



**espol**

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**

**Facultad de Ciencias Sociales y Humanísticas**

**ESTIMANDO EL CUSTOMER LIFETIME VALUE: UNA MÉTRICA  
ALTERNATIVA PARA LA SEGMENTACIÓN DE CLIENTES  
APLICADO AL SECTOR RETAIL**

**PROYECTO INTEGRADOR**

Previo la obtención del Título de:

**Economía**

Presentado por:

Melissa Gabriela Montero Montero

Diana Jomira Suárez Villamar

**GUAYAQUIL - ECUADOR**

Año: 2019



## **AGRADECIMIENTOS**

A Dios quien nos fortaleció y orientó en este proceso. Su fidelidad y misericordia nos alcanza en todo tiempo.

*Él es el camino, la verdad y la vida.*

Y aquellos que fueron parte del camino nuestros padres, amigos y maestros que nos brindaron su apoyo.

.

## **DEDICATORIA**

Dedicado para todos aquellos que tienen una meta, los que luchan incansablemente para alcanzarla, los que no se rinden.

Este trabajo es producto de una carrera que ha requerido esfuerzo, sacrificio y perseverancia pero que al fin da su fruto.

## DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; a Melissa Montero y Diana Suárez quienes damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



---

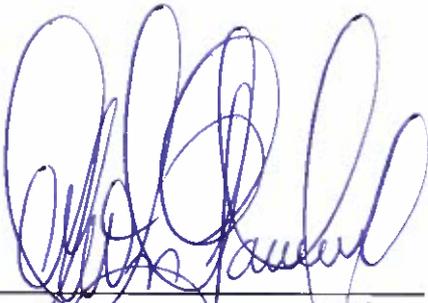
Melissa Montero  
Montero



---

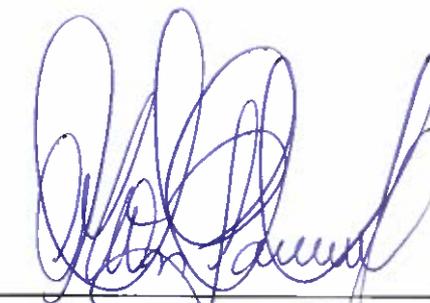
Diana Suárez  
Villamar

## EVALUADORES



---

MSc. Milton Paredes Aguirre  
PROFESOR DE LA MATERIA



---

MSc. Milton Paredes Aguirre  
PROFESOR TUTOR

## RESUMEN

El éxito que logre alcanzar una empresa depende de su capacidad para construir y mantener relaciones a largo plazo con sus compradores más valiosos. Las empresas ecuatorianas se enfrentan al reto de poder identificar, diferenciar y clasificar a sus clientes. En respuesta a esta necesidad, el presente estudio implementó un modelo probabilístico que se desprende de la metodología de Pareto/NBD con un enfoque de jerarquía bayesiana para estimar el valor del cliente (CLV), basados en una estructura de datos RFM que considera la recencia, frecuencia y monto de compra de los clientes. Después de realizar la estimación, se realizó la respectiva segmentación de clientes donde se halló que el 24% de ellos generan el 68% de las ventas, por lo tanto, las tres cuartas partes de clientes son considerados no rentables. Esta investigación propone una alternativa de segmentación más eficiente en función del valor del cliente en el contexto de una empresa de retail.

**Palabras Clave:** Customer Lifetime Value, Segmentación, Pareto NBD, MCMC.

## **ABSTRACT**

*The success of a company depends on its ability to build and maintaining long-term relationships with its most valuable buyers. Ecuadorian companies face the challenge of being able to identify, differentiate and classify their customers. In response to this need, this study implemented a probabilistic model that is derived from the Pareto/NBD methodology with a Bayesian hierarchy approach to estimating customer value (CLV), based on a RFM data structure that considers the customer's reluctance, frequency and purchase amount. After the estimate, the respective segmentation of customers was made where 24% of them were found to generate 68% of sales, therefore, three-quarters of customers are considered unprofitable. This research proposes a more efficient segmentation alternative based on customer value in the context of a retail company.*

**Keywords:** *Customer lifetime value, Segmentation, Pareto NBD, MCMC.*

# ÍNDICE

1. INTRODUCCIÓN.....	1
1.1 Descripción del problema .....	3
1.2 Justificación del problema .....	4
1.3 Objetivos .....	4
1.3.1 Objetivo General .....	4
1.3.2 Objetivos Específicos .....	5
1.4 MARCO TEÓRICO.....	5
1.4.1 Customer Relationship Management .....	5
1.4.2 Customer Lifetime Value .....	6
1.4.3 Estructura de datos RFM .....	8
1.4.3.1 Recencia.....	8
1.4.3.2 Frecuencia .....	8
1.4.3.3 Valor Monetario.....	8
1.4.4 Modelo de Pareto/NBD .....	9
CAPÍTULO 2 .....	12
2. METODOLOGÍA.....	12
2.1. Enfoque de la metodología general .....	13
2.2. Métodos de recolección de datos.....	13
2.3. Periodo de calibración y reserva .....	14
2.4. Modelando la Frecuencia de Compra .....	14
2.4.1. Introduciendo variables latentes .....	16
2.4.2. Especificaciones previas.....	17
2.5. Modelando MCMC .....	18
2.6. Validación de resultados Ajuste y Error.....	18
2.6.1. Errores de Predicción a Nivel Individual.....	18
2.6.2. Errores de Predicción a Nivel Agregado .....	20
CAPÍTULO 3 .....	23
3. RESULTADOS Y ANÁLISIS.....	23
3.1. Predicción del Customer Lifetime Value (CLV) Pareto NBD extensión Abe como una herramienta de segmentación.....	23
3.2. Cálculo del CLV .....	24
3.3. Caracterización por Segmento.....	27

3.4. Ranking de los clientes top 20 .....	31
3.5. Análisis de sensibilidad .....	33
CAPÍTULO 4 .....	36
4. CONCLUSIONES.....	36
4.1. Limitaciones .....	38
4.2. Recomendaciones.....	38
4.3. Contribuciones .....	39
BIBLIOGRAFÍA .....	1
APÉNDICES.....	5

## **ABREVIATURAS**

BG/NBD	Beta Geometric Negative Binomial Distribution
BTYD plus	Buy 'Till You Die plus
CLV	Customer Lifetime Value
CRM	Customer Relationship Management
FvA	Forecast vs Actual
GGG	Gamma-Gamma-Gamma
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCMC	Markov Chain Monte Carlo
MSLE	Mean Squared Logarithmic Error
NBD	Negative Binomial Distribution
RFM	Recency, Frequency and Monetary Value
RMSE	Root Mean Squared Error
TIC	Tecnologías de la Información y Comunicación

## ÍNDICE DE FIGURAS

Figura 2.1: Proceso de Modelado .....	12
Figura 2.2: Periodo.....	13
Figura 2.3: Línea de tiempo transaccional .....	15
Figura 3.1 Gráfico de Caja CLV general .....	25
Figura 3.2 Gráfico de Caja CLV por segmento de clientes .....	26
Figura 3.3 Gráfico de barras distribución de segmentos por frecuencia .....	28
Figura 3.4 Composición de cesta de compra por segmento de clientes .....	30

## ÍNDICE DE TABLAS

Tabla 1.1 Técnicas de modelado .....	6
Tabla 2.1: Errores de Predicción a Nivel Individual .....	19
Tabla 3.1 Métricas de error a nivel individual .....	23
Tabla 3.2 CLV general .....	25
Tabla 3.3 Descripción CLV por Segmento de clientes .....	26
Tabla 3.4 Contribución por segmento de clientes .....	27
Tabla 3.5 Distribución segmentos de clientes por frecuencia de compra.....	28
Tabla 3.6 Distribución de segmentos por sexo.....	29
Tabla 3.7 Distribución de segmentos por Estado Civil y Rango de Edad.....	29
Tabla 3.8 Distribución de segmentos por forma de pago .....	31
Tabla 3.9 Clientes Top 20 .....	32
Tabla 3.10 Analisis de sensibilidad por segmentos de clientes valiosos.....	33

# CAPÍTULO 1

## 1. INTRODUCCIÓN

En la actualidad, el mundo de los negocios se está digitalizando y cada día las empresas enfrentan un mayor nivel de competencia, haciendo más difícil la supervivencia. La economía ecuatoriana persiste en explorar nuevas estrategias para mejorar su competitividad, a paso lento, las empresas buscan mejorar sus modelos de negocios, cambiando la forma tradicional de tomar decisiones. El progreso de las tecnologías de información ha hecho posible que las compañías conozcan información de las preferencias de sus clientes. Sin embargo, estas constantemente fallan en el uso de esta información y en la definición eficiente de criterios de segmentación, debido a que no van acorde a sus objetivos.

Contar con información sobre las preferencias de los clientes es importante mantener una relación a largo plazo (Gilchrist, 2016). En el contexto del CRM, la relación de una empresa con sus compradores es el foco principal. Por ello, los gestores deben usar herramientas que le permitan identificar los clientes que tienen mayor probabilidad de permanecer con la empresa y reconocer los patrones de comportamientos futuros. De esta manera, las empresas contarán con información necesaria para poder gestionar las decisiones de marketing y valorar su base de clientes con el fin de mejorar la planificación estratégica (Rust & Chung, 2006).

El CLV es una métrica clave para identificar a los clientes con mayor valor en términos monetarios (Kumar & Pansari, 2016). Una segmentación a base de esta métrica es esencial para gestionar adecuadamente la relación con diferentes tipos de clientes (Yoo, Bai, & Singh, 2018). Empresas como Harrah's, IBM, Capital One, LL Bean, Netflix, entre otras, usan CLV de manera rutinaria como una herramienta para administrar y medir el éxito de sus negocios.

Esta investigación centra su análisis en empresas de comercio minorista. La importancia de este sector radica en su acelerado crecimiento gracias a la

evolución tecnológica y la mejora en los procesos de distribución brindando una mayor disponibilidad de múltiples productos (Kumar & Shah, 2015), además de ser un sector generador de empleo (Statista, 2017). Una característica importante de esta industria es que se desarrolla en entornos no contractuales, debido a que sus clientes son principalmente consumidores finales quienes no declaran en ningún momento su inactividad, sino que simplemente dejan de concurrir (Jasek, Vrana, Sperkova, Smutny, & Kobulsky, 2018).

Según datos del INEC, para el 2016 el comercio en el Ecuador representó el 51,16% del total de empresas, participación importante debido a que este mercado engloba todas las empresas especializadas en la comercialización masiva de productos y/o servicios. Específicamente, en el comercio del retail las empresas con mayor representatividad son supermercados. En el año 2018, las 5 cadenas de supermercados más grandes generaron una cifra superior a los 4.2 mil millones de dólares, el equivalente al 4% del PIB ecuatoriano.

El éxito de los supermercados radica en la posibilidad de adquirir múltiples productos y servicios bajo un mismo techo. La oferta de productos en este sector es muy variada desde productos básicos de consumo diario hasta productos de compra ocasional como electrodomésticos, ropa, zapatos, entre otros. A pesar de su gran alcance, estas empresas se enfrentan a un conjunto de consumidores con comportamiento bastante heterogéneo. Por lo tanto, es un gran desafío para ellos lograr identificar los grupos de clientes que visitan sus tiendas, entender sus necesidades y poder ofrecer servicios que logren satisfacerlos.

Un criterio eficiente de segmentación se convierte en una alternativa esencial para manejar una gran variedad de clientes con preferencias variadas (Hong & Kim, 2012). Hoy en día, las empresas no deben limitarse a tener un establecimiento atractivo con productos accesibles, sino ir un paso más allá generando estrategias a corto y largo plazo que contribuyan a la fidelización de los clientes a fin de garantizar su continuidad en un entorno cada vez más competitivo.

En la siguiente sección del capítulo, se realiza una breve contextualización del estudio, la descripción de los objetivos y una concreta revisión de literatura

considerando los tópicos necesarios para la comprensión de la propuesta. El capítulo 2 realiza una descripción minuciosa del procedimiento estadístico seguido en la aplicación del modelo seleccionado para la estimación. El capítulo 3 describe los principales hallazgos del cálculo del CLV y por último el capítulo 4 muestra las conclusiones e implicaciones extraídas en el estudio.

## **1.1 Descripción del problema**

La globalización y el entorno competitivo actual exigen a las empresas gestionar su portafolio de clientes como un activo estratégico, no obstante, en el Ecuador se ha incursionado poco en una gestión eficaz de relaciones con los clientes. Dentro de la economía ecuatoriana, uno de los sectores más robustos es el comercio, con el 9.5% de aporte al PIB, constituyéndose como un sector de gran diversidad y dinamismo.

En el mercado de comercio minorista ecuatoriano, las principales empresas son supermercados, estos son visitados por miles de compradores cada día. En consideración a ello, estas empresas realizan diferentes esfuerzos en estrategias de marketing para atraer a más compradores y así obtener mayores ganancias comerciales. Sin embargo, en la gran mayoría de los casos la asignación de estos recursos no es óptimo.

