

**UNIVERSIDAD NACIONAL AGRARIA LA MOLINA**

**CICLO OPTATIVO DE PROFESIONALIZACIÓN EN  
MARKETING Y FINANZAS**



**“SEGMENTACIÓN DE CLIENTES DE UNA TIENDA DE  
ELECTRODOMÉSTICOS UTILIZANDO EL ANÁLISIS DE  
CONGLOMERADOS”**

**Presentado por:**

**FELIX AUGUSTO TANG BEDOYA**

**CECILIA VARGAS CUYAN**

**Trabajo de Titulación para Optar el Título de:  
INGENIERO ESTADÍSTICO E INFORMÁTICO**

**Lima – Perú**

**2016**

## **DEDICATORIA**

"Dedicado a Cecilia, por su comprensión y amor; a mis padres y hermanos, por todo su apoyo."

*Felix Augusto Tang Bedoya*

## **DEDICATORIA**

"Dedicado a Felix, por todo su amor y cuidados; a mis padres y hermano, por su apoyo y aliento constante."

*Cecilia Vargas Cuyan*

## **AGRADECIMIENTO**

Agradecemos a Dios por habernos guiado y darnos la fortaleza para seguir adelante en los momentos de debilidad.

Agradecemos de manera especial al Ms. Jaime Carlos Porras Cerrón por creer en nosotros y brindarnos la oportunidad de desarrollar este trabajo, gracias por su tiempo, amistad y dedicación.

# ÍNDICE

I.	RESUMEN.....	1
II.	INTRODUCCIÓN.....	2
2.1.	El problema.....	3
2.1.1.	Problema Principal.....	3
2.1.2.	Problema Secundario.....	3
2.2.	Objetivos.....	3
2.2.1.	Objetivo general.....	3
2.2.2.	Objetivos específicos.....	3
III.	REVISIÓN DE LITERATURA.....	4
3.1.	Segmentación de Mercados.....	4
3.2.	Análisis Exploratorio de Datos (AED).....	6
3.2.1.	Diagrama de barras.....	7
3.2.2.	Diagrama de sectores o circular.....	7
3.2.3.	Histograma.....	7
3.2.4.	Diagrama de tallo y hojas.....	8
3.2.5.	Diagrama de cajas (Box plot).....	8
3.2.6.	Detección de outliers multivariantes.....	9
3.3.	Análisis de Conglomerados.....	10
3.4.	Análisis de Conglomerados Bietápico.....	12
3.4.1.	Consideraciones del conjunto de datos.....	13
3.4.2.	Medida de la distancia para variables cualitativas y cuantitativas.....	13
3.4.3.	Características de Conglomerados o CF.....	15
3.4.4.	CF Tree o Árbol de características de conglomerados.....	15
3.4.5.	Algoritmo de conglomerado en 2 etapas.....	16
3.5.	Estrategias de marketing mix.....	19
3.6.	Aplicaciones de análisis de conglomerados en segmentación de mercados....	23

IV.	MATERIALES Y MÉTODOS.....	25
4.1.	Diseño y tipo de la investigación .....	25
4.2.	Formulación de hipótesis .....	25
4.3.	Identificación de variables .....	25
4.4.	Procedimiento y análisis de datos .....	26
4.4.1.	Análisis exploratorio y descriptivo.....	26
4.4.2.	Limpieza de datos .....	26
4.4.3.	Obtención de conglomerados .....	27
4.4.4.	Definición de perfiles .....	27
4.4.5.	Plantear estrategia de marketing.....	27
V.	RESULTADOS Y DISCUSIÓN.....	28
5.1.	Análisis exploratorio y descriptivo .....	28
5.1.1.	Análisis exploratorio y descriptivo de variables cuantitativas .....	28
5.1.2.	Análisis exploratorio y descriptivo de variables cualitativas .....	30
5.2.	Limpieza de datos .....	31
5.2.1.	Limpieza de datos de variables cuantitativas .....	31
5.2.2.	Limpieza de datos de variables cualitativas .....	33
5.3.	Obtención de conglomerados.....	37
5.4.	Propuesta de perfiles de los conglomerados .....	40
5.4.1.	Perfil del conglomerado 1: Los cazaofertas .....	40
5.4.2.	Perfil del conglomerado 2: Los estacionales .....	40
5.4.3.	Perfil del conglomerado 3: Consumidores cautos .....	41
5.5.	Propuesta de estrategia de marketing.....	41
5.5.1.	Estrategia de marketing del conglomerado 1: Los cazaofertas .....	41
5.5.2.	Estrategia de marketing del conglomerado 2: Los estacionales .....	42
5.5.3.	Estrategia de marketing del conglomerado 3: Consumidores cautos .....	42
VI.	CONCLUSIONES.....	44
VII.	RECOMENDACIONES .....	46
VIII.	REFERENCIA BIBLIOGRÁFICA .....	47
IX.	ANEXOS .....	50

## ÍNDICE DE CUADROS

Cuadro 1: Variables cuantitativas.....	25
Cuadro 2: Variables cualitativas.....	26
Cuadro 3: Estadísticos descriptivos de variables cuantitativa del conjunto de datos inicial .....	28
Cuadro 4: Estadísticos descriptivos de Distancia de Mahalanobis.....	31
Cuadro 5: Estadísticos descriptivos de variables cuantitativas luego de la limpieza de datos.....	33
Cuadro 6: Distribución de cada conglomerado .....	37
Cuadro 7: Puntuaciones medias de las variables cuantitativas en cada conglomerado..	38

## ÍNDICE DE FIGURAS

Figura 1: Diagrama de cajas .....	8
Figura 2: Esquema gráfico del Conglomerado Bietápico.....	18
Figura 3: Modelo básico del proceso de marketing.....	20
Figura 4: Diagrama de cajas de variable cuantitativa: <i>Monto de Capital</i> .....	28
Figura 5: Diagrama de cajas de variable cuantitativa: <i>Edad</i> .....	29
Figura 6: Diagrama de cajas de variable cuantitativa: <i>Ingresos</i> .....	29
Figura 7: Gráfico de sector de las variables cualitativas .....	30
Figura 8: Diagrama de cajas de la Distancia de Mahalanobis con la presencia de <i>ouliers</i> multivariados .....	31
Figura 9: Diagrama de cajas de la Distancia de Mahalanobis sin la presencia de <i>ouliers</i> multivariados .....	32
Figura 10: Histograma de la variable <i>Monto capital</i> antes de la limpieza de datos .....	34
Figura 11: Histograma de la variable <i>Monto capital</i> después de la limpieza de datos... 34	
Figura 12: Histograma de la variable <i>Edad</i> antes de la limpieza de datos .....	35
Figura 13: Histograma de la variable <i>Edad</i> después de la limpieza de datos.....	35
Figura 14: Histograma de la variable <i>Ingreso</i> antes de la limpieza de datos .....	36
Figura 15: Histograma de la variable <i>Ingreso</i> después de la limpieza de datos.....	36
Figura 16: Tamaños de conglomerados.....	37
Figura 17: Comparación de conglomerados.....	39



## I. RESUMEN

El aumento continuo de grandes volúmenes de datos y la importancia de la utilización de éstos, junto con la búsqueda de satisfacción de los clientes se han vuelto un gran reto que las grandes empresas hoy en día quieren superar. En la actualidad las empresas conocen la importancia que tiene el almacenamiento, la captura de datos y el beneficio que le puede resultar si se explotan correctamente.

Este estudio presenta una técnica estadística para segmentar clientes de una tienda de *retail* en el Perú. El conjunto de datos está conformado por clientes (personas naturales) que han adquirido un artículo para su hogar mediante un crédito en el primer trimestre del año 2013 de una tienda de *retail*. El conjunto de datos inicial estaba compuesto por 6284 clientes. Luego de proceder con el análisis exploratorio y la limpieza de datos, correspondiente a la eliminación de datos *outliers* y faltantes; se trabajó un conjunto de datos compuesto por 4980 clientes.

Se aplicó análisis de conglomerados bietápico, que es una técnica de segmentación que permite trabajar con variables cuantitativas y categóricas. El resultado de la aplicación de la técnica brindó 3 conglomerados: el primero con 1817 clientes (36.5%), el segundo conglomerado con 1390 clientes (27.9%) y el tercer conglomerado con 1773 clientes (35.6%). Después de esto se procedió a describir los perfiles de cada conglomerado y se propusieron estrategias de marketing mix para cada uno de ellos, con el objetivo de fidelizar al cliente, aumentar las ventas y el posicionamiento de la tienda de *retail*.

Palabras claves: análisis de conglomerados bietápico, marketing mix, estrategias de marketing, perfiles de clientes.

## II. INTRODUCCIÓN

Hoy en día las empresas están enfocadas hacia un principal objetivo que es satisfacer las necesidades de sus consumidores.

Esto consiste en determinar las principales inquietudes de los clientes, ya que a medida que transcurre el tiempo sus necesidades, y problemas con el entorno van variando, lo que implica que los consumidores se van volviendo cada vez más exigentes; sus necesidades ya no son las mismas y siempre van a querer un producto mejor. Entonces a las organizaciones no les queda otra opción que determinar qué es lo que necesitan los clientes para así cumplir con sus expectativas. (Renart, 2004) y (Kotler & Armstrong, 2007)

Es importante recordar que no todas las personas son iguales, son seres con diferentes opiniones, necesidades, gustos y pensamientos; por lo tanto cada uno de ellos debe ser tratado como eso: un ser único. Aunque en muchas ocasiones se puede encontrar personas que comparten similares características. Es decir, un grupo de personas con el que se está identificado y más que eso relacionado porque tienen hábitos o necesidades similares. Es aquí cuando se puede discutir sobre segmentación de mercado. (Fernández & Aqueveque, 2001)

Las empresas son organizaciones dedicadas a satisfacer una necesidad. Para determinar qué es lo que necesita la población se debe empezar con la elaboración de un estudio de mercado y así encontrar el problema. Posteriormente se puede segmentar la población según los gustos, zonas, necesidades, edades, pensamientos, etc.

Este proyecto se enfoca en la segmentación de clientes para una tienda de *retail* de electrodomésticos, para identificar grupos homogéneos dentro de ellos y heterogéneos entre sí. Cada grupo de clientes tendrá características específicas que los harán distintos a los otros y a la empresa de *retail* le facilitará la implementación de estrategias para cada conglomerado encontrado. De esta forma se podrá satisfacer al cliente logrando que estos se fidelicen con la empresa.

La técnica estadística a emplear en el presente trabajo es el Análisis de Conglomerados Bietápico, el cual presenta una metodología apropiada ya que esta se puede aplicar a variables cuantitativas y categóricas a la vez.

## **2.1. El problema**

### *2.1.1. Problema Principal*

¿Se podrán segmentar a clientes de una tienda de electrodomésticos, para posteriormente aplicar a cada segmento estrategias de marketing?

### *2.1.2. Problema Secundario*

- ¿Con las variables que se considera en el estudio se podrá encontrar conglomerados con una proporción equilibrada?
- ¿Se podrá encontrar perfiles distintos entre cada conglomerado para poder aplicar a cada uno de ellos estrategias de marketing?

## **2.2. Objetivos**

En cuanto al objetivo general y objetivos específicos se encuentran:

### *2.2.1. Objetivo general*

Aplicar la metodología de Análisis de Conglomerados Bietápico a clientes de una tienda de *retail* para posteriormente poder proponer estrategias de marketing mix a cada conglomerado encontrado.

### *2.2.2. Objetivos específicos*

- Determinar perfiles a los segmentos o conglomerados encontrados.
- Determinar estrategias de marketing según el perfil de cada segmento, cuyo fin será fidelizar al cliente.
- Proponer nuevas variables que ayuden a mejorar una futura segmentación.

### **III. REVISIÓN DE LITERATURA**

En esta sección se define la teoría obtenida de diferentes estudios que ayudan a mejorar el entendimiento del presente trabajo, pasando por la segmentación de mercados, las técnicas de conglomerados, las estrategias de marketing y las investigaciones referentes a segmentación de mercados que sirvieron como referencia para este trabajo.

#### **3.1. Segmentación de Mercados**

La segmentación de mercados es el proceso de dividir un mercado en grupos uniformes más pequeños que tengan características y necesidades semejantes. Esto lleva a la conclusión de que el mercado se encuentra dividido por segmentos, siendo estos grupos homogéneos. Esta homogeneidad dentro de cada grupo, se relaciona a diferentes tipos de estrategias de marketing que se puede usar en ellos (Valderrey, 2011).

También se puede afirmar que la segmentación de mercados constituye la piedra angular de una empresa con una gestión orientada al mercado y es una herramienta fundamental para conseguir mejoras en la rentabilidad y productividad del marketing. Supone identificar grupos específicos de clientes, con necesidades y comportamientos de compra únicos, y definir las características demográficas relevantes que los identifican. (Best, 2007).

Según Valderrey (2011), una correcta segmentación debe tener los siguientes requisitos:

- Homogeneidad en el segmento.
- Heterogeneidad entre los segmentos.
- Estabilidad de segmentos.
- Los segmentos deben ser identificables y medibles.
- Los segmentos deben ser accesibles y manejables.
- Los segmentos deben ser suficientemente grandes para ser rentables.

Asimismo se indica que las variables utilizadas para segmentar pueden ser de diferentes tipos:

- Variables geográficas:
  - Región del mundo o del país.
  - Tamaño del país.
  - Clima.

- Variables demográficas;
  - Edad.
  - Género.
  - Orientación sexual.
  - Tamaño de la familia.
  - Ciclo de vida familiar.
  - Ingresos.
  - Profesión.
  - Nivel educativo.
  - Estatus socioeconómico.
  - Religión.
  - Nacionalidad.
- Variables psicográficas:
  - Personalidad.
  - Estilo de vida.
  - Valores de vida.
  - Actitudes.
- Variables de comportamiento:
  - Búsqueda del beneficio.
  - Tasa de utilización del producto.
  - Fidelidad de la marca.
  - Utilización del producto final.
  - Nivel de “listo-para-consumir”.
  - Unidad de toma de decisión.

Cuando muchas variables se combinan para proporcionar un conocimiento profundo del segmento, se le conoce como segmentación profunda. Cuando se tiene suficiente información para crear una imagen clara del miembro típico del segmento, se llama perfil del comprador. Una técnica estadística utilizada habitualmente en determinar un perfil es el análisis cluster o análisis de conglomerados.

El proceso de segmentación de mercados da muchos beneficios para la empresa, tales como identificar las necesidades específicas de cada segmento, focalizar la estrategia de marketing, hacer publicidad efectiva, identificar nichos donde no haya competencia

directa y aumentar las posibilidades de crecer rápidamente en segmentos de mercado sin competidores (Valderrey, 2011).

(Pérez, 2011) Define la segmentación de mercado como “El proceso de dividir un todo (población, consumidores, etc.) en grupos uniformes más pequeños que tengan características semejantes denominados segmentos. La segmentación suele realizarse según los valores de determinadas variables que son las que fijan sus características”.

(Lamb & McDaniel, 2006) Señalan que "El propósito de la segmentación de mercado es que el mercadólogo ajuste las mezclas de marketing a la medida de las necesidades de uno o más segmentos específicos. La segmentación de mercados juega un papel clave en la estrategia de marketing de casi todas las empresas exitosas, ya que representa una poderosa herramienta de marketing, que ayuda a los especialistas en marketing a definir con más precisión las necesidades y deseos de los consumidores”.

