y existen notorias diferencias entre ellos. El Quintil 1 tiene clientes que visitan el supermercado en promedio 1 vez al mes. Por otra parte, los clientes más valiosos del Quintil 5 tienen una frecuencia promedio de 7 visitas al mes, por lo que es probable que los clientes compren más de una vez a la semana.

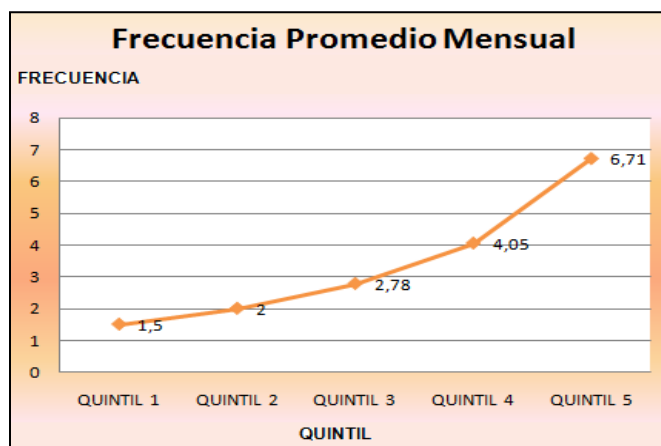


Gráfico 21: Frecuencia Promedio Mensual por Quintil de Clientes

Fuente: Elaboración propia

Si se observa el margen promedio que aportan los clientes de cada Quintil, es evidente una mayor contribución de los clientes con menor valor, lo que no es tan directo de entender. Pero al observar el Gráfico N°23 que representa el peso de cada categoría de productos por Quintil de clientes, este efecto se atribuye a que los consumidores con menor CLV compran más productos que marginan mayores porcentajes, a pesar de que compran menores montos.

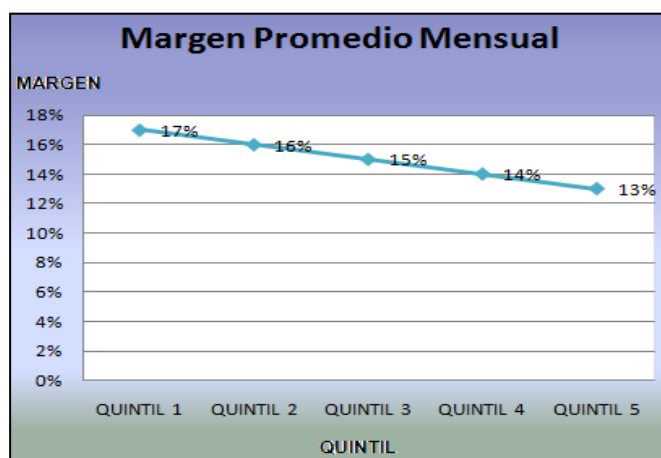


Gráfico 22: Margen Promedio Mensual por Quintil de Clientes

Fuente: Elaboración propia

El consumo de los clientes corresponde en su mayoría a Alimentos, según el Gráfico N°23¹³. Pero se pueden ver algunas diferencias entre los Quintiles, como por ejemplo un mayor consumo de Vestuario, Casa y Entretenimiento por parte de los clientes con menor valor. En cambio, los más valiosos consumen mayor porcentaje de Alimentos y Consumibles en comparación con el resto de los clientes.

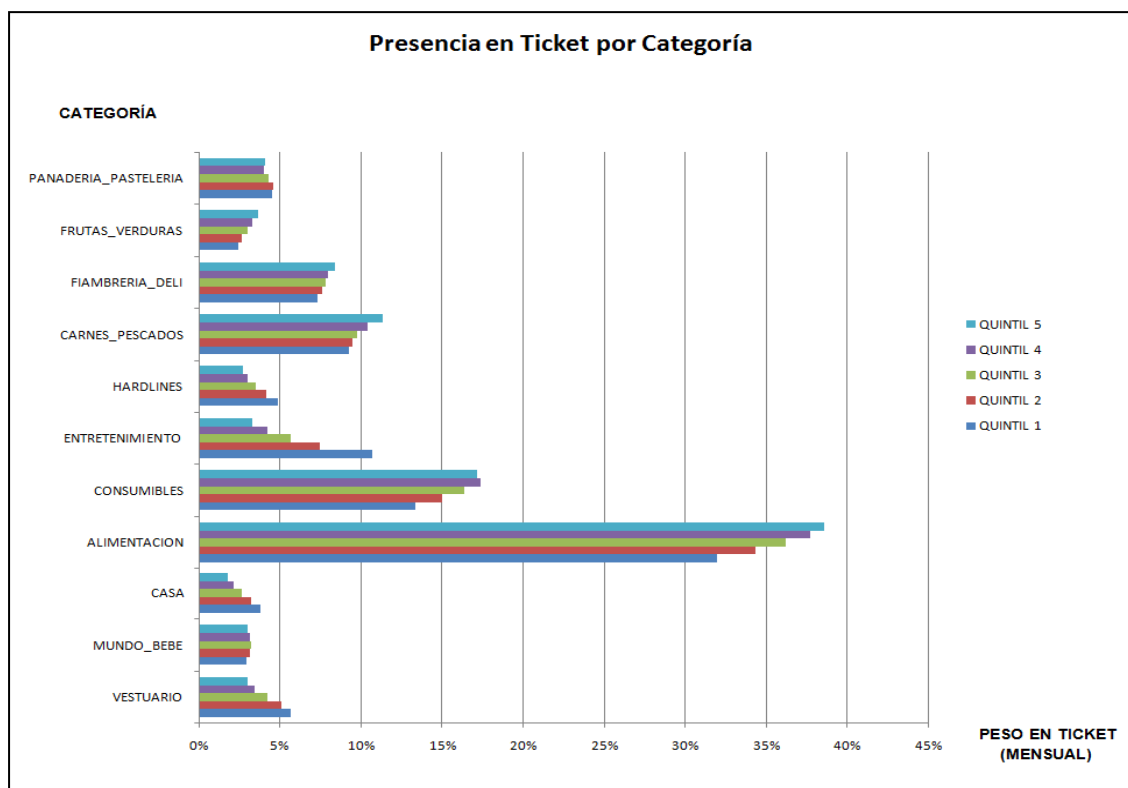


Gráfico 23: Presencia en Ticket por Categoría, a Nivel Mensual

Fuente: Elaboración propia en base a información proporcionada por la Empresa

La siguiente tabla muestra el promedio de los descriptivos más importantes, calculados para cada Quintil según el CLV de sus clientes:

	CLV PROMEDIO	MARGEN	FRECUENCIA
CLV Quintil 1	\$75.456	17%	1,5
CLV Quintil 2	\$188.187	16%	2,0
CLV Quintil 3	\$334.815	15%	2,8
CLV Quintil 4	\$537.293	14%	4,5
CLV Quintil 5	\$911.375	13%	6,7

Tabla 16: Estadísticos CLV Clientes por Quintil

Fuente: Elaboración propia

¹³ Para ver la presencia del resto de las categorías, ver Anexo E.

El primer Quintil corresponde a aquel 20% de clientes con menor valor, y se espera que compren bajos Montos durante el año en estudio, o incluso que no visiten el supermercado en algunos meses. Tienen una frecuencia promedio de compra mensual igual a 1, por lo que convierte a estos consumidores más bien esporádicos. Al observar las categorías que compran en mayor medida, es notorio un alto consumo de productos comestibles al igual que el resto de los clientes, sin embargo se aprecia un consumo mayor de las categorías Vestuario, Entretenimiento y Casa. Esto explica el hecho de que es el Quintil que mayor porcentaje margina (17%), dado que las categorías antes mencionadas son consideradas por la empresa como aquellas que más marginan.

El Quintil 2 tiene una frecuencia bastante similar al grupo anterior, ya que en promedio visitan el supermercado 2 veces al mes. El margen que contribuyen es un poco más bajo (16%), pero aún así sigue siendo mejor que el aportado por los clientes más valiosos. Este grupo también consume en su mayoría productos comestibles (Alimentos, Consumibles, Panadería, etc.), pero también tienen un alto interés por productos como Vestuario y Casa con respecto al resto de los clientes. Esto afecta positivamente el porcentaje que marginan, tal como se puede observar en el Gráfico N°22.

Los clientes del tercer Quintil claramente son más fieles que los anteriores, dada la frecuencia promedio de compra. Por lo mismo, son consumidores bastante potenciales, y con acciones de fidelización se podría lograr un aumento en su ticket promedio mensual. Las categorías que consumen en mayor medida son los comestibles, al igual que la mayoría de los clientes.

El Quintil 4 corresponde a aquellos consumidores que superan un CLV de \$500.000, y superan las 4 visitas por mes, lo que los hace clientes bastante fieles al supermercado. Al igual que el resto de los clientes, consumen en su mayoría productos como Alimentos, Consumibles y Carnes, pero se comienza a ver una disminución en el consumo de categorías como Casa y Vestuario. Esta situación explica también la disminución en el margen que rentabilizan a la empresa, ya que aumenta el consumo de productos que dejan menor utilidad, y disminuye el consumo de productos que marginan más.

El último Quintil lo conforman los clientes de mayor valor para la empresa, ya que tienen un CLV de \$911.375 cercano al millón de pesos. Este grupo también lo constituye un 20% de los clientes, lo que se traduce en una gran cantidad de clientes rentables. Este grupo de consumidores debe ser tratado de manera especial, con el objetivo de maximizar su relación con la empresa, y potenciar su valor hacia el futuro. Por otra parte, la frecuencia de compra promedio es cercana a las 7 visitas por mes, lo que evidencia una fidelidad alta hacia la empresa. Se espera que este grupo de clientes siga realizando sus compras con alta frecuencia, pero se le debe realizar acciones con el objetivo de maximizar el monto que consuman en el futuro. La categoría que compran en mayor medida son los productos comestibles, lo que explica el bajo margen que rentabilizan a la empresa. El consumo de categorías como Casa y Vestuario pasa a segundo plano, ya que no tienen mayor presencia en el ticket promedio. Sin embargo, se podría potenciar el consumo de dichas categorías, para así aumentar la contribución.