Este tipo de comercio se caracteriza por contar con una gran variedad de clientes cuyo comportamiento de compra es heterogéneo, haciendo que se convierta en un verdadero desafío identificar a los clientes más valiosos. Si se parte del principio de Pareto, donde se establece que el 20% de los clientes son capaces de generar el 80% de las ventas ¿Las empresas ecuatorianas pueden identificar ese 20% de clientes más valiosos? ¿Están en condiciones de estimar cuánto les van a comprar sus clientes los próximos meses?

Por las razones expuestas, la presente investigación emprende el análisis del comportamiento de compra de los clientes a través de sus actividades transaccionales, considerando su dinamismo durante un determinado período de tiempo y así estimar el valor que representa uno o un segmento de clientes para la empresa.

## **1.2 Justificación del problema**

El cliente es el activo clave de toda organización. Sin embargo, no todos los clientes son iguales; sólo una pequeña proporción generan la mayor rentabilidad para una empresa. Por ello, mantener a los clientes más rentables se convierte en un objetivo esencial para las empresas. Teniendo en cuenta esta premisa y considerando los siempre escasos recursos de una organización, impera la necesidad de mejorar las relaciones con clientes valiosos y potenciar el valor de los clientes menos valiosos. Para ello, se utiliza la estrategia de CRM, cuya finalidad es realizar una gestión eficaz en las relaciones con los clientes. No obstante, en la gran mayoría de empresas, no es conocido el término de las relaciones con los clientes, por ende, aumenta la complejidad para estudiar el comportamiento de estos.

En el desarrollo del estudio se sugiere un modelo de estimación del valor del cliente capaz de modelar el CLV de manera individual y general, de esta manera se logran estimaciones más precisas con resultados reales, superiores a formas de cálculo alternativas que basan su proyección en valores promedios. Desde una perspectiva práctica, se propone un criterio de segmentación alternativo basado en un modelo probabilístico aplicable a diferentes contextos de empresas cuyo cliente es el consumidor final del producto.

La segmentación en función al CLV es una alternativa más eficiente para modelar el comportamiento de compra, este criterio permite a las empresas optimizar sus recursos de marketing. Reconocer a los clientes como activos intangibles que aumentan el rendimiento empresarial, permitirá administrar mejor la creación de valor para la empresa y el proceso de gestión de flujo de efectivo, de esta manera mejorar su crecimiento y su sostenibilidad en el mercado. Adicionalmente, la información que otorga sobre el cliente permitirá establecer mecanismos para aumentar su nivel de satisfacción y la retención del comprador.

## **1.3 Objetivos**

### **1.3.1 Objetivo General**

Proponer una metodología de segmentación en función de la importancia del cliente para la empresa basados en el Valor Monetario Esperado de su compra.

### **1.3.2 Objetivos Específicos**

- Identificar las variables transaccionales relevantes en la estimación del valor del cliente.
- Sugerir un modelo probabilístico para estimar adecuadamente el CLV.
- Definir un ranking de clientes más valiosos y rentables para la empresa.
- Describir las características de los clientes de mayor valor para la empresa y de esta forma generar mejores estrategias de fidelización tanto focalizadas como masivas.

## **1.4 MARCO TEÓRICO**

### **1.4.1 Customer Relationship Management**

El Customer Relationship Management (CRM) es un modelo de negocios y una estrategia cuyo eje central es el cliente; se define como un proceso dedicado a establecer, mantener y mejorar las relaciones con clientes de alto valor a largo plazo (Parvatiyar & Sheth, 2001; Payne & Frow, 2005). Para lograr una gestión eficaz, el CRM debe tener en cuenta que no todos los clientes son iguales, por ende, deben ser tratados de manera distinta (Reinartz, Krafft, & Hoyer, The customer relationship management process: Its measurement and impact on performance, 2004). En virtud de ello, es necesario determinar una asignación de recursos según los diferentes niveles de clientes. Una métrica clave para la gestión efectiva de los clientes es el Customer Lifetime Value (CLV), el cual permite identificar a los clientes de alto valor (Blattberg, Kim, & Neslin, 2008; Thakur & Workman, 2016).

A medida que la competencia en el mercado se vuelve más severa, mantener a clientes valiosos se vuelve un desafío para las empresas. Blattberg, Gary y Thomas (2001) resaltan la importancia de identificar los tipos de cliente ya que estos no generan la misma rentabilidad. De acuerdo con el principio 80/20, también llamada Ley de Pareto, el 20% de los clientes aportan el 80% del beneficio a la empresa (Berry, 1995). Por lo tanto, identificar y retener ese 20% de los clientes es fundamental para una planificación estratégica de las compañías (Wu & Li, 2011).

### 1.4.2 Customer Lifetime Value

El CLV es definido como el valor presente de los beneficios futuros obtenidos de un cliente durante su vida en relación con la empresa (Nikkhahan, Badrabadi, & Tarokh, 2011). La versatilidad de esta métrica ha permitido su aplicación a diferentes niveles como: a) el cálculo y la predicción del CLV (Rust, Kumar, & Venkatesan, 2011); b) los efectos de estrategias de mercadotecnia basados en CLV (Rust, Lemon, & Zeithaml, 2004) y c) la optimización de la asignación de recursos de marketing a los clientes basados en CLV (Venkatesan & Kumar, 2004). De manera general, la estimación del CLV ayuda a las empresas a gestionar efectivamente las relaciones con sus clientes en términos de cómo iniciarlas, mantenerlas y mejorarlas (Romero, Lans, & Wierenga, 2013).

El desafío clave en la medición del CLV es el desarrollo de un modelo confiable para pronosticar el flujo futuro de ganancias que cada cliente proporcionará a la empresa. Los académicos han desarrollado una variedad de modelos que se aplican a contextos específicos. Estos generalmente se dividen en contextos de configuraciones contractuales y no contractuales (Fader & Hardie, 2009; Reinartz & V., 2000), tal como lo muestra la **Tabla 1.1**. En una configuración contractual, la empresa observa la deserción del cliente, mientras que en una configuración no contractual, la deserción no se observa (Romero, Lans, & Wierenga, 2013).

**Tabla 1.1 Técnicas de modelado**

	<b>Ajustes No Contractuales</b>	<b>Ajustes Contractuales</b>
<b>Compras Continuas</b>	<b>Modelos Probabilísticos</b>	<ul style="list-style-type: none"> <li>• Modelo Gamma Exponencial</li> <li>• Modelo Gamma-Weibull (Morrison &amp; Schmittlein, 1980)</li> </ul>
	<ul style="list-style-type: none"> <li>• Pareto/NBD (Schmittlein &amp; Colombo, 1987)</li> <li>• BG/NBD (Fader, Hardie, &amp; Lee, 2005)</li> <li>• Pareto GGG (Platzer &amp; Reutterer, 2016)</li> <li>• Modelo Gamma Gompertz GG/NBD (Glady, Baesens, &amp; Croux, 2009)</li> <li>• Bayesiano Jerárquico (Abe, 2009)</li> </ul>	
	<b>Modelos Machine Learnings</b>	
	<ul style="list-style-type: none"> <li>• CART + Regresión Logit/Lineal (Jamal &amp; Zhang, 2009)</li> </ul>	
	<b>Modelos de Markov</b>	
	<ul style="list-style-type: none"> <li>• Modelo oculto de Markov (Romero, Lans, &amp; Wierenga, 2013)</li> </ul>	

Estos modelos de estimación del CLV generalmente se enfocan en la frecuencia, recencia y el valor monetario promedio en cada compra. Varios enfoques relacionan estas variables con factores gestionables desde el punto de vista administrativo que permiten analizar el impacto de ellas en la rentabilidad esperada (Venkatesan & Kumar, 2004). Otros enfoques modelan estas variables utilizando supuestos del comportamiento de compra, sin embargo, omiten covariables explicativas que podrían mejorar las estimaciones (Fader, Hardie, & Lee, 2005). La ventaja principal de los modelos estocásticos mencionados es que se basan en un simple conjunto de datos sobre el comportamiento del cliente, esta estructura se adapta a casi todas las empresas y permite una comprensión más profunda de cómo se relaciona dicho comportamiento con el valor del cliente.

Investigaciones anteriores han demostrado que estos modelos estocásticos son adecuados para abordar una serie de problemas de gestión. Por ejemplo, la selección de clientes para la asignación de recursos y la medición de CLV durante períodos cortos y largos de tiempo. Uno de los primeros modelos en realizar este tipo de análisis es el modelo Pareto / NBD propuesto por Schmittlein, Morrison y Colombo (1987). A partir de su trabajo pionero, otros estudios han presentado extensiones mejoradas del modelo original (Schmittlein & Peterson, 1994; Colombo & Jiang, 1999; Reinartz W. &., 2003; Fader, Hardie, & Lee, 2005; Abe, 2009; Fader, Hardie, & Shang, 2010).

Sin embargo, han existido críticas en cuanto a la finalidad del CLV. Kumar y George (2007) y Homburg, Droll y Totzek (2008) han argumentado que esta estrategia puede tener efectos negativos en las relaciones de una empresa con sus clientes de bajo valor; esto debido a que los coloca en un nivel de bajo prioridad, por lo que, se puede caer en el error de dejarlos insatisfechos (Gerstner & Libai, 2006). Esta insatisfacción podría desencadenar una negativa publicidad boca a boca, dando lugar a una reducción en ventas y los beneficios a largo plazo (Hogan, Lemon, & Libai, 2003). Paralelamente, la teoría apunta un efecto positivo a largo plazo debido a la retención de clientes de alto valor.

Un hallazgo común es que los clientes de alto valor no reciben suficiente atención y que algunas empresas gastan demasiado en clientes de bajo valor (Zeithaml, 2001). La orientación de CLV debe mejorar la satisfacción de los clientes de alto valor a través de un tratamiento preferencial respecto al producto ofrecido, el precio, la atención y la comunicación con el cliente (Homburg, Droll, & Totzek, 2008). Por lo tanto, las estrategias basadas en el CLV provocan reacciones positivas en los clientes de alto valor que mejoran los ingresos y los beneficios a través de un incremento en los volúmenes de ventas.

### **1.4.3 Estructura de datos RFM**

Varios modelos probabilísticos parten de una estructura de datos con variables de recencia, frecuencia y valor monetario, RFM por sus siglas en inglés. El análisis bajo esta estructura es una técnica robusta y ampliamente utilizada en el marketing de bases de datos, su aplicación se ajusta a varios escenarios que implican un gran número de clientes como las compras en línea, ventas minoristas, etc (Christy, Umamakeswari, & Priyatharsini, 2018). Así mismo, es utilizado para clasificar a los clientes según su historial de compra en base a tres dimensiones:

#### **1.4.3.1 Recencia**

Se refiere al tiempo transcurrido desde la última transacción del cliente. Un menor valor de recencia implica que el cliente visitará la empresa en un corto período de tiempo. Por otro lado, un mayor valor supone que es menos probable que el cliente visite la empresa en breve.

#### **1.4.3.2 Frecuencia**

Es el número de transacciones que un cliente realizó durante un período específico. Cuanto mayor sea el valor de la frecuencia, los clientes presentan ser más fieles a la empresa.

#### **1.4.3.3 Valor Monetario**

Se define como la cantidad de dinero gastada por el cliente durante un período determinado. Cuanto mayor sea la cantidad de dinero que el cliente gasta, esto genera más ingresos a la empresa.

La estructura RFM tiene la capacidad de predecir el comportamiento de los clientes en el futuro, por lo tanto, existe una relación intrínseca en la estimación del CLV (Fader, Hardie, & Lee, 2005; Kumar & George, 2007). Sin embargo, la estructura como tal no provee información sobre la posibilidad de que los clientes realicen futuras compras, tampoco proporciona alguna estimación económica que refleje el valor de los clientes. En virtud de ello, es necesario un modelo que complemente el poder predictivo de las variables RFM.

#### **1.4.4 Modelo de Pareto/NBD**

Uno de los primeros modelos en abordar la probabilidad de que un cliente seguirá siendo activo en el futuro y sea capaz de describir su comportamiento de compra es el modelo de Pareto/NBD. Este modelo fue desarrollado por Schmittlein, Morrison y Colombo (1987) el cual describe el flujo de transacciones de los clientes en entornos no contractuales.

Basado en este modelo inicial, Fader, Hardie y Lee (2005) construyeron el modelo BG / NBD-GG aplicable en un entorno amplio de industrias. Sin embargo, este modelo se basa en el supuesto de deserción permanente del cliente y no toma en consideración la existencia de heterogeneidad en el comportamiento de los compradores. Estas limitaciones pueden conducir a predicciones inexactas debido a que estas suposiciones no se cumplen en la mayoría de los entornos (Romero, Lans, & Wierenga, 2013).

Por otro lado, estos modelos generalmente son estimados bajo máxima verosimilitud (MLE) presentando soluciones únicas, dichas soluciones singulares generan restricciones al momento de intentar relajar algunos supuestos subyacentes de comportamiento de compra de los clientes. Para resolver estas desventajas un método de estimación alternativo aplicado en modelos probabilísticos son las simulaciones de Markov-Chain-Monte-Carlo (MCMC). Esta metodología realiza cálculos más complejos lo que deriva a costos altos de implementación. Sin embargo, este método tiene ventajas superiores dado que permite realizar estimaciones de parámetros a nivel individual y general, además genera mayor flexibilidad en los supuestos restrictivos de manera que proporciona estimaciones más realistas.