### **3.2. Análisis Exploratorio de Datos (AED)**

El Análisis Exploratorio de Datos (AED) es un conjunto de técnicas estadísticas, que tiene como objetivo describir los datos, organizándolos y presentándolos de forma que pueda entenderse fácilmente su estructura. (Vargas, 1995)

La finalidad del Análisis Exploratorio de Datos es examinar los datos previamente a la aplicación de cualquier técnica estadística. (Martín & De Paz, 2007) Nos recomiendan que en la realización de un Análisis Exploratorio de Datos conviene seguir las siguientes etapas:

- a) Preparar los datos para hacerlos accesibles a cualquier técnica estadística.
- b) Realizar un examen gráfico de la naturaleza de las variables individuales a analizar y un análisis descriptivo numérico que permita cuantificar algunos aspectos gráficos de los datos.
- c) Realizar un examen gráfico de las relaciones entre las variables analizadas y un análisis descriptivo numérico que cuantifique el grado de interrelación existente entre ellas.
- d) Evaluar, si fuera necesario, algunos supuestos básicos subyacentes a muchas técnicas estadísticas como, por ejemplo la normalidad, linealidad y homocedasticidad.

- e) Identificar los posibles datos discordantes (*outliers*) y evaluar el impacto potencial que puedan ejercer en análisis estadísticos posteriores.
- f) Evaluar, si fuera necesario, el impacto potencial que pueden tener los datos ausentes (*missing*) sobre la representatividad de los datos analizados.

Para describir la distribución de una variable se empieza con un gráfico. Los diagramas de barras y los diagramas de sectores describen la distribución de variables cualitativas. Los histogramas, los diagramas de tallo y hojas además del diagrama de cajas representan gráficamente las distribuciones de variables cuantitativas.

### *3.2.1. Diagrama de barras*

Es una representación gráfica adecuada para variables cualitativas.

En el eje de abscisas (X) se dispone los valores de la variable, siendo la amplitud de la barra arbitraria pero la misma para todas ellas. El espacio entre las barras también es arbitrario pero ha de ser constante entre ellas. La altura de cada barra se refleja en el eje de las ordenadas (Y) y es proporcional a la frecuencia (absoluta o relativa) que hay en cada uno de los valores de las variables que se ha representado en el eje de las abscisas. (Icart, et al, 2012)

Para variables cuantitativas discretas se utiliza un gráfico muy parecido denominado diagrama de líneas.

### *3.2.2. Diagrama de sectores o circular*

Es una representación en forma de círculo en las que éste se divide en secciones cuya superficie es proporcional a la frecuencia de la modalidad correspondiente. El ángulo total del círculo, representa el número total de observaciones, y para determinar el ángulo de los sectores de cada modalidad se multiplica la frecuencia relativa por 360, que es el número de grados de una circunferencia. (Garriga, et al, 2010)

### *3.2.3. Histograma*

Es una gráfica de barras para datos numéricos agrupados en los que las frecuencias o porcentajes de cada grupo de datos numéricos están representados por barras individuales. La variable que nos interesa se despliega a lo largo del eje horizontal (X); el eje vertical (Y) representa la frecuencia o el porcentaje de los valores por intervalo de clase. (Levine, 1996)

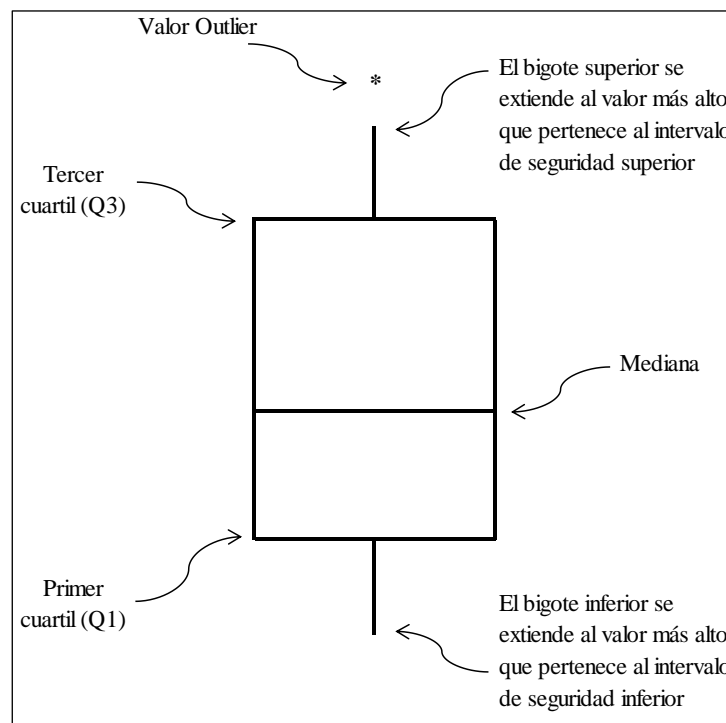
### 3.2.4. Diagrama de tallo y hojas

El diagrama de tallo y hojas es un tipo de representación gráfica propuesta por el estadístico y científico norteamericano J.W. Turkey. Este tipo de diagrama permite obtener simultáneamente tanto la distribución de frecuencias como su representación gráfica y tiene la ventaja de no perder la visualización de los datos. El diagrama está formado por una parte que es el tallo y otra parte que son las hojas. Los datos aparecen representados en una tabla de dos columnas. En la columna de la derecha (hojas) se representa las unidades de los valores, mientras que la columna de la izquierda (tallo) se representa las decenas, centenas de los datos. (Sarabia & Pascual, 2005)

### 3.2.5. Diagrama de cajas (Box plot)

Es una técnica gráfica que nos permite determinar: los valores extremos (*outliers*), valores atípicos o datos discordantes, la asimetría de la distribución que corresponde al 50% central de los datos y la variabilidad del 50% central de dos o más conjuntos de datos. (Miranda & Salinas, 2012)

**Figura 1: Diagrama de cajas**



Fuente: (Miranda & Salinas, 2012)



La caja contiene información del 50% de los datos ya que la línea inferior de la caja es el primer cuartil ( $Q_1=P_{25}$ ), y la línea superior de la caja es el tercer cuartil ( $Q_3=P_{75}$ ). La línea que divide a la caja es la mediana y su posición dentro de ella nos brindará información sobre la asimetría de la distribución del 50 % central de datos. Los bigotes son las líneas que se extienden de la parte inferior y superior de la caja hasta la observación más alta o más bajas dentro de los intervalos de seguridad. Estos intervalos son zonas donde teóricamente la variabilidad de los datos que forman parte de ella es similar a la que corresponde al 50% central de datos.

Intervalo de Seguridad Inferior (ISI):  $Q_1-1.5(Q_3-Q_1)$

Intervalo de Seguridad Superior (ISS):  $Q_3+1.5(Q_3-Q_1)$

Los *outliers* son observaciones fuera de los límites inferiores y/o superiores. (Miranda & Salinas, 2012)

### 3.2.6. Detección de outliers multivariantes

Una estadística que se usa para identificar *outliers* multivariantes es la distancia de Mahalanobis, dicha distancia se utiliza como la medida entre el centroide de cada caso al conjunto de los datos, donde el centroide es la intersección de las medias de todas las variables en el espacio multidimensional. (Weiner, et al, 2013)

Para datos multivariantes distribuidos normalmente, los valores de la distancia de Mahalanobis tienen aproximadamente una distribución Chi-cuadrado con tantos grados de libertad como variables consideradas. En consecuencia, aquellas observaciones con una distancia de Mahalanobis grande se indican como valores *outliers*. (Muñoz & Amón, 2013)

$$MSD_i = \sqrt{(x_i - \bar{x})^T - S_n^{-1}(x_i - \bar{x})}$$

Donde:

- El súper índice  $T$ , denota la matriz transpuesta
- $\bar{x}$  expresa la media del vector muestral
- $S_n$  la matriz de covarianza muestral:  $S_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)^T$

Los efectos de enmascaramiento y empantanamiento juegan un rol importante en la adecuación de la distancia de Mahalanobis como criterio para la detección de valores *outliers*. Es decir, los efectos de enmascaramiento podrían disminuir la distancia de Mahalanobis de un valor *outliers*. Esto puede ocurrir, por ejemplo, cuando un pequeño grupo de *outliers* atrae a  $x_n$  e infla  $S_n$  hacia su dirección. Por otra parte, los efectos de empantanamiento podrían aumentar la distancia de Mahalanobis de las observaciones que no son *outliers*. Por ejemplo, cuando un pequeño grupo de valores *outliers* atrae  $x_n$  e infla  $S_n$  lejos del patrón de la mayoría de las observaciones. (Quaglino & Merello, 2012)

Los problemas de enmascaramiento y empantanamiento pueden resolverse usando estimaciones robustas, como el estimador  $M$  multivariado, el estimador  $S$  bicuadrático, el estimador de covarianza de mínimo determinante ( $MCD$ ), entre otros, los cuales por definición son menos afectados por *outliers*, siendo menos probable que influyeran los parámetros usados en la  $MSD$ . (Muñoz & Amón, 2013)

### **3.3. Análisis de Conglomerados**

(McDaniel & Gates, 2005); señalan que: “El término Análisis de Conglomerados generalmente se refiere a los procedimientos estadísticos utilizados para identificar objetos o personas que son similares en lo que concierne a ciertas variables o mediciones. El propósito del Análisis de Conglomerados es clasificar objetos o personas en cierto número de conglomerados que se excluyen mutuamente y son exhaustivos, de manera que los que están dentro del conglomerado sean tan similares entre sí como sea posible. En otras palabras, los conglomerados deben ser homogéneos internamente (dentro del conglomerado) y heterogéneos externamente (fuera de los conglomerados).”

(Pérez, 2011) Proporciona la siguiente clasificación de los métodos de análisis de conglomerados basada en los algoritmos de agrupación de individuos:

- **Métodos Aglomerativos-Divisivos:** el método aglomerativo es aquel que considera tantos grupos como elementos y sucesivamente va agrupando los dos grupos más similares, hasta llegar a una clasificación determinada; mientras que el método divisivo parte de un solo grupo formado por todos los individuos, y en

cada etapa posterior va apartando individuos de los grupos establecidos anteriormente, formando nuevos grupos.

- Métodos Jerárquicos-No Jerárquicos: el método jerárquico consiste en una secuencia de  $g+1$  cluster:  $G_0, \dots, G_g$  en la que  $G_0$  es la participación disjunta de todos los elementos y  $G_g$  es el conjunto partición. Progresivamente, el número de partes de cada una de las particiones disminuye, lo que hace que estas sean cada vez más amplias y menos homogéneas. Por el contrario, en el método no jerárquico se forman grupos homogéneos sin establecer relaciones de orden o jerárquicas entre dichos grupos. En este método se conoce de antemano el número de conglomerados que se desea formar.
- Métodos Solapados-Exclusivos: el método solapado acepta que un elemento pueda pertenecer a dos grupos simultáneamente en alguna de las etapas de clasificación. El método exclusivo si prohíbe a los elementos pertenecer simultáneamente a dos grupos en la misma etapa.
- Métodos Secuenciales-Simultáneos: en el método secuencial se aplica el mismo algoritmo en forma recursiva a cada grupo, mientras que los métodos simultáneos son aquellos en los que la clasificación se logra por una simple y no reiterada operación sobre los elementos.
- Métodos Monotéticos-Politéticos: el método monotético está basado en una característica única de los objetos a clasificar; mientras que el politético no exige que todos los objetos posean las mismas características, aunque sí las suficientes como para poder justificar la analogía entre los miembros de una misma clase.
- Métodos Directos-Iterativos: el método directo utiliza algoritmos en los que una vez establecido un elemento a un grupo ya no se saca del mismo, mientras que los métodos iterativos comprueban en cada iteración si la asignación de un elemento a un conglomerado es óptima llevando a cabo un nuevo reagrupamiento de los elementos si es necesario.

- Métodos Ponderados-No Ponderados: el método no ponderado establece el mismo peso a todas las características (o variables) de los elementos a clasificar; mientras que el ponderado hace recaer mayor peso en determinadas características.
- Métodos Adaptativos-No Adaptativos: en el método no adaptativo, el algoritmo utilizado se dirige hacia una solución en la que el procedimiento de formación de conglomerados es fijo y está predeterminado, mientras que el adaptativo es aquel que de alguna manera aprende durante el proceso de formación de los grupos y cambia el criterio de optimización o la medida de similitud a utilizar.

Los métodos de análisis de conglomerados que más se usan son los que son a la vez secuenciales, aglomerativos, jerárquicos y exclusivos, y que reciben el acrónimo, en lengua inglesa, de S.A.H.N. (Sequential, Agglomerative, Hierarchic y Nonoverlapping). Entre los diferentes métodos de análisis de conglomerados de tipo S.A.H.N. tenemos los siguientes:

- Método de Unión Simple, entorno o vecino más cercano o método del mínimo.
- Método de la distancia máxima, entorno o vecino más lejano o método del máximo.
- Método de la media o de la distancia promedio no ponderado.
- Método de la media ponderada o de la distancia promedio ponderado.
- Método de la mediana o de la distancia mediana.
- Método del centroide o de la distancia prototipo.
- Método de Ward o de mínima varianza.

### **3.4. Análisis de Conglomerados Bietápico**

El Análisis de Conglomerados Bietápico tiene este nombre porque es un proceso que se ejecuta en 2 etapas. Esta técnica nos permite descubrir grupos o conglomerados de un conjunto de datos, que de otra manera no sería posible determinar. La particularidad de este algoritmo es que nos permite trabajar con variables categóricas y continuas (asumiendo independencia entre ellas), nos da la posibilidad de seleccionar automáticamente el número de conglomerados y nos da la facilidad de trabajar con conjuntos de datos grandes.

### 3.4.1. Consideraciones del conjunto de datos

- Como se mencionó anteriormente, este algoritmo trabaja con variables cuantitativas y variables categóricas.
- El orden de los datos es un factor a tomar en cuenta, el resultado final de la técnica puede depender del orden de los casos. Para reducir este impacto, se debe ordenar aleatoriamente los casos.
- Se asume que las variables continuas tienen distribución normal y las variables categóricas tienen distribución multinomial, así mismo se asume que las variables del modelo son independientes. Sin embargo, el análisis de conglomerados bietápico es bastante robusto frente a las violaciones del supuesto de las distribuciones y de independencia. (IBM\_SPSS\_Statistics\_22)

### 3.4.2. Medida de la distancia para variables cualitativas y cuantitativas

La medida de la distancia sirve para calcular similitudes entre 2 conglomerados. Debido a que la técnica trabaja en base a similitudes, es necesario encontrar una medida de distancia que nos permita juntar el efecto de las variables cualitativas y cuantitativas. Según un estudio de (Ming-Yi, 2010), el algoritmo *k-prototype* es uno de los métodos más famosos y usados para medir la distancia entre atributos mixtos. Este algoritmo hereda las ideas de *k-medias*. Para los atributos numéricos utiliza la distancia euclidiana y para los datos categóricos aplica la distancia binaria, siendo este último una desventaja para esta medida.