La siguiente tabla muestra la distribución de los Segmentos de Valor actuales por cada Quintil identificado:

	Quintil 1	Quintil 2	Quintil 3	Quintil 4	Quintil 5	TOTAL
PLATINO	0%	0%	0%	0%	100%	100%
ORO	2%	2%	5%	20%	71%	100%
PLATA	7%	6%	27%	39%	21%	100%
COBRE	30%	30%	22%	13%	5%	100%

Tabla 17: Distribución Segmento de Valor por Quintil

Fuente: Elaboración propia

4.12 Análisis de Sensibilidad CLV

La Tasa de Retención y Tasa de Descuento utilizados para el cálculo original, fueron decididos en base a la experiencia de expertos, pero corresponden sólo a supuestos. Por lo mismo, surge la necesidad de analizar la sensibilidad de los resultados al hacer variaciones en los valores de estos factores.

4.12.1 Tasa de Retención

Primero se analizó la sensibilidad de los resultados, al hacer pequeñas variaciones en la Tasa de Retención que originalmente era 100%. Se disminuye dicho factor en deltas de 5% hasta llegar a una del 80%, logrando 5 distintos escenarios.

La siguiente tabla muestra la variación de los promedios del CLV para los distintos Quintiles¹⁴:

	CLV (80%)	CLV (85%)	CLV (90%)	CLV (95%)	CLV (100%)
CLV Quintil 1	\$60.365	\$64.138	\$67.910	\$71.683	\$75.456
CLV Quintil 2	\$150.550	\$159.959	\$169.368	\$178.778	\$188.187
CLV Quintil 3	\$267.852	\$284.593	\$301.334	\$318.074	\$334.815
CLV Quintil 4	\$429.834	\$456.699	\$483.564	\$510.428	\$537.293
CLV Quintil 5	\$729.100	\$774.669	\$820.238	\$865.806	\$911.375

Tabla 18: Valor CLV por Variación de Tasa de Retención

Fuente: Elaboración propia

¹⁴ Para ver la variación del CLV entre un escenario y otro, revisar Anexo H.

Al hacer una variación pareja de la Tasa de Retención para todos los clientes, no afectaría el ranking de clientes, ya que el orden de los clientes según el CLV será el mismo, independiente de la tasa utilizada. Lo que podría ocurrir es una menor o mayor diferencia entre los CLV de los clientes, porque afectaría en mayor medida a los consumidores con alto valor.

Un análisis interesante observar los cambios del CLV al considerar tasas de retención específicas para cada segmentos de valor que actualmente posee la empresa. Dichas tasas están especificadas en la Tabla N°19¹⁵.

Segmento	Tasa Retención
PLATINO	82%
ORO	75%
PLATA	69%
COBRE	62%

Tabla 19: Tasas de Retención por Segmento “Valor”

Fuente: Elaboración propia

Los clientes PLATINO suelen ser bastante fieles al supermercado, y se ve reflejado en la alta tasa de retención. Esto se debe a que son clientes con alta frecuencia de compra mensual, y sería muy extraño que dejen de visitar el supermercado de un periodo a otro. En cambio los clientes COBRE son consumidores más bien esporádicos, y no necesariamente compran todos los meses, lo que se ve reflejado en su baja tasa de retención.

Los promedios de CLV por Quintil, luego de haber realizado las variaciones de Tasa de Retención por segmento¹⁶ se muestran en la Tabla N°20, donde se compara con los promedios calculados originalmente.

	Promedio CLV	Promedio CLV Original	Porcentaje Variación
CLV Quintil 1	\$65.378	\$75.456	13.36%
CLV Quintil 2	\$162.506	\$188.187	13.65%
CLV Quintil 3	\$278.797	\$334.815	16.73%
CLV Quintil 4	\$416.523	\$537.293	22.48%
CLV Quintil 5	\$693.223	\$911.375	23.94%

Tabla 20: Sensibilidad CLV por variación Tasa Retención por Segmento “Valor”

Fuente: Elaboración propia

¹⁵ Para ver el proceso de obtención de las Tasa de Retención por segmento de valor, revisar Anexo A.

¹⁶ El análisis se basó en el segmento al que pertenece el cliente el mes Abril del 2011.

Se puede ver que los Quintiles más afectados son el 4 y 5, que corresponden a los clientes más valiosos. La alta variación del promedio del CLV en los Quintiles de alto valor, se debe a la alta presencia de clientes PLATA, quienes tienen la tasa de retención más baja (69%). En cambio, los Quintiles 1, 2 y 3 poseen alto porcentaje de clientes COBRE, que tienen una tasa de retención muy alta (92%).¹⁷

4.12.2 Tasa de Descuento

El segundo análisis de sensibilidad que se realiza es en base a la variación de la tasa de descuento utilizada (11%.) Se vuelve a estimar el CLV, pero variando el factor de interés en deltas de 1%, llegando inferiormente a una tasa del 6%, y superiormente a una tasa de 13%. La tasa de retención utilizada fue la original (100%), para así ver netamente la sensibilidad de la tasa de descuento.

Las tablas N°21 y N°22 muestran la variación del CLV promedio para cada Quintil, considerando los distintos escenarios inferiores y superiores al 11% respectivamente:

	Variación 11%-8%	Variación 11%-9%	Variación 11%-10%	Promedio Variación
CLV Quintil 1	13%	8%	4%	13%
CLV Quintil 2	12%	8%	4%	13%
CLV Quintil 3	13%	8%	4%	13%
CLV Quintil 4	13%	8%	4%	13%
CLV Quintil 5	13%	8%	4%	13%

Tabla 21: Sensibilidad CLV por Variación Inferior Tasa de Descuento

Fuente: Elaboración propia

	Variación 11%-12%	Variación 11%-13%	Variación 11%-14%	Promedio Variación
CLV Quintil 1	-4%	-7%	-9%	-7%
CLV Quintil 2	-4%	-7%	-9%	-7%
CLV Quintil 3	-4%	-7%	-9%	-7%
CLV Quintil 4	-4%	-7%	-9%	-7%
CLV Quintil 5	-4%	-7%	-10%	-7%

Tabla 22: Sensibilidad CLV por Variación Superior Tasa de Descuento

Fuente: Elaboración propia

¹⁷ Para ver gráficamente la distribución de los Segmentos de Valor por Quintil, ver Anexo I.

Como se ve en las tablas resumen, una variación de la tasa de descuento afecta prácticamente igual en todos los Quintiles. Dado que el objetivo del CLV en este estudio es para términos comparativos entre clientes, se deduce que la tasa de descuento no afecta mayormente en las decisiones, ya que independiente del valor utilizado, este afecta en la misma medida a todos los clientes. Por lo tanto, los clientes más valiosos seguirán siendo los mismos, independiente de la tasa de descuento utilizada para calcular el CLV.

4.13 Ranking de Clientes

Una de las herramientas que se pueden crear en base al nivel del cliente es el 'ranking'¹⁸. Bajo el mismo criterio de la segmentación de valor actual, donde un cliente Premium es aquel que gasta al menos \$200.000 mensuales, se estimó el mínimo CLV que debe tener un cliente que gastará en los meses futuros al menos ese monto. El CLV que se obtuvo bajo este criterio fue de \$1.441.303, por lo que se identificaron a los clientes que tuvieran al menos ese valor estimado. El grupo de clientes que cumplen con esta condición son en total 9.963 consumidores, que corresponden a los clientes más rentables para la empresa.

Según los resultados obtenidos, los top 20 clientes en CLV están representados en la siguiente tabla:

¹⁸ Definición Wikipedia, palabra clave 'Ranking': "Es una relación entre un conjunto de elementos tales que, para uno o varios criterios, el primero de ellos presenta un valor superior al segundo"

ID	CLV	MONTO	MARGEN	FREC.	PROM.	CICLO VIDA	1RA CATEG	2DA CATEG	RANKING
1767764	\$ 3,257,115	\$ 307,636	2%	10	22	FAM HIJOS IGUAL O MENOR A 3	Aliment	Consum	1
1199938	\$ 3,032,274	\$ 185,809	5%	11	19	FAM HIJOS IGUAL O MENOR A 3	Aliment	Consum	2
1221045	\$ 2,965,293	\$ 153,530	-4%	11	18	FAM HIJOS ENTRE 4 Y 12	Aliment	Consum	3
2538164	\$ 2,941,418	\$ 220,529	5%	10	18	FAM HIJOS ENTRE 4 Y 12	Aliment	Consum	4
665078	\$ 2,923,469	\$ 310,630	8%	9	18	FAM HIJOS ENTRE 4 Y 12	Aliment	Consum	5
2496570	\$ 2,921,256	\$ 329,563	8%	12	18	FAM HIJOS ENTRE 4 Y 12	Aliment	Consum	6
1183520	\$ 2,920,678	\$ 264,060	5%	6	20	FAM HIJOS ENTRE 4 Y 12	Aliment	Consum	7
1931898	\$ 2,863,013	\$ 305,976	6%	10	20	FAM HIJOS ENTRE 4 Y 12	Aliment	Carn_Pesc	8
1612838	\$ 2,815,114	\$ 293,874	-6%	10	18	FAM HIJOS ENTRE 4 Y 12	Carn_Pesc	Aliment	9
2888897	\$ 2,800,536	\$ 279,827	8%	6	18	FAM HIJOS IGUAL O MENOR A 3	Aliment	Consum	10
1873288	\$ 2,796,678	\$ 239,619	7%	13	15	FAM HIJOS ENTRE 4 Y 12	Aliment	Consum	11
2116754	\$ 2,789,137	\$ 306,567	9%	11	16	FAM HIJOS ENTRE 4 Y 12	Aliment	Entret	12
2708219	\$ 2,773,888	\$ 317,307	8%	13	16	FAM HIJOS IGUAL O MENOR A 3	Aliment	Consum	13
2602592	\$ 2,760,026	\$ 411,115	2%	12	16	FAM HIJOS ENTRE 4 Y 12	Aliment	Carn_Pesc	14
2599838	\$ 2,756,597	\$ 212,772	-2%	8	19	FAM HIJOS ENTRE 4 Y 12	Aliment	Carn_Pesc	15
1903697	\$ 2,701,374	\$ 261,023	2%	10	16	FAM HIJOS ENTRE 4 Y 12	Aliment	Consum	16
3211929	\$ 2,657,149	\$ 331,203	7%	8	19	FAM HIJOS ENTRE 4 Y 12	Aliment	Consum	17
1223312	\$ 2,646,126	\$ 273,320	7%	11	15	FAM HIJOS IGUAL O MENOR A 3	Aliment	Consum	18
2245751	\$ 2,639,147	\$ 241,631	5%	12	15	FAM HIJOS ENTRE 4 Y 12	Aliment	Carn_Pesc	19
1401427	\$ 2,628,961	\$ 148,509	7%	15	14	FAM HIJOS IGUAL O MENOR A 3	Aliment	Consum	20

Tabla 23: Ranking Clientes TOP 20

Fuente: Elaboración propia

Se puede ver que todos los clientes con los mejores 20 CLV poseen frecuencias mensuales bastante elevadas, por lo que efectivamente corresponden a consumidores fieles al supermercado. Otra característica común es el bajo Margen que representan todos, incluso algunos marginan negativo, lo que se debe a un alto consumo de productos que marginan bajos porcentajes. Lo anterior se justifica con el elevado número de promociones que aparecen en los tickets de estos clientes, por lo que suelen comprar productos en oferta. Esto se traduce en que los clientes más valiosos son consumidores “inteligente” para comprar, ya que buscan siempre los productos en oferta.