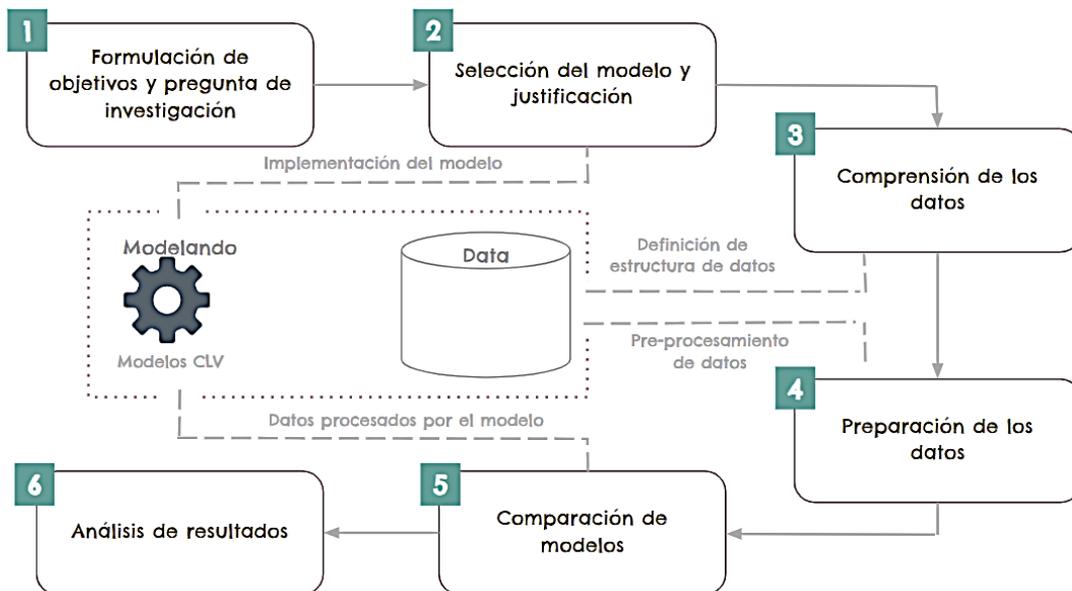
Bajo esta metodología de estimación se han desarrollado varias extensiones del modelo original propuesto por Schmittlein y Colombo (1987). Ma y Liu (2007) han desarrollado una variante jerárquica de Bayes, esta propuesta mejora los resultados de estimación, pero aún no aprovecha al máximo la flexibilidad obtenida, debido a que mantiene los supuestos restrictivos del modelo original. Mientras que, la propuesta de Abe (2009) de Pareto / NBD consigue aprovechar las ventajas de estimación relajando el supuesto de independencia en del proceso de compra y la tasa de abandono, además considera la heterogeneidad existente en la cohorte de clientes. Por último, Platzer y Reutterer (2016) proponen Pareto / GGG, el cual permite un grado variable de regularidad en los tiempos de transacción, sin embargo, este modelo mantiene el supuesto de independencia entre la frecuencia de compra y la tasa abandono, lo que no es realista en la mayoría de los casos.



# CAPÍTULO 2

## 2. METODOLOGÍA

En este estudio se aplicó la extensión del modelo de Pareto/NBD propuesto por Abe (2009) que utiliza un enfoque jerárquico bayesiano para predecir el comportamiento de compra de los clientes en contextos no contractuales y bajo un enfoque basado en la supervivencia la probabilidad de que un cliente realice compras en el futuro. El método de estimación utilizado es a través de simulaciones de Markov-Chain-Monte-Carlo (MCMC), con un algoritmo que realiza la estimación de los parámetros individuales y generales del modelo probabilístico, en base a registros de transacciones históricas del cliente. Finalmente, una evaluación de métricas de error permite explorar la idoneidad del modelo. Se usó el software estadístico R para el cálculo del CLV.



*Figura 2.1: Proceso de Modelado*

en un escenario donde la compra puede ocurrir en cualquier momento y el término de las relaciones transaccionales es desconocido.

## 2.1. Enfoque de la metodología general

Se realizó un análisis a nivel individual con datos transaccionales de múltiples clientes. Tomando como referencia artículos anteriores en el cálculo de CLV (Schmittlein & Peterson, 1994; Venkatesan et al., 2004; Fader et al., 2005), se aplicó un proceso de dos fases para la estimación que consisten en: 1) pronosticar el número de transacciones futuras de cada individuo y su probabilidad de mantener una relación activa con la empresa, 2) Calcular el beneficio promedio individual por transacción. De esta manera, con el producto de la cantidad futura de transacciones, el beneficio por transacción, la probabilidad de que se realice compras en el futuro y la tasa de descuento se obtuvo una aproximación del CLV para cada cliente.

Para realizar las estimaciones se dividió el conjunto de datos en un período de calibración y un período de reserva para poder validar las predicciones del modelo (Fader P. , Hardie, Lee, & K.L., 2005). Los datos a nivel de cliente se transformaron en una estructura RFM. El conjunto de datos preparado del período de calibración se utilizó para ajustar los parámetros de Pareto / NBD Abe. Las predicciones a nivel individual sobre el número futuro de compras, la probabilidad de estar activo y el valor monetario se utilizan para calcular el CLV y posteriormente se descuentan mediante una tasa. Finalmente, se estiman dos valores de CLV: el CLV predicho a través del modelo de Pareto / NBD Abe y el CLV real en el período de reserva.

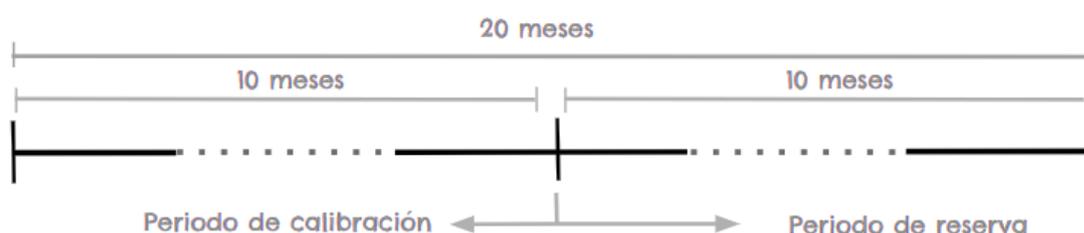


Figura 2.2: Periodo

## 2.2. Métodos de recolección de datos

La estructura de los clientes que pertenecen a las principales cadenas de supermercados como La Favorita, El Rosado y Tiendas Industriales y Asociados TIA se caracterizan por tener compradores con un comportamiento heterogéneo

tanto en frecuencia, recencia y monto de compra. Para este estudio, se seleccionó una sucursal que pertenece a una importante cadena de supermercados a nivel nacional, esta tienda opera en la ciudad de Guayaquil y aceptó participar en esta investigación bajo condición de absoluta confidencialidad

Se analizaron clientes con tarjeta de fidelización de la empresa y que hayan realizado transacciones por lo menos una vez durante el mes de septiembre del 2017. En cuanto a los productos, solo se incluyeron los ingresos generados por la venta de 10 secciones de la tienda debido a que éstas guardan la naturaleza de un supermercado, los cuales son: artículos de limpieza, bebidas, carnes, comestibles, congelados, embutidos, frutas y verduras, golosinas, lácteos y perfumería. Adicionalmente, como restricción se excluyeron productos cuyo PVP fuera mayor a 100 USD, para evitar la inclusión de datos atípicos lo cual generaría una sobreestimación en los montos de compra.

### **2.3. Periodo de calibración y reserva**

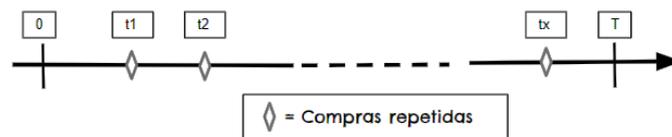
Para este estudio, el conjunto de datos suministrado por la cadena de supermercado comprende un total de 20 meses, desde el 1 de septiembre del 2017 hasta el 30 de abril del 2019. Tomando como referencia los estudios de Gladys, Baesens, & Croux (2009) se divide el conjunto de datos en dos partes iguales, los primeros 10 meses corresponden al período de calibración y los subsiguientes meses al período de reserva. Se evaluó un período poco extenso, considerando la necesidad gerencial es proporcionar información que permita una asignación de marketing basada en el valor de los clientes en un futuro próximo, dado que los flujos futuros de períodos más amplios se descuentan considerablemente.

### **2.4. Modelando la Frecuencia de Compra**

El modelo de Pareto/ NBD Abe analiza la relación entre la frecuencia y recencia, de esta manera provee la probabilidad de que un cliente individual esté activo en un momento determinado en el tiempo. En este modelo, el comportamiento del consumidor se caracteriza por: a) distribución de compra de Poisson con el parámetro  $\lambda$  que describe la tasa de compra, y además b) el tiempo de vida modelado como una función exponencial con  $\mu$  como parámetro para la tasa de

deserción (Abe, 2009). El presente modelo captura la heterogeneidad de los clientes a través de la estimación de parámetros específicos individuales con un enfoque jerárquico bayesiano. En particular, la propuesta mantiene los supuestos del comportamiento propuestos por Schmittlein, Morrison y Colombo (1987), pero tiene dos diferencias características: (1) complementa la parte analítica de la distribución global, considerando una distribución individual que en cuenta la heterogeneidad de los clientes con un método de simulación (Abe, 2009).

Los datos de frecuencia, recencia y tiempo están representados con las siguientes variables  $(x, t_x, T)$ . La vida comienza a el tiempo 0 (cuando se produce la primera transacción) y las transacciones de los clientes se monitorean hasta el tiempo  $T$ .  $x$  es el número de transacciones repetidas observadas en el período de tiempo  $(0, T]$ , con la última compra ( $x$ th repeticiones) ocurriendo en  $t_x$ . Por lo tanto, la recencia se define como  $T - t_x$ .



**Figura 2.3: Línea de tiempo transaccional**

Usando estas notaciones matemáticas, los supuestos de este modelo se pueden expresar de la siguiente manera:

**Proceso de compra de Poisson:**

$$P[x|\lambda] = \begin{cases} \frac{(\lambda T)^x}{x!} e^{-\lambda T} & \text{if } \tau > T \\ \frac{(\lambda \tau)^x}{x!} e^{-\lambda \tau} & \text{if } \tau > T \end{cases} \quad x = 0, 1, 2, \dots, \quad (1)$$

Donde  $\tau$  es la vida del cliente, variable no observada.

**Tiempo de vida con una distribución exponencial:**

$$f(\tau) = \mu e^{-\mu\tau} \quad \tau \geq 0, \quad (2)$$

## Incorporando Covarianzas

$$\begin{bmatrix} \log(\lambda) \\ \log(\mu) \end{bmatrix} \sim MVN \left( \theta_0 = \begin{bmatrix} \theta_\lambda \\ \theta_\mu \end{bmatrix}, \Gamma_0 = \begin{bmatrix} \sigma_\lambda^2 & \sigma_{\mu\lambda} \\ \sigma_{\mu\lambda} & \sigma_\mu^2 \end{bmatrix} \right) \quad (3)$$

donde MVN denota una distribución normal multivariada.

### 2.4.1. Introduciendo variables latentes

Para estimar  $\lambda$  y  $\mu$  de forma individual dadas las funciones a priori, se incorporan las variables latentes y así obtener las funciones posteriori. Es decir, que no se puede aprovechar la simple actualización bayesiana si se desconocen las variables no observadas (si el cliente  $i$  está activo en  $T_i$  y, si no, el tiempo de deserción  $y_i < T_i$ ). Esta es la verdadera razón del complejo proceso de estimación asociado con el modelo Pareto / NBD. Por ello, se incorporan variables no observables como variables latentes en el presente modelo. Para reducir la notación, el subíndice  $i$  se elimina.  $z$  se define como 1 si un cliente está activo en el momento  $T$  y 0 de lo contrario. Otra variable latente es el tiempo de deserción  $y$  que aparece cuando  $z = 0$  (el cliente se encuentra inactivo). Si se conocen  $z$  y  $y$ , implica que la función de probabilidad para los datos de frecuencia y recencia  $(x, t_x, T)$  se convierte en una expresión simple para  $x > 0$ .