A continuación presentamos otra medida para atributos mixtos, el cual se usará en el presente estudio.

- Distancia Log-Verosimilitud: Esta medida de distancia consiste en la disminución del logaritmo de verosimilitud como resultado de asociarlos a dos grupos conjuntamente. La distancia entre 2 cluster  $i$  y  $j$  se define como:

$$d(i, j) = \xi_i + \xi_j - \xi_{\langle i, j \rangle}$$

Donde:

$$\xi_v = -N_v \left\{ \sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right\}$$

Para  $v = i, j$

$$\hat{E}_{vk} = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v}$$

Las notaciones son:

$N_v$ : Número de datos registrados en el cluster  $v$ .

$K^A$ : Número de variables continuas usadas en el procedimiento.

$K^B$ : Número de variables categóricas usadas en el procedimiento.

$\hat{\sigma}_k^2$ : Varianza estimada de la  $k$ -ésima variable continua.

$\hat{\sigma}_{vk}^2$ : Varianza estimada de la  $k$ -ésima variable continua en el cluster  $v$ .

$L_k$ : Número de categorías para la  $k$ -ésima variable cualitativa.

$N_{vkl}$ : Número de datos registrados en el cluster  $v$  cuya  $k$ -ésima variable cualitativa forma parte de la  $l$ -ésima categoría.

Los conceptos y notaciones de distancia Log-Verosimilitud fueron tomados de (Salinas, 2010) e (IBM\_SPSS\_Statistics\_22).

Como paso previo a la descripción del algoritmo, definiremos 2 conceptos claves para su entendimiento.

Los siguientes conceptos están basados en la investigación de (Zhang, et al, 1996) del algoritmo BIRCH.

### 3.4.3. Características de Conglomerados o CF

Tenemos un conjunto de datos de  $N$  casos con  $d$  dimensiones en un cluster definido como  $\{\vec{X}_i\}$  donde  $i = 1, 2, \dots, N$ . El CF representa un vector compuesto con las siguientes medidas:

$$CF = (N, \overrightarrow{LS}, SS)$$

Donde  $N$  es el número de casos en un cluster,  $\overrightarrow{LS}$  es la suma lineal de los  $N$  casos en un cluster y  $SS$  es la suma de cuadrados de los  $N$  casos en un cluster.

Una propiedad del CF es su aditividad, esto quiere decir que se cumple lo siguiente:

Dado  $CF_1 = (N_1, \overrightarrow{LS}_1, SS_1)$  y  $CF_2 = (N_2, \overrightarrow{LS}_2, SS_2)$ , se cumple que  $CF_1 + CF_2 = (N_1 + N_2, \overrightarrow{LS}_1 + \overrightarrow{LS}_2, SS_1 + SS_2)$ .

De la definición del CF y su propiedad de aditividad se desprende que los vectores CF pueden almacenar estadísticas de resumen de cada grupo, con lo cual este método nos sirve para poder trabajar con grandes conjuntos de datos, ya que no asocia o agrupa caso por caso, sino las medidas estadísticas de resumen  $(N, \overrightarrow{LS}, SS)$  de cada grupo. Con esto tenemos suficiente información para poder realizar decisiones de agrupamiento usando el algoritmo *BIRCH*.

### 3.4.4. CF Tree o Árbol de características de conglomerados

El CF Tree se construye con la información proporcionada del CF. Es un árbol balanceado que depende de 2 parámetros:  $B$  que es un factor de ramificación y un umbral  $T$ .

Inicialmente se tiene un *nodo sin hojas* con la siguiente estructura:  $[CF_i, \text{hijo}_i]$ , donde  $i = 1, 2, \dots, B$ ; además  $\text{hijo}_i$  es un indicador para el  $i$ -ésimo nodo hijo y  $CF_i$  es el CF del subgrupo representado por este  $\text{hijo}_i$ . Por lo tanto un *nodo sin hojas* representa un cluster hecho de todos los representados por sus entradas.

Un *nodo hoja* está compuesto de  $L$  entradas de la siguiente forma  $[CF_i]$ , donde  $i = 1, 2, \dots, L$ . Cada nodo hoja tiene 2 indicadores o punteros llamados “prev” y “next” que son usados para juntar todos los nodos hojas. Un *nodo hoja* representa un cluster hecho de todos los representados por sus entradas. Pero las entradas de los nodos hojas deben satisfacer el parámetro umbral  $T$ .

Cabe agregar que el tamaño del árbol es una función  $T$ ; cuando más grande es el umbral, menor será el tamaño del árbol.

Cada nodo del CF tree indica el número de grupos que pertenecen a esa ramificación y sus respectivos CF.

#### 3.4.5. Algoritmo de conglomerado en 2 etapas

Como su nombre lo indica, el análisis cluster bietápico tiene este nombre debido a que el algoritmo empleado en su ejecución tiene 2 pasos:

- Paso 1: Esta parte es conocida como etapa de pre-agrupamiento, ya que consiste en la construcción de pre-conglomerados. Este proceso comienza con la construcción del CF Tree, el cual coloca la primera observación en un *nodo hoja* que contiene información sobre la misma. A continuación se añade cada caso a un *nodo hoja* existente o se forma uno nuevo, si la medida de similitud (distancia log verosimilitud) es cercana a un nodo hoja que ya existe se le asocia a él, de lo contrario se crea un nuevo nodo hoja.

Una vez terminado el proceso anterior tendremos construidos los pre-cluster (a través del CF tree), los cuales contienen almacenadas las estadísticas de resumen  $CF = (N, \vec{LS}, SS)$ .

- Paso 2: Una vez construido el CF tree en el paso anterior, los *nodos hojas* se agrupan utilizando un algoritmo aglomerativo de cluster. Este algoritmo aglomerativo nos da una gama de soluciones o modelos. Para determinar la “mejor” solución se puede comparar las diferentes soluciones con el Criterio Bayesiano de Schwarz (BIC) o el Criterio de Información de Akaike (AIC).

El criterio que se utilizó en el presente estudio fue el BIC, el cual nos permitirá determinar automáticamente el número óptimo de conglomerados.

La relación es la siguiente:

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N)$$



Donde:

$$m_J = J \left\{ 2K^A + \sum_{k=1}^{K^B} (L_k - 1) \right\}$$

$N$ : Número total de datos.

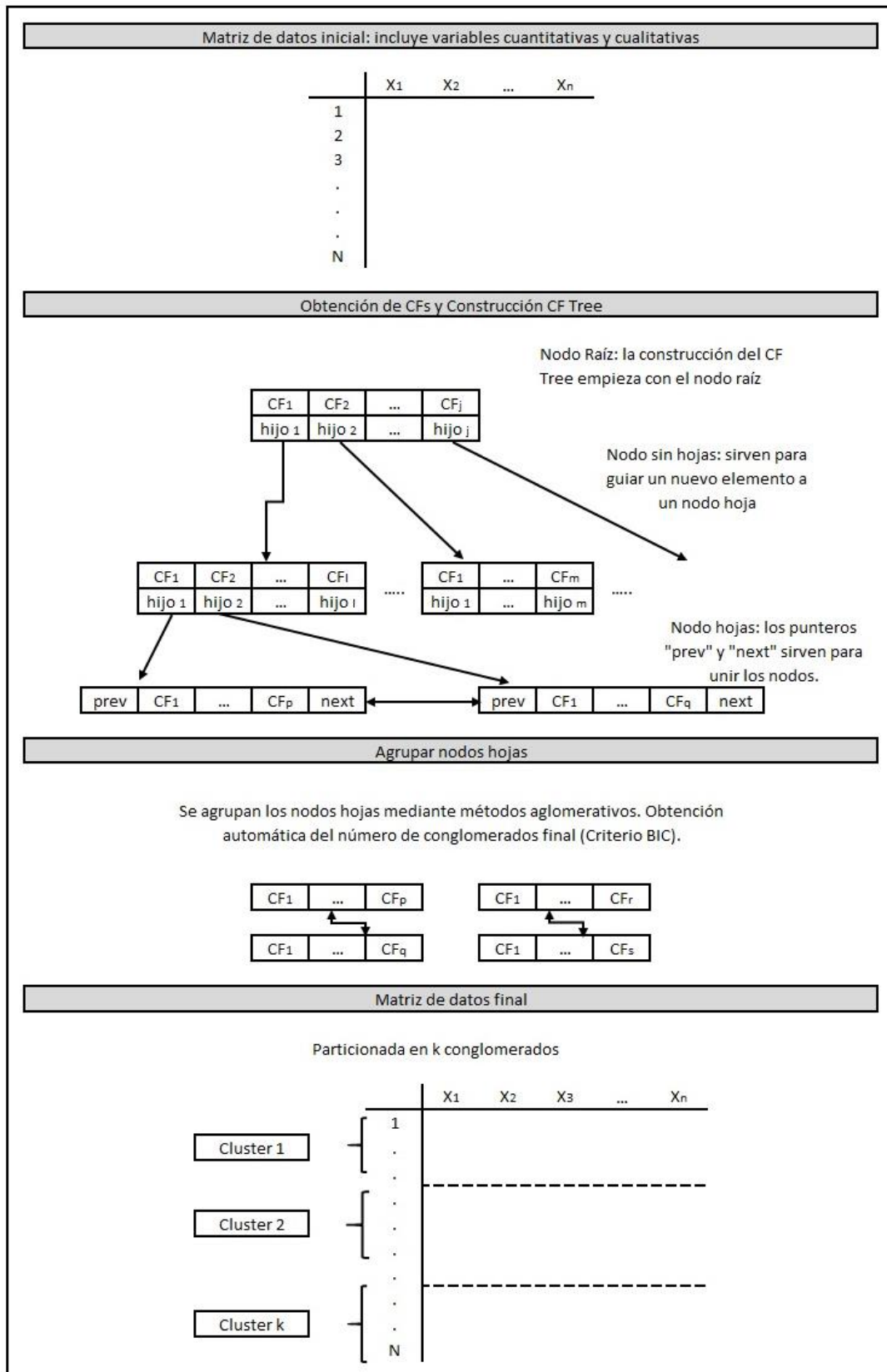
$K^A$ : Número de variables continuas usadas en el procedimiento.

$K^B$ : Número de variables categóricas usadas en el procedimiento.

$L_k$ : Número de categorías para la  $k$ -ésima variable cualitativa.

Sin embargo, hay que considerar que el procedimiento para el análisis de conglomerados es bastante robusto frente a las violaciones tanto del supuesto de independencia como de las distribuciones (Universidad\_de\_Granada).

**Figura 2: Esquema gráfico del Conglomerado Bietápico**



Fuente: Elaboración propia

### **3.5. Estrategias de marketing mix**

Uno de los objetivos de segmentar una cartera de clientes, es aplicarles diferentes tipos de estrategias de marketing mix a cada grupo encontrado, de modo que se le puede dar un valor al proceso de segmentación e incrementar el valor de las ganancias de la empresa.

De acuerdo a las características y perfiles encontrados para cada segmento, se le puede asignar una estrategia a fin de obtener la mayor rentabilidad posible.

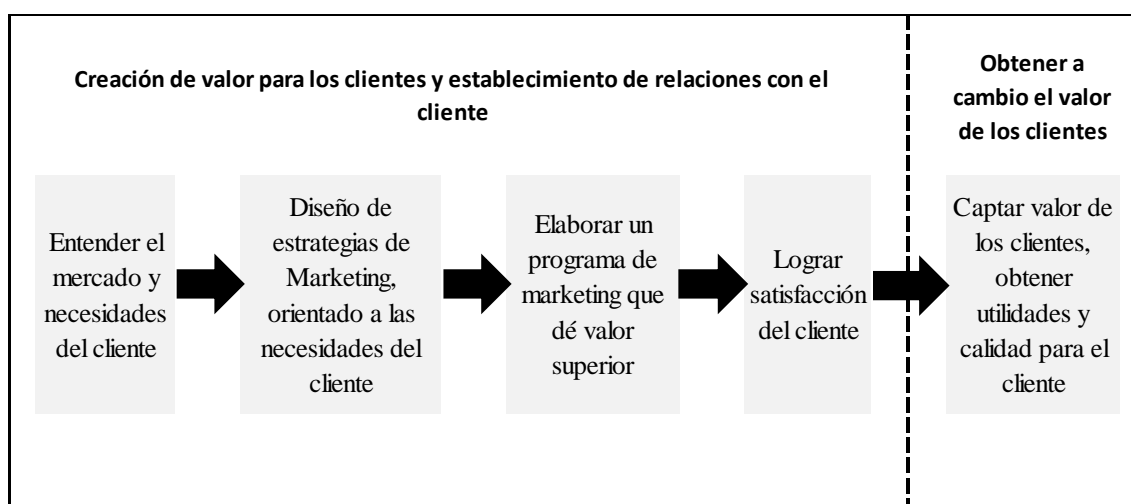
(Kotler & Armstrong, 2007); señalan que el marketing tiene doble meta: por un lado, atraer nuevos clientes prometiéndoles un valor superior y, por otro, mantener y hacer crecer a los clientes actuales al entregarles satisfacción.

El marketing se aplica en todo ámbito, para empresas o compañías con fines de lucro y también lo aplican organizaciones no lucrativas.

(Kotler & Armstrong, 2007); definen al marketing como el sentido de satisfacer las necesidades del cliente. Si el gerente de marketing entiende bien las necesidades del consumidor; si desarrolla productos que ofrezcan un valor superior; y si fija sus precios, distribuye y promueve de manera eficaz, sus productos se venderán con mucha facilidad. De manera que las ventas y la publicidad son únicamente parte de una “mezcla de marketing” mayor. Es decir, del conjunto de herramientas de marketing que funcionan en conjunto para satisfacer las necesidades del cliente y para establecer relaciones con éste.

La figura 2 presenta un modelo sencillo de los cinco pasos del proceso de marketing. En los primeros cuatro pasos, las compañías trabajan para entender a los consumidores, crear valor para el cliente y establecer relaciones firmes con éste. En el último paso, las compañías obtienen las recompensas de crear un valor superior en el cliente. Al crear valor para los consumidores, ellas, a la vez, captan el valor de los consumidores en forma de ventas, ingresos y utilidades a largo plazo del cliente.

**Figura 3: Modelo básico del proceso de marketing**



*Fuente: (Kotler & Armstrong, 2007)*

Una vez entendido los conceptos básicos y el proceso de marketing, se puede hablar de las estrategias de marketing. Se define una estrategia de marketing como el arte y la ciencia de elegir mercados meta y diseñar relaciones beneficiosas con ellos. La meta del gerente de marketing es encontrar, atraer, mantener y cultivar clientes meta mediante la creación, la entrega y la comunicación de valor superior para el cliente. (Kotler & Armstrong, 2007).

El marketing busca diseñar estrategias que construyan relaciones redituables con sus consumidores meta. Según (Kotler & Armstrong, 2007); hay cinco conceptos alternativos que las organizaciones podrían utilizar para diseñar y poner en práctica como estrategias de marketing: la producción; el producto; las ventas; el marketing y el marketing social.

- La producción: Señala que los consumidores favorecen los productos que están disponibles y que son altamente costeables. Por lo tanto, la organización debería concentrarse en mejorar la eficacia de la producción y la distribución.
- El producto: Establece que los consumidores favorecen los productos que ofrecen la calidad, el desempeño y las mejores características innovadoras. Desde este concepto, la estrategia de marketing se enfoca en mejorar el producto de forma continua.

