

**SEGMENTACIÓN DE CLIENTES EN UNA COOPERATIVA DE AHORRO Y
CRÉDITO MEDIANTE ANÁLISIS DE CONGLOMERADOS**

DIEGO JULIÁN DELGADO



**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2016**

**SEGMENTACIÓN DE CLIENTES EN UNA COOPERATIVA DE AHORRO Y
CRÉDITO MEDIANTE ANÁLISIS DE CONGLOMERADOS**

DIEGO JULIÁN DELGADO

**Proyecto de grado para optar al título de
Especialista en Estadística**

Director

Dr. HENRY LAMOS DIAZ

Ph. D. en Matemáticas - Física

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE MATEMÁTICAS

BUCARAMANGA

2016

AGRADECIMIENTOS

A mi madre, que siempre me ha inspirado a mejorar, y con su amor y dedicación me ha ayudado a superar obstáculos.

A mi esposa por su comprensión y apoyo incondicional.

A La Cooperativa que me brindó la oportunidad de seguir creciendo profesionalmente.

A mi profesor Henry Lamos, por sus enseñanzas, la paciencia y dedicación en la dirección de este proyecto.

CONTENIDO

	Pág.
INTRODUCCIÓN	17
CAPÍTULO 1. GENERALIDADES SEGMENTACIÓN DE CLIENTES	20
1.1 ANTECEDENTES.....	20
1.2 MARCO TEÓRICO	23
1.2.1 Generalidades.	23
1.2.2 Medidas de distancia.	24
1.2.3 Tipos de datos.	26
1.2.4 Estandarización de los datos.	27
1.2.5 Algoritmos de clasificación.....	28
1.2.6 Tipos de segmentación.....	34
CAPÍTULO 2. SELECCIÓN, DEPURACIÓN Y DESCRIPCIÓN DE LAS VARIABLES	36
2.1 SELECCIÓN DE LAS VARIABLES	36
2.1.1 Definición de las variables	36
2.1.1.1 Variables demográficas y socioeconómicas	36
2.1.1.2 Variables conductuales	37
2.2 DEPURACIÓN DE LAS VARIABLES	39
2.3 ANÁLISIS DESCRIPTIVO	41
2.3.1 Variables cuantitativas	41
2.3.1.1 Edad	41
2.3.1.2 Antigüedad	43
2.3.1.3 Total Ingresos.....	44

2.3.1.4	Total Activos	46
2.3.1.5	Saldo Aportes	48
2.3.1.6	Saldo Ahorros	49
2.3.1.7	Saldo PAP	51
2.3.1.8	Saldo CDAT.....	52
2.3.1.9	Saldo Crédito.....	54
2.3.2	Variables cualitativas	55
2.3.2.1	Género	55
2.3.2.2	Estrato	56
2.3.2.3	Ocupación	57
2.3.2.4	Tiene Ahorros	58
2.3.2.5	Tiene PAP	58
2.3.2.6	Tiene CDAT.....	59
2.3.2.7	Tiene Crédito.....	59
2.3.2.8	Clasificación Crédito.....	60
2.3.2.9	Escolaridad.....	61
2.3.2.10	Diversificación.	61
2.3.2.11	Perfil Asociado.....	62
2.4	CATEGORIZACIÓN DE VARIABLES.....	63
2.4.1	Categorización de variables cuantitativas.....	63
2.4.1.1	Antigüedad.	63
2.4.1.2	Total Ingresos.....	64
2.5	RESUMEN VARIABLES SELECCIONADAS.....	64
CAPÍTULO 3. ANÁLISIS DE CONGLOMERADOS		66
3.1	SELECCIÓN DEL ALGORITMO DE CLASIFICACIÓN	66
3.2	PROCEDIMIENTO PARA EL ANÁLISIS DE CONGLOMERADOS UTILIZANDO LA HERRAMIENTA SPSS.....	68
3.3	RESULTADOS DEL ANÁLISIS DE CONGLOMERADOS	75

CAPÍTULO 4. DESCRIPCIÓN DE LOS CONGLOMERADOS.....	84
4.1 DESCRIPCIÓN CONGLOMERADO 1.....	84
4.2 DESCRIPCIÓN CONGLOMERADO 2.....	84
4.3 DESCRIPCIÓN CONGLOMERADO 3.....	85
CONCLUSIONES	86
BIBLIOGRAFÍA.....	88

LISTA DE TABLAS

	Pág.
Tabla 1. Categorías variable Antigüedad.....	63
Tabla 2. Categorías variable Total Ingresos.	64
Tabla 3. Resumen variables seleccionadas.....	65
Tabla 4. Frecuencias y porcentajes de participación por conglomerado	76
Tabla 5. Perfil de conglomerados. Variable Tiene Ahorros.....	77
Tabla 6. Perfil de conglomerados. Variable Tiene PAP	77
Tabla 7. Perfil de conglomerados. Variable Tiene CDAT.....	78
Tabla 8. Perfil de conglomerados. Variable Tiene Crédito.....	78
Tabla 9. Perfil de conglomerados. Clasificación de Crédito.....	78
Tabla 10. Perfil de conglomerados. Variable Diversificación.	79
Tabla 11. Perfil de conglomerados. Variable Perfil Asociado.	79
Tabla 12. Perfil de conglomerados. Variable Ocupación.	80

Tabla 13. Perfil de conglomerados. Variable Antigüedad (categorizada).	80
Tabla 14. Perfil de conglomerados. Variable Estrato.	81
Tabla 15. Perfil de conglomerados. Variable Total Ingresos (categorizada).	82
Tabla 16. Perfil de conglomerados. Variable Escolaridad.	82
Tabla 17. Resumen de los resultados del análisis.	83

LISTA DE FIGURAS

	Pág.
Figura 1. Segmentación de Bancolombia según necesidades y momentos de vida del cliente.....	22
Figura 2. Métodos de interdependencia en Análisis Multivariado	23
Figura 3. Técnicas para proceso de segmentación.	24
Figura 4. Evaluación de distancias (vecino más cercano, vecino más lejano, centroides y promedio inter-grupos).	30
Figura 5. Dendrograma. Representación gráfica de una clasificación jerárquica. .	31
Figura 6. Tabla estadísticos descriptivos y diagrama de caja. Variable Edad.	42
Figura 7. Tabla de frecuencias e histograma. Variable Edad.	42
Figura 8. Tabla estadísticos descriptivos y diagrama de caja. Variable Antigüedad.	43
Figura 9. Tabla de frecuencias e histograma. Variable Antigüedad.....	44
Figura 10. Tabla estadísticos descriptivos y diagrama de caja. Variable Total Ingresos.	45

Figura 11. Tabla de frecuencias e histograma. Variable Total Ingresos.	46
Figura 12. Tabla estadísticos descriptivos y diagrama de caja. Variable Total Activos.	47
Figura 13. Tabla de frecuencias e histograma. Variable Total Activos.	47
Figura 14. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo Aportes.	48
Figura 15. Tabla de frecuencias e histograma. Variable Saldo Aportes.	49
Figura 16. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo Ahorros.	50
Figura 17. Tabla de frecuencias e histograma. Variable Saldo Ahorros.	51
Figura 18. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo PAP.	51
Figura 19. Tabla de frecuencias e histograma. Variable Saldo PAP.....	52
Figura 20. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo CDAT.	53
Figura 21. Tabla de frecuencias e histograma. Variable Saldo CDAT.	54
Figura 22. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo Crédito.	54

Figura 23. Tabla de frecuencias e histograma. Variable Saldo Crédito.	55
Figura 24. Tabla de frecuencias y diagrama de barras. Variable Género.	56
Figura 25. Tabla de frecuencias y diagrama de barras. Variable Estrato.	56
Figura 26. Tabla de frecuencias y diagrama de barras. Variable Ocupación.	57
Figura 27. Tabla de frecuencias y diagrama de barras. Variable Tiene Ahorros. ...	58
Figura 28. Tabla de frecuencias y diagrama de barras. Variable Tiene PAP.	58
Figura 29. Tabla de frecuencias y diagrama de barras. Variable Tiene CDAT.	59
Figura 30. Tabla de frecuencias y diagrama de barras. Variable Tiene Crédito. ...	60
Figura 31. Tabla de frecuencias y diagrama de barras. Variable Clasificación Crédito.	60
Figura 32. Tabla de frecuencias y diagrama de barras. Variable Escolaridad.	61
Figura 33. Tabla de frecuencias y diagrama de barras. Variable Diversificación. ...	62
Figura 34. Tabla de frecuencias y diagrama de barras. Variable Perfil Asociado. .	62
Figura 35. Ruta para realizar un cluster Bietápico.	69
Figura 36. Cuadro de diálogo de análisis de cluster Bietápico.	70
Figura 37. Cuadro de diálogo “Opciones” en conglomerado bietápico.	72

Figura 38. Cuadro de diálogo “Resultados” en conglomerado Bietápico.	73
Figura 39. Cuadro de diálogo análisis de cluster bietápico con las variables y opciones seleccionadas.	75
Figura 40. Resumen del modelo y calidad del cluster.	76

RESUMEN

TÍTULO

SEGMENTACIÓN DE CLIENTES EN UNA COOPERATIVA DE AHORRO Y CRÉDITO MEDIANTE ANÁLISIS DE CONGLOMERADOS*

AUTOR

DIEGO JULIÁN DELGADO**

PALABRAS CLAVES

SEGMENTACIÓN, ANÁLISIS DE CONGLOMERADOS, MÉTODOS MULTIVARIADOS, COOPERATIVA DE AHORRO Y CRÉDITO

DESCRIPCIÓN

Buscando un mayor conocimiento de sus asociados, la Cooperativa decide iniciar un proceso de segmentación de clientes como apoyo al desarrollo de los procesos comerciales y de mercadeo.

Este tema cobra cada día mayor importancia, si tenemos en cuenta que el mercado se mueve en un entorno cada vez más complejo, exigente y competitivo, al cual se le debe adicionar la llegada de entidades financieras grandes y fuertes al país. El conocer a sus clientes, saber seleccionarlos y establecer diferenciales se convierte en “un deber ser” en este tipo de organizaciones.

Todos los asociados son diferentes, al igual que sus necesidades. Atrás quedó ofrecer el mismo producto para todo el mundo.

La segmentación se fundamenta en la existencia de grupos de personas con características similares, pero al mismo tiempo diferentes a otros grupos. El determinar esas características, permitiría a La Cooperativa definir estrategias dirigidas y entregar ofertas de valor a cada segmento.

Para lo anterior, se aplica una técnica estadística, llamada análisis de conglomerados, mediante la selección, depuración y utilización de variables sociodemográficas y conductuales, así como el empleo de diferentes algoritmos de clasificación.