**Caso  $z = 0$  (el cliente está activo en el tiempo  $T$ ).**  $P(xth$  compra en  $t_x$  y está activo hasta  $T$  y no compra entre  $[t_x, T])$

$$\begin{aligned} &= \frac{\lambda^x t_x^{x-1}}{\Gamma(x)} e^{-\lambda t_x} \times e^{-\lambda(T-t_x)} \\ &= \frac{\lambda^x t_x^{x-1}}{\Gamma(x)} e^{-(\lambda+\mu)T} \end{aligned}$$

**Caso  $z = 1$  (el cliente está inactivo en el tiempo  $T$ ).**  $P(xth$  compra en  $t_x$  y no compra entre  $[t_x, T]$  y abandona en  $y \leq T$ )

$$= \frac{\lambda^x t_x^{x-1}}{\Gamma(x)} e^{-\lambda t_x} \times e^{-\lambda(y-t_x)} \times \mu e^{-\mu y}$$

$$= \frac{\lambda^x t_x^{x-1}}{\Upsilon(x)} \mu e^{-(\lambda+\mu)y} (t_x \leq y \leq T)$$

Combinando estos dos casos, se obtuvo una notación más compacta para la función de probabilidad:

$$L(x, t_x, T | \lambda, \mu, z, y) = \frac{\lambda^x t_x^{x-1}}{\Upsilon(x)} \mu^{1-z} e^{-(\lambda+\mu)\{zT+(1-z)y\}} \quad (4)$$

Para  $x = 0$ , no hay compras repetidas y  $t_x = 0$ . Por lo tanto,  $\Gamma(x = 0)$  y  $t_x^{x-1}$  no están definidas. La función de probabilidad apropiadas es  $e^{-(\lambda+\mu)T}$  para  $z = 1$ , y  $\mu e^{-(\lambda+\mu)y}$  para  $z = 0$ . Por lo tanto, la ecuación (4) se convierte en

$$L(x = 0, t_x, T | \lambda, \mu, z, y) = \mu^{1-z} e^{-(\lambda+\mu)\{zT+(1-z)y\}} \quad (5)$$

Sin embargo, como no se observa  $z$  ni  $y$ , son considerados como datos faltantes, por ello se aplica una técnica de aumento de datos (Tanner & Wong, 1987). Para simular  $z$  en nuestro procedimiento de estimación de MCMC, se usa la siguiente expresión para la probabilidad de que un cliente esté activo en  $T$ , o de manera equivalente  $z = 1$ .

$$\begin{aligned} & P[\tau > T | \lambda, \mu, T, t_x] \\ &= P[z = 1 | \lambda, \mu, T, t_x] \\ &= \frac{1}{1 + (\mu/(\lambda + \mu))[e^{(\lambda+\mu)(T-t_x)} - 1]} \end{aligned} \quad (6)$$

#### 2.4.2. Especificaciones previas

Para estimar los parámetros específicos por cliente denotamos  $\theta_i = [\log \lambda_i, \log \mu_i]$ , el cual se distribuye normalmente con media  $\beta' d_i$  y la matriz de varianza-covarianza  $\Gamma_0$ , donde el subíndice  $i$  indica los clientes individuales. El objetivo será estimar los parámetros  $\{\theta_i, y_i, z_i, \forall i; \beta, \Gamma_0\}$  en base a las variables observadas de frecuencia y recencia  $\{x_i, t_{x(i)}, T_i; \forall i\}$ .

Los parámetros  $\lambda_i$  y  $\mu_i$  se eligen como lognormal con la función prior especificada en la ecuación (3). A su vez los hiperparametros  $\beta$  y  $\Gamma_0$  se estiman respectivamente de forma bayesiana con un prior multivariado normal y un prior Wishart inverso.

$$\beta \sim MVN(\beta_0, \Sigma_0), \quad \Gamma_0 \sim IW(v_{00}, \Gamma_{00})$$

## 2.5. Modelando MCMC

Dadas las especificaciones previas se puede estimar los parámetros  $\{\theta_i, y_i, z_i, \forall i; \beta, \Gamma_0\}$  utilizando el método MCMC. Para estimar la densidad conjunta, se genera secuencialmente cada parámetro, desde su distribución condicional hasta lograr la convergencia. El procedimiento se describe a continuación:

1. Establecer el valor inicial para  $\theta_i^{(0)} \forall i$ .
2. Para cada cliente  $i$ ,
  - 2.1. Generar  $\{z_i | \theta_i\}$  de acuerdo con la ecuación (6).
  - 2.2. Si  $z_i = 0$ , se genera  $\{y_i | z_i, \theta_i\}$  usando una distribución exponencial truncada.
  - 2.3. Obtener  $\{\theta_i | z_i, y_i\}$  utilizando la ecuación (4).
3. Generar  $\{\beta, \Gamma_0 | \theta_i, \forall i\}$  usando una actualización de regresión normal multivariada estándar.
4. Iterar los pasos [2] y [3] hasta que se logre la convergencia.

## 2.6. Validación de resultados Ajuste y Error

### 2.6.1. Errores de Predicción a Nivel Individual

Para evaluar el desempeño de los modelos a nivel individual se aplicaron varias métricas de error que sugiere la literatura: MAE, MSLE, Media RMSE, Mediana RMSE y las medidas de sensibilidad fueron calculadas y analizadas en este estudio. Las métricas se definen a continuación.

**Tabla 2.1: Errores de Predicción a Nivel Individual**

Nombre	Evaluación de los autores	Ejemplos de estudios
<b>Error Absoluto Medio (MAE)</b>	Es una medida sencilla utilizada por múltiples autores	(Jasek, Vrana, Sperkova, Smutny, & M., 201; Platzer & Reutterer, 2016); Schwartz, Bradlow and Fader, 2014; Platzer, 2008)
<b>Error Cuadrático Medio (RMSE Medio)</b>	Es sensible a los datos atípicos	(Platzer, 2008; Wübben & Wangenheim, 2008)
<b>Error Cuadrático de la Mediana (Mediana RMSE)</b>	Presenta menor sensibilidad a los datos atípicos	(Wübben & Wangenheim, 2008)
<b>Sensibilidad</b>	El porcentaje de clientes que están correctamente clasificados para pertenecer al 10% superior	(Jasek et al., 2018; Malthouse and Blattberg, 2005; Wübben and Wangenheim, 2008)
<b>Correlación</b>	Proposed by Fader, Hardie, & Lee, (2005) y cuestionado por Platzer (2008).	Fader, Hardie, & Lee, (2005); Platzer (2008).

El Error Absoluto Medio (MAE) es una medida sencilla que describe el tamaño del error promedio (Jasek et al., 2018). Se utiliza para medir la diferencia entre dos variables continuas, en este caso, el  $CLV_i$  real vs el  $CLV_i$  estimado. Se representa a través de la siguiente ecuación:

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - F_i|$$

Donde,

$A_i$ : es el CLV real del cliente durante el período de reserva.

$F_i$ : es el CLV estimado del cliente durante el período de reserva.

$n$ : es el número total de clientes.

La media del Error Cuadrático (RMSE medio) indica la desviación estándar de los errores de predicción, sin embargo, esta medida puede verse afectada por valores atípicos (Hyndman & Koehler, 2006).

$$Mean\ RMSE = \sqrt{mean(A_i - F_i)^2}$$

El error cuadrático de la mediana (RMSE median) es similar al RMSE medio, pero difiere en que presenta menor sensibilidad en los datos atípicos y utiliza el error cuadrado de la mediana (Hyndman & Koehler, 2006):

$$MedianRMSE = \sqrt{\text{median}(A_i - F_i)^2}$$

Los estados de sensibilidad indican el porcentaje de clientes que están correctamente clasificados para pertenecer al 10% superior, es decir cuántos del 10% de los clientes más rentables se asignaron correctamente a la clase del 10%, por lo tanto, resultados cercanos a 1 o 100% son los más deseables (Jasek et al., 2018):

$$Sensibilidad = \frac{TP}{TP + TN}$$

Donde *TP* está representado por el número de clientes que se clasificaron correctamente y el *TN* representa el número de clientes que se clasificó incorrectamente.

Con la finalidad de evaluar la idoneidad del modelo Pareto / NBD Abe como una herramienta de gestión, no se ha estudiado la correlación entre el *CLV<sub>i</sub> estimado* y el *CLV<sub>i</sub> real*, debido a que la correlación tan solo describe la relación lineal entre los dos.

### **2.6.2. Errores de Predicción a Nivel Agregado**

Con el objetivo de evaluar el desempeño del modelo probabilístico aplicado, se compararon los resultados de las métricas de error con estudios similares realizados por varios autores en el pasado. A nivel agregado, la literatura sugiere dos métricas de error para comparar los CLV: Forecast vs Actual (FvA) y el Error porcentual absoluto medio (MAPE). El MAPE es una de las medidas de error de mayor uso, sin embargo, presenta limitaciones cuando la data contiene valores extremos o ceros. Por esta razón, no es una medida aplicable en este estudio debido a la existencia de clientes desertores que generan CLV de cero. En virtud de ello, la evaluación del FvA será suficiente para comprender la precisión del modelo a nivel agregado (Jasek et al., 2018). Esta métrica FvA describe si el pronóstico sobrestima o subestima el CLV y se define como:

$$FvA = \frac{F_t}{A_t} * 100$$

Donde:

$F_t$ : Es la suma del CLV forecast para todos los clientes en el período de reserva

$A_t$ : Es la suma del CLV actual para todos los clientes en el período de reserva

Finalmente, se evaluó cuántos del 10% de los clientes más rentables están identificados correctamente por el modelo. Para ello, se realizó un análisis de sensibilidad como lo sugieren (Fader P. , Hardie, Lee, & K.L., 2005). No se utilizó la correlación como una medida de error, debido a que solo indica en qué medida las variables cambian al unísono como lo cuestionó Platzer (2008).



# CAPÍTULO 3

## 3. RESULTADOS Y ANÁLISIS

### 3.1. Predicción del Customer Lifetime Value (CLV) Pareto NBD extensión Abe como una herramienta de segmentación.

Para evaluar la competencia del modelo de Pareto/NBD Abe (2009), se analizan las métricas del error tanto a nivel individual y como a nivel agregado.

A nivel individual se estimó el MAE, la media y la mediana del RMSE, así también se analizaron los parámetros de sensibilidad. Con el objetivo de evaluar el desempeño del modelo probabilístico aplicado, se compararon los resultados de las métricas de error con estudios similares realizados por varios autores en el pasado. El MAE del modelo Abe Pareto/NBD es de 40% del CLV promedio, un valor más bajo comparando con el estudio de Bernant, Koning y Fok (2018) con un MAE de 85,6% en una empresa de retail. La media del RMSE es aproximadamente el 72% del promedio del CLV y la mediana RMSE un 19%, estos resultados son comparables con el estudio de Do Ruibin y Vintilescu Borglöv (2018) aplicado en un contexto similar que presentan un 120% y 34% para la media y mediana RMSE respectivamente, mostrando la superioridad del modelo de Pareto/NBD (Abe, 2009) frente al modelo original Pareto/NBD (Schmittlein & Colombo, 1987).

Respecto a la sensibilidad, el modelo también presenta una mejor clasificación respecto al original. En su estudio, Wübben y Wangheim (2008) encontraron que el modelo de Pareto/NBD clasificaba correctamente cerca del 62% del 10% de los clientes reales y al nivel del 20% clasificaba adecuadamente un 67%. Mientras que, el modelo de Pareto/NBD (Abe) presenta una mejor clasificación tanto al 10% como al 20%, tal como lo muestra la tabla 3.1.

**Tabla 3.1 Métricas de error a nivel individual**

		Pareto/ NB (Abe)
Nivel Individual	MAE	39,83
	Media RMSE	72,17

	<b>Mediana</b>	
	<b>RMSE</b>	18,7
	<b>Sensibilidad</b>	
	<b>10%</b>	70
	<b>Sensibilidad</b>	
	<b>20%</b>	78
<b>Nivel</b>	<b>FvA</b>	108,5
<b>Agregado</b>		

A nivel agregado, se realiza la evaluación en base al Forecast vs Actual (FvA), que es de 108,5% para el período de calibración, lo cual indica un error pequeño del modelo Pareto/NBD (Abe) en comparación con el Pareto/NBD de 128% (Do Ruibin & Vintilescu Borglöv, 2018), ya que sobreestima solo un 8,5% el CLV promedio.

### 3.2. Cálculo del CLV

Una vez analizada la capacidad predictiva del modelo, se predice el monto en un período de un año iniciando en el mes de mayo del 2019. En base a la frecuencia de compra se realizó un muestreo aleatorio estratificado, donde se tomó una muestra representativa de 500 clientes para realizar el análisis. La tasa de descuento utilizada para estimar el valor actual de los flujos futuros de los clientes fue de 9%. Esta tasa se determinó con la colaboración del departamento del área de finanzas de la empresa que nos proporcionó la información.

La fórmula matemática para descontar los flujos del CLV se define de la siguiente forma:

$$CLV_i = \sum_{t=0}^T \frac{V_{it}}{(1+r)^t} \quad (3.1)$$

Donde:

$V_{it}$  = Compra futura del cliente  $i$  en la semana  $t$ .

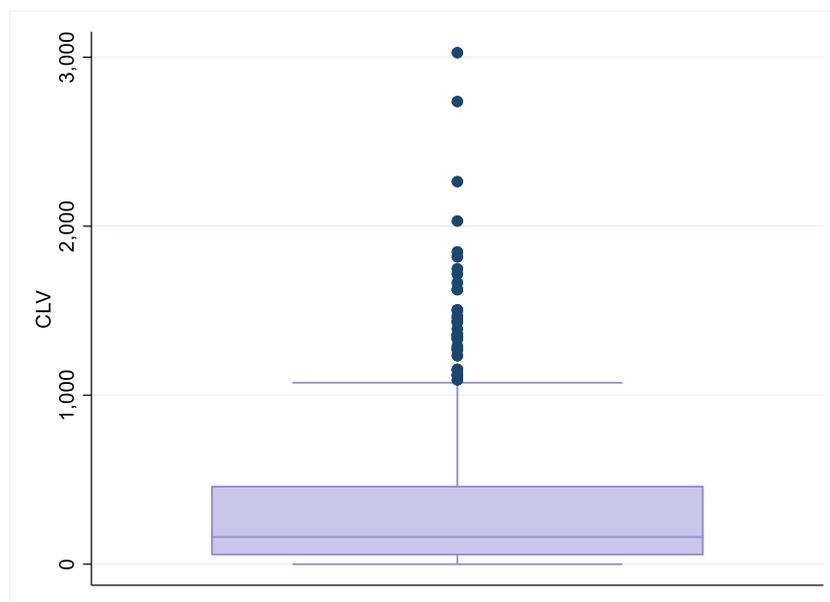
$r$  = tasa de descuento semanal.