- Las ventas: Señala que los consumidores no comprarán el número suficiente de productos de la empresa a menos que ésta realice un esfuerzo de promoción y ventas a gran escala.
- El marketing: Establece que el logro de las metas organizacionales depende de conocer las necesidades y los deseos de los mercados meta, así como de proporcionar las satisfacciones deseadas, mejor que la de los competidores.
- El marketing social: Establece que una compañía debería tomar buenas decisiones de marketing, considerando los deseos de los consumidores, los requerimientos de la compañía, así como los intereses a largo plazo de los consumidores y de la sociedad.

Para que una estrategia de segmentación resulte exitosa en su aplicación debe serlo en todos los elementos de su marketing mix. La estrategia de posicionamiento del segmento puede que incluya las estrategias de producto y precio. Una estrategia completa de marketing mix debe incluir también las estrategias de comunicación y de lugar (ventas y distribución). Si los clientes no conocen de una forma adecuada la proposición de valor o no pueden adquirir el producto en los lugares que a ellos les gustaría, la ejecución de la estrategia de segmentación será un fracaso.

A continuación revisaremos las definiciones de estrategias para producto, precio, distribución y promoción.

- Producto: El producto es un conjunto de atributos tangibles e intangibles que abarcan empaque, color, precio, calidad y marca, además del servicio y la reputación del vendedor; el producto puede ser un bien, un servicio, un lugar, una persona o una idea. Los clientes compran mucho más que un conjunto de atributos cuando adquieren un producto: compran satisfacción en la forma de los beneficios que esperan recibir del producto.

En muchas ocasiones una empresa se ve en la necesidad de crear un nuevo producto, además es necesario reconocer tres categorías de estos: innovadores, significativamente diferentes e imitadores. Al decidir si se agrega o no un nuevo producto, un productor o un intermediario deben considerar si hay adecuada demanda de mercado. El producto debe asimismo ajustarse a los recursos de marketing, producción y financieros de la compañía. (Stanton & Walker, 2007).

- **Precio:** Es un factor significativo para lograr el éxito en marketing; y en muchas situaciones de compra puede ser de gran importancia para los consumidores. Antes de poner el precio base de un producto, la administración debe identificar su objetivo de asignación de precios. Los objetivos principales de la asignación de precios son: 1) percibir una retribución meta sobre la inversión o en ventas netas, 2) maximizar utilidades (ganancias), 3) incrementar ventas, 4) mantener o captar una participación meta del mercado, 5) estabilizar los precios y 6) hacer frente a los precios de la competencia. La asignación de precios a los niveles prevalecientes en el mercado tiene sentido para las empresas que venden productos muy conocidos y estandarizados, y en ocasiones para las compañías individuales en un oligopolio. Dos variantes de la asignación de precios al nivel del mercado son la puesta del precio por encima o debajo de los niveles de los competidores principales. (Stanton, et al, 2007)
- **Distribución:** La función de la distribución es hacer que un producto llegue a su mercado meta. Un canal de distribución realiza esta asignación, en el que los intermediarios hacen algunas de las tareas. Un intermediario es una empresa de negocios que presta servicios directamente relacionados con la compra o venta de un producto conforme éste pasa del productor al consumidor. Se puede eliminar a los intermediarios de un canal, pero alguna organización o individuo tiene que realizar las funciones esenciales de los intermediarios. Un canal de distribución es el conjunto de las personas y compañías que participan en el paso de los derechos de propiedad del producto a medida que éste pasa del productor al consumidor último o al usuario de negocios. Un canal incluye al productor, cliente final y a los intermediarios que participen en el proceso. El diseño de un canal de distribución para un producto sigue una secuencia de cuatro decisiones: 1) delinear la función de la distribución dentro de la mezcla de marketing; 2) seleccionar el tipo apropiado de canal de distribución; 3) determinar la intensidad idónea de la distribución, y 4) elegir a los miembros específicos del canal. Es necesario considerar numerosos factores al seleccionar un canal de distribución. La consideración principal es la naturaleza del mercado meta. Otras se relacionan con el producto, los intermediarios y la compañía misma. (Stanton, et al, 2007).

- **Promoción:** Es el cuarto componente de la mezcla de marketing total de una compañía. En términos de marketing, significa informar, persuadir y servir de recordatorio entre los clientes actuales o prospectos. Los métodos primarios de promoción son: las ventas personales, la publicidad, la promoción de ventas y las relaciones públicas. La promoción es comunicación. En lo fundamental, el proceso de comunicación consiste en enviar un mensaje por un canal a un receptor. El éxito de la comunicación depende de lo bien que se codifique el mensaje, con qué facilidad y claridad se pueda decodificar, y si cualquier “ruido” (interferencia) obstaculiza su transmisión. La retroalimentación, que es la respuesta creada por un mensaje, es una medida de la eficacia de una comunicación. (Stanton, et al, 2007)

### **3.6. Aplicaciones de análisis de conglomerados en segmentación de mercados**

En este punto se presenta los diferentes trabajos de investigación que han realizado el proceso de segmentación de mercados y han utilizado el análisis de conglomerados como herramienta para dicho objetivo.

Por ejemplo, el estudio publicado por la Revista Española de Investigación de Marketing llamado “El comportamiento de cambio de canal del telespectador en Prime Time: Análisis a través de un panel de audimetría”. El cual consiste en estudiar el comportamiento en el cambio de canal de un televidente. Para ello estudiaron las principales variables que explican dicho comportamiento, luego se aplicó un Análisis Conglomerados Bietápico para conocer las características de los individuos más propensos a cambiar de canal. Los resultados obtenidos sugieren que las variables más influyentes en el comportamiento del cambio de canal (*zapping*) del individuo son el número de canales visionados que emiten publicidad en la franja analizada y las variables sociodemográficas relacionadas con el individuo y el tamaño del hogar. (Gázquez, et al, 2010)

Otro estudio basado en Análisis de Conglomerados es “La incidencia de los estilos de vida en la segmentación de mercado turístico: Aplicación a la población urbana de Castilla y León”, el cual tiene como objetivo determinar diferentes grupos de turistas residentes en las poblaciones urbanas de Castilla y León en España, mediante criterios de segmentación de mercados. El método de agrupación usado fue k-means. Esto les

permitió encontrar 5 segmentos, en los que según el estilo de vida de cada uno de ellos se pueden distinguir los siguientes comportamientos diferenciados: familiares, idealistas, autónomos, hedonistas y conservadores. (González, et al.)

También hay trabajos que demuestran el uso de diferentes programas estadísticos para el proceso de agrupación y clasificación. Tal como el caso del estudio “SPSS orientado a la gestión de mercados”. Esta investigación se encarga de mostrar la utilidad de las diferentes herramientas del programa estadístico SPSS para la realización de investigaciones de mercado, principalmente la segmentación de mercado. Abarca desde los métodos jerárquicos, no jerárquicos (k-means) y conglomerado en dos etapas. (Guevara & Cortés, 2006)

Se encontró otro proyecto que utiliza el programa de análisis SAS para el análisis de conglomerados, se llama “The Science and Art of Market Segmentation Using PROC FASTCLUS”. Como su nombre lo indica, tiene como objetivo describir un proceso para desarrollar una estrategia de segmentación de mercados multidimensional usando el procedimiento PROC FASTCLUS que está incluido en el software SAS. Para este estudio se utilizó un conjunto de datos de 60 000 negocios en California. Este trabajo comprende la captura del conjunto de datos y sus variables, estimación de conglomerados y revisión de resultados, y finalmente la determinación e interpretación de los perfiles encontrados. (Thompson)



## IV. MATERIALES Y MÉTODOS

En este capítulo se explicará brevemente los procedimientos y/o métodos que se utilizaron para la solución del problema planteado, explicando cómo se realizó la investigación y la técnica estadística a utilizar.

### 4.1. Diseño y tipo de la investigación

Según, (Sampieri, et al, 2006), el presente trabajo de investigación es no experimental, transversal del tipo descriptivo ya que tiene como propósito encontrar segmentos de los clientes de la tienda de electrodomésticos a través de las diversas variables recolectadas en un tiempo establecido para la correcta aplicación de estrategias de marketing.

### 4.2. Formulación de hipótesis

La correcta segmentación de los clientes mediante el análisis de conglomerados permitirá enfocar estrategias de marketing para cada grupo encontrado.

### 4.3. Identificación de variables

El conjunto de datos a utilizar procede de una empresa de *retail* que vende electrodomésticos para el hogar. Este conjunto almacena datos de sus clientes (personas naturales), las cuales han adquirido un artículo para su hogar mediante un crédito en el primer trimestre del año 2013. El conjunto de datos cuenta con 6284 registros de clientes, es decir se trabajó inicialmente con toda esta población.

El conjunto de datos cuenta con 3 variables cuantitativas y 6 variables cualitativas. Estas 9 variables miden las características propias del cliente a quien se le otorgó el crédito por el electrodoméstico para el hogar adquirido.

**Cuadro 1: Variables cuantitativas**

Variable	Descripción
Monto capital	Valor del producto sin contar los intereses (en soles)
Edad	Edad del cliente (en años)
Ingreso mensual	Ingreso mensual del cliente (en soles)

*Fuente: Elaboración propia*

**Cuadro 2: Variables cualitativas**

<b>Variable</b>	<b>Descripción</b>	<b>Valores</b>
Género	Género del cliente	1. Femenino 2. Masculino
Estado civil	Estado civil del cliente	1. Casado/Conviviente 2. Soltero
Compra	Motivo de la compra	1. Para aprovechar una promoción / descuento 2. Por una ocasión especial (cumpleaños, matrimonio, etc.) 3. Porque tuve ingresos adicionales / aumento de sueldo
Tipo de vivienda	Tipo de vivienda del cliente	1. Alquilada 2. Familiar 3. Propia
Tipo de actividad económica	Tipo de actividad económica que desempeña el cliente	1. Dependiente 2. Independiente
Línea	Línea del producto adquirido por el cliente	1. Audio 2. Blanca 3. Video

*Fuente: Elaboración propia*

#### **4.4.Procedimiento y análisis de datos**

##### *4.4.1. Análisis exploratorio y descriptivo*

El primer paso a realizar es un análisis exploratorio y descriptivo de las variables cuantitativas y cualitativas que permitió observar el comportamiento de dichas variables.

##### *4.4.2. Limpieza de datos*

Posteriormente se realiza la limpieza del conjunto de datos, se eliminaron los datos inconsistentes como datos mal ingresados. En el caso de las variables cualitativas

aquellas que su valor era diferente a los ya mencionados, datos vacíos o nulos y datos con valores outliers.

#### *4.4.3. Obtención de conglomerados*

Se aplicó la técnica estadística de Análisis de Conglomerados Bietápico para la obtención de dichos conglomerados. El programa estadístico empleado para procesar dichos datos fue la versión *Demo -IBM SPSS Statistics 22*.

#### *4.4.4. Definición de perfiles*

Se definió un perfil para cada conglomerado encontrado en base a las características de las variables.

#### *4.4.5. Plantear estrategia de marketing*

Para cada perfil (conglomerado) se definió una estrategia de marketing.

## V. RESULTADOS Y DISCUSIÓN

### 5.1. Análisis exploratorio y descriptivo

#### 5.1.1. Análisis exploratorio y descriptivo de variables cuantitativas

El análisis de conglomerados es muy sensible a la presencia de elementos con valores extremos u *outliers*.

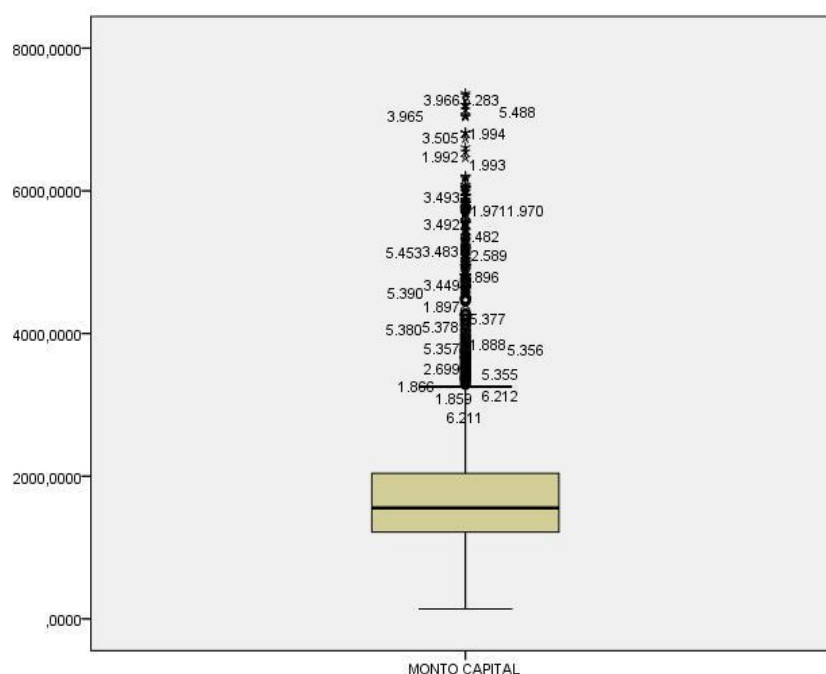
Se presenta a continuación los resultados obtenidos con el conjunto de datos inicial.

**Cuadro 3: Estadísticos descriptivos de variables cuantitativa del conjunto de datos inicial**

	N	Mínimo	Máximo	Media	Desv. típ.
<b>MONTO CAPITAL</b>	<b>6284</b>	<b>138.69</b>	<b>7366.53</b>	<b>1845.80</b>	<b>1091.75</b>
<b>EDAD</b>	<b>6284</b>	<b>18</b>	<b>71</b>	<b>35.20</b>	<b>12.414</b>
<b>INGRESOS</b>	<b>6284</b>	<b>100</b>	<b>80000</b>	<b>1274.32</b>	<b>1421.51</b>

*Fuente: Elaboración propia*

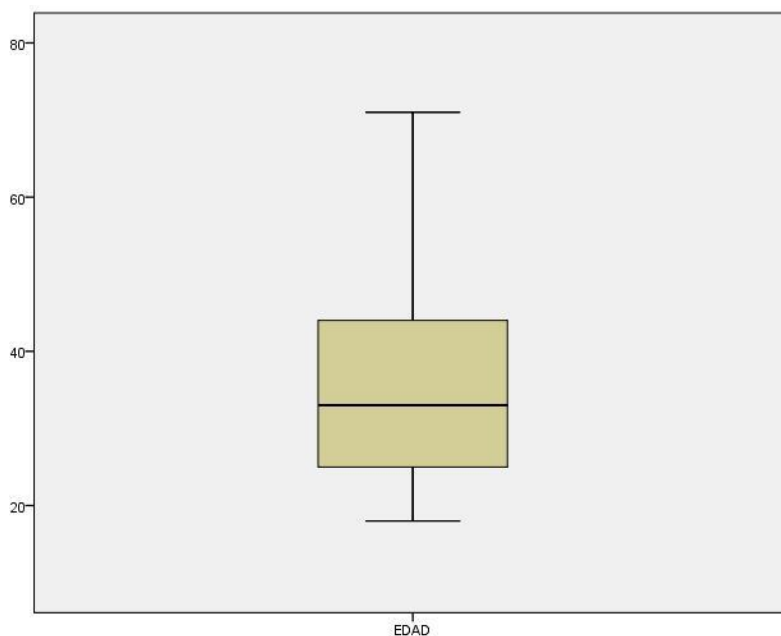
**Figura 4: Diagrama de cajas de variable cuantitativa: Monto de Capital**



*Fuente: Elaboración propia*

Del gráfico anterior se puede observar que los datos provenientes de la variable *Monto capital* presenta *outliers*, llega hasta un valor máximo de 7366.53 soles y un valor mínimo de 138.69 soles.