Al final, se presenta el procedimiento ejecutado y una descripción de cada uno de los conglomerados hallados.

* Trabajo de grado.

** Facultad de Ciencias. Escuela de Matemáticas. Especialización en Estadística.

ABSTRACT

TITLE

CUSTOMER SEGMENTATION IN A SAVINGS AND CREDIT COOPERATIVE THROUGH CLUSTER ANALYSIS.*

AUTHOR

DIEGO JULIÁN DELGADO**

KEYWORDS

SEGMENTATION, CLUSTER ANALYSIS, MULTIVARIATE METHODS, SAVINGS AND CREDIT COOPERATIVE

DESCRIPTION

In order to obtain a better understanding of its associates, The Cooperative decides to carry out a customer segmentation process as a support to the development of the commercial and marketing processes.

This issue is of increasing importance, if we consider that the market moves in an increasingly complex, challenging and competitive environment, to which must be added the arrival of large and strong financial institutions in the country. The know their customers, know select and establish differential becomes "a duty to be" in such organizations.

All partners are different, as are their needs. Gone was offering the same product for everyone.

The segmentation is based upon the existence of groups of people with similar characteristics, but at the same time, different to other groups. Determining these differences will allow The Cooperative to define addressed strategies and to give value offers to each cluster.

For doing so, a statistical technique called cluster analysis, by selection, purification and use of sociodemographic and behavioral variables, as well as the use of different classification algorithms applied.

At the end, the fulfilled procedure and a description of each of the identified clusters are presented.

* Grade work.

** Science Faculty. School of Mathematics. Specialization in Statistics.

INTRODUCCIÓN

En 1962 se crea una cooperativa multiactiva de trabajadores en el departamento de Santander (Colombia), que por razones de reserva de información en el presente análisis se llamará La Cooperativa.

Después de 38 años, La Cooperativa por motivos estrictamente legales acordes con la ley 454 de 1998, en donde se exigía una especialización de actividades, se convierte en una cooperativa especializada en ahorro y crédito.

Desde el año 2003, La Cooperativa es catalogada por un estudio de ASCOOP - Asociación Colombiana de Cooperativas- como la primera cooperativa del país en número de asociados y la primera en activos y patrimonio entre las cooperativas de ahorro y crédito.

El mercado objetivo de La Cooperativa son las personas naturales, mayores de edad, de estratos 2, 3 y 4, principalmente empleados, pensionados e independientes, ubicados en la zona de influencia de sus 51 agencias, las cuales están distribuidas en los departamentos: Santander, Cesar, Norte de Santander, Boyacá, Atlántico y Bogotá D.C., atendiendo a un número aproximado de 390.000 asociados, mediante un catálogo de productos y servicios que incluye ahorro, crédito e inversión.

Su principal competencia está compuesta por bancos, compañías de financiamiento, cooperativas financieras, microfinancieras y otras cooperativas de ahorro y crédito que hacen presencia en el centro y nororiente del país; entidades que en su mayoría, a lo largo de los últimos años, han demostrado solidez y constante crecimiento dentro del sistema financiero colombiano.

Precisamente, una cantidad cada vez mayor de entidades financieras atendiendo el mismo mercado de la cooperativa, obliga a una constante revisión de los procesos al interior de la organización, para asegurar su permanencia en el tiempo.

El tener claros y definidos los segmentos a los cuales se desea llegar con una propuesta de valor, hace parte del conocimiento que La Cooperativa debe tener de sus asociados y del compromiso con los mismos para cumplir su misión.

El objetivo principal de este proyecto es construir una segmentación de los clientes de La Cooperativa, mediante el uso de herramientas de clustering, para mejorar los procesos comerciales y de mercadeo buscando impactar los resultados de la empresa.

En este punto es importante diferenciar la segmentación de clientes de la segmentación de mercados¹. Esta última describe los mercados en función del tipo de clientes que lo componen, clientes no identificados, es decir, no se tiene información relevante de los mismos. Por otra parte, la segmentación de clientes tiene como base todos los datos e información que la empresa posea o pueda poseer de sus clientes (interna o externa), con la cual los puede segmentar, perfilar y describir mediante el uso de métodos estadísticos.

Existen diversos métodos estadísticos con los cuales se podría realizar una segmentación; sin embargo, por el objetivo del presente trabajo y las características de las bases de datos con que se cuenta, la segmentación se desarrollará por medio de técnicas de análisis de conglomerados (clusters), las cuales “tiene por objeto agrupar elementos en grupos homogéneos en función de

¹ “El proceso de dividir un mercado en distintos grupos de compradores con diferentes necesidades, características y comportamientos, y quienes podrían requerir productos o programas de marketing separados, se llama segmentación de mercado”. Kotler y Armstrong (2013).

las similitudes o similaridades entre ellos”. Peña (2002). Este método también se conoce como método de clasificación automática o no supervisada.

En el documento se presenta los conceptos de distancia, medidas de similitud, que se usan en los diferentes métodos que se discuten. La selección de las variables y el uso de la estandarización son decisiones que se analizan en el documento ya que es uno de los pasos importantes en el proceso de “clustering”. Bien es sabido que incluir una variable inadecuada en la segmentación tiene serias repercusiones en las posibles estrategias a implementar. Aunque no existe un criterio único para la selección, se comenzará tipificando las variables según establece Kotler y Armstrong (2013) en variables: geográficas, demográficas, sicográficas y conductuales.

Una vez seleccionadas y descritas estas variables, se procede a realizar la depuración de los datos y construir una base única con la información necesaria. Posteriormente, para la elección del número de segmentos e identificación de los mismos se utilizarán diversos métodos de clustering.

Finalmente, se describen cada uno de los segmentos encontrados con base en las variables que los identifican y al mismo tiempo los diferencian con los demás segmentos.

Vale la pena aclarar que para segmentar datos existen en el mercado una gran cantidad de herramientas (software estadístico) ofreciendo esta solución, entre los cuales se encuentran: Minitab, SAS, R, Statgraphics, SPSS, etc. Precisamente este último (SPSS) es ampliamente usado en el mercado y del cual actualmente La Cooperativa posee una licencia del software, razón por la cual para el desarrollo de la segmentación se utilizará esta herramienta.

CAPÍTULO 1. GENERALIDADES SEGMENTACIÓN DE CLIENTES

1.1 ANTECEDENTES

Antes de finalizar la década de los setenta, las entidades financieras manejaban su mercado como un todo, es decir, ofrecían sus productos y servicios por igual a sus clientes, sin importar las necesidades o características de los mismos, lo que hoy en día se conoce como marketing indiferenciado (o masivo)².

En España, por ejemplo, esta situación cambió con la llegada de la banca americana al mercado español. Según Bello (2012), esta penetración hizo que por primera vez la banca española llevase a cabo una política comercial basada en estrategias de segmentación, distinguiendo inicialmente dos:

- Por una parte, el Banco Bilbao (quien posteriormente se fusionó con el Banco Viscaya para crear el BBVA) realizó una segmentación de sus clientes en función a sus características básicas: actividad laboral, nivel de renta, lugar de residencia, edad y sexo; tanto es así, que llegó a crear el primer banco exclusivo para un segmento, el “Banco de la mujer”³.
- Por otra parte, el Banco Vizcaya optó por llevar a cabo una segmentación (especialización) de su red de oficinas en función de la naturaleza de la población objeto de atención.

² La empresa decide ignorar las diferencias de segmento y dirigirse a todo el mercado con una oferta. Kotler y Armstrong (2013).



³ A nivel internacional se tiene el caso del Women Banks en Nueva York. A nivel local el caso de la Fundación de la Mujer y el Banco Mundo Mujer.

Los criterios para segmentación pueden ser variados y dependen del objetivo del análisis; es así como Beerli y García (1992) mediante investigación de mercados y posterior Análisis de Conglomerados para la segmentación del mercado bancario concluyen: “son justamente los criterios orientados hacia el mercado los que proporcionan las verdaderas razones en las que los clientes fundamentan sus decisiones en la utilización de los servicios bancarios, siendo las ventajas buscadas el enfoque de segmentación que mejor define los motivos que inducen a comportamientos diferenciados”.

Actualmente podemos decir que todas las entidades financieras realizan de una u otra forma una segmentación de mercados, dirigiendo sus productos y servicios a uno, varios o todos los segmentos. Este último caso es representado por los grandes bancos colombianos como Bancolombia o Banco de Bogotá, quienes atienden prácticamente a todos los segmentos de mercado; sin embargo existen otras entidades como Banco Popular que dirige sus esfuerzo a algunos segmentos como los empleados y pensionados. Por otra parte, existen entidades financieras concentradas en un solo segmento o nicho de mercado, es el caso de Bancamía o Banco WWB, atendiendo a microempresarios o la compañía de financiamiento Tuya atendiendo tarjetahabientes, y el Fondo Nacional del Ahorro con productos enfocados a la adquisición de vivienda.

No obstante lo anterior, desde unos años se observa un cambio en la segmentación de mercado que realizan algunas de estas entidades. Con clientes cada vez más exigentes, ahora la segmentación se enfoca a atender directamente las necesidades de sus clientes, incluso se tienen en cuenta los momentos de vida o “estilos de vida” de los mismos, en función de variables como: edad, ingresos, gastos y composición familiar (ver figura 1).

Figura 1. Segmentación de Bancolombia según necesidades y momentos de vida del cliente.

Necesidades		Momentos de vida	
 Administra tu dinero	 Compra lo que te gusta	Bancolombia y los niños	
 Estudio	 Empieza tu negocio	Bancolombia y los jóvenes	
 Vacaciones	 Organiza tus créditos	Bancolombia y la familia	
 Carro o moto	 Planea tu retiro laboral	Colombianos en el exterior	
 Casa			

Fuente: <http://www.grupobancolombia.com/wps/portal/personas/>

En este momento La Cooperativa no cuenta con una segmentación de clientes. En general, la gestión comercial de los productos y servicios se ha enfocado principalmente basada en la actividad económica de sus asociados (independiente, empleado, pensionado, microempresario, estudiante...), y en otros casos en la tenencia y uso de los productos y servicios.