$t$  = indicativo del período en semanas.

El resultado de  $CLV_i$  corresponde al valor presente de todos los montos que gastará el cliente en el futuro. Los resultados a nivel general son:

**Tabla 3.2 CLV general**

	Promedio	Mínimo	Máximo	Suma Total
CLV	\$ 338,88	\$ -	\$ 3.026,78	\$ 169.438,25



**Figura 3.4 Gráfico de Caja CLV general**

El promedio del CLV es aproximadamente \$339 el cual no es tan bajo considerando un periodo corto de pronóstico de un año. Además, solo se incluyeron los ingresos generados por 10 secciones de la tienda, aquellos guardan la naturaleza de un supermercado, y se excluyeron productos cuyo PVP fuera su superior a 100 USD.

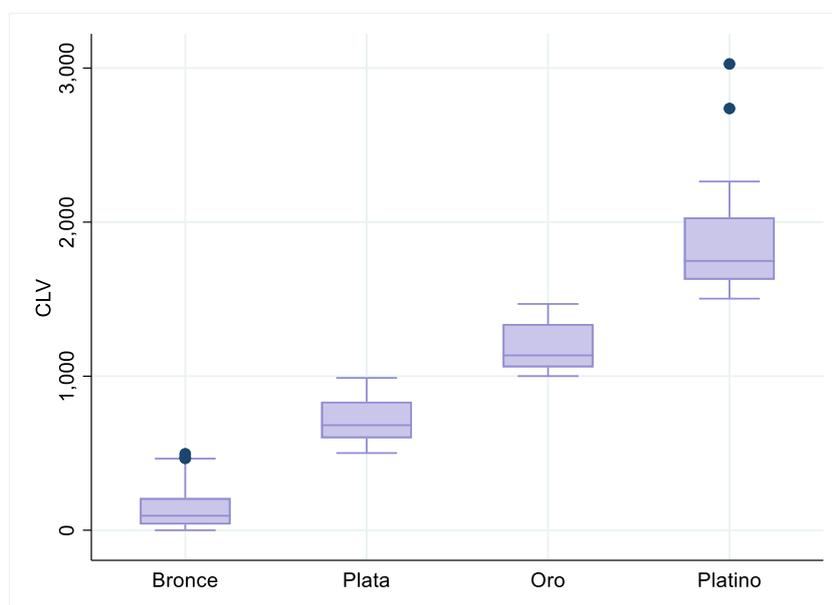
Dado que se obtuvieron las estimaciones del CLV de todos los clientes de la muestra representativa, se procede a la segmentación de los consumidores en base al valor económico que le generan a la empresa. Al segmentar a los clientes, la empresa puede comprender mejor la importancia de cada cliente para el beneficio total de la empresa. La segmentación basada en los aportes del cliente, generará estrategias de marketing más rentables.

Finalmente se clasifican los clientes en cuatro segmentos descritos como: a) los clientes platino, son aquellos clientes con mayor valor para la compañía con un valor estimado de compra superior a \$1500; b) oro se refiere a aquellos que generan entre \$1000 y \$1500; c) plata quienes mantiene compras entre \$500 y

\$1000 y por último d) los clientes bronce para aquellos con compras menores a \$500. La siguiente tabla resume un breve análisis descriptivo por grupo.

**Tabla 3.3 Descripción CLV por Segmento de clientes**

	CLV				%
	Promedio	Mínimo	Máximo	Suma Total	
<b>Platino</b>	\$ 1.931,63	\$ 1.503,27	\$ 3.026,78	\$ 25.111,13	2,6%
<b>Oro</b>	\$ 1.193,29	\$ 1.000,41	\$ 1.468,67	\$ 38.185,28	6,4%
<b>Plata</b>	\$ 708,13	\$ 500,94	\$ 988,58	\$ 52.401,41	14,8%
<b>Bronce</b>	\$ 141,05	\$ -	\$ 496,16	\$ 53.740,44	76,2%
<b>Total</b>	\$ 338,88	\$ -	\$ 3.026,78	\$ 169.438,25	100,0%



**Figura 3.5 Gráfico de Caja CLV por segmento de clientes**

Los clientes Platino, los más valiosos para la compañía, cuentan con un valor promedio de \$1.931,63, estos consumidores representan apenas el 2,6% del total de clientes, un incremento de clientes en este segmento sería una gran oportunidad para la empresa, debido a que les permitiría tener mayor confianza frente al futuro incierto, así mismo obtener mayores niveles de ingresos. Los clientes Oro también tienen poca representatividad un 6,4%, a continuación, los clientes Plata con un CLV promedio de \$703,87 una participación 14,8% y por

último los clientes Bronce tienen una mayor participación de 76,2% con un escaso valor económico promedio de \$141,05.

Como se menciona anteriormente, el grupo de clientes analizados presenta una alta concentración en el segmento de clientes Bronce, esta situación podría ser interpretada como un error en el muestreo, sin embargo, cabe recordar que para obtener la muestra se realizó un tipo de muestreo aleatorio estratificado clasificándolos en función a su concurrencia, producto de ello se obtuvo un grupo de clientes representativos a la población.

Por otro lado, la contribución económica que generan los grupos, muestra la verdadera rentabilidad que genera cada segmento a la compañía. A pesar de que la representatividad física de los clientes platino (2,6%), oro (6,4%) y plata (14,8%) es apenas un 33%, estos grupos generan una contribución económica superior al 60%, es decir un tercio de los clientes aporta más del 68% de las ventas.

**Tabla 3.4 Contribución por segmento de clientes**

	<b>% Según Contribución</b>
<b>Platino</b>	14,82%
<b>Oro</b>	22,54%
<b>Plata</b>	30,93%
<b>Bronce</b>	31,72%

### **3.3. Caracterización por Segmento**

Una vez realizada la segmentación de los clientes se procedió a analizar su comportamiento de compra durante los últimos 20 meses donde se identificaron patrones de conducta útiles para definir acciones futuras.

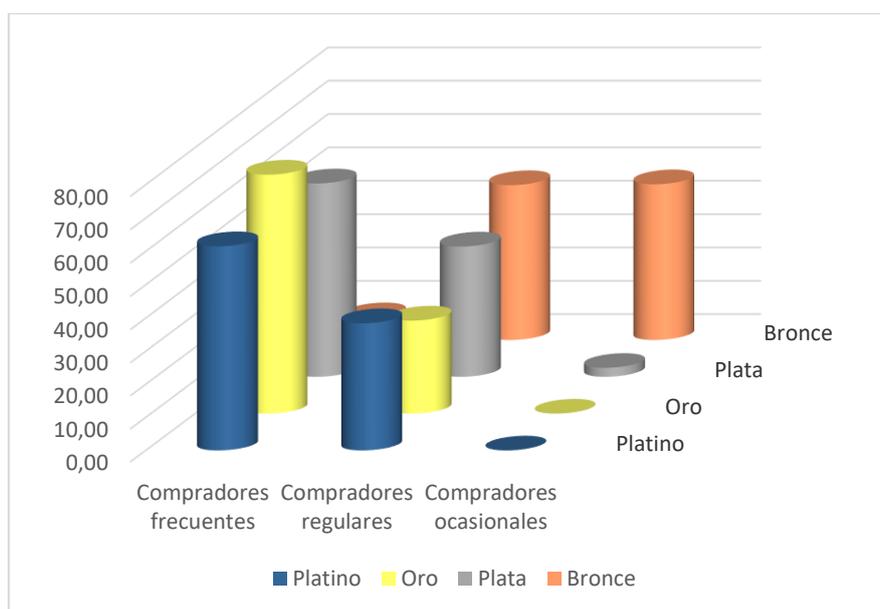
En primer lugar, se analizó la concurrencia de los clientes por segmento. La tabla 3.5 muestra una clara tendencia de frecuencia en los clientes que poseen mayor valor económico. Un 64% de los clientes platinos son considerados consumidores frecuentes, estos visitan la tienda más de dos veces por mes, mientras que, un 27% son compradores regulares con una frecuencia promedio de 2 veces al mes.

Los clientes Oro y plata se distribuyen de forma similar, más del 60% corresponden a consumidores frecuentes, destacándose el segmento Oro con clientes más fieles, y el segundo grupo más representativo corresponden a clientes regulares.

Por otro lado, se encuentran el segmento Bronce que presenta una mayor concentración en compradores ocasionales con una concurrencia promedio de una ocasión de compra mensual y menos del 30% con clientes regulares o frecuentes. La distribución de este segmento respecto a su frecuencia explica sutilmente la menor contribución económica que genera este grupo.

**Tabla 3.5 Distribución segmentos de clientes por frecuencia de compra**

	Compradores frecuentes	Compradores regulares	Compradores ocasionales	Total
<b>Platino</b>	61,54	38,46	0,00	2,6
<b>Oro</b>	71,88	28,13	0,00	6,4
<b>Plata</b>	58,11	39,19	2,70	14,8
<b>Bronce</b>	6,82	46,46	46,72	76,2
<b>Total</b>	20,00	44,00	36,00	100



**Figura 3.6 Gráfico de barras distribución de segmentos por frecuencia**

Respecto a las características demográficas de los clientes por segmento, es notable la contribución femenina. Debido a que en el contexto ecuatoriano es más común la participación de mujeres en compras relacionadas al hogar, circunstancia que reflejan los segmentos Oro, Plata y Bronce con un porcentaje cercano al 60%. No obstante, es curiosa la mayor participación masculina para

el grupo de clientes más valiosos, esto es un indicativo de que para grupos con niveles de ingresos mayores hay una mayor equidad respecto a las funciones del hogar.

**Tabla 3.6 Distribución de segmentos por sexo**

	Femenino	Masculino
<b>Platino</b>	38,46	61,54
<b>Oro</b>	56,25	43,75
<b>Plata</b>	59,46	40,54
<b>Bronce</b>	57,78	42,22
<b>Total</b>	57,43	42,57

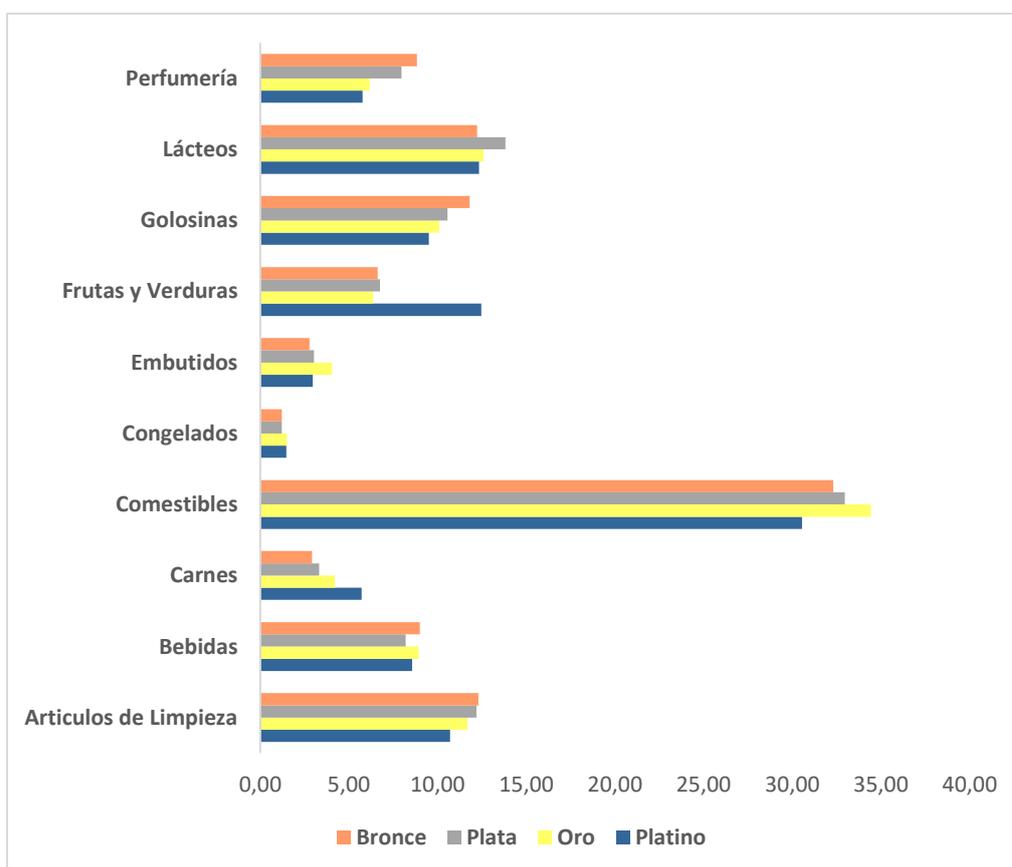
Otra característica que comparten los segmentos es la proporción alta de clientes Adultos Maduros con edades entre 35 a 65 años, tal como lo muestra la tabla 3.7. Lo cual es una consideración importante al momento de realizar las promociones debido a que en esta una edad el cliente ya tiene un criterio definido, marcas de referencia, por lo que, la estrategia de ventas debe ser dirigida a las características psicológicas de este tipo de clientes.