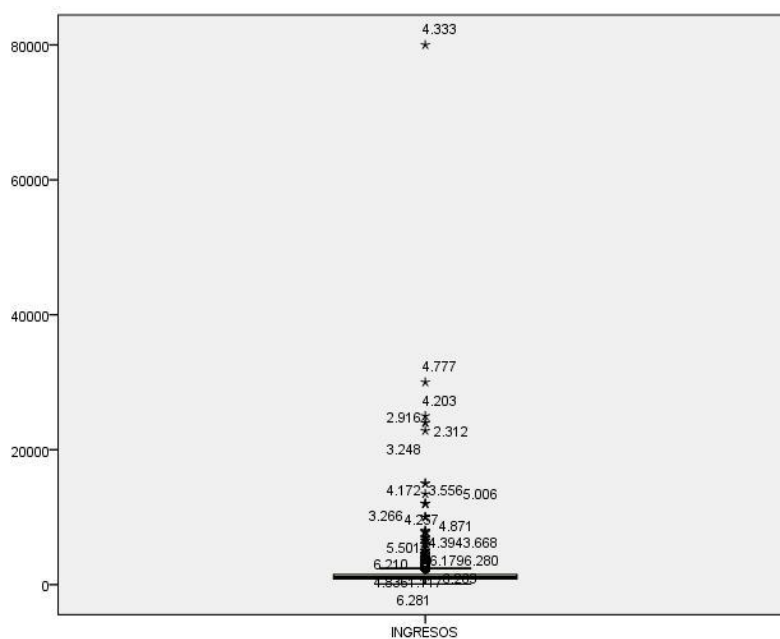
**Figura 5: Diagrama de cajas de variable cuantitativa: *Edad***



*Fuente: Elaboración propia*

Los datos de la variable *Edad* no presenta *outliers*, llega hasta un valor máximo de 71 años de edad y un valor mínimo de 18 años de edad.

**Figura 6: Diagrama de cajas de variable cuantitativa: *Ingresos***

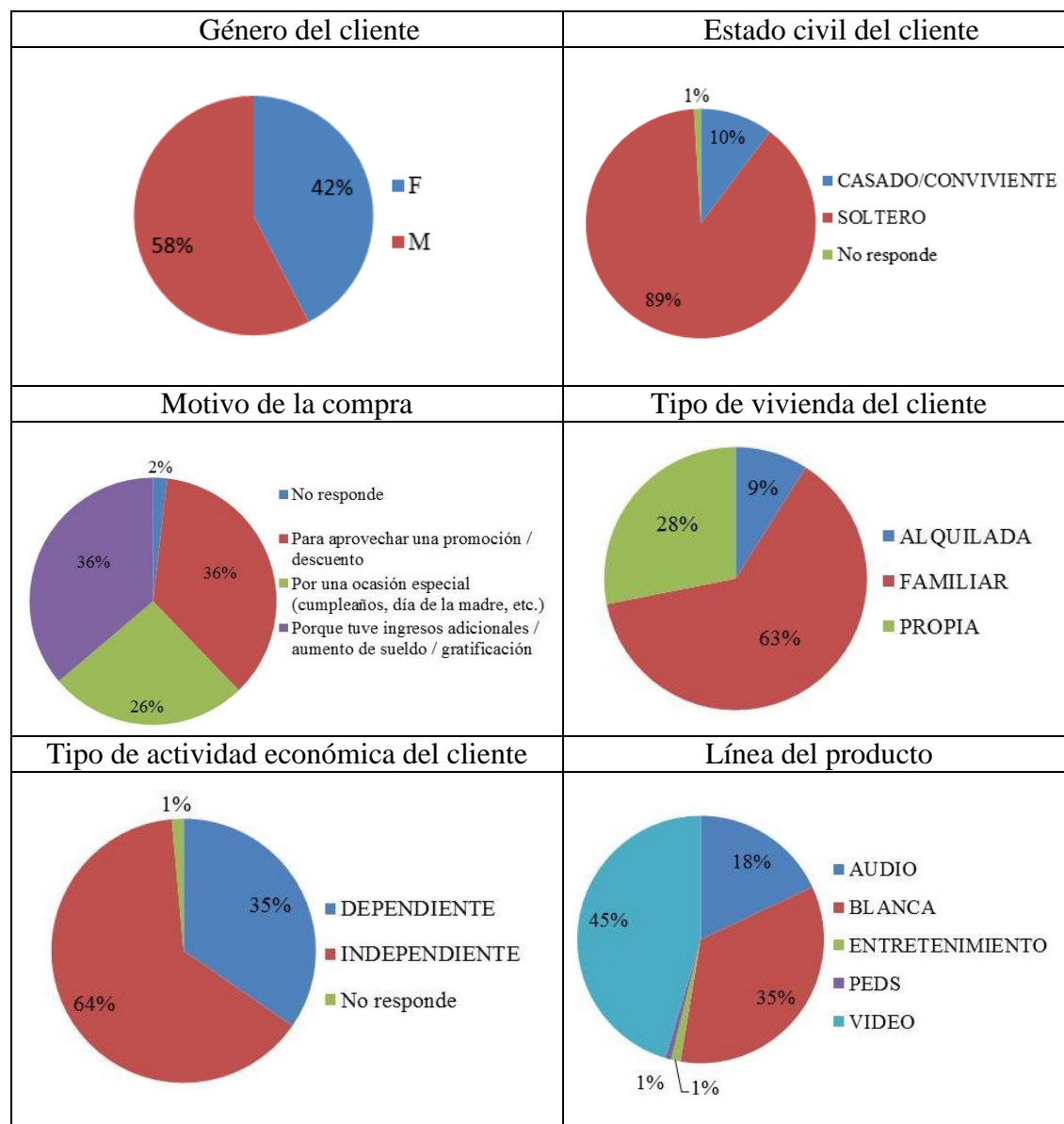


*Fuente: Elaboración propia*

Del gráfico anterior se puede observar que los datos de la variable *Ingresos* presenta *outliers*, llega hasta un valor máximo de 80 000 soles y un valor mínimo de 100 soles.

### 5.1.2. Análisis exploratorio y descriptivo de variables cualitativas

**Figura 7: Gráfico de sector de las variables cualitativas**



Fuente: Elaboración propia

Se identificó la ausencia de datos en variables cualitativas, etiquetada con **No responde**, en la variable *Estado civil* tiene un 1% de no respuestas, la variable *Motivo de compra* presenta un 2% de datos perdidos y la variable *Tipo de actividad económica* un 1%. Para el caso de la variable *Línea de producto* se encontró dos categorías adicionales Entretenimiento y PEDS cada una con 1% respectivamente.

## 5.2. Limpieza de datos

### 5.2.1. Limpieza de datos de variables cuantitativas

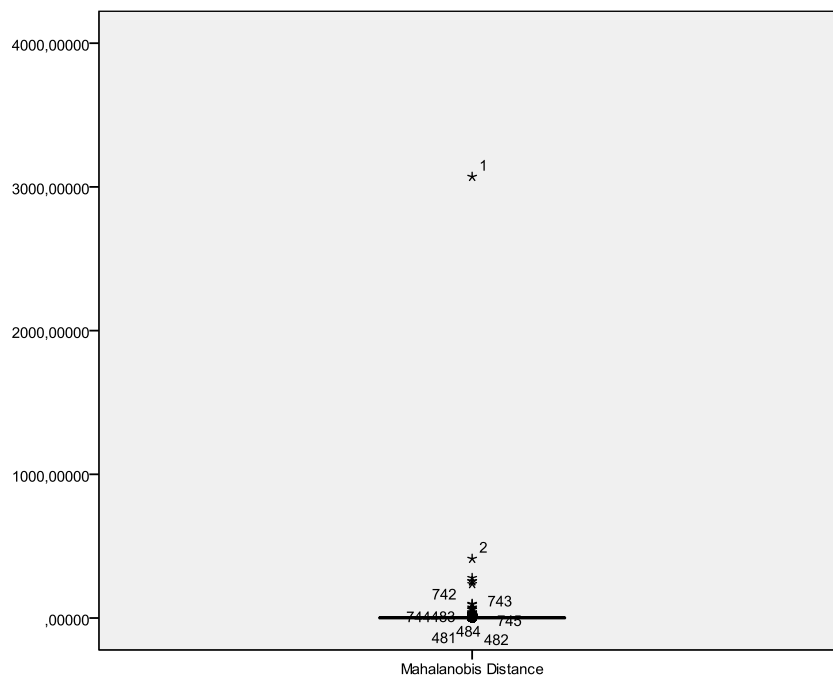
Como primer paso en la limpieza de las variables cuantitativas (Monto de capital, Edad e Ingresos) se realizó la detección de *outliers* multivariantes mediante la distancia de Mahalanobis. Se obtuvo como valor mínimo de dicha distancia 0.00253 y como valor máximo 3070.5 (Ver Cuadro 4).

**Cuadro 4: Estadísticos descriptivos de Distancia de Mahalanobis**

	N	Mínimo	Máximo	Media	Desv. típ.
Distancia de Mahalanobis	6284	0.0025	3070.5117	2.9995	39.7603

Fuente: Elaboración propia

**Figura 8: Diagrama de cajas de la Distancia de Mahalanobis con la presencia de *outliers* multivariados**

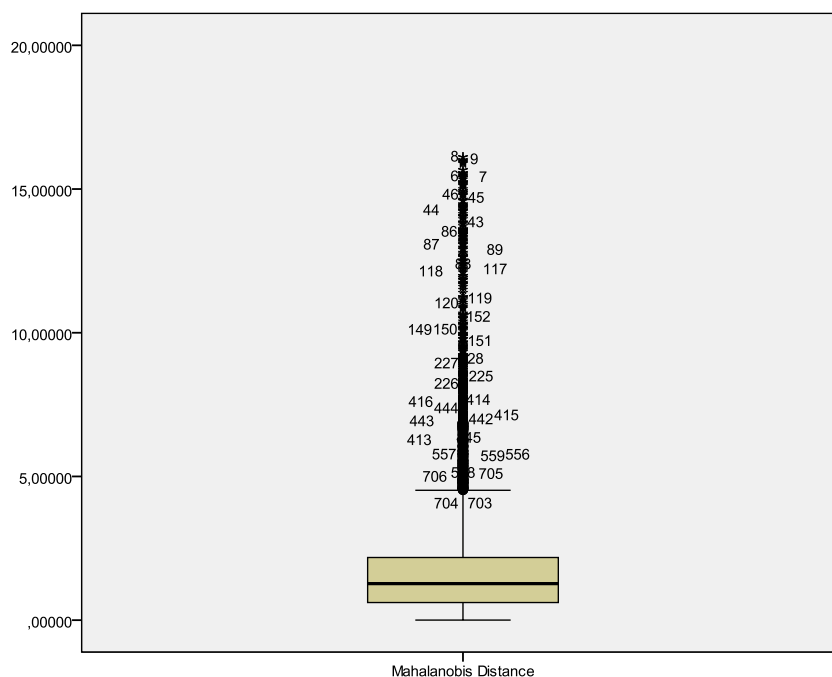


Fuente: Elaboración propia

En el diagrama de caja de la distancia de Mahalanobis se aprecia la presencia de *outliers* para identificarlos se calculó la probabilidad superior de la distribución Chi-Cuadrada con 3 grados de libertad (se analizaron 3 variables). Los *outliers* multivariados son aquellos valores cuya probabilidad sea inferior a 0.001.

El valor del umbral de 0.001 fue sugerido por (Tabachnick & Fidell , 2000), quienes afirman que es una estimación de probabilidad muy conservadora para la identificación de valores *outliers* apropiada para la Distancia de Mahalanobis. Se encontró 54 registros con *outliers* multivariados. Retirando los *outliers* multivariados se obtuvo el siguiente diagrama de cajas (Ver Figura 8).

**Figura 9: Diagrama de cajas de la Distancia de Mahalanobis sin la presencia de *outliers* multivariados**



*Fuente: Elaboración propia*

Para determinar los registros de clientes con los que finalmente se trabajó se consideró un estudio de la consultora IPSOS Perú (Perú21, 2013), el cual establece los siguientes ingresos familiares mensuales para los niveles socioeconómicos. El segmento A, con ingresos familiares mensuales de S/.10,622, en promedio; el B, con S/.5,126; el C, con S/.3,261; el D, con S/.1,992, y el E, con S/.1,027.

Debido a que la empresa de *retail* está enfocado en los sectores B, C, D y E; y considerando las reglas del negocio. Además para la limpieza de datos de las variables cuantitativas también se consideró los siguientes criterios:

- Para la variable *Monto de capital* nos quedamos con los registros que no sobrepasen los S/. 4000 Nuevos soles.
- Con la variable *Edad* se trabajó con los clientes de 18 a 60 años de edad.



- Para la variable *Ingreso* nos quedamos con valores mayor igual a S/. 750 y menor igual a S/. 5000 Nuevos soles.

Luego de la limpieza de datos de las variables cuantitativas se retiró en total 1029 registros de los cuales 54 presentan *outliers* multivariados y 975 registros que no cumplen con la regla del negocio.

#### 5.2.2. Limpieza de datos de variables cualitativas

- Para las variables cualitativas se eliminaron los registros con la etiqueta *No responde*, en la variable *Estado civil* con 1%, la variable *Motivo de compra* con un 2% y la variable *Tipo de actividad económica* con 1%.
- Para la variable *Línea de producto* se eliminó los registros de *Entretenimiento* y *PEDS* con un 2%.

Luego de la limpieza de datos de las variables cualitativas se encontró 325 registros de los cuales 50 de ellos estaban considerados dentro de los registros a eliminar por no cumplir con los criterios de las variables cuantitativas.