4.14 Propuesta Acciones de Marketing

Luego de tener un diagnóstico de los 5 quintiles de clientes según el CLV, se deben proponer acciones de marketing que ayuden a potenciar la relación con los clientes más valiosos.

En base a reuniones con expertos de la empresa pertenecientes al área de Marketing, se decide enfocar los esfuerzos en sólo 3 grupos (Quintiles 1, 4 y 5). Los clientes del primer Quintil compran muy esporádicamente, y mediante acciones focalizadas se puede potenciar su frecuencia de visita. Los Quintiles 4 y 5 corresponden a los consumidores con alta frecuencia y monto gastado, y por ende se debe potenciar la relación con estos clientes más valiosos. En cambio, los Quintiles 2 y 3 son consumidores que asisten mensualmente al supermercado, pero a pesar de ser más fieles que el primer grupo, resulta más difícil potenciarlos a gastar mayores montos del que están acostumbrados.

Los clientes del primer segmento corresponden mayormente a adultos jóvenes hombres y mujeres. Por su parte, los del Quintil 4 son aquellos clientes con buena frecuencia de compra, pero que podrían aumentar su ticket promedio a través del consumo de otras categorías de productos. Finalmente, el último segmento está formado por los clientes con mejor valor, pero muy propensos a captar promociones. En base a estas características, los Quintiles seleccionados quedan definidos con los siguientes nombres:

- **QUINTIL 1:** “Joven”
- **QUINTIL 4:** “Potencial”
- **QUINTIL 5:** “Gold Inteligente”

Las acciones propuestas fueron analizadas en conjunto con el área de Marketing, y surgen desde el diagnóstico realizado previamente en base a los resultados obtenidos.

4.14.1 Acciones para Segmento “Joven”

Este segmento es interesante de potenciar dado el bajo monto que consumen, y el alto margen que reportan a la compañía. El CLV promedio de estos clientes es más bien bajo, y por lo mismo se debe enfocar la acción en aumentar este valor. Por lo mismo, se proponen las siguientes acciones de Marketing:

- Envío de catálogos CASA y VESTIMENTA al hogar:

La acción consiste en potenciar el consumo de estas categorías, dado el alto consumo de ellas con respecto al resto de los quintiles. El objetivo es seguir

incentivando la compra de productos para vestimenta y hogar, para que así aumenten los montos gastados mensualmente, y como consecuencia aporten mayor contribución a la empresa. Paralelamente, el objetivo es potenciar la relación de un cliente joven potencial, que en el futuro podría elegir este supermercado como su “favorito”.

- Descuentos en caja para categoría frecuentes:

Esta acción surge al identificar la baja frecuencia de compra que presentan los clientes del segmento “Joven”. El objetivo es mantenerlos atraídos al supermercado, entregándole descuentos estructurales¹⁹ en caja que tengan como periodo de validez 2 semanas. De esta manera se lograría aumentar la frecuencia de compra mensual y el monto total consumido en el periodo.

Esta acción se justifica al observar el efecto de la variable Cantidad Promociones en la predicción de montos, ya que tiene una fuerte incidencia positiva a medida que el cliente capta más descuentos.

- Descuentos en productos potenciales:

Este tipo de clientes consumen bajos niveles de productos Alimentos y Consumibles, por lo que existe una oportunidad de venta cruzada que aumente el consumo de éstos últimos.

Mediante un análisis de canasta, se puede identificar ciertas categorías potenciales donde el cliente presenta un consumo bajo, pero que los datos históricos muestran alguna tendencia positiva en su compra. Dicha acción pretende incentivar al cliente a elegir este supermercado para empezar a comprar productos que normalmente están comprando en la competencia. De esta manera se logra aumentar la probabilidad de que el cliente prefiera este supermercado como el principal para realizar las compras mensuales.

4.14.2 Acciones para Segmento “Potencial”

A diferencia del grupo anterior, estos clientes compran mayores montos, pero siguen siendo inferiores a los consumidos por los clientes más valiosos. El objetivo de las acciones que se proponen para este grupo es precisamente incentivarlos a subir el escalón del CLV a un nivel superior.

- Descuentos en categorías más consumidas:

Los clientes “Potencial” marginan mayor porcentaje que los clientes más valiosos, sin embargo consumen menor monto en cada periodo. Una manera de aumentar este monto es entregándoles descuentos en las categorías más compradas, pero que aseguren un pequeño aumento en la cantidad consumida. Por ejemplo, si actualmente están consumiendo una mayonesa de 200 grs, el descuento debe ser enfocado en el mismo producto, pero de tamaño 350 grs.

¹⁹ Descuentos Estructurales: Corresponde a descuentos entregados al finalizar una transacción, los que tienen un tiempo de validez, aumentando la probabilidad de que el cliente vuelva a comprar en ese periodo.

Además, el descuento tendrá una duración de validez de 2 semanas desde el momento que se entrega, lo que se traduce en una mantención o aumento de la frecuencia de compra.

Durante la metodología se comprobó que la Cantidad de Promociones que capta un cliente en su boleta corresponde a una variable de fuerte incidencia positiva en los montos que gasta, y un descuento en la categoría más consumida, resultaría altamente atractivo para un cliente que suele captar este tipo de promociones.

- Envío de descuentos a domicilio:

Esta acción está enfocada en aquellos clientes que presenten una ausencia al supermercado superior a 2 semanas. El objetivo es volver a aumentar su frecuencia de compra que posiblemente se vea afectada dada esta inasistencia. Se sigue utilizando el descuento como la principal estrategia, dado que corresponde al mayor incentivo para estos clientes, pero ligado a este caso específico donde se espera atacar una ausencia poco común en los consumidores “Potencial”.

4.14.3 Acciones para Segmento “Gold Inteligente”

Este grupo es el más valioso según los resultados obtenidos, pero son interesante dado el bajo margen que reportan a la compañía. Podrían considerarse clientes perfectos, sin embargo es necesario buscar una acción que logre aumentar el margen, además de potenciar la relación con ellos hacia el futuro.

- Sampling²⁰ de Marcas Propias “Selección”:

El supermercado en estudio posee productos de Marcas Propias para 2 tipos de categorías: “Básicas” y de “Selección”²¹. Estos últimos compiten con las mejores marcas dado su alto estándar de calidad, pero con la diferencia que marginan mayores porcentajes.

La acción consiste en enviar Sampling de productos “Selección” a domicilio, para asegurarse de que el cliente pruebe esta nueva marca de alta calidad, y que no está acostumbrado a comprar. El objetivo es lograr que un porcentaje de los clientes empiece a comprar éstas marcas, reemplazando los productos “caros” que habitualmente compran en cada categoría de Alimentos. De esta manera, se logra aumentar el margen que reportan a la compañía, sin ver afectado otros factores como el monto o la frecuencia.

- Muestra de marcas propias “Selección” en locales Sector Oriente:

La acción consiste en promocionar en sala los productos “Selección”, aumentando la probabilidad de que el cliente pruebe la calidad éstos, y que perfectamente podrían reemplazar la marca “cara” que compran habitualmente en cada categoría. El objetivo principal es seguir satisfaciendo la necesidad del cliente,

²⁰ Samplig: Pequeña muestra de un producto, la que se puede entregar en sala o a domicilio, y que tiene como objetivo presentar un producto al consumidor.

²¹ Los productos “Selección” corresponde a marcas propias con estándares de calidad superiores al básico, que compiten con las mejores marcas en algunas de las categorías de Alimentos.

pero aumentando en alguna medida el margen que actualmente contribuyen a la compañía.

Esta actividad se recomienda para locales del sector oriente, donde existe un mayor consumo en productos de mayor calidad, y donde el cliente tiene mayor disposición a pagar.

- Descuentos estructurales en caja:

Una de las principales características positiva de estos clientes es la alta frecuencia de compra. Esta acción consiste precisamente en entregarle un descuento al momento de finalizar una transacción, con el objetivo de mantener o aumentar su cantidad de visitas.

Para lograr mayor efectividad en la acción, se propone realizar un análisis de canasta que logre identificar ciertos productos potenciales, principalmente aquellos que compran con mayor frecuencia.

5. CONCLUSIONES

5.1 Conclusiones del Trabajo

El presente trabajo tuvo por objetivo general la estimación del valor de clientes fidelizados en una cadena de supermercados, el cual fue logrado con éxito mediante una metodología novedosa. El principal factor que era necesario considerar en la estimación de los montos era precisamente la heterogeneidad que existe en el comportamiento de los clientes, razón por la cual se utiliza el Modelo Jerárquico Bayesiano. Este modelo es capaz de capturar tendencias y dinámicas particulares para cada individuo, obteniendo estimaciones a nivel individual.

El estudio surge de la necesidad de la empresa por estimar un valor del cliente que considerara proyecciones hacia el futuro, ya que actualmente se le asigna un valor en base al monto gastado en los últimos 3 meses. Luego de pronosticar los montos futuros bajo ese mismo criterio (Método Simple), se concluye que el error de la predicción es cercana al 200%, lo que motiva a seguir buscando nuevos métodos de estimación.