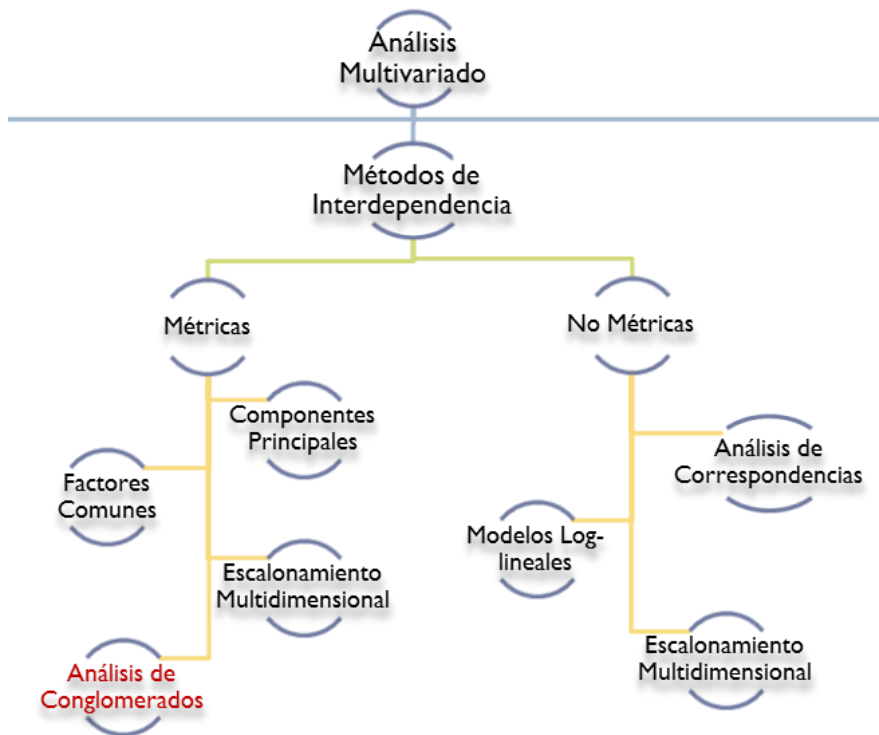
Dentro de la Dirección de Mercadeo de La Cooperativa existe el área de CRM⁴ (o Fidelización), relativamente reciente y con la cual se han realizado esfuerzos hacia el perfilamiento de clientes, más no hacia la segmentación de los mismos. Precisamente, la teoría del CRM tiene tres componentes: CRM Analítico, CRM Operativo y CRM Colaborativo. Para el CRM Analítico lo más importante es el proceso de conocimiento del cliente y es allí donde se plantea desarrollar el proyecto de grado.

⁴ La administración de las relaciones con los clientes (CRM, por sus siglas en inglés) consiste en software y herramientas analíticas sofisticadas que integran la información de los clientes proveniente de todas las fuentes, la analizan a fondo y aplican los resultados para crear relaciones más sólidas con los clientes.

1.2 MARCO TEÓRICO

1.2.1 Generalidades. El análisis de conglomerados es una técnica descriptiva de análisis de datos diseñada para determinar los agrupamientos (conglomerados) dentro de un conjunto de datos. Hace parte de los métodos multivariados (ver figura 2), los cuales en general tienen como objetivo principal el resumir grandes cantidades de datos por medio de pocos parámetros, y en otras ocasiones el de encontrar relaciones entre 1) las variables, 2) las unidades u observaciones, y 3) tanto las variables como las unidades.

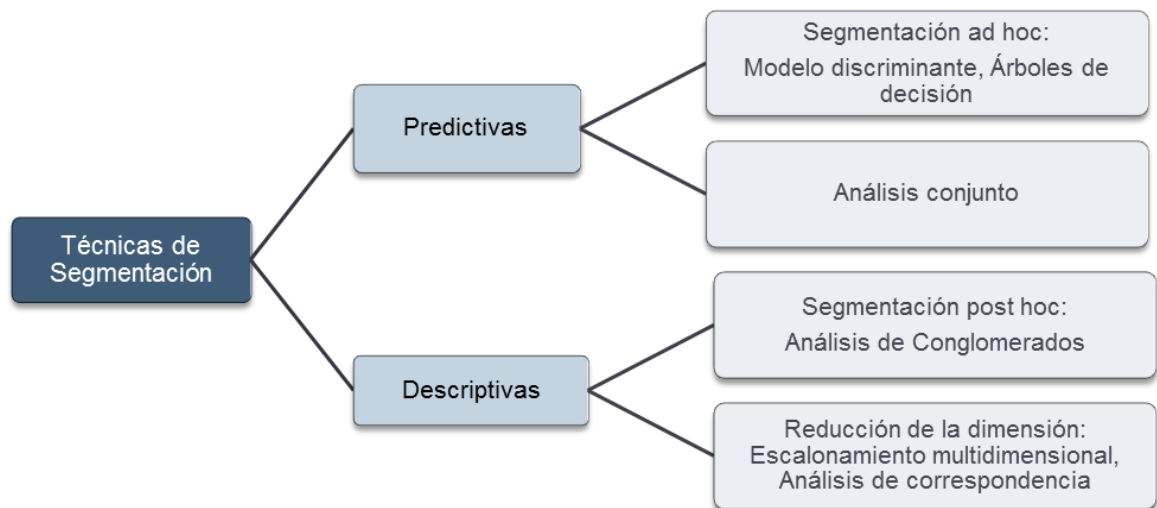
Figura 2. Métodos de interdependencia en Análisis Multivariado



Fuente: Elaboración propia.

Dentro de los métodos multivariados se encuentran diferentes técnicas para realizar segmentación (ver figura 3). La decisión de emplear una u otra depende del objetivo del estudio y de la información que se tenga. Para el proceso de segmentación de clientes, objetivo de este análisis, se utilizará la técnica descriptiva llamada Análisis de Conglomerados (o clustering en inglés).

Figura 3. Técnicas para proceso de segmentación.



Fuente: Elaboración propia.

A nivel fundamental, el análisis de conglomerados se basa en el simple concepto de dividir los individuos en grupos basados en su proximidad (cercanía) entre sí; distancias derivadas de las mediciones tomadas en las observaciones o variables consideradas.

1.2.2 Medidas de distancia. Uno de los aspectos importantes del análisis de conglomerados es la elección de la medida que se desea utilizar para cuantificar la

distancia entre los elementos. Estas medidas se diferencian por el tipo de distancia evaluada: similaridad o disimilaridad.

Las medidas de *similaridad* (proximidad o semejanza) evalúan el grado de cercanía o proximidad existente entre dos elementos. Cuanto mayor es su valor, mayor es el grado de similaridad o cercanía. Cuando dos elementos se encuentran juntos, el valor de la medida es máximo y es más probable que se clasifique en un mismo grupo. Un ejemplo de medida de similaridad son los coeficientes de congruencia y de correlación:

- **Coeficiente de Congruencia (c).** El coeficiente se apoya en el coseno de dos vectores. Si X e Y son dos vectores, $\text{Cos}(X, Y)$ es mayor si los valores de las pendientes de los vectores son similares. Puede variar de -1 a +1.

$$c = \frac{\sum_{j=1}^P X_{rj} Y_{sj}}{\sqrt{\sum_{j=1}^P X_{rj}^2} \sqrt{\sum_{j=1}^P Y_{sj}^2}}$$

- **Coeficiente de Correlación de Pearson (r).** El coeficiente r puede variar de -1 a +1, donde el signo indica la dirección de la correlación y el valor numérico la magnitud de la correlación; esto es, (-1) correlación negativa (o inversa) perfecta, (0) no existe correlación entre los elementos y (+1) es un correlación positiva (o directa) perfecta.

$$\text{CORRELATION}(X, Y) = \frac{\sum_i^n Z_{x_i} Z_{y_i}}{n-1}$$

donde n es el tamaño de la muestra y Z_x Z_y son las puntuaciones tipificadas del sujeto i en las variables X y Y, que son las variables entre las que se calcula la distancia.

Las medidas de *disimilitud* (o distancia) hacen énfasis sobre el grado de diferencia o lejanía existente entre dos elementos. Cuanto mayor es su valor, mayor la diferencia o lejanía entre los elementos comparados.

A continuación se relacionan las medidas de distancia más usadas:

- **Distancia Euclídea.** Medida utilizada por defecto para datos de intervalo. Raíz cuadrada de la suma de los cuadrados de las diferencias entre los valores de las variables:

$$\text{EUCLID}(X,Y) = \sqrt{\sum_i (X_i - Y_i)^2}$$

- **Distancia Euclídea al Cuadrado.** Suma de los cuadrados de las diferencias entre los valores de las variables:

$$\text{SEUCLID}(X,Y) = \sum_i (X_i - Y_i)^2$$

- **Chebychev.** Diferencia más grande en valor absoluto entre los valores de dos variables:

$$\text{CHEBYCHEV}(X,Y) = \max_i |X_i - Y_i|$$

- **Chi-cuadrado.** Medida utilizada por defecto para datos categóricos. Se base en la divergencia existente entre las frecuencias observadas y el modelo de independencia. La magnitud de esta medida depende del tamaño muestral. Los valores esperados se obtienen asumiendo independencia entre las variables:

$$\text{CHISQ}(X,Y) = \sqrt{\sum_i [X_i - E(X_i)]^2 / E(X_i) + \sum_i [Y_i - E(Y_i)]^2 / E(Y_i)}$$

1.2.3 Tipos de datos. Existen dos tipos básicos de datos: no métricos (cualitativos) y métricos (cuantitativos).

- Los *datos no métricos* son atributos, características o propiedades categóricas que identifican o describen a un individuo. Describen diferencias en tipo o clase indicando la presencia o ausencia de una característica o propiedad. Las medidas no métricas pueden tener escalas nominales u ordinales.

La medida con una escala nominal asigna números que se usan para etiquetar o identificar sujetos u objetos, sin ningún significado cuantitativo ya que sólo indican la presencia o ausencia o ausencia del atributo o característica bajo investigación.

Las escalas ordinales representan un nivel superior de precisión de la medida. Estas variables son ordenadas o clasificadas con relación a la cantidad del atributo poseído.

- Las medidas de *datos métricos* están constituidas de tal forma que los sujetos pueden ser identificados por diferencias entre grado o cantidad.

Las variables medidas métricamente reflejan cantidades relativas o grado, y proporcionan el nivel más alto de medida de precisión, permitiendo realizar con ellas todas las operaciones matemáticas.

1.2.4 Estandarización de los datos. El uso de datos no estandarizados puede implicar inconsistencias en el análisis de conglomerados cuando cambia la escala de las variables. Este es un problema al que se enfrentan todas las medidas de distancia. Por lo tanto, siempre que sea posible debería emplearse la estandarización de las variables de aglomeración, convirtiendo las medidas en unidades adimensionales, y por lo tanto comparables.