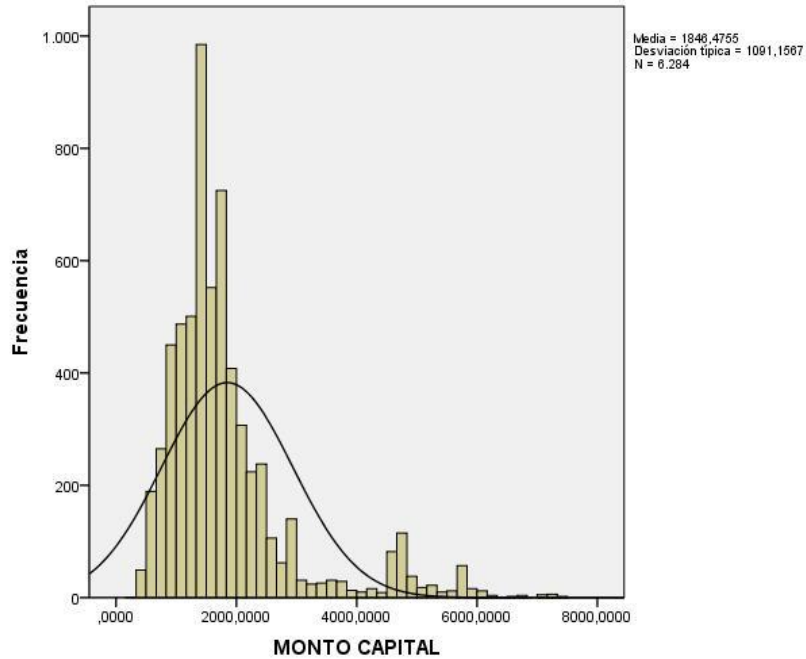
Al concluir la limpieza de datos (tanto de las variables cuantitativas y cualitativas) se retiraron en total 1304 registros. El conjunto de datos final que se trabajó para el análisis de conglomerados bietápico es de 4980 registros.

**Cuadro 5: Estadísticos descriptivos de variables cuantitativas luego de la limpieza de datos**

	Nº	Mínimo	Máximo	Media	Desv. típ.
<b>MONTO CAPITAL</b>	4980	180.69	3995.50	1623.00	644.76
<b>EDAD</b>	4980	18	60	36.56	10.46
<b>INGRESOS</b>	4980	750	5000	1263.51	532.58

Fuente: *Elaboración propia*

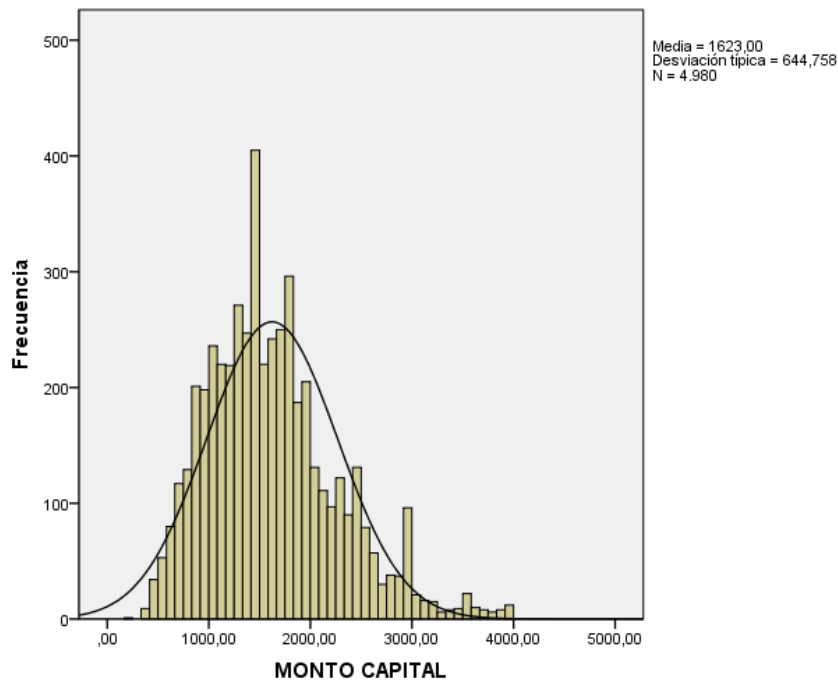
**Figura 10: Histograma de la variable *Monto capital* antes de la limpieza de datos**



*Fuente: Elaboración propia*

El conjunto de datos para la variable *Monto de capital* presenta una distribución asimétrica hacia la derecha y su pico está alrededor de los 1800 soles.

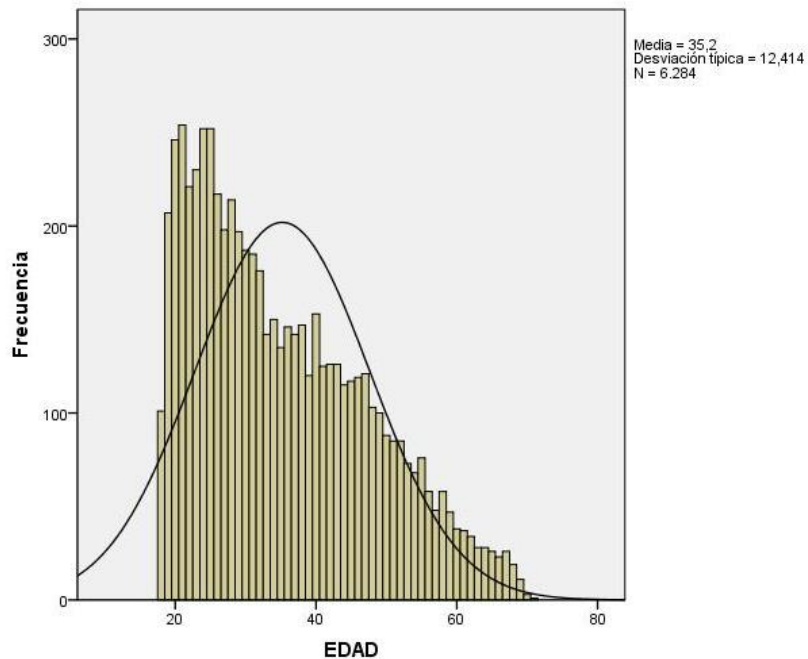
**Figura 11: Histograma de la variable *Monto capital* después de la limpieza de datos**



*Fuente: Elaboración propia*

Luego de la limpieza del conjunto de datos se observa que la asimetría presentada anteriormente disminuye.

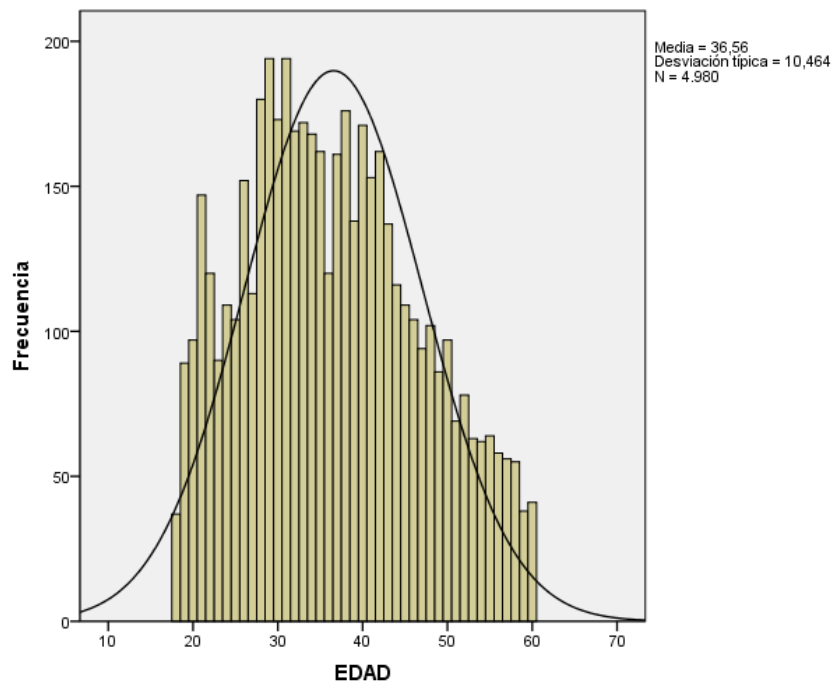
**Figura 12: Histograma de la variable *Edad* antes de la limpieza de datos**



*Fuente: Elaboración propia*

El conjunto de datos para la variable *Edad* presenta una distribución asimétrica hacia la derecha y sus valores más frecuentes se sitúan en un rango de 25 a 30 años.

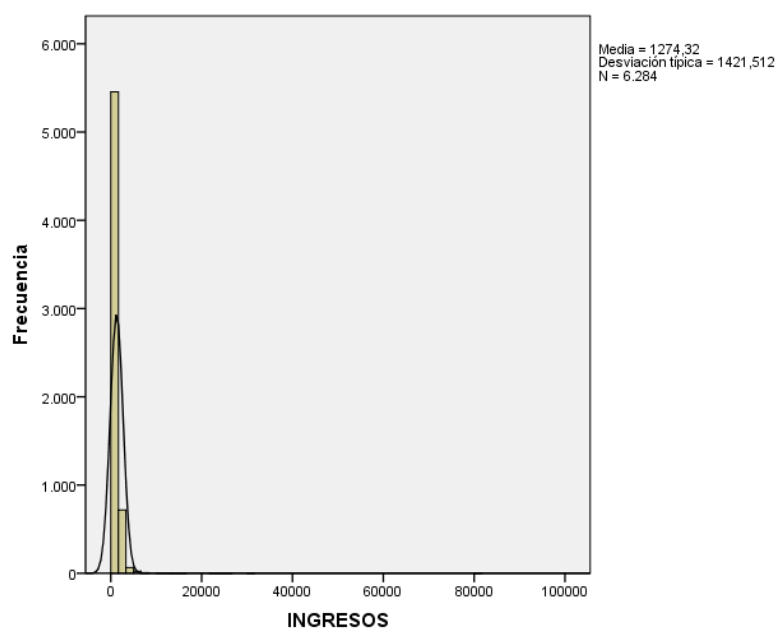
**Figura 13: Histograma de la variable *Edad* después de la limpieza de datos**



*Fuente: Elaboración propia*

Luego de la limpieza del conjunto de datos se observa que hay simetría para la variable *Edad*.

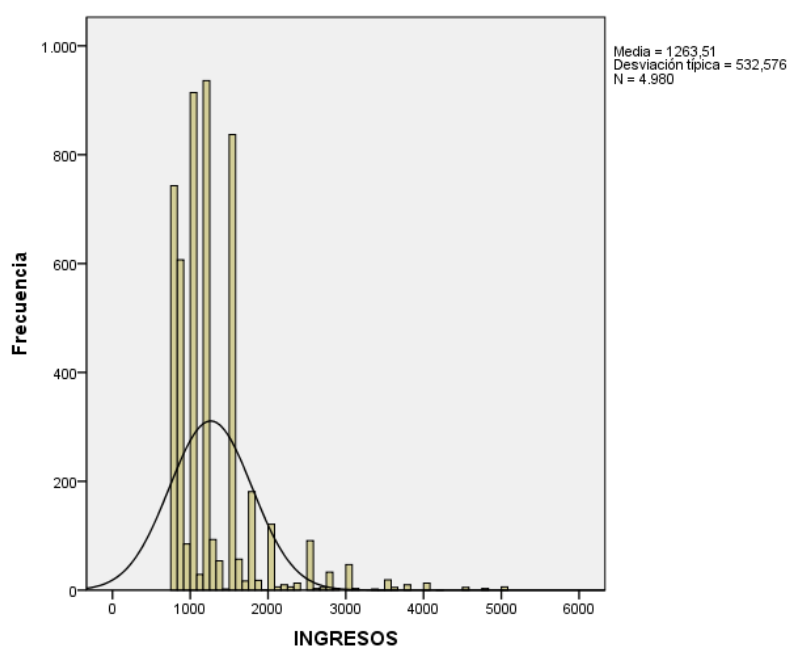
**Figura 14: Histograma de la variable *Ingreso* antes de la limpieza de datos**



*Fuente: Elaboración propia*

El conjunto de datos para la variable *Ingreso* presenta un sesgo muy marcado hacia la derecha, esto se debe a que hay valores *outliers*.

**Figura 15: Histograma de la variable *Ingreso* después de la limpieza de datos**



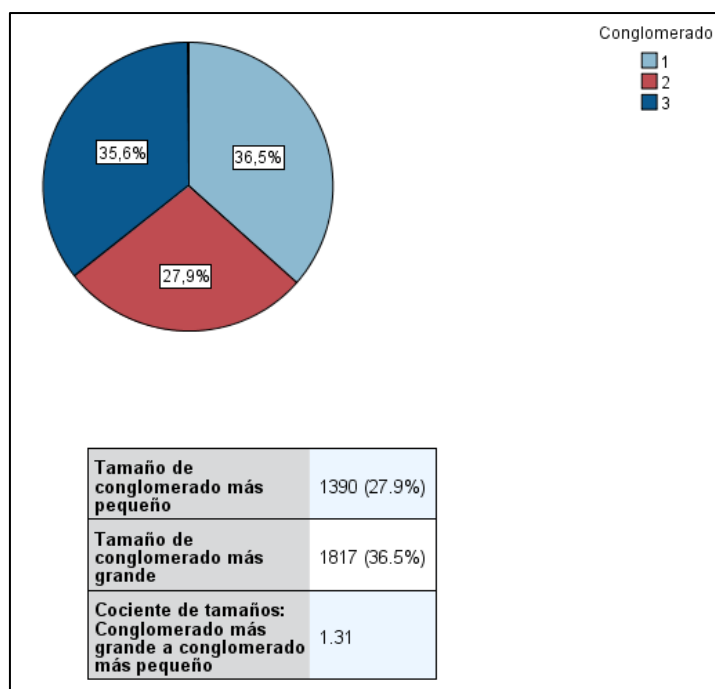
*Fuente: Elaboración propia*

Luego de la limpieza del conjunto de datos se observa que hay asimetría hacia la derecha, sin embargo no es tan marcada como antes de la limpieza. Los valores más frecuentes se sitúan entre los 1000 y 1800 soles.

### 5.3. Obtención de conglomerados

Con los 4980 datos restantes se aplicó el análisis de conglomerados bietápico usando el programa *IBM SPSS Statistics 22* (Versión Demo), la medida de distancia utilizada fue Log-verosimilitud, el criterio de conglomeración utilizado fue el BIC (criterio bayesiano de Schwarz) y los resultados fueron los siguientes:

**Figura 16: Tamaños de conglomerados**



*Fuente: Elaboración propia*

Se obtuvieron tres conglomerados el primer con 1817 clientes (36.5%), el segundo conglomerado con 1390 clientes (27.9%) y el tercer conglomerado con 1773 clientes (35.6%) (Ver Cuadro 6).