De la misma manera, los clientes casados son los que tienen mayor presencia en los tres primeros segmentos, principalmente los clientes platino con un 72%. Un análisis más profundo permitió conocer que los clientes platino y oro son principalmente casados y tienen edades de 35 a 50 años. Mientras que los clientes bronce son principalmente solteros.

**Tabla 3.7 Distribución de segmentos por Estado Civil y Rango de Edad**

		Platino	Oro	Plata	Bronce	Total
<b>Estado Civil</b>	<b>Casado(a)</b>	69,23	56,25	47,30	39,74	42,69
	<b>Divorciado(a)</b>	0,00	6,25	2,70	4,47	4,21
	<b>Soltero(a)</b>	23,08	31,25	37,84	46,84	43,89
	<b>Unión Libre</b>	7,69	6,25	9,46	7,63	7,82
	<b>Viudo(a)</b>	0,00	0,00	2,70	1,32	1,40
<b>Rango de Edad</b>	<b>Adulto Joven (20 a 35)</b>	15,38	9,38	16,22	23,62	21,40
	<b>Adulto Maduro (35 a 50)</b>	61,54	43,75	45,95	43,31	44,20
	<b>Adulto Maduro (50 a 65)</b>	23,08	37,50	28,38	24,67	26,00
	<b>Adulto Mayor (&gt; 65)</b>	0,00	9,38	9,46	8,40	8,40

En cuanto a canasta de compra, los diferentes segmentos muestran proporciones muy similares en la composición de su cesta. La principal sección de consumo son los comestibles con una participación promedio del 35%, seguido de los lácteos, artículos de limpieza y bebidas. Sin embargo, existen varias diferencias interesantes en los segmentos de clientes con mayor valor respecto al segmento de valor más bajo. Por ejemplo, los clientes platino tienen un mayor consumo en carnes que embutidos, de la misma manera compran más frutas que golosinas, mientras que los clientes bronce consumen más embutidos y golosinas que carnes y frutas, este comportamiento refleja la brecha en el nivel de ingresos de ambos grupos debido a que las carnes y las frutas, productos naturales y perecederos, tienden a tener un precio mayor respecto a los embutidos empaquetados y las golosinas.



**Figura 3.7 Composición de cesta de compra por segmento de clientes**

Finalmente, analizamos cómo influyen los segmentos a la liquidez de la empresa. La tabla 3.8 muestra el pago frecuente en efectivo por parte de los clientes con un promedio de **81%**, lo cual es coherente considerando que los promedios de compra no son tan altos. En consecuencia, un aumento de la facturación

resultará en un incremento casi inmediato de la liquidez, situación que es favorable en cierta medida. Sin embargo, la empresa debe aprovechar la oportunidad de generar incentivos para el uso de tarjetas y de esta forma poder alcanzar a clientes que optan por medios de pagos diferentes al efectivo.

**Tabla 3.8 Distribución de segmentos por forma de pago**

	Efectivo	Tarjeta
<b>Platino</b>	77,08	22,92
<b>Oro</b>	84,93	15,07
<b>Plata</b>	80,06	19,94
<b>Bronce</b>	83,07	16,93
<b>Total</b>	81,78	18,22

### 3.4. Ranking de los clientes top 20

Una de las ventajas en la estimación del CLV es que proporciona información de los clientes a nivel individual. Lo cual genera una gran utilidad para la compañía, en virtud de que se puede analizar las características demográficas y el comportamiento de compra de los clientes más rentables para estar en la capacidad de una estratégica asignación de recursos basándose en el valor del cliente para la empresa.

En la **Tabla 3.9** se presenta el top de los 20 clientes más valiosos. En el primer lugar en el ranking se posiciona un hombre casado de 38 años con un valor promedio de compra de \$19.48 que frecuenta la tienda alrededor de 3 o 4 veces a la semana, su primera sección de preferencia son los comestibles. En el segundo lugar se ubica una mujer casada con un promedio de visita de 15 veces mensuales. Por otro lado, en el tercer lugar en el ranking, se encuentra un hombre casado de 43 años, con 2 visitas mensuales, no obstante, su monto promedio de compra asciende los \$100 USD por cada transacción.

Adicionalmente, se puede notar el respaldo de los clientes hacia los productos que ofrece el supermercado en cuanto a su línea propia de productos, debido a que se muestra que destinan una proporción de la composición de su cesta a la compra de productos en marcas blancas.

**Tabla 3.9 Clientes Top 20**

Ranking	Código de Cliente	CLV	Edad	Estado Civil	Sexo	Valor promedio por visita	Frecuencia Mensual	1º Sección de preferencia	2º Sección de preferencia	% Compra en marca blanca
1	18328	\$ 3.026,78	38	Casado(a)	Masculino	\$ 19,48	14	Comestibles	Artículos de limpieza	1%
2	9344	\$ 2.737,56	61	Casado(a)	Femenino	\$ 14,27	15	Comestibles	Artículos de limpieza	14%
3	4224	\$ 2.263,64	43	Casado(a)	Masculino	\$ 114,98	2	Comestibles	Frutas y verduras	5%
4	5710	\$ 2.030,60	42	Casado(a)	Masculino	\$ 12,56	13	Comestibles	Frutas y verduras	4%
5	1645	\$ 1.847,99	39	Casado(a)	Masculino	\$ 15,09	8	Comestibles	Frutas y verduras	5%
6	8592	\$ 1.817,93	51	Casado(a)	Femenino	\$ 23,10	8	Comestibles	Frutas y verduras	10%
7	2384	\$ 1.747,93	32	Casado(a)	Masculino	\$ 31,39	3	Comestibles	Lácteos	5%
8	15079	\$ 1.715,88	38	Soltero(a)	Femenino	\$ 9,78	15	Comestibles	Bebidas	6%
9	1019	\$ 1.664,85	42	Soltero(a)	Femenino	\$ 69,99	3	Comestibles	Artículos de limpieza	10%
10	14222	\$ 1.626,13	33	Casado(a)	Masculino	\$ 21,66	4	Comestibles	Artículos de limpieza	1%
11	10216	\$ 1.623,80	38	Casado(a)	Masculino	\$ 22,76	6	Comestibles	Lácteos	5%
12	882	\$ 1.504,76	36	Casado(a)	Masculino	\$ 46,83	2	Frutas y verduras	Comestibles	6%
13	13971	\$ 1.503,27	53	Soltero(a)	Femenino	\$ 9,34	16	Comestibles	Golosinas	5%
14	17716	\$ 1.468,67	42	Unión Libre	Masculino	\$ 24,87	6	Comestibles	Bebidas	4%
15	2712	\$ 1.454,42	57	Soltero(a)	Femenino	\$ 26,87	6	Comestibles	Artículos de limpieza	4%
16	11470	\$ 1.434,55	64	Casado(a)	Masculino	\$ 11,72	9	Comestibles	Lácteos	8%
17	16234	\$ 1.432,36	71	Casado(a)	Femenino	\$ 17,07	6	Comestibles	Carnes	4%
18	17282	\$ 1.391,79	35	Soltero(a)	Masculino	\$ 30,89	4	Comestibles	Lácteos	10%
19	10103	\$ 1.357,49	63	Casado(a)	Masculino	\$ 16,39	8	Comestibles	Artículos de limpieza	5%
20	128	\$ 1.356,92	49	Casado(a)	Masculino	\$ 33,75	4	Comestibles	Artículos de limpieza	9%

### 3.5. Análisis de sensibilidad

Esta segmentación conduce a realizar un análisis de los potenciales incrementos en ventas, utilidad y liquidez de la empresa, tan solo si incrementa el nivel de los clientes en cada uno de los distintos segmentos. En vista de ello, se presenta la **Tabla 3.10** donde se describen los efectos del incremento en el nivel de clientes en un 1% y 5%.

**Tabla 3.10 Análisis de sensibilidad por segmentos de clientes valiosos**

Incremento	Segmento	Indicadores	Variación
1% +	Platino	$\Delta$ + Ventas	6,07%
		$\Delta$ + Utilidad	1,61%
		$\Delta$ + Liquidez	1,24%
5% +	Oro	$\Delta$ + Ventas	18,06%
		$\Delta$ + Utilidad	4,78%
		$\Delta$ + Liquidez	4,06%
5% +	Plata	$\Delta$ + Ventas	10,24%
		$\Delta$ + Utilidad	2,71%
		$\Delta$ + Liquidez	2,17%

En primer lugar, se observa el segmento de los clientes platino, los más valiosos para la empresa, si el nivel de sus clientes incrementa en un 1% esto contribuye a un aumento del 6,1% en las ventas, esta alta proporción se debe a que la canasta promedio de compra de este segmento está compuesta por una gran cantidad de productos; por otro lado, la utilidad de la empresa crece en un 1,6% y la liquidez en un 1,2%, el menor incremento en la liquidez se presenta en vista de que el 23% de los clientes de este segmento, realizan sus pagos con tarjeta.

Por otra parte, el incremento en un 5% de los clientes oro y plata provoca un aumento en ventas en un 18,1% y 10,2% respectivamente, estos resultados muestran que el aporte de los clientes oro es mayor, pese a contar con un menor número de clientes en comparación al segmento de clientes plata. Los efectos en la utilidad se ven reflejados en un aumento del 4,8% proveniente de los clientes oro y 2,7% de los clientes plata, cumpliéndose así, que los clientes de mayor valor generan mayores rentabilidades empresariales.

Finalmente, el incremento del 5% de clientes oro genera un aumento en la liquidez del 4%; por su parte el incremento en la misma proporción para los clientes plata provoca un

aumento del 2,2% en la liquidez. La diferencia en estas contribuciones se presenta debido a que los clientes plata optan por realizar sus compras por pago con tarjeta, en cambio los clientes oro se inclinan más por realizar pagos en efectivo, por ello, el mayor aporte en la liquidez está dado por el segmento de los clientes oro.

Estos hallazgos destacan las ventajas de realizar una segmentación en función del valor de los clientes para la empresa, mostrando que los clientes de mayor valor generan mayores beneficios empresariales en términos de ventas, utilidad y liquidez. Este análisis permite realizar una evaluación a los tomadores de decisión de los beneficios económicos que se obtienen si la empresa aplica estrategias que logren aumentar el valor de sus clientes en cada segmento.



# CAPÍTULO 4

## 4. CONCLUSIONES

Hoy en día compañías se benefician constantemente de los avances en tecnológicos. Con la evolución de las TICs y las estrategias de marketing uno a uno, CRM, marketing de bases de datos han permitido que las empresas puedan generar mejores tácticas para crear relaciones estables con sus clientes. Sin embargo, las empresas frecuentemente fallan en identificar los receptores adecuados para sus estrategias debido a que segmentan sus clientes en base a criterios poco tangibles como sus características demográficas, geográficas, psicográficas, actitudinales, uso, etc. Esta situación ha generado incertidumbre para las empresas en cuanto a la efectividad de su inversión en publicidad. En respuesta a ello, este estudio propuso una metodología de segmentación en base al valor monetario esperado del cliente calculado mediante el CLV.

Se estimó el CLV bajo la estructura de datos RFM, utilizada ampliamente en la literatura para clasificar clientes según su historial de compras. Esta estructura identifica como variables relevantes la frecuencia, recencia y valor monetario para proporcionar información relevante que permita conocer el comportamiento de compra de los clientes y la probabilidad de que se mantengan vivos. Es un conjunto simple pero eficiente para encontrar diferencias y patrones comunes en los clientes, que posteriormente ayudará a resolver importantes problemas de marketing, como el perfil demográfico de los mejores clientes y si los clientes más fieles compran más.

El modelo probabilístico usado en la estimación fue la extensión Pareto/NBD propuesta por Abe (2009) que captura la heterogeneidad de los clientes y muestra mayor flexibilidad en los supuestos. Este modelo tiene cualidades superiores al modelo original Pareto NBD (1987) en las estimaciones tanto a nivel global como a nivel individual, ya que posee menores tasas de error respecto a otros estudios realizados en campos similares.

La segmentación del cliente se puede realizar utilizando una variedad de características únicas del cliente para ayudar a los empresarios a personalizar los planes de marketing, identificar tendencias, planificar el desarrollo de productos, campañas publicitarias y entregar productos relevantes.

Para este estudio, se realizó la segmentación de clientes con diferentes niveles de CLV en categorías de platino, oro, plata y bronce. Los clientes de platino y oro son muy leales y altamente rentables. Mientras que, los clientes de plata y bronce son considerados como clientes de bajo valor. Una comprensión informada de los clientes clave, ayudaría a los gerentes a formular y cambiar el enfoque estratégico hacia la cartera de clientes que se ajuste a los objetivos de la organización.

Los resultados muestran que los clientes tienen un comportamiento heterogéneo en cuanto a la frecuencia de compra, pero presentan características demográficas similares. Esta información es útil para realizar un seguimiento de los cambios en el comportamiento de los clientes a lo largo del tiempo. Los hallazgos de este estudio promueven el uso del CLV a distintas áreas de aplicación.