- Variables con una misma escala (razón) pueden tener ámbitos muy distintos de medida. Ejemplo: Ingresos en miles de pesos y edad en años. En estos casos conviene la estandarización.
- No siempre la estandarización es necesaria, a veces los resultados obtenidos sin estandarizar son idénticos a los que se obtienen estandarizando. Por supuesto, lo anterior depende de cada variable que se quiera analizar.
- La estandarización permite dar a todas las variables el mismo peso, pero puede que para el interés del investigador, no todas tengan la misma importancia.
- El problema de escalas es solucionable por conversión en otras escalas.

A menudo se realiza una estandarización usando variables de *puntuaciones z* normalizadas, para tener una media cero y una desviación estándar 1; sin embargo, existen otros métodos de estandarización, por ejemplo, las variables pueden normalizarse para ir de -1 a +1, o de 0 a 1, o de manera que el máximo sea 1, etc.

1.2.5 Algoritmos de clasificación. Para la agrupación de elementos o variables similares, se debe aplicar un conjunto de reglas, a las cuales se les denomina algoritmos de clasificación.

La primera distinción amplia en los algoritmos de clasificación de da entre los métodos *Jerárquicos* y *No Jerárquicos*. La idoneidad y eficacia del algoritmo empleado dependerá en gran medida de la propia estructura de los datos.

a. Métodos Jerárquicos. A continuación se describen cinco maneras comunes de evaluar la distancia en con este método:

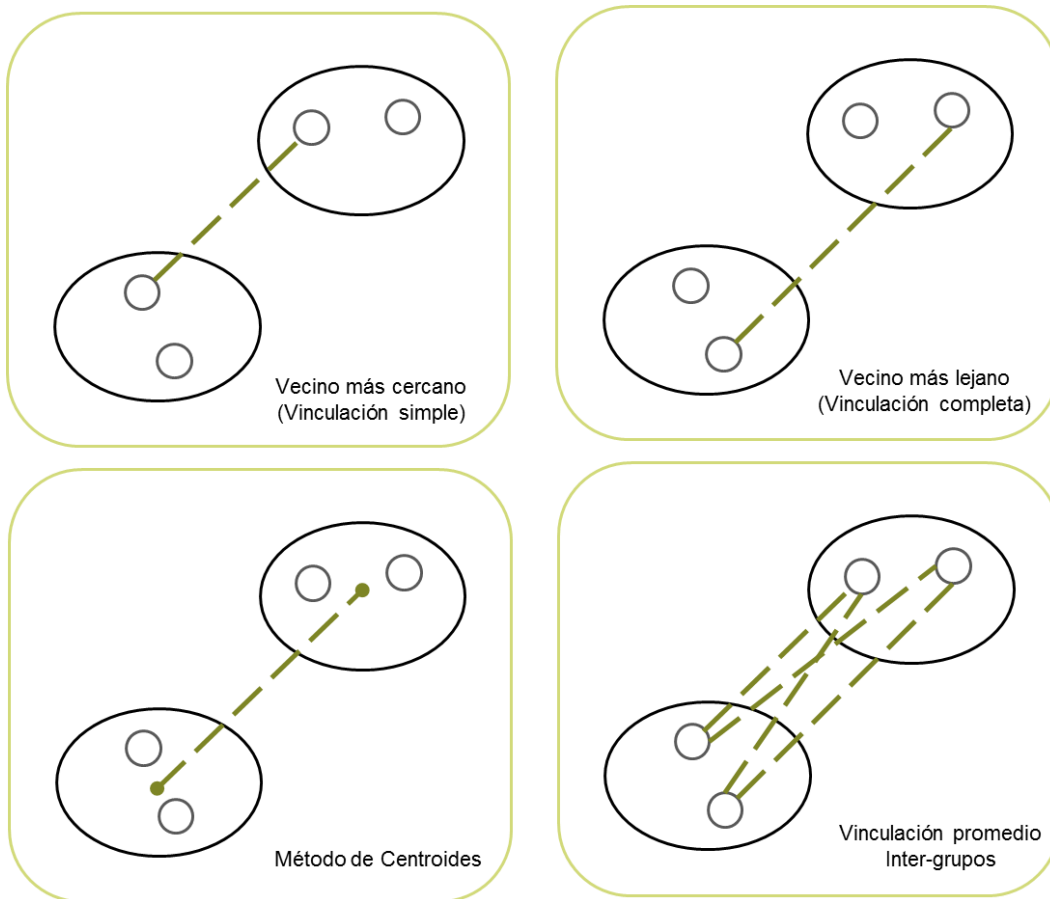
- *Vecino más cercano (vinculación simple).* Calcula la distancia entre los conglomerados como la distancia más corta entre todos los pares de puntos, uno de cada grupo.
- *Vecino más lejano (vinculación completa).* Este método también es relativamente insensible a los outliers y, como su nombre lo sugiere, los dos miembros más lejanos dentro de los conglomerados determinan la distancia entre dos conglomerados.
- *Centroides.* Evalúa las distancias entre los conglomerados en base a las distancias entre sus medias. Este método es también relativamente robusto frente a la presencia de outliers, pero no funciona bien con datos ruidosos⁵.
- *Vinculación promedio inter-grupos.* Evalúa la distancia para todos los pares de puntos posibles entre conglomerados, y luego toma la distancia entre los conglomerados como el promedio de estas distancias. Funciona bien con los datos ruidosos y en una variedad de condiciones. Sin embargo, su debilidad es que los outliers influyen sobre el procedimiento.
- *Método de Ward.* El proceso comienza con m grupos, cada uno de los cuales está compuesto por un solo individuo. En el primer paso, el primer cluster es formado por la selección de dos de esos m grupos. En cada paso o etapa se considera la posibilidad de la unión de cada par de grupos y se opta por la fusión de aquellos dos grupos que menos incrementen la

⁵ Los datos ruidosos son datos que contienen errores, que pueden provenir de errores en los aparatos de medida, de medidas reales pero altamente improbables o de casos excepcionales.

suma de los cuadrados de las desviaciones al unirse. Es bueno con los datos ruidosos, y sensible a los outliers. Otra propiedad de este método, es que tiende a crear conglomerados de aproximadamente el mismo tamaño.

La figura 4 representa gráficamente cuatro formas mencionadas en la evaluación de las distancias.

Figura 4. Evaluación de distancias (vecino más cercano, vecino más lejano, centroides y promedio inter-grupos).



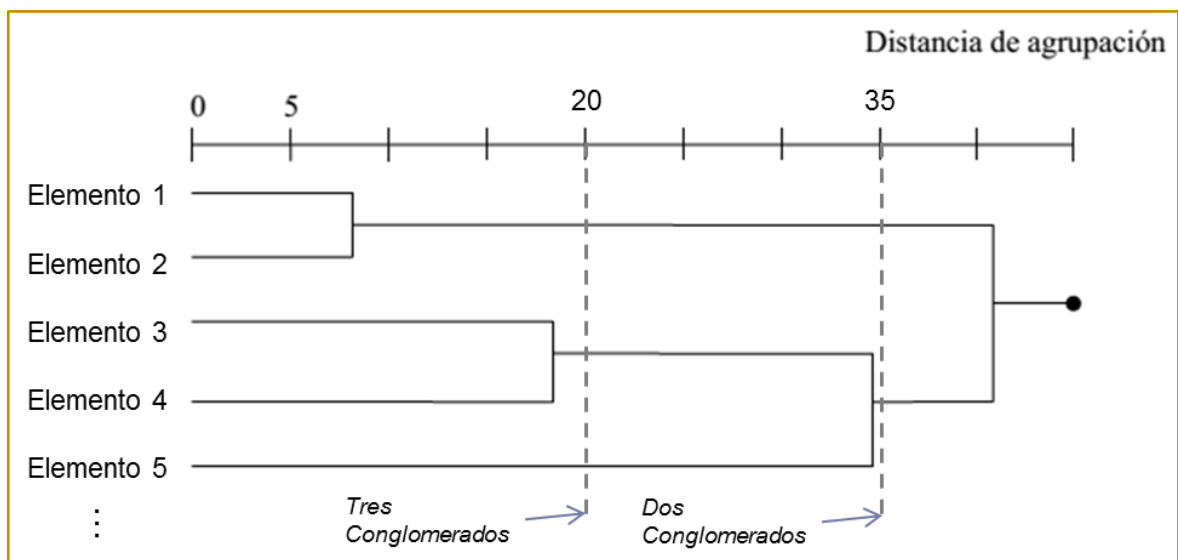
Fuente: Elaboración propia.

La representación gráfica del resultado de la agrupación jerárquica es el dendrograma (ver figura 5). En él, los conglomerados están representados mediante trazos horizontales y las etapas de la fusión mediante trazos

verticales. La separación entre etapas de la fusión es proporcional a la distancia a la que se están fundiendo los elementos en esa etapa. Por ejemplo, a una distancia de “35 unidades” se crearían dos conglomerados (uno con los elementos 1 y 2, y otro con los elementos 3, 4 y 5)

El dendrograma es útil para evaluar la homogeneidad de los conglomerados y facilita la decisión sobre el número óptimo de conglomerados para una cantidad de elementos pequeña (menos de 100), dado que para un mayor número de elementos (miles), la representación gráfica del dendrograma se vuelve inmanejable.

Figura 5. Dendrograma. Representación gráfica de una clasificación jerárquica.



Fuente: Elaboración propia.

b. Métodos No Jerárquicos. Este tipo de método no requiere que dos observaciones en el mismo conglomerado permanezcan unidas; es decir, se elige una partición inicial de individuos y después se intercambian los miembros de estos clusters para obtener una mejor partición. De esta

manera imponen una estructura menos rígida en los datos que los métodos jerárquicos. Los métodos no jerárquicos conviene utilizarlos cuando los datos a clasificar son muchos (cientos o miles) y/o para refinar una clasificación obtenida utilizando un método jerárquico. Supone que el número de grupos es conocido a priori.

El método de conglomerados no jerárquicos más popular es el *K-medias* (o algoritmo K-medias). La “K” es el nombre que se deriva del hecho de que para una ejecución específica el analista/investigador debe definir el número de conglomerados (K) adecuado. Generalmente se realizan varios intentos, cada uno con un número diferente de conglomerados, y los resultados son evaluados con criterios como separación, tamaño del grupo, patrón de medidas y validación. La porción “medias” en el nombre se refiere al hecho de que la media (o el centroide) de observaciones en un conglomerado representa al conglomerado.

Cuando se conglomeran muchas observaciones, a menudo el método k-medias es el elegido, aunque se puede tomar una muestra de un archivo grande y aplicar un método jerárquico, o ejecutar un *Algoritmo Bietápico*, que también funciona eficientemente con grandes cantidades de observaciones/elementos.