**Cuadro 6: Distribución de cada conglomerado**

		Frecuencia absoluta	Frecuencia relativa
Conglomerados	1	1817	36.5%
	2	1390	27.9%
	3	1773	35.6%
	Total	4980	100.0%

*Fuente: Elaboración propia*

**Cuadro 7: Puntuaciones medias de las variables cuantitativas en cada conglomerado**

Número de conglomerados en dos fases	MONTO CAPITAL	INGRESOS	EDAD
	Media	Media	Media
1	<i>1621</i>	<i>1264</i>	<i>37</i>
2	<i>1703</i>	<i>1144</i>	<i>37</i>
3	<i>1562</i>	<i>1357</i>	<i>35</i>

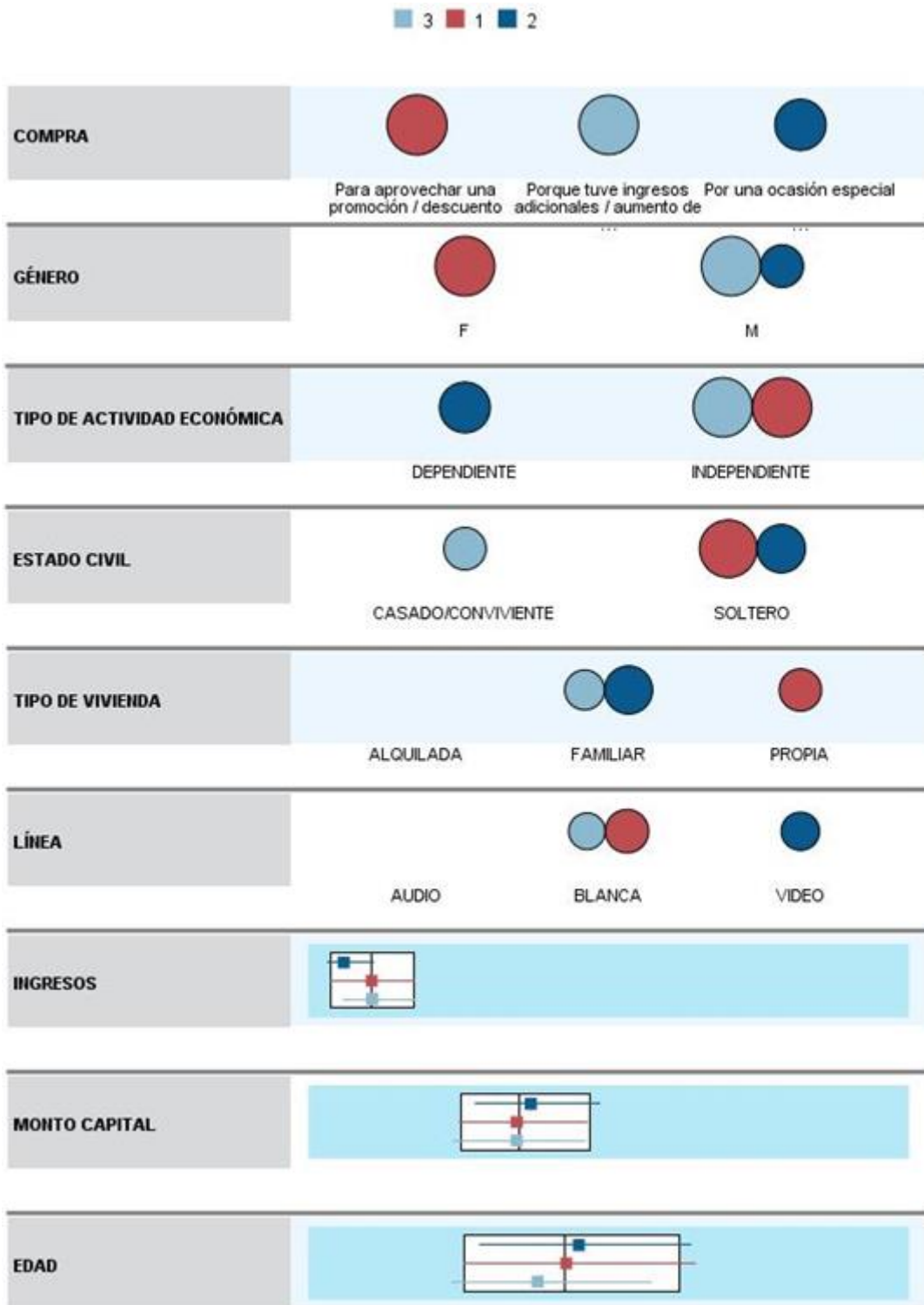
*Fuente: Elaboración propia*

El segundo conglomerado presenta una mayor media en la variable *Monto de Capital* con 1621 soles mientras que el tercer conglomerado presenta la menor media para dicha variable con 1562 soles.

El primer conglomerado tiene un *Ingreso* promedio de 1264 soles, el segundo conglomerado es el que presenta menor *Ingreso* promedio con 1144 soles y el tercer conglomerado tiene el *Ingreso* promedio más alto con 1357 soles.

Para la variable *Edad* el primer y segundo conglomerado tiene el mismo comportamiento, con una *Edad* promedio de 37 años mientras que el tercer conglomerado presenta la *Edad* promedio menor con 35 años.

**Figura 17: Comparación de conglomerados**



*Fuente: Elaboración propia*

El primer conglomerado está representado en un 100% por aquellos clientes que su motivo de la compra fue aprovechar una promoción/ descuento, en un 98.7% son del género femenino, 98.2% trabajan de manera independiente, 92.4% son solteras, 49.5% su tipo de vivienda es propia y 52.6 % adquirieron un electrodoméstico de la línea blanca.

El segundo conglomerado está representado en un 95% por aquellos clientes que el motivo de su compra fue una ocasión especial (cumpleaños, matrimonio, etc.), 66.5% son de género masculino, 93.5% tienen un trabajo dependiente, 85.1% son solteros, 84.3% su tipo de vivienda es familiar y 53.1% adquirieron un electrodoméstico de la línea de video.

El tercer conglomerado está representado en un 100% por aquellos clientes que su motivo de compra se dio porque tuvieron ingresos adicionales/aumento de sueldo o recibieron su gratificación, 100% son de género masculino, 99.7% tienen un trabajo independiente, 51.1% son casados o convivientes, 45% su tipo de vivienda es familiar y 38.2% adquirieron un electrodoméstico de la línea blanca.

#### **5.4. Propuesta de perfiles de los conglomerados**

##### *5.4.1. Perfil del conglomerado 1: Los cazaofertas*

Este tipo de conglomerado agrupa a las personas que compran cuando hay algún descuento o promoción especial. Mayormente son mujeres solteras, con trabajo independiente y viven en casa propia o familiar, y prefieren productos de línea blanca y video. La mitad de las personas en este conglomerado tienen más de 36 años y el 25% de tienen ingresos menores a 909 soles mensuales.

##### *5.4.2. Perfil del conglomerado 2: Los estacionales*

Este conglomerado se caracteriza por agrupar a los clientes que compran mayormente línea de video (televisores, blu ray, cámaras), son solteros y trabajadores dependientes. La mayoría son hombres, sin embargo no difieren demasiado de las mujeres. Son personas que compran cada vez que se acerca alguna fecha importante para ellos (día del padre, día de la madre, cumpleaños). La mitad de ellos tienen ingresos mensuales



por encima de los 1200 soles y tienen más de 36 años. El 25% de ellos tienen antecedentes de compra en la tienda por menos de 1200 soles.

#### *5.4.3. Perfil del conglomerado 3: Consumidores cautos*

En este grupo destacan las personas de sexo masculino, que tienen trabajo independiente (taxistas, negocio propio pequeño, micro empresarios), habitan en vivienda familiar, son casados o convivientes (tienen familia), tienen un ingreso mensual promedio de 1350 soles y el 75% de ellos tienen menos de 42 años. Suelen hacer compras cuando hay mayor movimiento de dinero en el mercado, ya sea por fiestas patrias o celebraciones de fin de año. No hay marcadas diferencias en los productos que compran, siendo su principal adquisición artículos para el hogar (línea blanca).

### **5.5. Propuesta de estrategia de marketing**

#### *5.5.1. Estrategia de marketing del conglomerado 1: Los cazaofertas*

Esta estrategia consiste en dar descuentos del 15%, si la compra es con la tarjeta de crédito de la tienda, y del 10% con otro medio de pago, a productos grandes en línea blanca y video (cocinas, refrigeradoras, lavadoras y televisores). Los productos que forman parte de esta campaña son mayormente aquellos que no tuvieron salida en la temporada anterior (stock antiguo), de esta forma la empresa no pierde cuando estos se deprecien con el pasar del tiempo.

Esta promoción se dará en los meses de enero, febrero, agosto y setiembre; ya que en estos meses la actividad comercial no es tan fuerte como en los demás meses del año.

Por ejemplo, un televisor que vale 2000 soles en todas las tiendas por departamentos del Perú, costaría hasta 1600 soles en nuestra tienda.

La promoción de la oferta se hará a través de redes sociales, correo electrónico y envío de mensajes de texto al celular.

De esta forma, se podrá motivar el uso de la tarjeta de crédito, además de aumentar la colocación de éstas en el mercado.

Con esto busca aprovechar las compras por impulso, y captar la atención del público que busca comprar cada vez que hay descuentos especiales.

#### *5.5.2. Estrategia de marketing del conglomerado 2: Los estacionales*

Esta estrategia consiste en dar bonos de descuento de 100 soles para una próxima compra en línea de audio (para productos seleccionados y cuyo precio sea mayor a los 800 soles), por cada 1500 soles de compra en línea de video. Tanto la línea de audio y video deja un margen de ganancias de aproximadamente 20%, siendo esta promoción un riesgo que la empresa está dispuesta a enfrentar.

Esta promoción se efectuará en los meses de julio y diciembre de cada año.

La publicidad se hará a través de envío de folletos y redes sociales.

Debido a que en este grupo sobresalen las personas de sexo masculino y estado civil soltero, se le da mayor énfasis a la línea de audio y video.

Otra estrategia consiste en ofrecer al cliente 5% de descuento en el mes de su cumpleaños para productos de línea blanca, audio y video. Se ofrecerá esta promoción a través de correo electrónico y redes sociales. Con esto se busca fidelizar al cliente y mantenerlo satisfecho con las promociones que la tienda le ofrece.

#### *5.5.3. Estrategia de marketing del conglomerado 3: Consumidores cautos*

Ofrecer paquetes de 3 productos de línea blanca, audio y video para el hogar, al precio total se le puede hacer un descuento del 10%. Estos paquetes pueden ser armados a gusto del cliente asesorado por algún representante de la tienda. El límite será de 1000 soles, es decir, el precio total no debe sobrepasar este monto, y no están incluidos refrigeradoras, lavadoras y televisores. Con esto se busca mayor actividad de venta a los productos menores de línea blanca (licuadoras, microondas, blu ray, minicomponentes, etc.). Esta campaña se aplicará los meses de marzo, julio y diciembre de todos los años.

Por ejemplo, se tiene el siguiente paquete cuyo precio total es de 700 soles: “horno microondas + licuadora + reproductor blu ray”. Al hacer efectivo el descuento, el cliente estaría pagando 630 soles por el paquete de 3 productos.

La estrategia de comunicación se hará a través de envíos de propagandas y llamadas al domicilio o negocio del cliente.

La idea de esta estrategia es aprovechar que este segmento suele invertir en productos de línea blanca y video, ya que la mayoría de personas de este cluster son hombres de estado civil casados o convivientes y tienen un hogar conformado. Este grupo además aprovecha los meses de mayor actividad comercial, donde las personas tienen ingresos extra (gratificación de julio y diciembre, cobro de utilidades) para hacer compras.

## VI. CONCLUSIONES

Las conclusiones según lo desarrollado en las secciones de Resultados y Discusión son las siguientes:

1. Las variables cualitativas consideradas en el estudio fueron Género, Estado Civil, Motivo de Compra, Tipo de Vivienda, Tipo de Actividad Económica y Línea del Producto; y las variables cuantitativas fueron Ingresos, Edad y Monto Capital. Se eligieron estas variables debido a que la empresa de *retail* nos brindó esta información de sus clientes por criterios de confidencialidad del mismo negocio.
2. El conjunto de datos utilizado en un inicio fue de 6284, luego de la limpieza de datos se eliminaron en total 1304 (20.8%) registros. Para la limpieza de las variables cuantitativas (Monto de Capital, Edad e Ingresos) se identificaron 54 registros que presentan *outliers* multivariados mediante el método de Distancia de Mahalanobis y 975 registros que no cumplen con los criterios de la regla de negocio. Mientras que para la limpieza de las variables cualitativas (Estado civil, Motivo de compra, Tipo de actividad económica y Línea) se identificaron en total 325 registros con ausencia de datos y en el caso de la variable *Línea* se eliminaron los registros de las categorías PEDS y Entretenimiento ya que ambas contaban con 1% de registros respectivamente. De estos 325 registros identificados en la limpieza de las variables cualitativas 50 de ellos ya estaban considerados dentro de los registros a eliminar por no cumplir los criterios de las variables cuantitativas.
3. Se encontraron tres grandes grupos de clientes con diferentes perfiles cada uno. Siendo la variable *Motivo de compra* la que mayor decisión hace en cada uno de ellos. De una muestra de 4980 clientes se obtuvo el primer conglomerado de 1817 clientes (36.5%), el segundo conglomerado con 1309 clientes (27.9%) y el tercer conglomerado con 1773 clientes (35.6%), la proporción es equilibrada entre los conglomerados.

4. Para cada conglomerado encontrado se estableció un perfil de acuerdo a sus características más resaltantes. Se obtuvo el perfil de los cazaofertas, los estacionales y los compradores cautos.
  
5. Para cada perfil encontrado se propusieron estrategias de marketing mix, el cual tendrá como objetivo fidelizar al cliente e incrementar las colocaciones de productos de la empresa de *retail*.

## VII. RECOMENDACIONES

A partir del estudio realizado, se muestran las siguientes recomendaciones para mejorarlo:

1. Mejorar toma de datos, con el fin de no perder muchos registros que pueden aportar información importante al objetivo del estudio.
2. Incorporar más variables de estudio que permitan mejorar la calidad de cada conglomerado. En base a criterios de expertos se propone las siguientes variables cualitativas: la variable *Nivel socioeconómico* (según filtro de NSE del APEIM<sup>1</sup>), la variable *Decisor de compra* (si la persona que realizó el pago de la compra, tomó o no la decisión de comprar dicho producto) y además la variable *Región* ya que la tienda de electrodomésticos se está expandiendo a nivel nacional. Además tomar en cuenta variables cuantitativas como *Antigüedad laboral*, *número de hijos del cliente*, *deuda en el sistema financiero*.
3. Comparar la técnica de conglomeración propuesta, con otras técnicas como K-medias para analizar las discrepancias en su segmentación.
4. Debido a que la finalidad de este estudio es proponer estrategias de marketing, se sugiere realizar un seguimiento frecuente de las estrategias propuestas en cada perfil.
5. Aplicar una futura segmentación ampliando el conjunto de datos, ya que la implementación de las estrategias de marketing cambiarán el comportamiento de los clientes en el tiempo.

---

<sup>1</sup> APEIM: Asociación Peruana de Empresas de Investigación de Mercado.