Los gerentes y tomadores de decisión deben determinar qué nivel de detalle es conveniente analizar en función a sus objetivos. Inicialmente existen tres niveles de detalle útiles de analizar los cuales son:

- **A nivel de cliente individual:** como se pudo observar en el ranking las características y el comportamiento del cliente puede ser meticulosamente analizado. El análisis a este nivel es de útil para determinar diferentes estrategias como la selección de campañas publicitarias o las preferencias de atención al cliente. La aplicación varía según los objetivos comerciales de la empresa en un determinado momento, la cual decidirá incluir o excluir a un grupo determinado de clientes.
- **Por segmento de clientes:** estos niveles de detalle analizan las características comunes que comparten los consumidores por segmentos de valor para así poder identificar los factores promotores de crecimiento del segmento de cliente.
- **A nivel general de clientes:** las estimaciones de CLV son útiles para una mejor perspectiva en la planificación de negocios y una adecuada gestión estratégica.

En resumen, el CLV es uno de los indicadores más importantes para las empresas en general. A pesar de su subestimación, esta métrica es de uso transversal puesto que contribuye a mejorar diversas actividades comerciales como: marketing, servicio al cliente, administración, finanzas, etc. Los tomadores de decisiones tienen a su alcance información valiosa que les permitirá crear oportunidades para extender la relación entre

las empresas y sus clientes siendo una herramienta viable para la satisfacción y retención de clientes.

#### **4.1. Limitaciones**

En este estudio se analizó el comportamiento de compra de los clientes usuarios de la tarjeta de fidelización de la empresa en un período superior al año. Dentro del tiempo de estudio existen varias temporadas donde las ventas se ven incrementadas por factores externos a la compañía, tales comportamientos de compra estacionales no fueron controlados en este estudio.

Para una evaluación completa del CLV, es necesario contar con datos respecto a los costos de distribución y comercialización que permitan proporcionar información sobre de ganancias a nivel de cliente. Esta información fue restringida en el estudio por políticas de confidencialidad de la empresa. Sin embargo, esto no afectó en gran medida la interpretación de las estimaciones debido a que fue posible una aproximación de dichos valores mediante información pública sobre costos de la sucursal objeto de estudio.

Otra limitante hallada fue la imposibilidad en examinar el efecto que ejercen las ofertas de productos en el comportamiento de compra. A pesar de que la información de las compras con oferta se proporcionó, no fue posible la normalización de la variable, puesto que, la empresa no lleva un registro continuo del volumen de productos que coloca en oferta ni el tiempo de duración. Asimismo, no se pudo incorporar el análisis otros factores que influyen el comportamiento de compra como son el lugar de residencia, la profesión, estabilidad laboral del cliente, etc.

#### **4.2. Recomendaciones**

En el modelo aplicado para la estimación del CLV se identificaron restricciones de longitud mínima de datos, lo cual no fue un problema en el estudio debido a la gran cantidad de datos proporcionados por la empresa. Sin embargo, es necesario que en futuros estudios se tenga presente este requerimiento al momento de iniciar las estimaciones, no obstante, aconseja tener prudencia al incorporar múltiples tipos de datos, debido a que, este podría ser un error que afecte el rendimiento predictivo.

Para un adecuado uso del CLV se recomienda efectuar una fase previa de experimentación que permita conocer la factibilidad de una campaña segmentada o

individualizadas para aprovechar el alto valor esperado de los clientes. Adicionalmente, para el uso del ranking se sugiere el uso de la métrica de sensibilidad debido a que permitirá evaluar la capacidad del modelo para realizar la correcta clasificación de los clientes platino, lo cual es necesario para las campañas de marketing individualizadas.

Dada la capacidad predictiva en la estimación, el CLV puede considerarse una medida de rendimiento efectiva para las ventas o ganancias. Por lo tanto, esta métrica podría ser de gran utilidad para las empresas al momento de evaluar la efectividad de las promociones, programas de fidelidad y otros tipos de esfuerzos de marketing. Sin embargo, la metodología empleada en el cálculo no permite determinar los posibles factores causales para el CLV. Por ello, se recomienda que las futuras investigaciones puedan analizar la relación entre la vida del cliente y su rentabilidad.

Por otro lado, debido a la existencia de múltiples metodologías para la estimación del CLV se sugiere otros estudios que puedan comparar diferentes modelos, otros más recientes y menos conocidos, en búsqueda de un rendimiento superior. Además, se agradecería realizar estudios empíricos similares en diferentes tipos de negocios dentro país del para poder analizar la convergencia o diferencia en los resultados.

### **4.3. Contribuciones**

Se presenta una alternativa innovadora para la segmentación de clientes, que no ha sido aplicado en las industrias ecuatorianas y servirá para elaborar estrategias dirigidas a clientes selectos. Debido a que se analiza el valor que representan los clientes para la empresa a nivel global, por segmentos e individual.

El uso de esta métrica incentiva la competencia en análisis de datos a nivel de negocios, a través del aprovechamiento de la información que ya dispone la empresa. De esta forma, es posible analizar el comportamiento del cliente, ahorrando recursos y esfuerzos, haciendo uso de la información disponible en lugar de realizar inversiones en nuevas investigaciones de mercado.

Con el uso del CLV se proporciona a los gerentes y tomadores de decisión información que les permitirá comprender qué tipo de estrategias de marketing se deben preferir para qué clientes, cuántas inversiones se deben hacer para ellos y qué campañas de marketing se deben seguir.



# BIBLIOGRAFÍA

- Abe, M. (2009). "Counting Your Customers" One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model. *Marketing Science*, 541–553.
- Bernat, J. R., Koning, A. J., & Fok, D. (2018). Modelling Customer Lifetime Value in a Continuous, Non-Contractual Time Setting. *Business Analytics & Quantitative Marketing*.
- Berry, L. (1995). Relationship marketing of services growing interest, emerging perspective. *Journal of the Academy of Marketing Science*, 236-245.
- Blattberg, R., Gary, G., & Thomas, J. S. (2001). Customer Equity: Building and Managing Relationships as Valuable Assets. *Harvard Business School Press*.
- Blattberg, R., Kim, P., & Neslin, S. (2008). Database marketing: analyzing and managing customers. *New York, NY: Springer*. Obtenido de [https://login.e.bibl.liu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat00\\_115a&AN=Ikp.850862&site=eds-live&scope=site](https://login.e.bibl.liu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat00_115a&AN=Ikp.850862&site=eds-live&scope=site)
- Borle, S. S., & Dipak, J. (2007). Customer Lifetime Value Measurement. *Management Science*.
- Christy, J., Umamakeswari, L., & Priyatharsini, A. (2018). RFM-Ranking And effective approach to customer segmentation. *Journal of King Saud University - Computer and Informacion Science*.
- Colombo, R., & Jiang, W. (1999). A Stochastic RFM Model, . *Journal of Interactive Marketing*, 2-12.
- Do Ruibin, K., & Vintilescu Borglöv, T. (2018). Predicting Customer Lifetime Value: Understanding its accuracy and drivers from a frequent flyer program perspective.
- Fader, P., & Hardie, B. (2009). Probability Models for Customer-Base. *Journal of Interactive Marketing*, 61-9.
- Fader, P., Hardie, B., & Lee, K. (2005). Counting your Customers the Easy Way: An Alternative to the Pareto/NBD Model. *Marketing Science*, 84-275.
- Fader, P., Hardie, B., & Lee, K. (2005). RFM and CLV: Using Iso Value Curves for Customer Base Analysis. *Journal of Marketing Research*, 30-415.
- Fader, P., Hardie, B., & Shang, J. (2010). Customer Base Analysis in a Discrete Time Noncontractual Setting. *Marketing Science*, 108-1086.
- Gerstner, E., & Libai, B. (2006). Why does poor service prevail? *Marketing Science*, 601-603.
- Gilchrist, A. (2016). *Industry 4.0: the industrial internet of things*. New York : Apress.

- Glady, N., Baesens, B., & Croux, C. (2009). A modified Pareto/NBD approach for predicting customer lifetime value. *Expert Systems with Applications*, 2062-2071.
- Hogan, J., Lemon, K., & Libai, B. (2003). What is the true value of a lost customer? . *Journal of Service Research*, 196–208.
- Homburg, C., Droll, M., & Totzek, D. (2008). Customer prioritization: Does it pay off, and how should it be implemented? *Journal of Marketing*, 110-130.
- Hong, T., & Kim, E. (2012). Segmenting customers in online stores based on factors that affect the customer's intention to purchase. *Expert Systems with Applications*, 39(2), 2127-2131.
- Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 679–688.
- Jamal, Z., & Zhang, A. (2009). DMEF Customer Lifetime Value Modeling. *Journal of Interactive Marketing*, 279-283.
- Jasek, P., Vrana, L., Sperkova, L., Smutny, Z., & Kobulsky, M. (2018). Modeling and Application of Customer Lifetime Value in online Retail. 2-22.
- Kumar, V., & George, M. (2007). Measuring and maximizing customer equity: A critical analysis. *Journal of the Academy of Marketing Science*, 157-171.
- Kumar, V., & Pansari, A. (2016). National culture, economy, and customer lifetime value: Assessing the relative impact of the drivers of customer lifetime value for a global retailer . *Journal of International Marketing*, 24(1), 1-21.
- Kumar, V., & Shah, D. (Edits.). (2015). *Handbook of research on customer equity in marketing*. Edward Elgar Publishing.
- Ma, S.-H., & Liu, J.-L. (2007). The MCMC Approach for Solving the Pareto/NBD Model and Possible Extensions. *Third International Conference on Natural Computation (ICNC 2007)*. Haikou, China.
- Morrison, D. G., & Schmittlein, D. C. (1980). Jobs, strikes, and wars: Probability models for duration. *Organizational Behavior and Human Performance*, 224-251.
- Nikkhahan, B., Badrabadi, A. H., & Tarokh, M. J. (2011). Customer lifetime value model in an online toy store. *Journal of Industrial Engineering International*, 19-31.
- Parvatiyar, A., & Sheth, N. (2001). Conceptual framework of customer relationship management. *Emerging concepts tools and applications*, 3-25.
- Payne, A., & Frow, P. (2005). A strategic framework for customer relationship management. *Journal of Marketing*, 167–176.
- Platzer, M., & Reutterer, T. (2016). Ticking Away the Moments: Timing Regularity Helps to Better Predict Customer Activity. *Marketing Science*, 35, 693-829.
- Reinartz, W. J., & Kumar, V. (2000). On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. *Journal of Marketing*, 17–35.

- Reinartz, W., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, 77-99.
- Reinartz, W., Krafft, M., & Hoyer, W. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of Marketing Research*, 293–305.
- Romero, J., Lans, R., & Wierenga, B. (2013). A Partially MARKov Model of Customer Dynamics for CLV Measurement. *Journal of Interactive Marketing*.
- Rust, R. T., & Chung, T. S. (2006). Marketing models of service and relationships. *Marketing Science*, 25(6), 560-580.
- Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, 27-109.
- Rust, R., Kumar, V., & Venkatesan, R. (2011). Will the frog change into a prince?: Predicting future customer profitability. *International Journal of Research in Marketing*, 94-281.
- Schmittlein, D. C., & Colombo, R. (1987). Counting your Customers: Who Are They and What Will They Do Next? *Management Science*, 1-24.
- Schmittlein, D., & Peterson, R. (1994). Customer Base Analysis: An Industrial Purchase Process Application. *Marketing Science*, 41-67.
- Statista. (2017). *Statista*. Obtenido de Retail & Trade: <https://www.statista.com/markets/423/retail-trade/>
- Tanner, A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (theory and methods). En J. A. Statist.
- Thakur, R., & Workman, L. (2016). Customer Portfolio Management (CPM) for Improved Customer Relationship Management (CRM): Are your Customer Platinum, Gold, Silver, or Bronze? *Journal of Business Research*, 4095-5102.
- Venkatesan, R., & Kumar, V. (2004). A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Marketing*, 25-106.
- Wu, S.-I., & Li, P.-C. (2011). The relationship between CRM, RQ, and CLV based on different hotel preferences. *International Journal of Hospitality Management*.
- Wübben, M., & Wangenheim, F. (2008). Instant Customer Base Analysis: Managerial Heuristics Often “Get It Right”. *Journal of Marketing*, 72(3), 82–93.
- Wübben, M., & Wangenheim, F. (2008). Instant Customer Base Analysis: Managerial Heuristics Often “Get It Right”. *Journal of Marketing*, 82–93.
- Yoo, M. M., Bai, B., & Singh, A. (2018). Loyalty program effectiveness: Investigating the evolution of behavioral loyalty and Customer Lifetime Value over time. In *8th advances in hospitality and tourism marketing and management (ahtmm) conference*, (pág. 262).

Zeithaml, V. R. (2001). The customer pyramid creating and serving profitable customers. *California Management Review*, 118-142.