En el *algoritmo bietápico* hace referencia a dos pasos (o fases), uno que se puede llamar pre-agrupamiento y el otro agrupamiento. Durante el primer paso los individuos se agrupan en pre-conglomerados en una “pasada simple” de datos. Después, en el paso de agrupamiento se utiliza un método jerárquico para “fundir progresivamente” los pre-conglomerados. Normalmente, en cada etapa de este paso se registra un estadístico de bondad de ajuste (entre los más usados están el AIC-Criterio de Información Akaike- y el BIC –Criterio Bayesiano de Schwartz-), indicando

cuánto se ajustan los datos al número actual de conglomerados. El algoritmo examina el tamaño relativo de los sucesivos cambios en la medición de los criterios (si el cambio es pequeño, significa que ocurre una mejora relativamente pequeña a medida que se agrega un conglomerado adicional), luego se realiza una búsqueda más refinada de conglomerados.

Adicional a lo anterior, comparado con otros algoritmos de clasificación, el bietápico tiene varias ventajas potenciales:

- Puede realizarse en una pasada única de datos y por lo tanto, no requerir una gran cantidad de memoria de almacenamiento.
- Puede seleccionar automáticamente el número de conglomerados en base a criterios estadísticos, a diferencia de otros algoritmos revisados.
- Puede incorporar variables categóricas y continuas en su algoritmo de conglomeración tomando en cuenta sus distintas propiedades.

No obstante, se debe tener especial cuidado con los supuestos, dado que el bietápico presume que las variables continuas del conglomerado no están correlacionadas y siguen distribuciones normales. Además presume que las variables categóricas siguen distribuciones multinomiales y que son independientes unas de otras, así como independientes de las predictoras continuas.

Como se mencionó anteriormente, para archivos con gran volumen de casos, los métodos de k-medias y bietápico son métodos eficientes a la hora de tratar estas bases desde el punto computacional.

Para archivos de datos más pequeños (especialmente menos de 100 casos), varios métodos jerárquicos (vinculación completa, promedio inter-grupos, método

de Ward) funcionan bastante bien y los resultados pueden mostrarse usando el dendrograma. En esta situación, el procedimiento bietápico ofrece menos medidas de distancia pero tiene la ventaja de seleccionar automáticamente el número de conglomerados.

La distancia *Euclídea cuadrada* es muy común en los estudios de conglomerados, aunque el bietápico usa la distancia log-verosimilitud para acomodar las variables categóricas y continuas. Si se están agrupando variables con escalas muy diferentes, entonces la estandarización de algún tipo (generalmente puntuaciones z) se realiza sobre las variables. Generalmente, el método de conglomerado influye sobre la solución más que la medida de distancia elegida. Si la escala (desviación estándar) varía sustancialmente en las variables, luego de la estandarización puede influir fuertemente en la solución (no es así cuando se usa la distancia log-verosimilitud en el bietápico).

1.2.6 Tipos de segmentación. Una de las tantas aplicaciones que ofrece la segmentación, se da en el campo del mercadeo, donde se pueden agrupar los clientes de tal forma que se diseñen productos y estrategias dirigidas (específicas para cada segmento) según sus características.

No existe un criterio único para segmentar los clientes, el investigador debe probar y analizar diferentes variables de segmentación; sin embargo, Kotler y Armstrong (2013) describen variables que pueden usarse en segmentación, agrupándolas en variables: geográficas, demográficas, sicográficas y conductuales.

- La **segmentación geográfica** requiere dividir al mercado en diferentes unidades geográficas, como naciones, regiones, departamentos, municipios, o incluso barrios.

- La **segmentación demográfica** divide el mercado en segmentos con base en variables como la edad, etapa del ciclo de vida, género, ingresos, ocupación, educación, religión, origen étnico y generación. Los factores demográficos son las bases más populares para la segmentación de grupos de clientes. Una razón es que las necesidades del consumidor, sus deseos y tasas de utilización a menudo varían estrechamente con las variables demográficas. Otra razón es que las variables demográficas son más fáciles de medir que la mayoría de los demás tipos de variables.
- La **segmentación sicográfica** divide a los compradores en diferentes segmentos con base en características como la personalidad, el estilo de vida o la clase social. Las personas del mismo grupo demográfico pueden tener características sicográficas muy distintas.
- La **segmentación conductual** divide a los compradores en segmentos basados en sus conocimientos, actitudes, usos o respuestas a un producto. Muchos investigadores creen que las variables de comportamiento son el mejor punto de partida para la creación de segmentos del mercado.

CAPÍTULO 2. SELECCIÓN, DEPURACIÓN Y DESCRIPCIÓN DE LAS VARIABLES

2.1 SELECCIÓN DE LAS VARIABLES

Como se mencionó en un principio, no existe un criterio único para la selección de las variables. Para el presente trabajo, la selección de variables obedece al conocimiento del mercado, la información que La Cooperativa tiene de sus clientes en las bases de datos y la experiencia en el manejo de las mismas.

Una de las preguntas que el investigador debe hacerse es si ha seleccionado las variables correctas y si éstas son relevantes para la agrupación que se busca.

Otra inquietud se presenta con el número de variables a tener en cuenta. Muchas variables complican el cálculo de las similitudes y pocas variables podrían aportar poco o nada a la solución final.

2.1.1 Definición de las variables

2.1.1.1 Variables demográficas y socioeconómicas

- **Género.** Se definen dos categorías: Masculino y Femenino
- **Estrato.** Corresponde al estrato socioeconómico en el cual se ubica la residencia del asociado.

- **Edad.** Corresponde al número de años del asociado según la fecha de nacimiento registrada en el documento correspondiente.
- **Ingresos.** Es el valor total en pesos que el asociado recibe mensualmente por su actividad económica, entre otras fuentes.
- **Activos.** Es el valor en pesos que el asociado manifiesta tener en el total de sus activos.
- **Ocupación.** Hace referencia a la principal actividad económica del asociado, de donde normalmente provienen los ingresos.
- **Escolaridad.** Corresponde al máximo nivel de preparación educativa que ha recibido el asociado. Ejemplo: Primaria, Secundaria, Universitaria, etc.

2.1.1.2 Variables conductuales

- **Saldo Aportes.** Es el valor total que el asociado tenía en su cuenta de aportes en el mes de marzo de 2016.
- **Saldo Ahorros.** Es el valor total que el asociado tenía en sus cuentas de ahorros en el mes de marzo de 2016. Se incluyen los siguientes productos: Ahorro a la vista⁶, Ahorro Especial⁷ y Comultracheque⁸.

⁶ Cuenta de ahorros que le permite al asociado consignar y retirar en efectivo o cheque, con disponibilidad inmediata de sus saldos.

⁷ Cuenta de ahorros que permite realizar transacciones a través de comprobantes de retiro especiales utilizados directamente o girados a un autorizado transitorio para el pago a proveedores o de nómina.

⁸ Cuenta de ahorros que ofrece los beneficios de una cuenta corriente, como realizar consignaciones, además de disponer del saldo mediante el giro de cheques.

- **Saldo PAP⁹**. Es el valor total que el asociado tenía en sus cuentas de Ahorro Programado en el mes de marzo de 2016.
- **Saldo CDAT¹⁰**. Es el valor total que el asociado tenía en sus Certificados de Depósito a Término –CDAT- en el mes de marzo de 2016.
- **Saldo Crédito**. Es el valor total pendiente por pagar del crédito o créditos que el asociado tenía con La Cooperativa en el mes de marzo de 2016.
- **Tiene Ahorros**. Especifica si el asociado tenía por lo menos una cuenta de ahorros (entre ahorro a la vista, ahorro especial y comultracheque) en el mes de marzo de 2016.
- **Tiene PAP**. Especifica si el asociado tenía por lo menos una cuenta de ahorro programado en el mes de marzo de 2016.
- **Tiene CDAT**. Especifica si el asociado tenía por lo menos un Certificado de Depósito a Término –CDAT- en el mes de marzo de 2016.
- **Tiene Crédito**. Especifica si el asociado tenía por lo menos un crédito vigente en el mes de marzo de 2016.
- **Antigüedad**. Cantidad de años que la persona tiene como asociada a La Cooperativa, contados a partir de la fecha de ingreso.

⁹ PAP. Es una modalidad de ahorro en la cual el asociado puede seleccionar el plazo, monto y periodicidad de consignación.

¹⁰ CDAT o certificado de depósito de ahorro a término. Es una opción de inversión a plazo fijo.

- **Diversificación.** Hace referencia a la diversificación en productos. Se identifican los asociados que solo tenían aportes y aquellos asociados que tenían uno, dos, tres o más tipos de productos en el mes de marzo de 2016.
- **Clasificación Crédito.** Dependiendo de la naturaleza del crédito, éste se clasifica en: Microcrédito, Consumo, Comercial y Vivienda.
- **Perfil Asociado.** Dependiendo del tipo y saldo de los productos que el asociado tenía a marzo de 2016 se clasifican en tres categorías: Solo aportes, Captadores y Deudores. Donde “Solo aportes” son aquellos asociados que no tiene vigente otro producto diferente a Aportes, “Captadores” son aquellos asociados cuya sumatoria de saldos en productos de captación es superior al saldo de los productos de crédito, y “Deudores” lo conforman aquellos asociados cuyo saldo de crédito es superior a la sumatoria de saldos de sus productos de captación.

2.2 DEPURACIÓN DE LAS VARIABLES

El número de asociados persona natural que tiene La Cooperativa es de 389.699 (a corte de marzo de 2016). Sin embargo, para el desarrollo de la segmentación se decidió realizar una depuración de las variables buscando una mayor consistencia en los datos.

Los siguientes fueron los “criterios filtro” o de eliminación empleados:

1. Menores de edad. La mayoría de información que aportan corresponde a los padres o tutores de los menores. Se eliminaron 29.925 casos.

2. Mayores de 80 años. Dado que no son mercado objetivo de La Cooperativa para la oferta de productos y servicios. Se eliminaron 8.593 casos.
3. Sin estrato. Asociados que no presentan información de estrato dentro de la base de datos. Se eliminaron 51.387 casos.
4. Residencia fuera de los departamentos. No se tiene en cuenta asociados cuyo lugar de residencia esté fuera de los seis departamentos donde la cooperativa hace presencia física. Se eliminaron 3.169 casos.
5. Empleados de La Cooperativa. Dado que no son mercado objetivo. Se eliminaron 1.088 casos.
6. Personas a cargo. Se toma como error de interpretación o digitación aquellos asociados con 10 o más personas a cargo. Se eliminaron 126 casos.
7. Ingreso y activos inferiores. Se filtraron los asociados que mostraban ingresos y/o activos entre \$1 y \$9.999, dado que esta cifra no tiene sentido. Se eliminaron 6.404 casos.
8. Ingresos y activos superiores. Se filtraron los asociados que registraran ingresos mensuales superiores a \$45.000.000 y/o activos superiores a \$520.000.000. Se toman como datos atípicos. Se eliminan 5.613 casos (menos del 1,5% del total de la base).
9. Saldo muy superior en los productos. Se filtraron asociados cuyo saldo en productos estuviera cinco desviaciones estándar por encima de la media en cada producto, de la siguiente forma: 1.503 casos por Aportes, 1.232 casos por Ahorros, 556 casos por PAP, 88 casos por CDAT y 529 casos por

crédito. En total con este filtro se eliminaron 3.908 casos (menos del 1% del total de la base).