Una de las etapas más importantes del trabajo fue el análisis de la base de datos disponibles, la cual contempla 2 años de data histórica a nivel mensual. Fue necesario realizar cruces entre las variables con el objetivo de identificar dinámicas o tendencias en el comportamiento de compra de los clientes. Finalmente se logran captar dos dinámicas y una tendencia, todas relacionadas con la variable del tiempo "Periodo". Las dinámicas identificadas corresponden a la estacionalidad, y por otra parte el aumento del monto gastado en algunos ciclos de vida. Con respecto a la tendencia, se logra identificar ciertos clientes que tienen a compra menores montos hacia el futuro, mientras que otros tienden a comprar más, factor que no logra ser capturado por una regresión Lineal a nivel grupal.

La primera estimación de montos futuros mediante modelos econométricos, fue utilizando regresiones Lineales a nivel agregado, es decir, considerando a todos los clientes dentro de un mismo grupo. A pesar de que el error de pronóstico se redujo en un 35% con respecto al método simple, la homogeneidad asumida empeora la estimación. Este método resulta significativo dada la cantidad de registros a nivel de grupo, pero no logra captar la heterogeneidad que existe entre sus individuos. Frente a esta necesidad, surge la opción de realizar varias regresiones lineales individuales, las que utilizan los registros particulares de cada cliente para estimar los parámetros del modelo. Sin embargo, sólo una variable predictiva resulta ser significativa, y tiende a existir efecto de sub y sobre ajuste en la curva de estimación. El error MAPE para el periodo de validación se reduce notoriamente, pero persiste el afán por encontrar un modelo significativo que logre captar la heterogeneidad existente en un grupo de clientes.

La Regresión Jerárquica Bayesiana utiliza la información global de la población para encontrar una distribución a priori de los datos, la que posteriormente se utiliza como inputs en la regresión a nivel individual. De esta manera se logra estimar los

parámetros particulares para cada cliente, pero mejorando la significancia de las variables predictivas. Por su parte, el error MAPE se reduce a un 82%, porcentaje notoriamente mejor que el obtenido en las predicciones realizadas previamente.

El CLV estimado a nivel de cliente corresponde al valor presente de los montos que éste gastará en el futuro, y según las predicciones realizadas por el modelo Jerárquico Bayesiano, se obtiene un resultado promedio de \$424.330. Este promedio es relativamente bueno, considerando que el horizonte definido es sólo un año. Existen clientes con muy bajo valor, en quienes no vale la pena esforzar recursos adicionales de marketing. Sin embargo, hay clientes que representan una alta rentabilidad para la compañía, y se debe potenciar una buena relación cliente-empresa hacia el futuro.

Para analizar el comportamiento de los consumidores según el CLV, se separaron en Quintiles desde el menos al más valioso. El primer grupo está formado principalmente por adultos jóvenes, quienes a pesar de gastar bajos montos, marginan un alto porcentaje a la compañía. Esto los convierte en un Quintil interesante, ya que se podría potenciar su compra con acciones focalizadas, y así aumentar su valor futuro. Los Quintiles 2 y 3 tienen un comportamiento relativamente parejo, y son los menos atractivos para destinar recursos de marketing adicionales. Por su parte, los últimos 2 grupos son los más valiosos, y se debe potenciar una relación Cliente-Empresa a largo plazo. Sin embargo, estos consumidores son muy propensos a captar promociones, y la utilidad que dejan se ve negativamente afectada. Esto se soluciona incentivando la compra de productos que marginen mayores porcentajes a la compañía.

El análisis de sensibilidad realizado para la Tasa de Retención, muestra un notorio efecto en el valor de los clientes, al considerar tasas particulares para cada Segmento de Valor que actualmente existe en la empresa. Por ejemplo, el Quintil 1 posee mayoritariamente clientes COBRE, quienes tienen una tasa de retención cercana al 90%, por lo que los resultados del CLV se ven afectados en un 13%. Por otra parte, el Quintil 5 que tiene a casi todos los clientes PLATINO con tasa de retención cercana al 80%, se ve más afectado con una variación promedio del CLV del 24%. Otra conclusión que se obtiene luego de realizar el análisis de sensibilidad de la Tasa de Descuento, es que los quintiles se ven afectados prácticamente en la misma medida. Además, como el objetivo del estudio es para temas comparativos entre los clientes, y como esta variación de tasa afecta de igual manera a todos, los clientes más valiosos seguirán siéndolo independiente de la tasa utilizada.

Finalmente, las predicciones de montos mediante una Regresión Jerárquica Bayesiana permiten estimar el valor de los clientes de manera individual, captando la heterogeneidad que existe en un grupo de clientes, y que logra reducir el error predicción considerablemente.

5.2 Trabajos Futuros

Una de las ventajas al realizar estudios del valor del cliente, es que existen muchas maneras de poder complementar los resultados.

Se sugiere realizar un análisis de canasta completo para cada cliente, que logre identificar productos ya sea potenciales, o que se compren con alta frecuencia. Este trabajo sería un complemento perfecto para algunas de las acciones de marketing propuestas. Por otra parte, sería interesante analizar a los clientes que se encuentre en un punto límite, ya sea porque están bajando su frecuencia de compra, como también aquellos que están próximos a ser clasificados como valiosos.

Una mejoría importante para el modelo desarrollado, sería realizar el estudio para la totalidad de los clientes del supermercado, dado que este trabajo se realizó sólo para una muestra de 50.000 clientes. Se podría investigar nuevos programas que logren desarrollar modelos Jerárquicos Bayesianos, que sean capaces de considerar mayor cantidad de observaciones, en un menor tiempo de ejecución.

Dado que se pueden realizar acciones de marketing basadas en el valor del cliente, se propone realizar experimentos promocionales y posterior análisis del CLV, para ver la evolución de los clientes. De esta manera se podrían identificar acciones específicas para lograr ciertos efectos específicos en el CLV de los clientes.

Otro trabajo futuro que se propone, sería realizar un modelo estadístico de proyección para las variables predictivas, que mejore la calidad de la estimación de los montos. Se podría analizar la data a nivel semanal, para así tener un comportamiento más específico de las variables utilizadas, y en base a esta información, desarrollar el modelo propuesto.

Finalmente, se propone realizar un modelo de proyección para la tasa de retención a nivel individual, acción que permitiría mejorar la calidad de la predicción de los montos futuros que actualmente se realiza considerando una tasa de retención del 100%.

5.3 BRIBLIOGRAFÍA

- [1] Brian Woolf. 2010. Marketing de Lealtad, El Segundo Acto.
- [2] Estrategia, El Diario de Negocios de Chile. 2005. [en línea]
Disponible en: http://www.estrategia.cl/detalle_noticia.php?cod=37907.
- [3] Feller-Rate Clasificadora de Riesgo. 2010. Informe de Clasificación, Distribución y Servicio D&S S.A: [9-11]
- [4] Garrido, Marcel. 2009. Estimación del Valor de Clientes de un Club de Fidelización de una Tienda de Retail. Memoria para optar al título de Ingeniero Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.
- [5] Luis G. Renart Cavas. 2004. CRM, tres estrategias de éxito: [5-19]
- [6] Marketing Directo. 2010. Palabra clave:” Marketig Directo”.[En Línea].
Disponible en: www.marketingdirecto.com
- [7] MDS Marketing de Servicios. 2008. Marketing Relacional, cada Cliente Único y Especial.
- [8] Peter E. Rossi, Greg M. Allenby, Robert McCulloch. 2005. Bayesian Statistics and Marketing: [70-75, 267-]
- [9] Sofía J Vallejos. 2006. Minería de Datos: [6-10]
- [10] Sunil Gupta, Dominique Hanssens y Bruce Hardie. 2006. Modeling Customer Lifetime. Volume 9.
- [11] Troncoso, Catalina. 2010. Determinación de Precios Óptimos en una Cadena de Supermercado Utilizando Modelos Jerárquicos Bayesianos. Memoria para optar al título de Ingeniero Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.
- [12] Urzúa, Pedro. 2007. Estimación de Customer Lifetime Value mediante Técnicas Supervisadas de Data Mining en una Empresa de Retail. Memoria para optar al título de Ingeniero Civil Industrial, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.
- [13] Walmart Chile (Derek Farren). 2010. Presentación Subgerente de CRM. [diapositivas]

5.4 ANEXOS

ANEXO A

Actualmente la variable Segmento “Valor” se utiliza para asignarle un valor a cada cliente, pero se basa en el promedio gastado por cada cliente en los últimos 3 meses. El criterio utilizado es el siguiente:

RANGO MONTO GASTADO	SEGMENTACIÓN
\$200.000 <= MONTO	PLATINO
\$100.000 <= MONTO < \$200.000	ORO
\$50.000 <= MONTO < \$100.000	PLATA
\$0 < MONTO < \$50.000	COBRE

Tabla 24: Condiciones de Segmentación de Valor según Monto

Fuente: Elaboración propia en base a información proporcionada por la Empresa

El siguiente gráfico representa los porcentajes de clientes que conforman cada segmento, según los meses Junio a Agosto del año 2010:

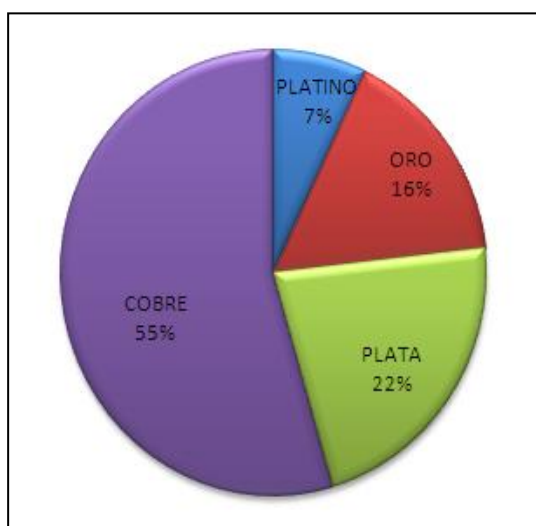


Gráfico 24: Distribución de Clientes por Segmentación de Valor

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Pero al hacer una comparación con el periodo siguiente desde Septiembre a Noviembre del mismo año, se observa una alta migración del 33% de los clientes totales, los que cambian de segmento entre dichos periodos continuos:

SEGMENTO	%CLIENTES QUE MIGRAN
PLATINO	32%
ORO	41%
PLATA	48%
COBRE	56%

Tabla 25: Porcentaje Clientes que Migran desde cierto Segmento “Valor”

Fuente: Elaboración propia en base a información proporcionada por la Empresa

ANEXO B

El siguiente gráfico representa el error MAPE por Quintil, pero en base al periodo con que se calibró el modelo:

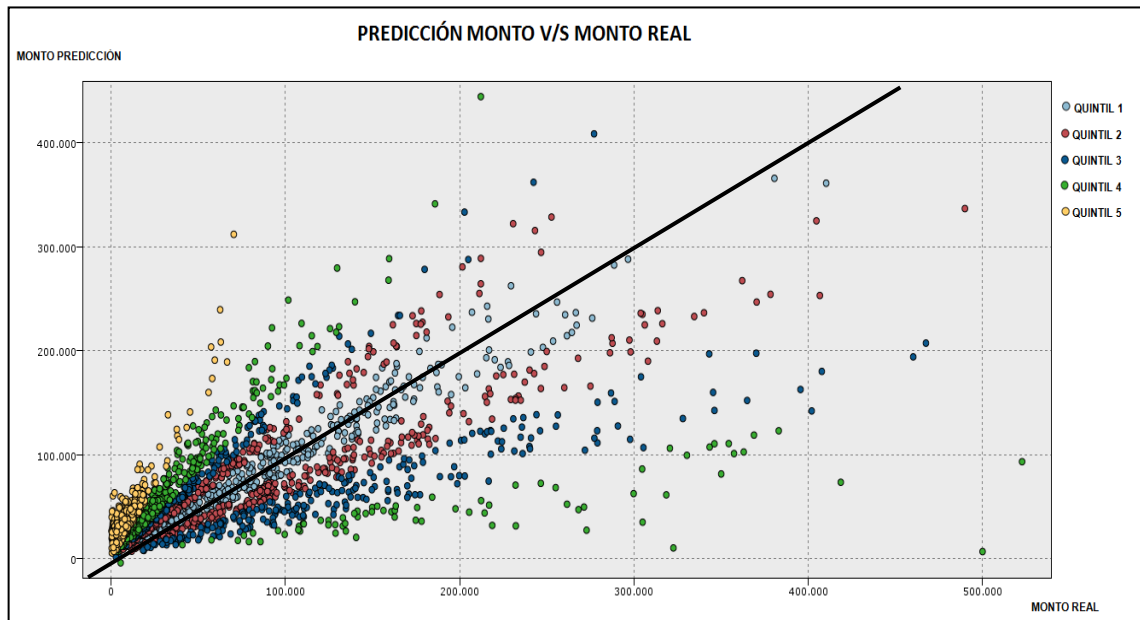


Gráfico 25: MAPE por quintil, Monto Real v/s Predicción (periodo calibración)

Fuente: Elaboración propia

Los siguientes gráficos muestran el mismo análisis del error, pero considerando el MAPE sobre el valor predicho ($MAPE_{PRED}$).

El Gráfico N°25 muestra quintiles por MAPE considerando el periodo de calibración. Mientras que el Gráfico N°26 considera el periodo de validación, que es el importante de analizar.

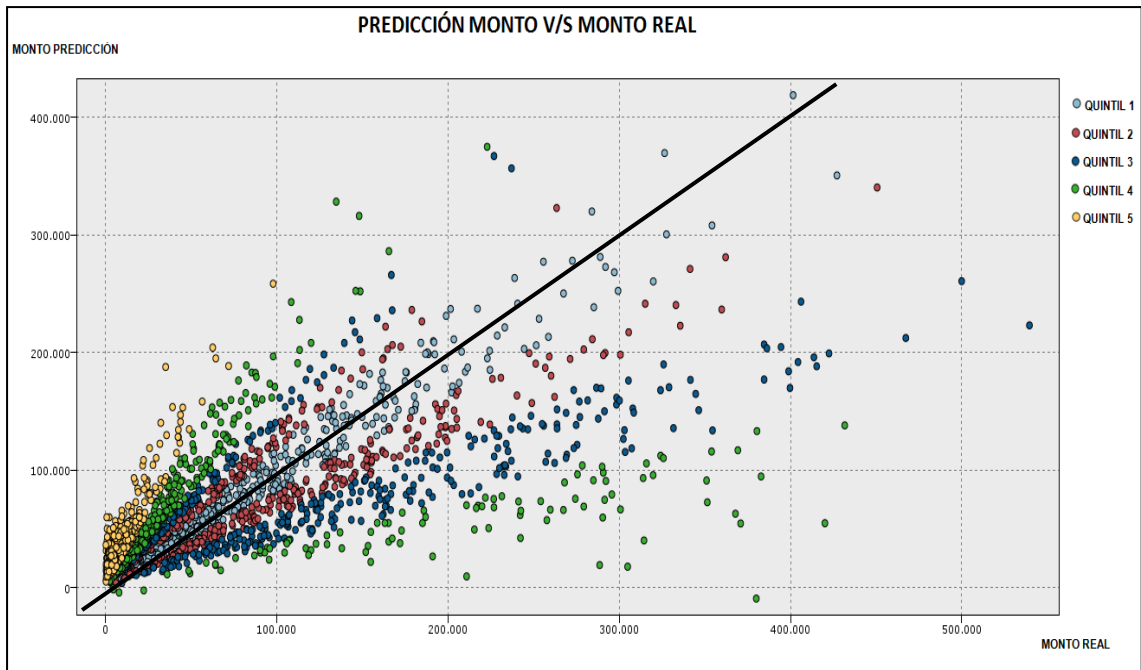


Gráfico 26: MAPE_{PRED} por quintil, Monto Real v/s Predicción (periodo calibración)

Fuente: Elaboración propia

Se observa el mismo comportamiento del análisis realizado con el MAPE sobre el valor real, ya que al observar el gráfico N°27, el mayor error se concentra en los montos altos (color amarillo).

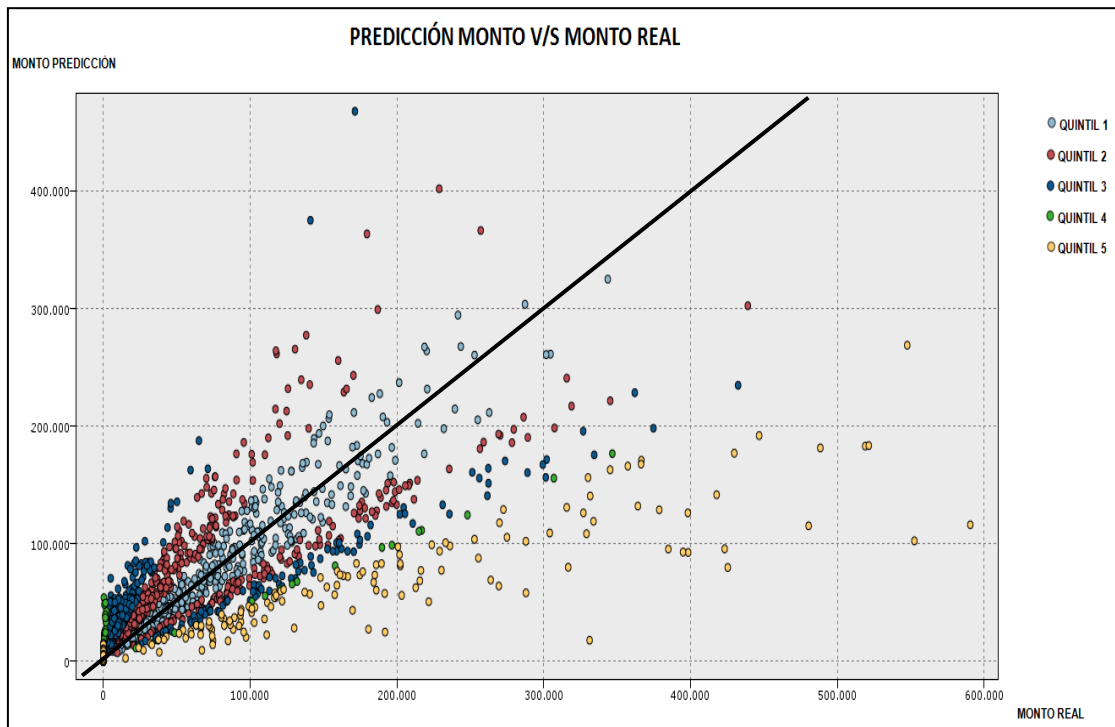


Gráfico 27: MAPE_{PRED} por quintil, Monto Real v/s Predicción (periodo validación)

Fuente: Elaboración propia

El Gráfico N°28 muestra el porcentaje de clientes que explica el error $MAPE_{real}$ cometido en la predicción, según el periodo de calibración.

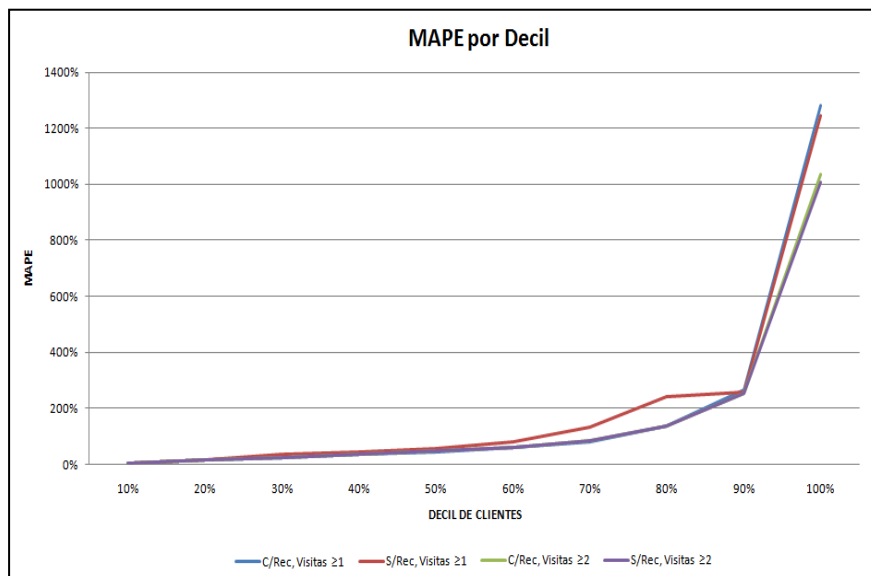


Gráfico 28: Error MAPE por decil de Clientes (periodo de calibración)

Fuente: Elaboración propia

El siguiente gráfico muestra los errores cometidos en los distintos modelos luego de trabajar la Recencia y la Frecuencia, pero esta vez considerando el MAPE sobre el valor predicho. En el gráfico N°28 se muestra el análisis en base al periodo de calibración, mientras que el Gráfico N°29 es con respecto al periodo de validación.