# APÉNDICES

## Participación por Segmentos clientes

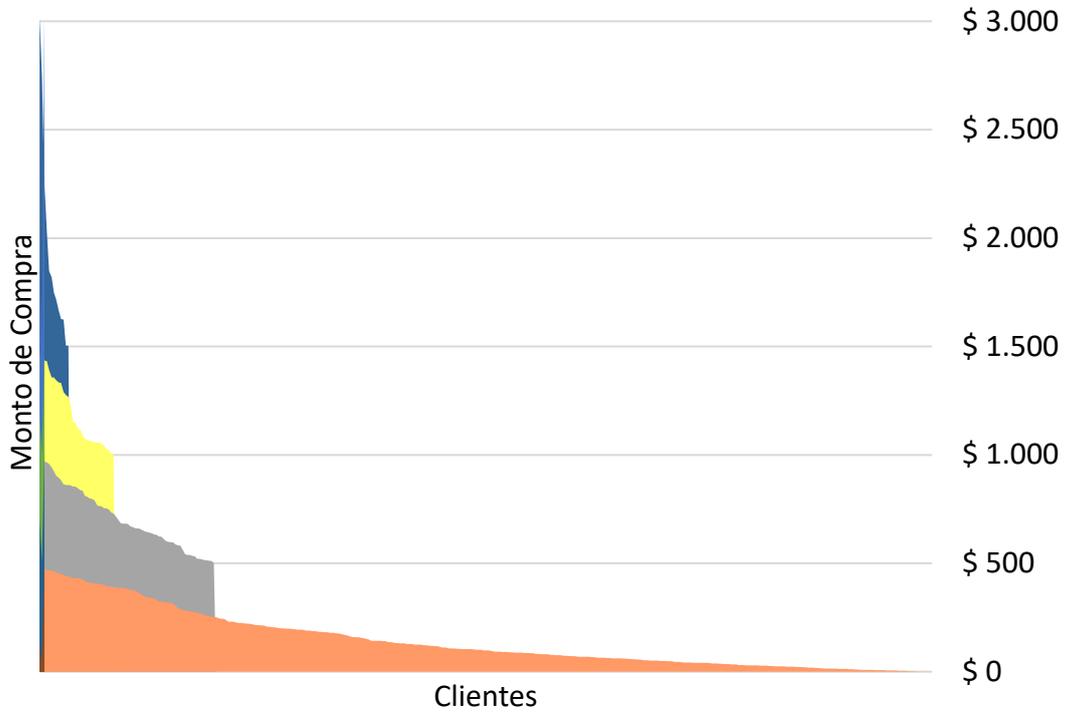


Figura 8 Monto promedio de compra

## Caracterización del cliente

### Cientes Platino

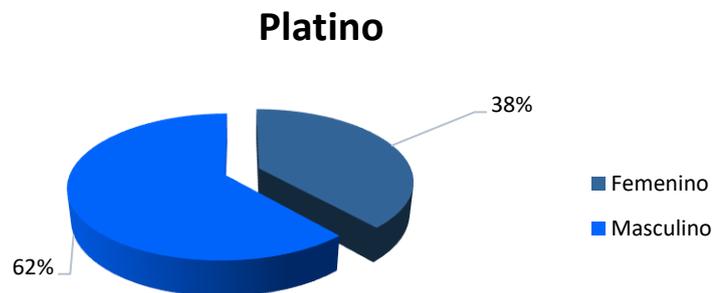


Figura 9 Distribución de clientes Platino según su sexo

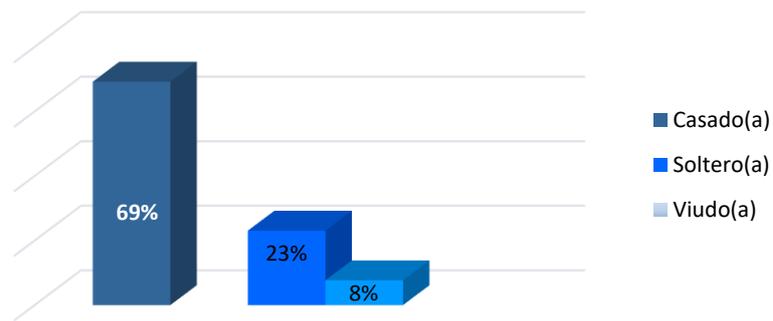


Figura 10 Distribución de clientes Platino según Estado Civil

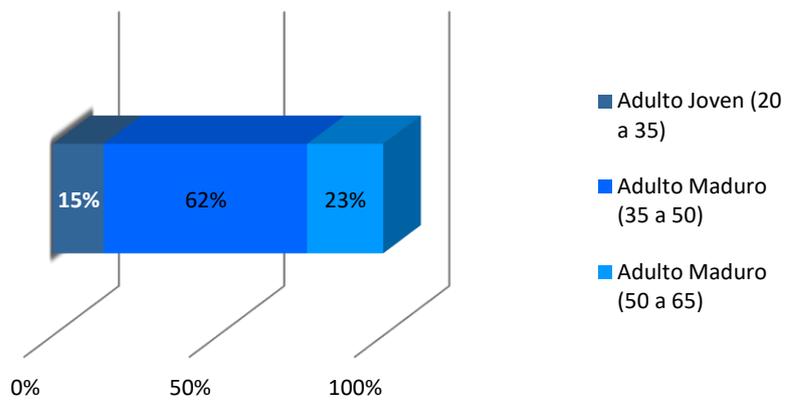


Figura 11 Distribución de clientes Platino según Rango de Edad

## Cientes Oro

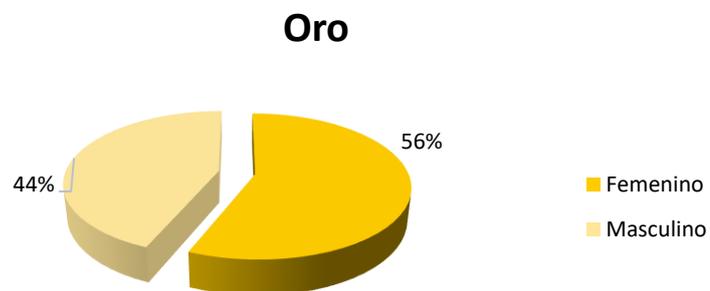
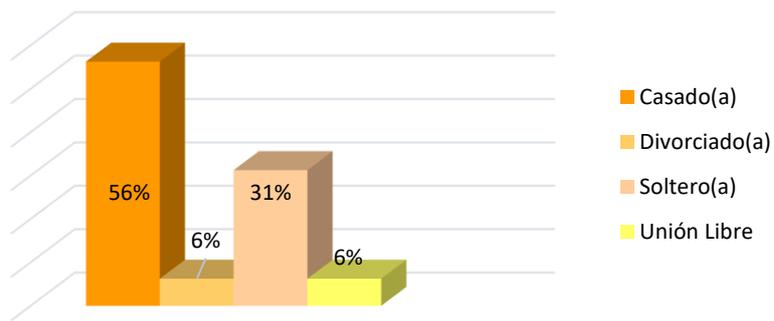
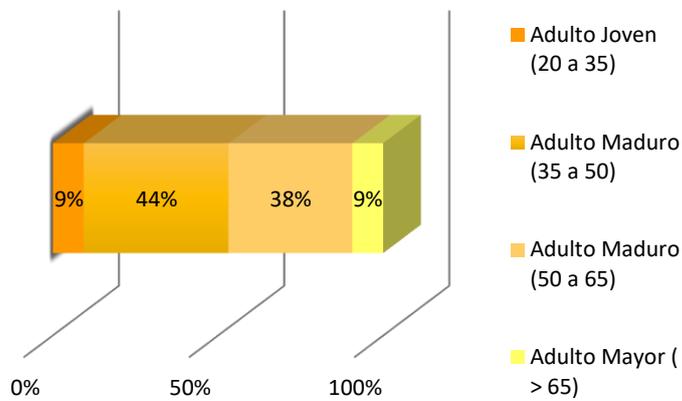


Figura 12 Distribución de clientes Oro según su sexo

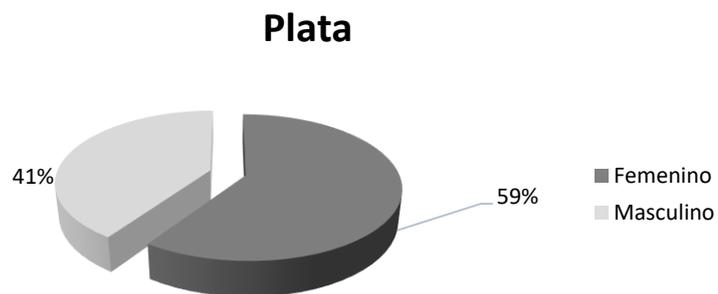


**Figura 13 Distribución de clientes Oro según Estado Civil**

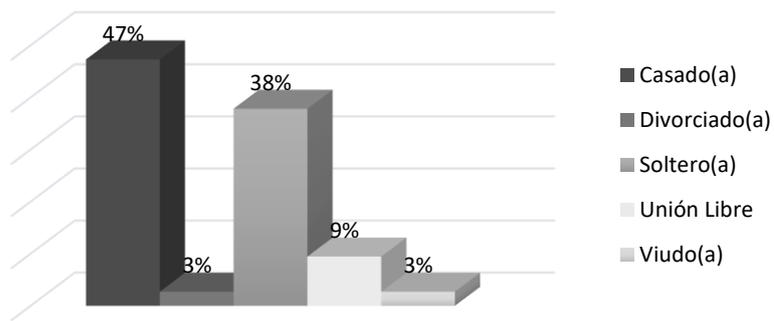


**Figura 14 Distribución de clientes Oro según Rango de Edad**

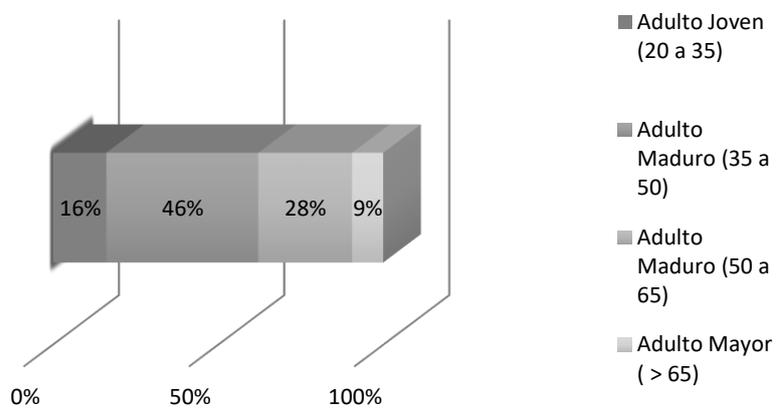
## Cientes plata



**Figura 15 Distribución de clientes Plata según su sexo**

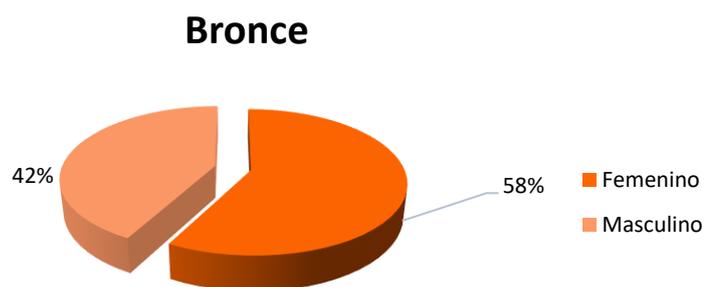


**Figura 16 Distribución de clientes Plata según Estado Civil**

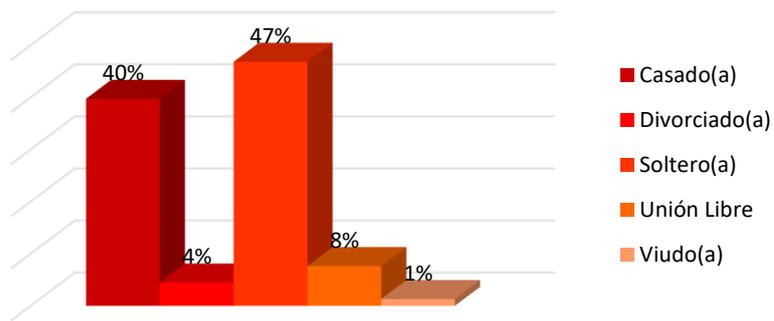


**Figura 17 Distribución de clientes Plata según Rango de Edad**

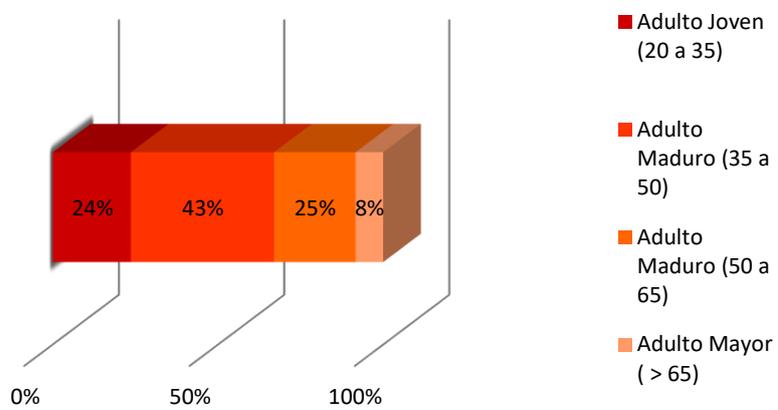
## Cientes bronce



**Figura 18 Distribución de clientes Bronce según su sexo**



**Figura 19 Distribución de clientes Bronce según Estado Civil**



**Figura 20 Distribución de clientes Bronce según Rango de Edad**