10. Desactualizados. Con el objetivo de tener datos más precisos y concisos, se filtraron aquellos asociados que al mes de marzo de 2016 presentaran más de un año de desactualización en la base de datos. Se eliminaron 154.531 casos.

Después de realizados los filtros y la depuración de los datos, la base de asociados con la cual se desarrollará el análisis de conglomerados es de: 124.955.

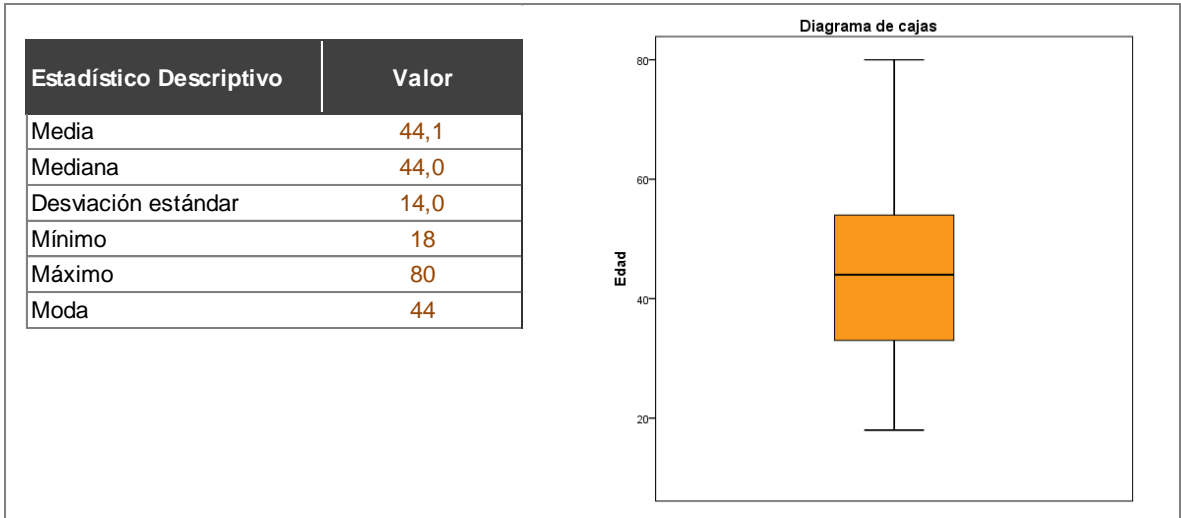
2.3 ANÁLISIS DESCRIPTIVO

Con la base de datos depurada, se procede a profundizar la información que se tiene de las variables mediante un análisis descriptivo de las mismas, utilizando tablas de frecuencia, diagramas de barras, diagramas de cajas y los principales estadísticos descriptivos. Para facilitar la visualización, el análisis se divide en variables cuantitativas y variables cualitativas.

2.3.1 Variables cuantitativas

2.3.1.1 Edad

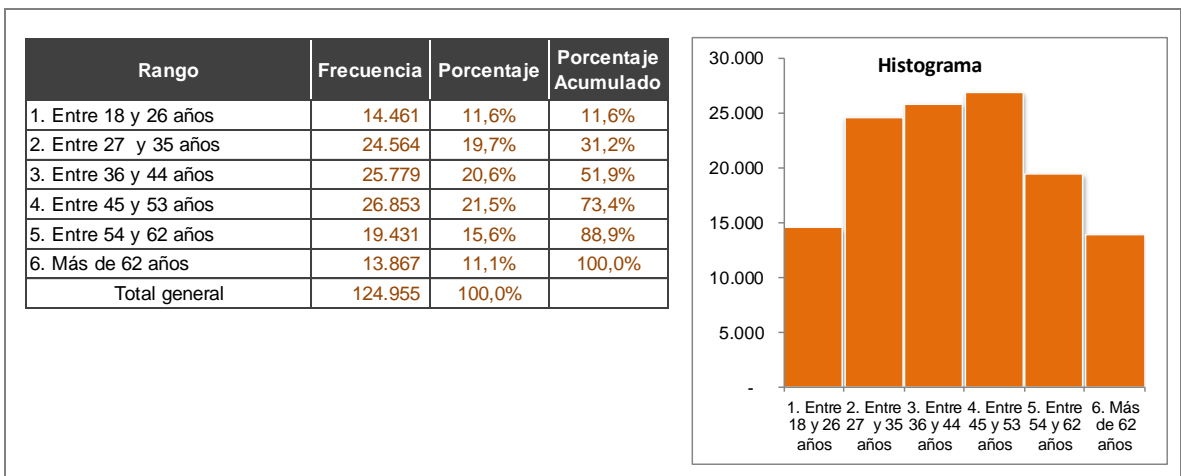
Figura 6. Tabla estadísticos descriptivos y diagrama de caja. Variable Edad.



Fuente: La Cooperativa, base de datos con corte a marzo de 2016

La edad de los asociados de La Cooperativa oscila entre 18 y 80 años con media de 44,1 años muy cercana a la mediana (44 años), lo que evidencia simetría. No se observan datos atípicos. Recordar que la edad fue una variable que se utilizó en la depuración (tomando mayores de edad y menores de 80 años).

Figura 7. Tabla de frecuencias e histograma. Variable Edad.

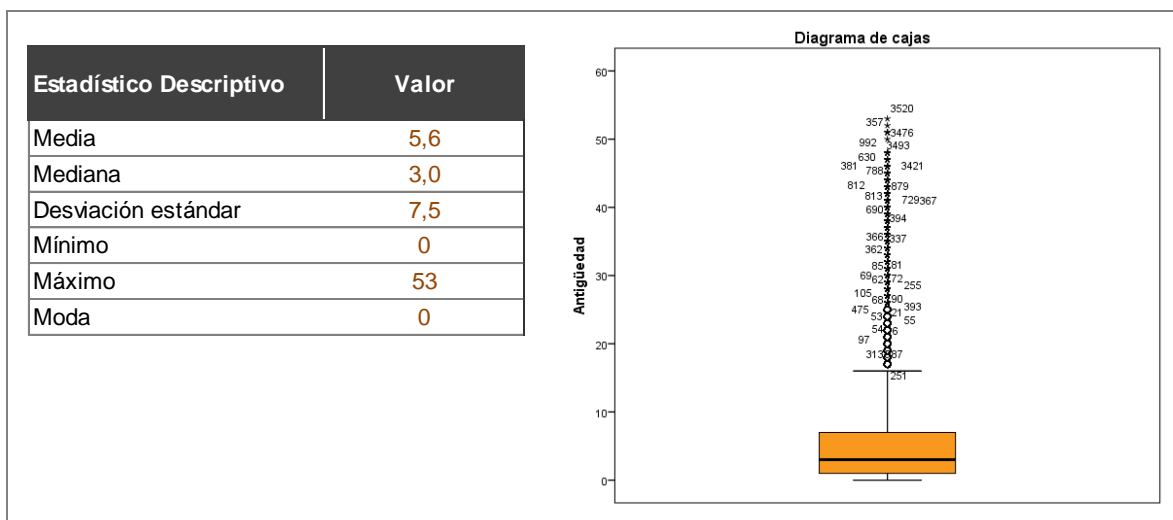


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 62% de los asociados tienen entre 27 y 53 años de edad. El valor de la asimetría es de 0,22; confirmando que la distribución es simétrica (tomando como simetría valores cercanos a +/- 0,5), aunque el coeficiente de curtosis es de -0,73 (indicando una distribución platicúrtica o “achatada”). El coeficiente de variación es de 0,3 lo que indica alta dispersión respecto a la media.

2.3.1.2 Antigüedad

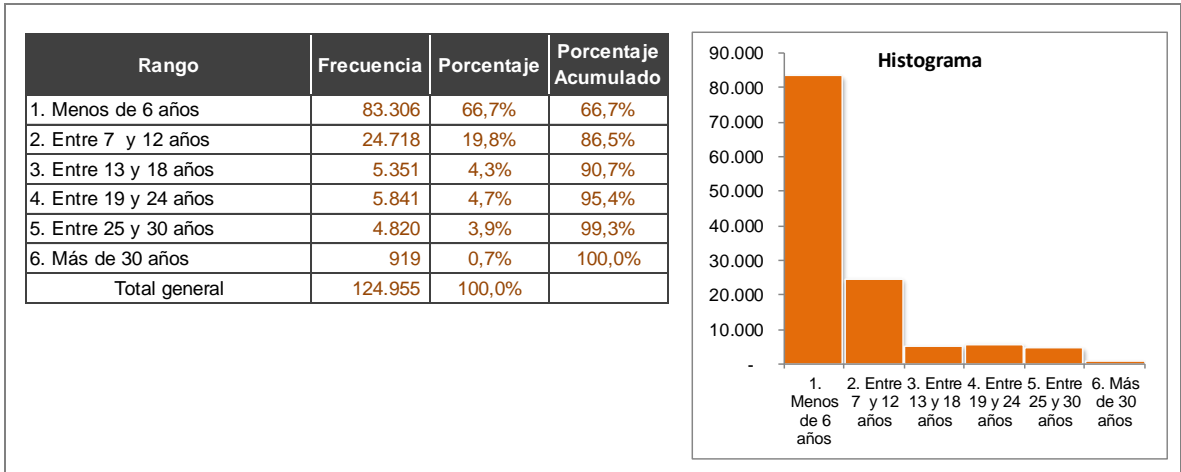
Figura 8. Tabla estadísticos descriptivos y diagrama de caja. Variable Antigüedad.



Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

Dado que La Cooperativa a marzo de 2016 tiene 53 años de creada, la antigüedad de los asociados oscila entre 0 años (menos de un año) y 53 años, con una media de 5,6 años y mediana de 3 años, es decir presenta una asimetría positiva. Aunque en el diagrama de caja se observan puntos atípicos, éstos son necesarios para el análisis.

Figura 9. Tabla de frecuencias e histograma. Variable Antigüedad.