## VIII. REFERENCIA BIBLIOGRÁFICA

1. *Asociación Peruana de Empresas de Investigación de Mercados*. (s.f.). Recuperado el Julio de 2015, de <http://apeim.com.pe/>
2. Best, R. (2007). *Marketing Estratégico* (4ta Edición ed.). Madrid: Pearson Educación.
3. Fernández, C., & Aqueveque, C. (2001). Segmentación de mercados: buscando la correlación entre variables psicológicas y demográficas. *Revista Colombiana de Marketing*.
4. Garriga Trillo, A., Lubin Pigouche, P., Merino Merino, J., Padilla Suárez, M., Recio Saboya, P., & Suárez Falcón, J. (2010). *Introducción al análisis de datos*. UNED.
5. Gázquez, J., Jimenez, D., Sáez, E., & Sánchez, M. (2010). El comportamiento de cambio de canal del telespectador en Prime Time. Análisis a través de un panel de audimetría. *Revista Española de Investigación de Marketing ESIC*.
6. González, A., Cervantes, M., & Muñiz, N. (s.f.). *La incidencia de los estilos de vida en la segmentación del mercado turístico: Aplicación a la población urbana de Castilla y León*.
7. Guevara, M., & Cortés, H. (2006). *SPSS orientado a la gestión de mercados*. Bogotá.
8. IBM\_SPSS\_Statistics\_22. (s.f.). *IBM Knowledge Center*. Recuperado el Julio de 2015, de [http://www-01.ibm.com/support/knowledgecenter/SSLVMB\\_22.0.0/com.ibm.spss.statistics.algorithms/alg\\_2step\\_precluster.htm](http://www-01.ibm.com/support/knowledgecenter/SSLVMB_22.0.0/com.ibm.spss.statistics.algorithms/alg_2step_precluster.htm)
9. Icart Isern, M., Garrido Aguilar, E., Pulpón Segura, A., & Delgado Hito, P. (2012). *Cómo elaborar y presentar un proyecto de investigación una tesina y una tesis*. Universidad de Barcelona.
10. Kotler, P., & Armstrong, G. (2007). *Marketing Versión para Latinoamérica*.
11. Lamb, C., Hair, J., & McDaniel, C. (2006). *Marketing*. Cengage Learning.

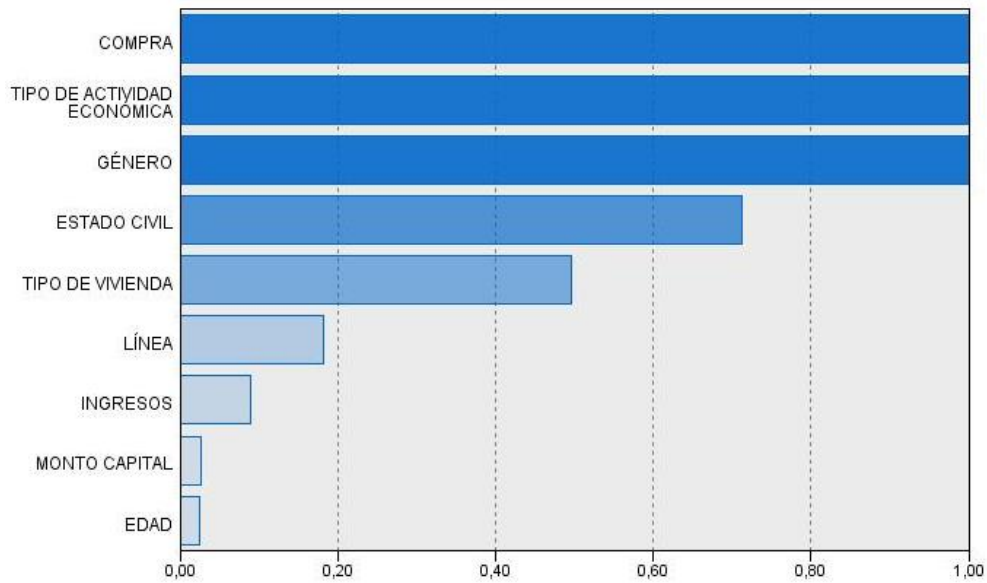
12. Levine, D. M. (1996). Estadística básica en administración: conceptos y aplicaciones. Pearson Educación.
13. Martín, Q. M., & De Paz Santana, Y. (2007). Tratamiento estadístico de datos con SPSS. Paraninfo.
14. McDaniel, C., & Gates, R. (2005). *Investigación de mercados* (6ta ed.). México: Thomson.
15. Ming-Yi Shih, J.-W. J.-F. (2010). A Two-Step Method for Clustering Mixed Categorical. *Tamkang Journal of Science and Engineering*.
16. Miranda Villagómez, F., & Salinas Flores, J. (2012). Estadística General. UNALM.
17. Muñoz García, J., & Amón Uribe, I. (2013). Técnicas para detección de outliers multivariantes. *Revista en Telecomunicaciones e Informática*.
18. Pérez López, C. (2011). *Técnicas de Segmentación. Conceptos, Herramientas y Aplicaciones*. Madrid: Garceta.
19. Perú21. (28 de Agosto de 2013). *peru21.pe*. Recuperado el Julio de 2014, de <http://peru21.pe/economia/55-hogares-urbanos-pertenece-clase-media-y-alta-2146622?href=mwapp>
20. Quaglino, M., & Merello, J. (2012). Métodos multivariados en estudios de vulnerabilidad. *Decimoséptimas Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística*.
21. Renart, L. (2004). *CRM: Tres estrategias de éxito*. PricewaterhouseCoopers & IESE.
22. Salinas Flores, J. W. (2010). Segmentación de mercados usando el análisis cluster bietápico. *Anales Científicos*.
23. Sampieri Hernández, R., Fernández Collado, C., & Batista Lucio, P. (2006). *Metodología de la investigación*. México, D.F: McGraw-Hill Interamericana.
24. Sarabia Alegría, J., & Pascual Sáez, M. (2005). Curso básico de estadística para economía y administración de empresas. Universidad de Cantabria.



25. Stanton, W., Etzel, M., & Walker, B. (2007). *Fundamentos de Marketing*.
26. Tabachnick, B., & Fidell, L. (2000). *Using Multivariate Statistics*. Allyn & Bacon.
27. Thompson, M. (s.f.). The Science and Art of Market Segmentation Using PROC FASTCLUS. *Statistics, Data Analysis, and Modeling-Paper 270*.
28. Universidad\_de\_Granada. (s.f.). *Estadística*. Recuperado el Julio de 2015, de <http://wdb.ugr.es/~bioestad/guia-spss/practica-8/#17>
29. Valderrey Sanz, P. (2011). *Segmentación de mercados* (1era ed.). Bogotá: Ediciones de la U.
30. Vargas Sabadías, A. (1995). *Estadística descriptiva e inferencial*. Servicio de Publicaciones de la Universidad de Castilla - La Mancha.
31. Weiner, I. B., Schinka, J. A., & Veli, W. F. (2013). *Handbook of Psychology, Research Methods in Psychology*. John Wiley & Sons.
32. Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases. *ACM SIGMOD*.

## IX. ANEXOS

### Anexo 1: Importancia del predictor

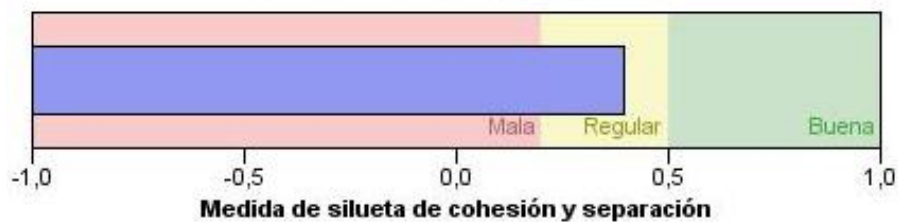


Fuente: Elaboración propia

### Anexo 2: Resumen de modelo

<b>Algoritmo</b>	Bietápico
<b>Entradas</b>	9
<b>Conglomerados</b>	3

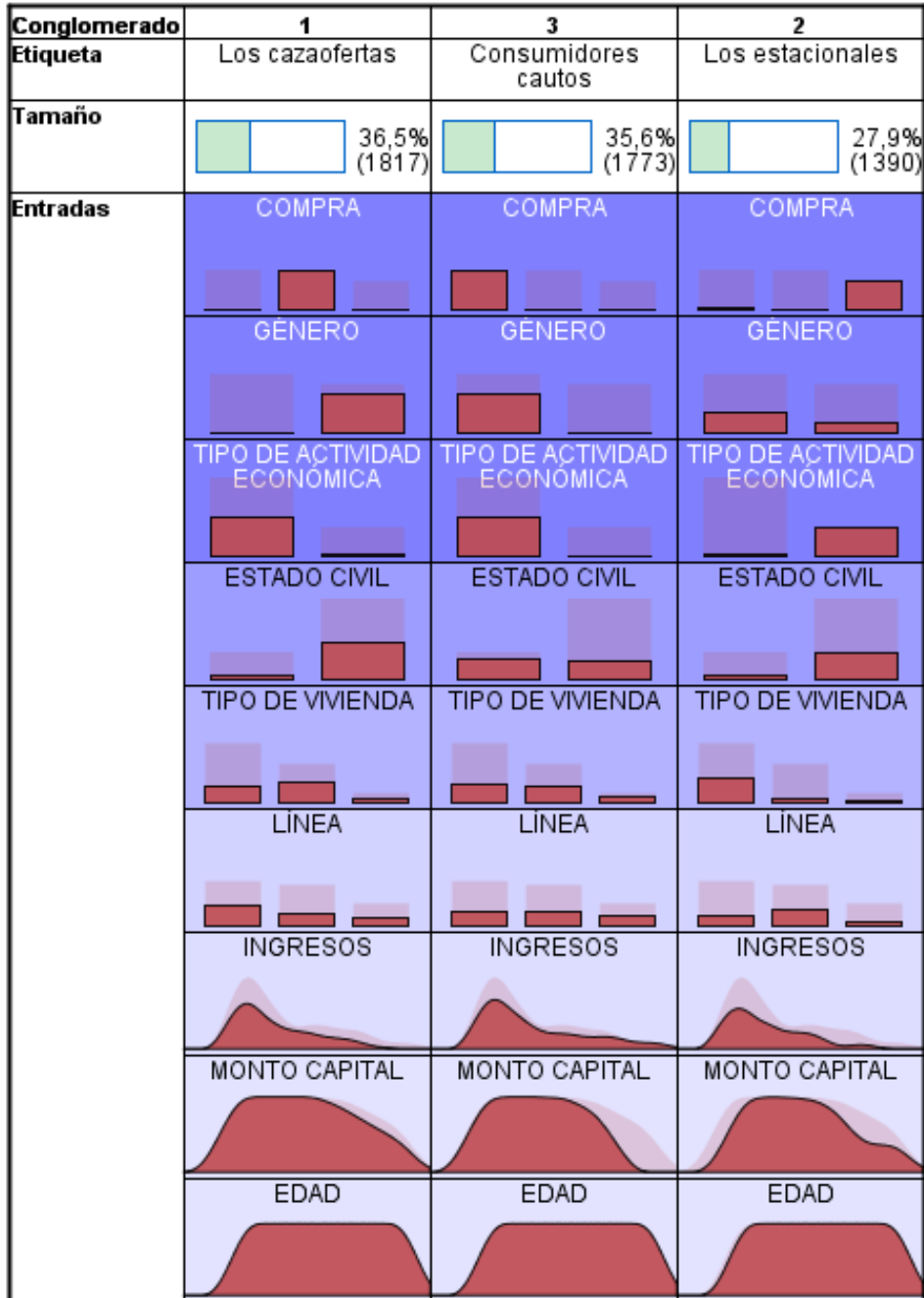
### Calidad de conglomerado



Fuente: Elaboración propia

### Anexo 3: Comportamiento de las variables en cada conglomerados

Importancia de entrada (predictor)  
 1,0 0,8 0,6 0,4 0,2 0,0



Fuente: Elaboración propia

## Anexo 4: Filtro de Nivel Socioeconómico

### FILTRO DE NSE

**(LEER)** Con la finalidad de agrupar sus respuestas con las de otras personas de similares características a las de usted, nos gustaría que responda a las siguientes preguntas referentes al jefe de hogar:

**JEFE DE HOGAR:** Aquella persona, hombre o mujer, de 18 a más, que aporta más económicamente en casa o toma las decisiones financieras de la familia, y vive en el hogar. **HOGAR:** conjunto de personas que, habitando en la misma vivienda, preparan y consumen sus alimentos en común.

**N1.** ¿Cuál es el último año o grado de estudios y nivel que aprobó el jefe de hogar? **(ACLARAR "COMPLETA O INCOMPLETA")**

Sin educación/ Educación Inicial	0	Superior Técnico Completa	3	Superior Univ. Completa	5
Primaria incompleta o completa/ Secundaria incompleta	1	Superior Univ. Incompleta	4	Post-Grado Universitario	7
Secundaria completa/ Superior Técnico Incompleta	2				

**N2.** ¿Cuál de estos bienes tiene en su hogar que esté funcionando?

	NO	SI
Computadora, laptop, tablet en funcionamiento	0	2
Lavadora en funcionamiento	0	2
Horno microondas en funcionamiento	0	2
Refrigeradora/ Congeladora en funcionamiento	0	2
<b>SUMAR PUNTAJE</b>		

**N3.** ¿Cuál de los siguientes bienes o servicios tiene en su hogar que esté funcionando?

	NO	SI
Auto o camioneta solo para uso particular <b>(NO TAXI NI AUTO DE LA EMPRESA)</b>	0	5
Servicio doméstico en el hogar pagado <b>(MÍNIMO QUE VAYA AL HOGAR UNA VEZ POR SEMANA)</b>	0	5
<b>SUMAR PUNTAJES</b>		

**N4.** ¿Cuál es el material predominante en los pisos de su vivienda? **(CONSIDERAR ÁREA CONSTRUIDA. RESPUESTA ÚNICA)**

Tierra / Otro material (arena y tablones sin pulir)	0	Laminado tipo madera, láminas asfálticas o similares	7
Cemento sin pulir o pulido / Madera (entablados)/ tapizón	3	Parquet o madera pulida y similares; porcelanato, alfombra, mármol	8
Losetas / terrazos, mayólicas, cerámicos, vinílicos, mosaico o similares	5		

**N5.** ¿A qué sistema de prestaciones de salud está afiliado el jefe de hogar? **(SI TIENE MÁS DE UNO CONSIDERAR EL DE MAYOR PUNTAJE. RESPUESTA ÚNICA)**

No está afiliado a ningún seguro/ Seguro Integral de Salud (SIS)	0	Seguro Salud FFAA/ Policiales	4
ESSALUD	2	Entidad prestadora de salud (EPS)/ Seguro privado de salud	6

**N6.** ¿Cuál es el material predominante en las paredes exteriores de su vivienda? **(NO REVESTIMIENTO, ES EL MATERIAL. RESPUESTA ÚNICA)**

Estera	0	Piedra o sillar con cal o cemento	4
Madera/ Piedra con barro/ Quincha (caña con barro)/ Tapia/ Adobe	2	Ladrillo o bloque de cemento	6

**N7.** El baño o servicio higiénico que tiene en su hogar está CONECTADO a:

NO TIENE O NO ESTÁ CONECTADO A UN DESAGÜE (SIN RED PÚBLICA)		SÍ ESTÁ CONECTADO AL DESAGÜE (CON RED PÚBLICA)	
No tiene baño	0	Baño compartido fuera de la vivienda. <i>(Ejem: quintas, corralones, cuartos con baño compartido, etc.)</i>	3
Baño que da a un pozo ciego, silo, letrina, pozo séptico, río, acequia o canal dentro o fuera del hogar	1	Baño dentro de la vivienda	5

N1	
N2	
N3	
N4	
N5	
N6	
N7	
Total	



.12 puntos o menos	NSE E	8
.De 13 a 19 puntos	NSE D	7
.De 20 a 22 puntos	NSE C2	6
.De 23 a 28 puntos	NSE C1	5

.De 29 a 33 puntos	NSE B2	4
.De 34 a 39 puntos	NSE B1	3
.De 40 a 47 puntos	NSE A2	2
.48 puntos a más	NSE A1	1

Fuente: <http://apeim.com.pe/>