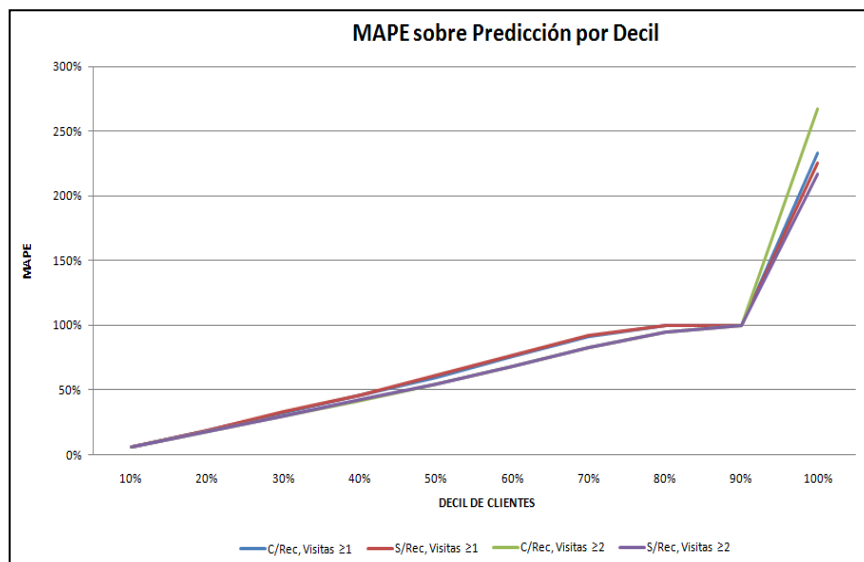


Gráfico 29: Error MAPE por decil de Clientes (Periodo de calibración)

Fuente: Elaboración propia

Sin embargo, el más interesante de analizar es el Gráfico N°30, que logra reflejar la calidad del modelo en un periodo en que los montos no se utilizaron para la calibración. Este análisis no difiere del realizado respecto al valor real, donde el mejor modelo sigue siendo con Frecuencia mayor a 1, y sin la variable Recencia dentro de las independientes (color morado).

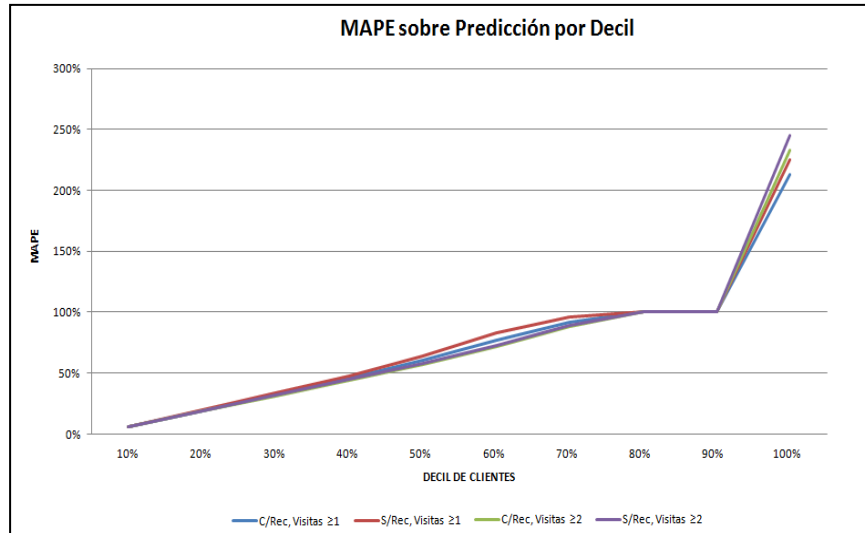


Gráfico 30: Error MAPE por decil de Clientes (Periodo de validación)

Fuente: Elaboración propia

ANEXO C

Para identificar los valores outliers, se realizaron gráficos de cajas de las variables transaccionales como Margen, Monto y Frecuencia. De esta manera, se logra identificar los datos muy escapados de la media, los que deben ser eliminados del estudio para evitar ruido. De esta manera se logra reducir el Error cometido en la estimación de los montos.

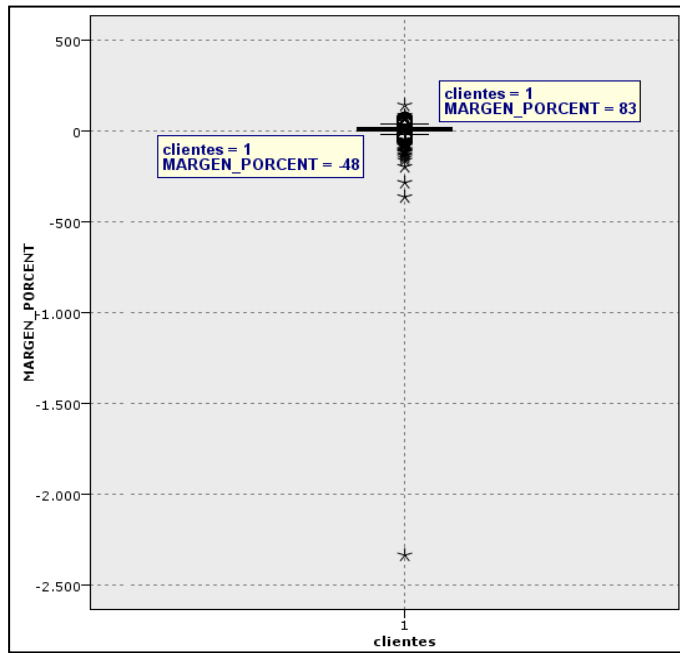


Gráfico 31: Gráfico de Caja para Porcentaje Margen

Fuente: Elaboración propia en base a información proporcionada por la Empresa

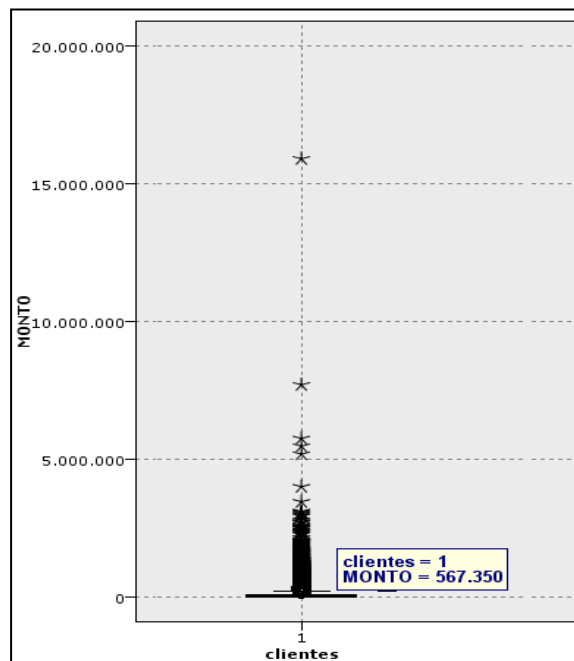


Gráfico 32: Gráfico de Caja para Monto

Fuente: Elaboración propia en base a información proporcionada por la Empresa

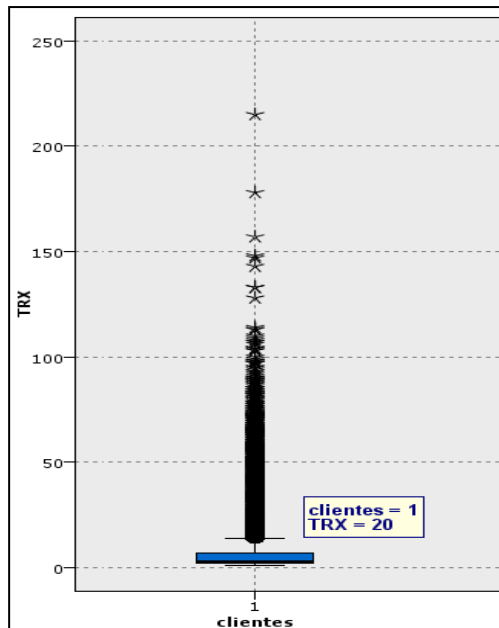


Gráfico 33: Gráfico de Caja para Frecuencia

Fuente: Elaboración propia en base a información proporcionada por la Empresa

ANEXO D

Los siguientes gráficos muestran el comportamiento de las variables Margen y Cantidad promociones, análisis realizado a nivel agregado, donde se observa un comportamiento prácticamente constante en el tiempo.