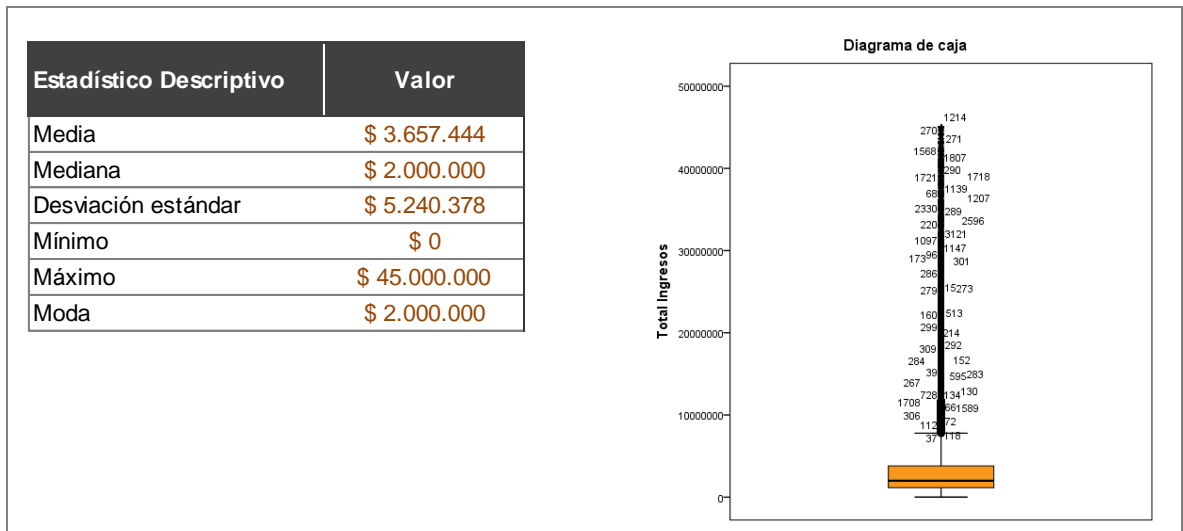


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 67% de los asociados tienen menos de 6 años de antigüedad, esto se debe principalmente a dos razones, por una parte la expansión de La Cooperativa en los últimos años (nuevos asociados) y por otra parte, los asociados antiguos no suelen actualizar datos (recordar que la actualización fue un criterio de selección).

2.3.1.3 Total Ingresos

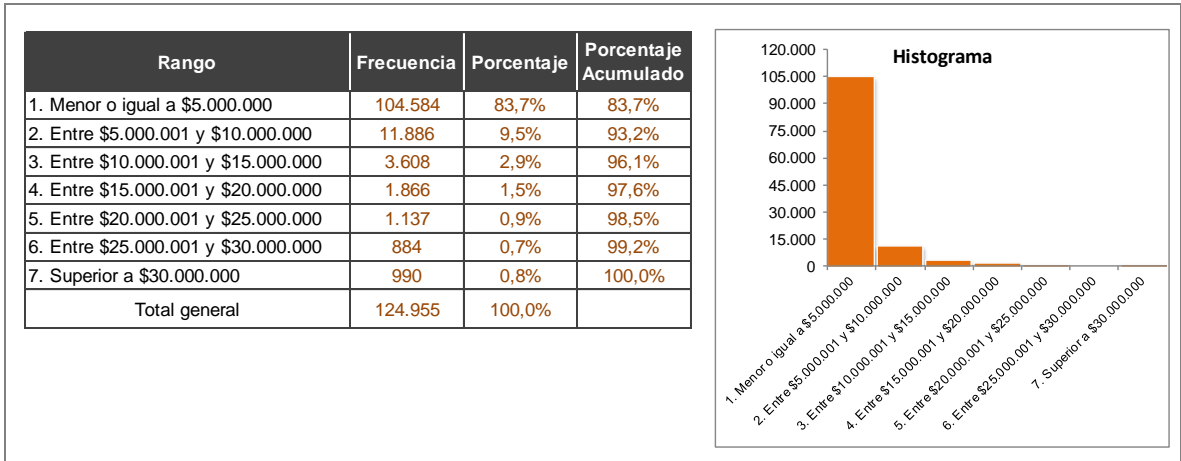
Figura 10. Tabla estadísticos descriptivos y diagrama de caja. Variable Total Ingresos.



Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

La media del total ingresos de los asociados es de \$3.657.000 con desviación estándar de \$5.240.000. Esta desviación es muy alta debido a las diferentes actividades económicas de los asociados, que van desde estudiantes y amas de casa (algunos con ingreso igual a \$0), hasta empleados, comerciantes y microempresarios. Tanto la mediana como la moda tienen el mismo valor de \$2.000.000.

Figura 11. Tabla de frecuencias e histograma. Variable Total Ingresos.

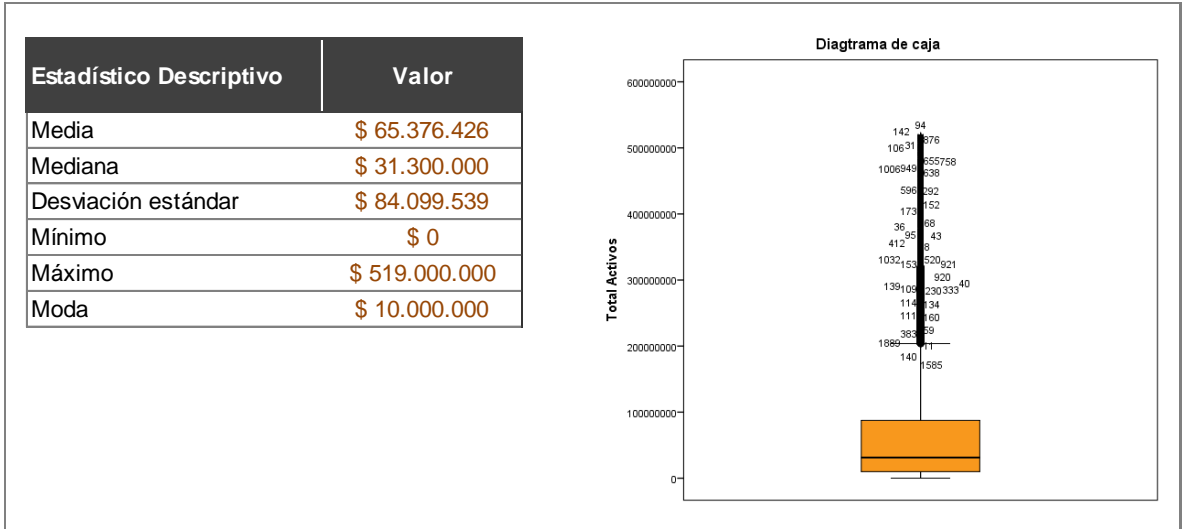


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 84% de los asociados tienen ingresos menores o iguales a \$5.000.000 mensuales, y tan solo el 7% superan los \$10.000.000 de ingresos mensuales (principalmente trabajadores independientes). El histograma muestra la asimetría positiva.

2.3.1.4 Total Activos

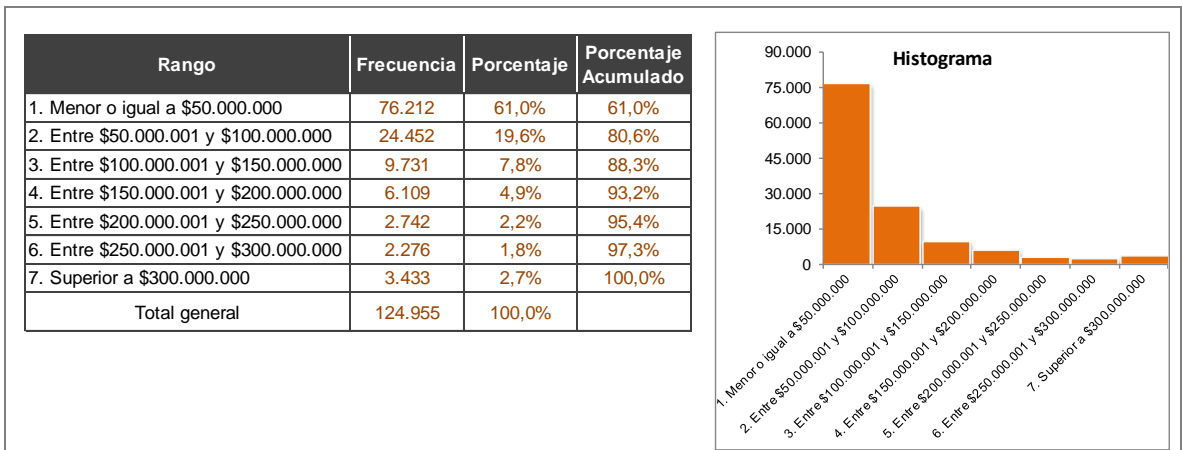
Figura 12. Tabla estadísticos descriptivos y diagrama de caja. Variable Total Activos.



Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

La media del Total Activos es de \$65.476.000 con una desviación estándar \$84.100.000. Esta desviación es alta por las mismas razones explicadas en la variable Total Ingresos. Recordar que para esta variable se aplicó un filtro de eliminación (Total Activos superiores a \$520.000.000).

Figura 13. Tabla de frecuencias e histograma. Variable Total Activos.

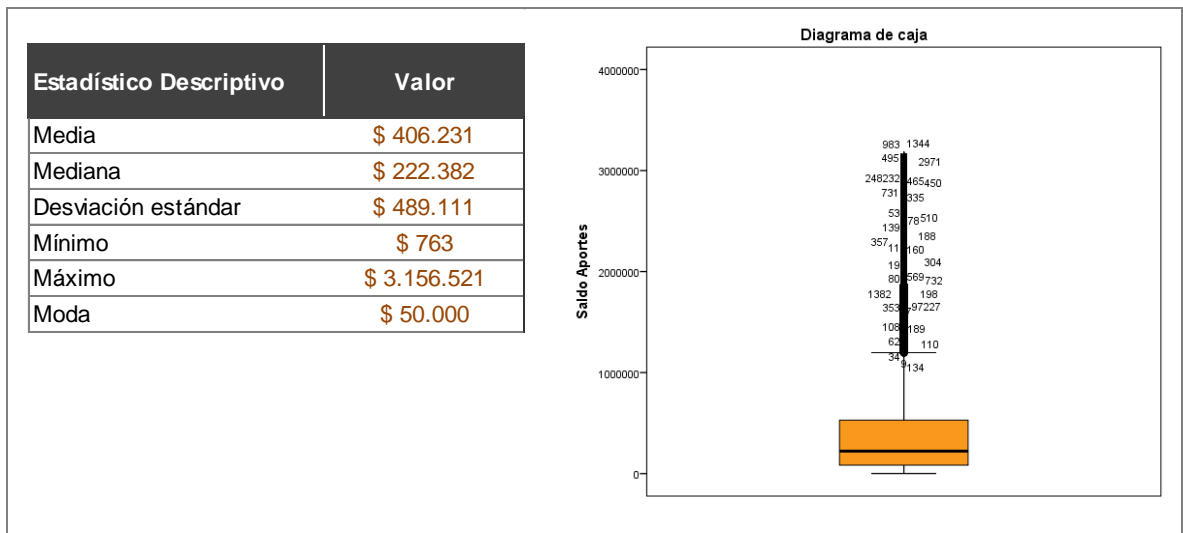


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 81% de los asociados tienen activos iguales o menores a \$100.000.000. Solo el 7% supera los \$200.000.000 en activos, en su mayoría trabajadores independientes (según base de datos de clientes).