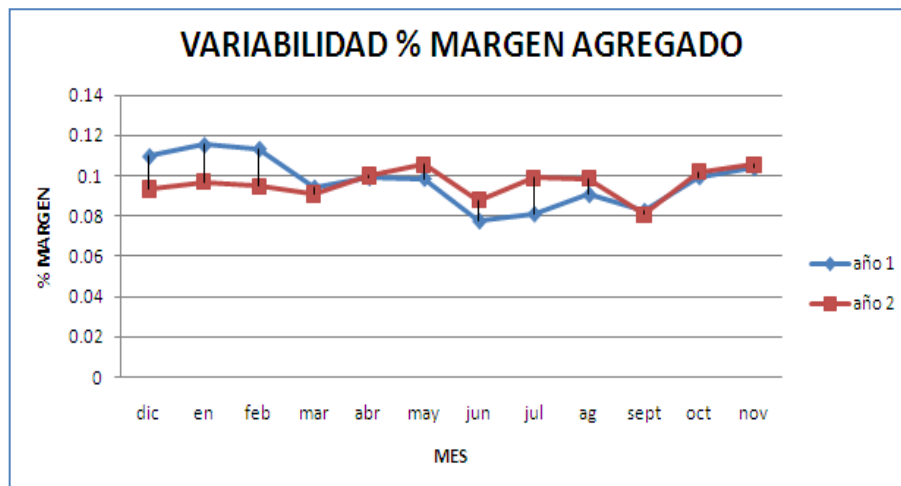


Gráfico 34: Variabilidad Margen Agregado, Año 1 – Año 2

Fuente: Elaboración propia en base a información proporcionada por la Empresa

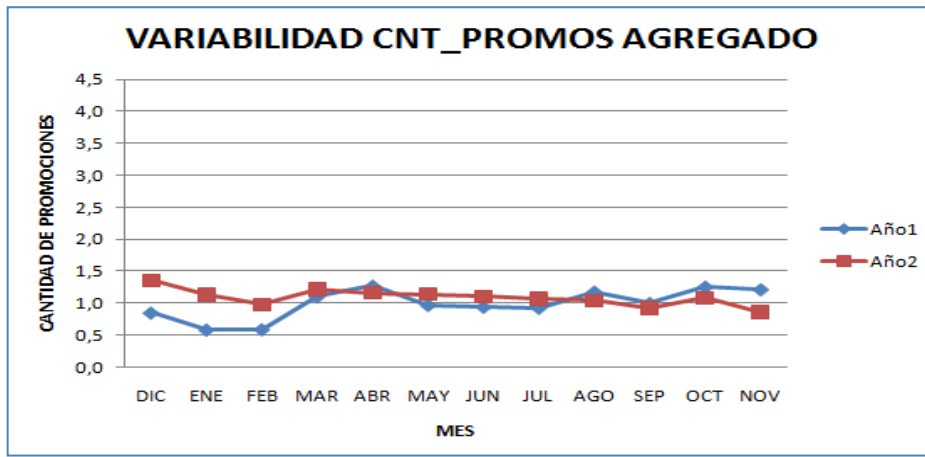


Gráfico 35: Variabilidad Promociones Agregado, Año 1 – Año 2

Fuente: Elaboración propia en base a información proporcionada por la Empresa

ANEXO E

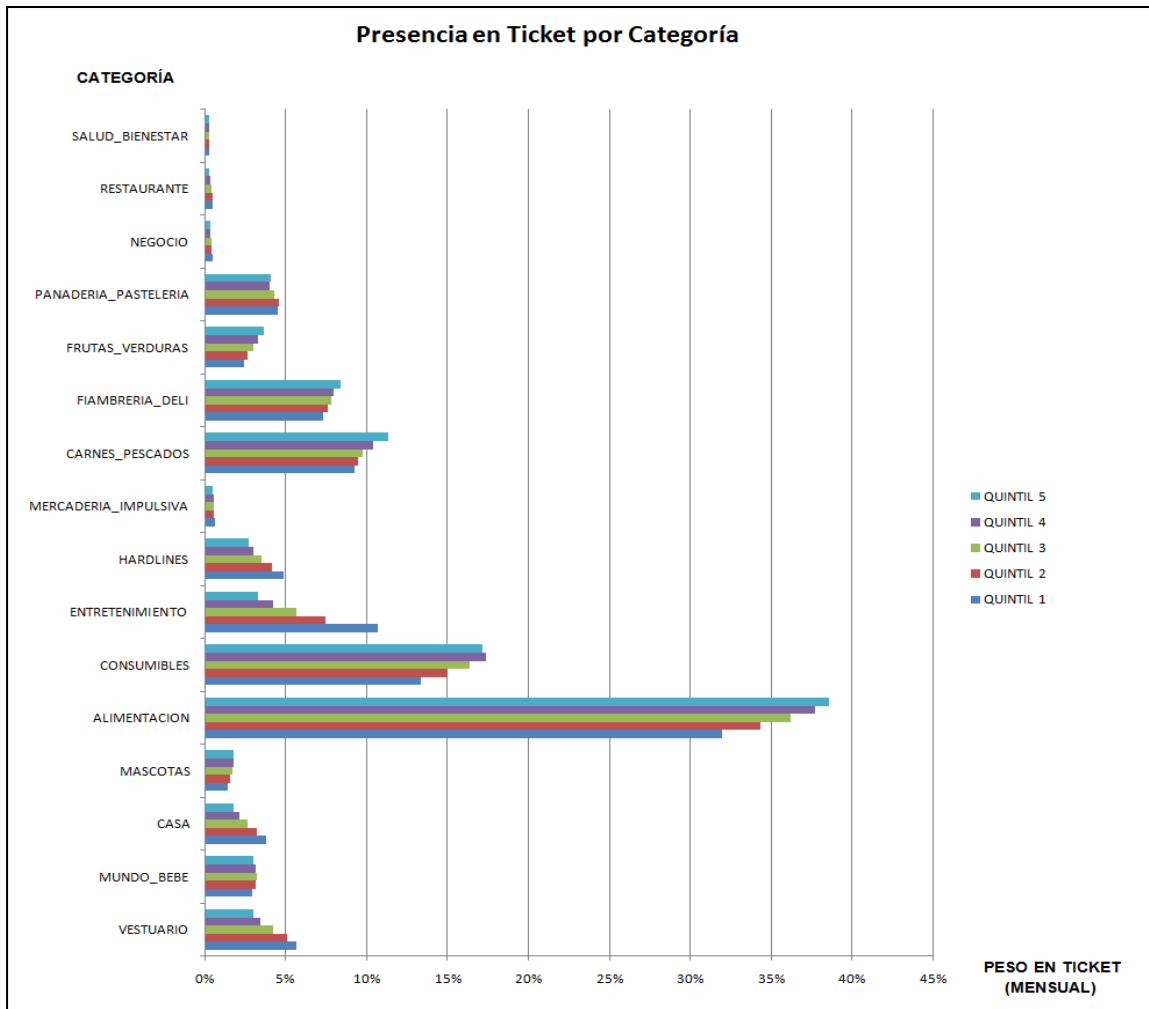


Gráfico 36: Presencia en Ticket por Categoría, a Nivel Mensual (completo)

Fuente: Elaboración propia en base a información proporcionada por la Empresa

Dado que el gráfico de Presencia en Ticket mostrado en el trabajo, fue realizado con un resumen de las categorías más importantes, el Gráfico N°36 muestra el mismo análisis pero considerando la totalidad de las categorías disponibles.

ANEXO F

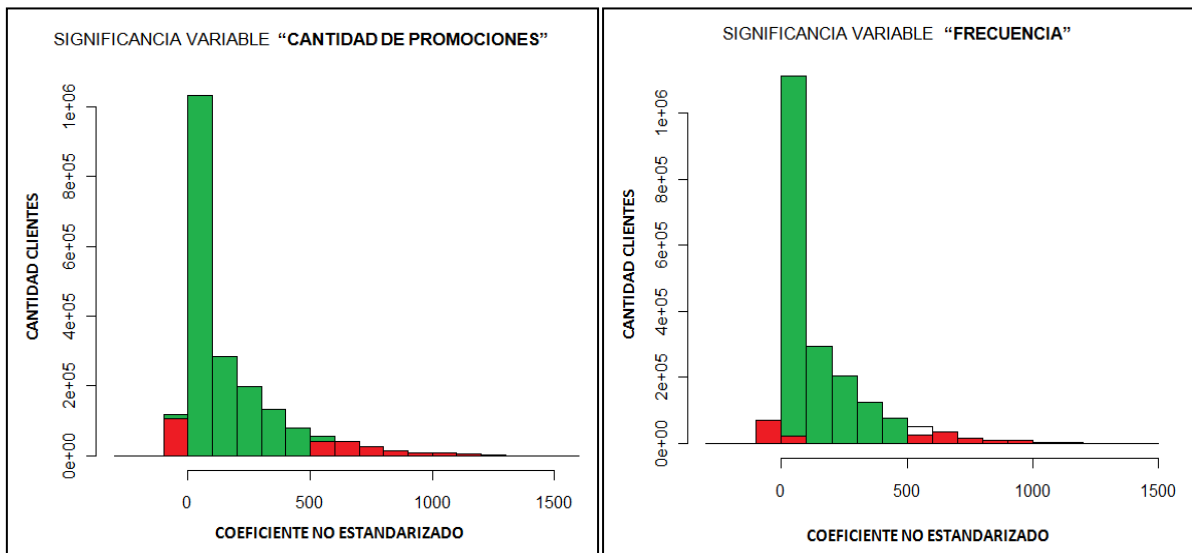


Gráfico 37: Significancia Variable Cantidad Promociones, R. J. Bayesiana
Gráfico 38: Significancia Variable Frecuencia, R. J. Bayesiana

Fuente: Elaboración propia

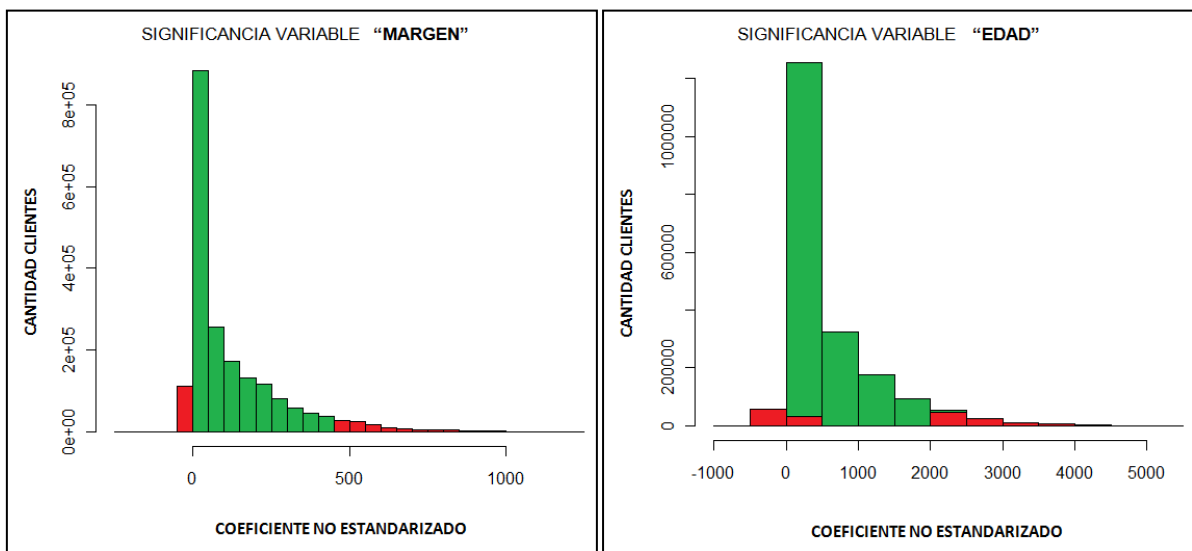


Gráfico 39: Significancia Variable Margen, R. J. Bayesiana
Gráfico 40: Significancia Variable Edad, R. J. Bayesiana

Fuente: Elaboración propia

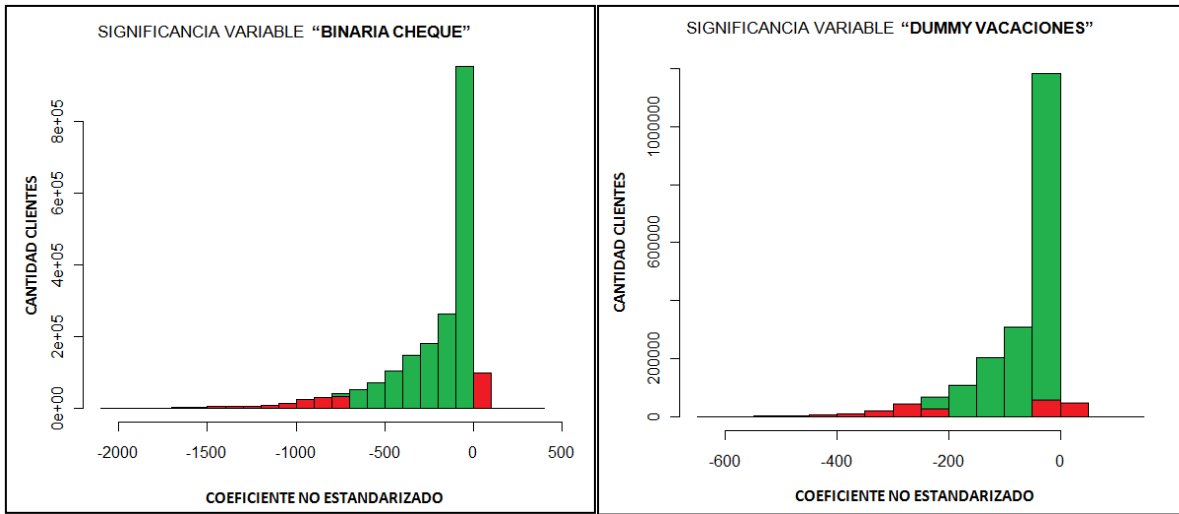


Gráfico 41: Significancia Variable Binaria Cheque, R. J. Bayesiana
Gráfico 42: Significancia Variable Dummy Vacaciones, R. J. Bayesiana

Fuente: Elaboración propia

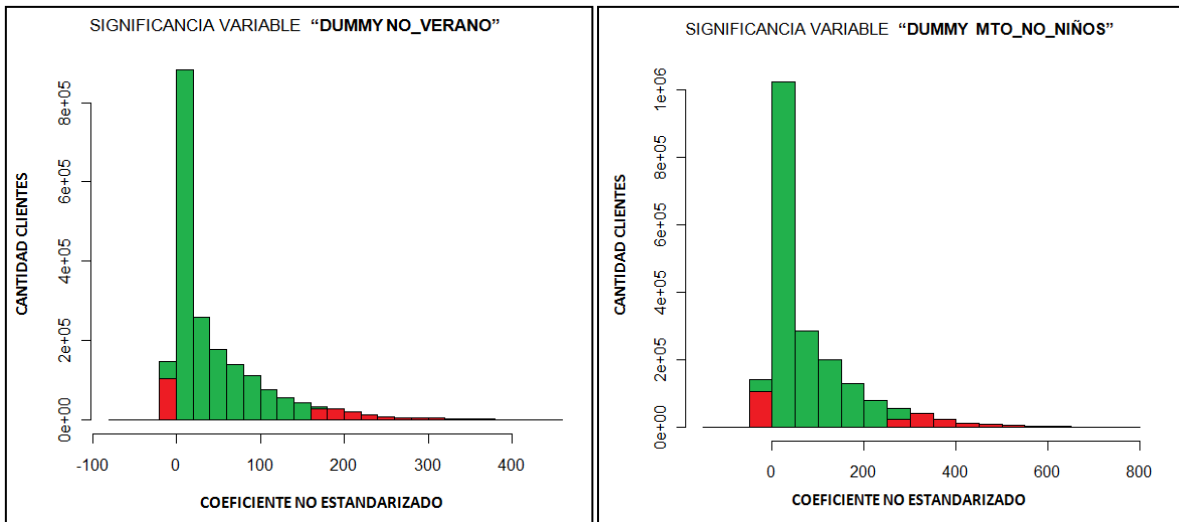


Gráfico 43: Significancia Variable Dummy No_Verano, R. J. Bayesiana
Gráfico 44: Significancia Variable Dummy Mto_No_Niños, R. J. Bayesiana

Fuente: Elaboración propia

ANEXO G

Se analizó la significancia para un nivel de confianza del 95%, para observar el comportamiento de las variables frente a un criterio más exigente. Sin embargo, el resultado fue el mismo, ya que la única variable significativa resulta ser la Frecuencia.

VARIABLE	% Clientes Significativa	% Clientes No Significativa	Conclusión
Constante	17%	83%	NO SIG
Periodo	1%	99%	NO SIG
Cantidad Promociones	22%	78%	NO SIG
Frecuencia	70%	30%	SIG
Margen	31%	69%	NO SIG
Edad	16%	84%	NO SIG
Binaria Cheque	11%	89%	NO SIG
Dummy VACACIONES	0%	100%	NO SIG
Dummy NO VERANO	15%	85%	NO SIG
Dummy VACACIONES	21%	79%	NO SIG
Ajuste R² Promedio 0,752			

Tabla 26: Resultados Regresión Lineal Individual

Fuente: Elaboración propia

ANEXO H

La tabla N°27 muestra los porcentajes en que varía el promedio de CLV para cada Quintil de clientes, al disminuir la tasa en 5% entre un escenario y otro:

	Variación 85%-80%	Variación 90%-85%	Variación 95%-90%	Variación 100%-95%	Promedio Variación
CLV Quintil 1	5.88%	5.55%	5.26%	5.00%	5.43%
CLV Quintil 2	5.88%	5.56%	5.26%	5.00%	5.43%
CLV Quintil 3	5.88%	5.56%	5.26%	5.00%	5.43%
CLV Quintil 4	5.88%	5.56%	5.26%	5.00%	5.43%
CLV Quintil 5	5.88%	5.56%	5.26%	5.00%	5.43%

Tabla 27: Sensibilidad CLV por Variación de Tasa de Retención

Fuente: Elaboración propia

La disminución promedio del CLV para los clientes de cada Quintil se observa en la siguiente tabla:

	Disminución Promedio CLV
CLV Quintil 1	-\$3.773
CLV Quintil 2	-\$9.409
CLV Quintil 3	-\$16.741
CLV Quintil 4	-\$26.865
CLV Quintil 5	-\$45.569

Tabla 28: Disminución Promedio CLV por Sensibilidad en Tasa de Retención

Fuente: Elaboración propia

Es evidente una mayor disminución del CLV en el Quintil 5 que posee a los clientes con más valor. Esto se debe a que los montos gastados por estos clientes son mayores.

ANEXO I

Los siguientes gráficos representan la distribución de clientes para cada Quintil, según el segmento al que pertenecen actualmente.

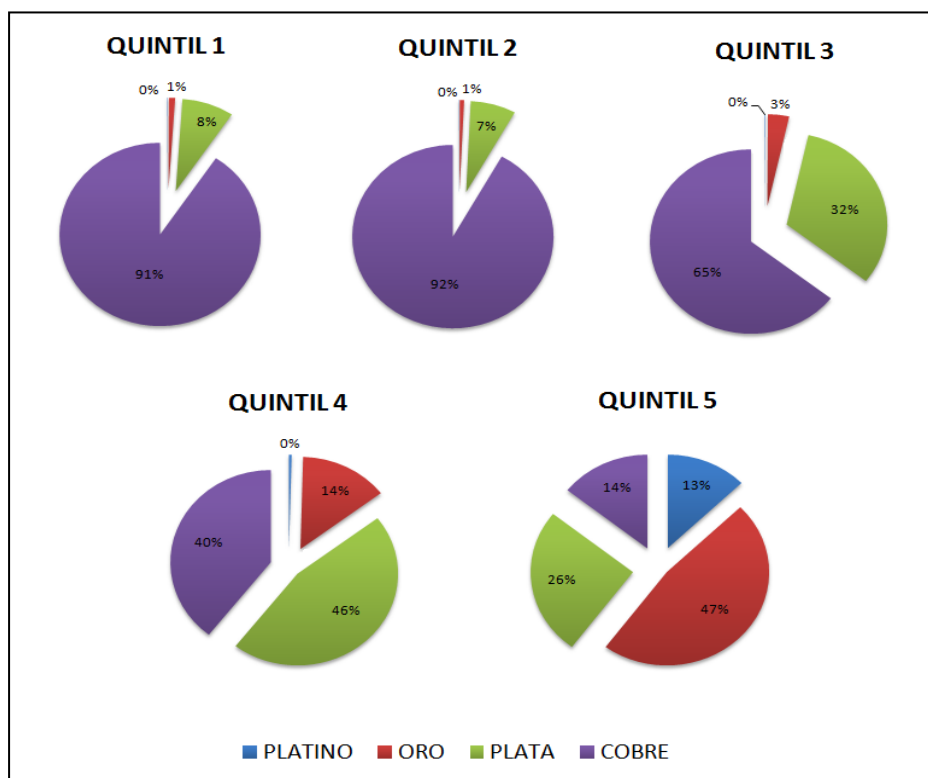


Gráfico 45: Distribución Segmento "Valor" por Quintil

Fuente: Elaboración propia