2.3.1.5 Saldo Aportes

Figura 14. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo Aportes.

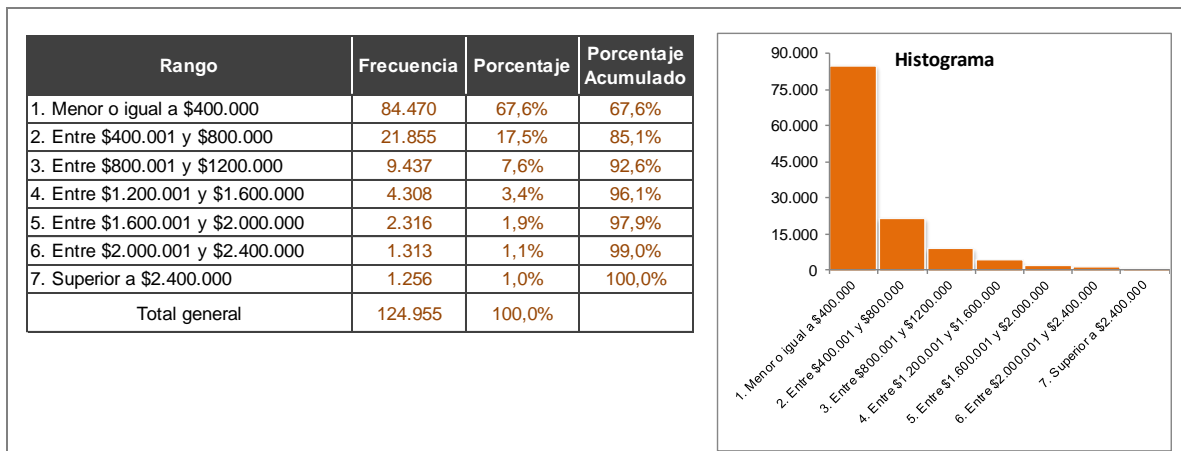


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

La media en la variable Saldo Aportes es igual a \$406.000, superior a la moda y a la mediana, es decir, tiene una asimetría negativa. La desviación estándar es de \$489.000, con un coeficiente de variación igual a 1,2 (de las cinco variables de saldo, los Aportes presentan el menor coeficiente de variación). Esta menor variación se debe en parte porque no es un producto a la vista; es decir, no es habitual que un asociado consigne y retire frecuentemente afectando el saldo del producto.

La moda es igual a \$50.000, que por estatutos de La Cooperativa es el valor mínimo en aportes para ser un “asociado hábil”. El valor máximo de esta base de asociados es de \$3.157.000.

Figura 15. Tabla de frecuencias e histograma. Variable Saldo Aportes.

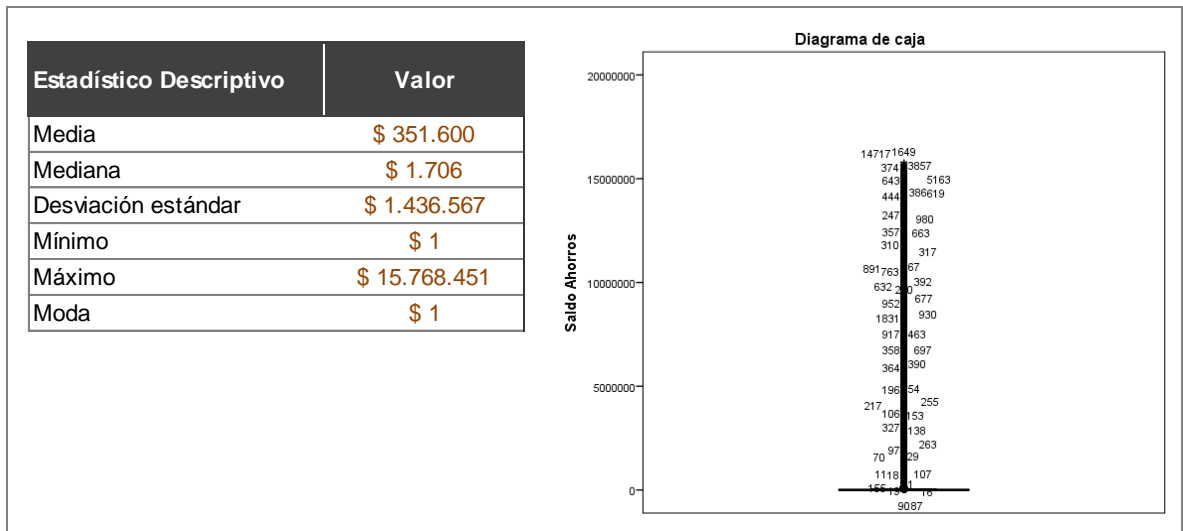


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 85% de los asociados tienen saldo de aportes inferior a los \$800.000. Solo el 2% tienen un saldo superior a los \$2.000.000.

2.3.1.6 Saldo Ahorros

Figura 16. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo Ahorros.

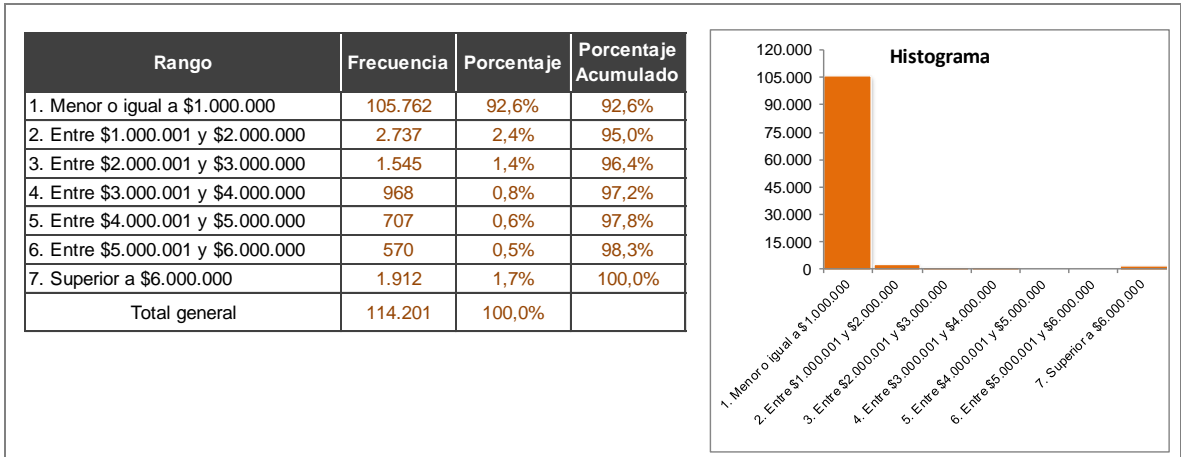


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

De las variables analizadas, el Saldo Ahorros es la que mayor coeficiente de variación presenta (4,1), es decir, la desviación estándar es muy alta comparada con la media. Una explicación de esta variación es que básicamente el producto es ahorro a la vista, esto quiere decir, que el asociado en cualquier momento puede disponer del dinero, retirando y consignado por diferentes canales transaccionales.

Por otra parte, tanto el valor mínimo como la moda es igual a \$1. Lo anterior como resultado de una práctica no adecuada por parte de algunas agencias, en las cuales para no cancelar la cuenta de ahorros del asociado, ésta se deja con valores muy pequeños de saldo.

Figura 17. Tabla de frecuencias e histograma. Variable Saldo Ahorros.

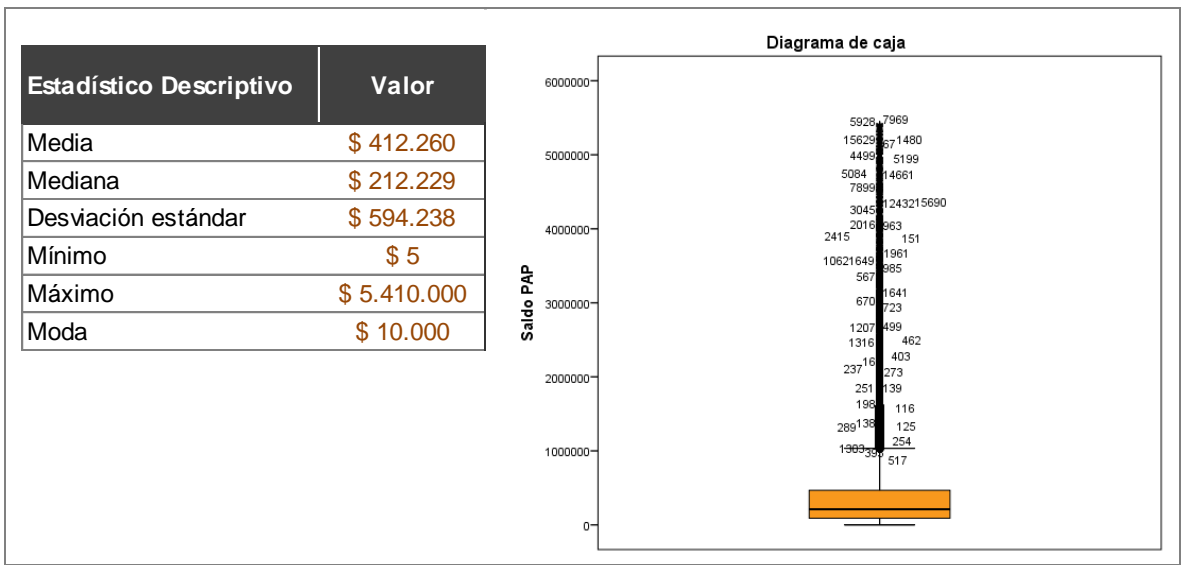


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 93% de los asociados manejan un saldo de ahorros menor o igual a \$1.000.000. Para rangos superiores a \$5.000.000, el porcentaje de participación del saldo no sobrepasa el 2,4%.

2.3.1.7 Saldo PAP

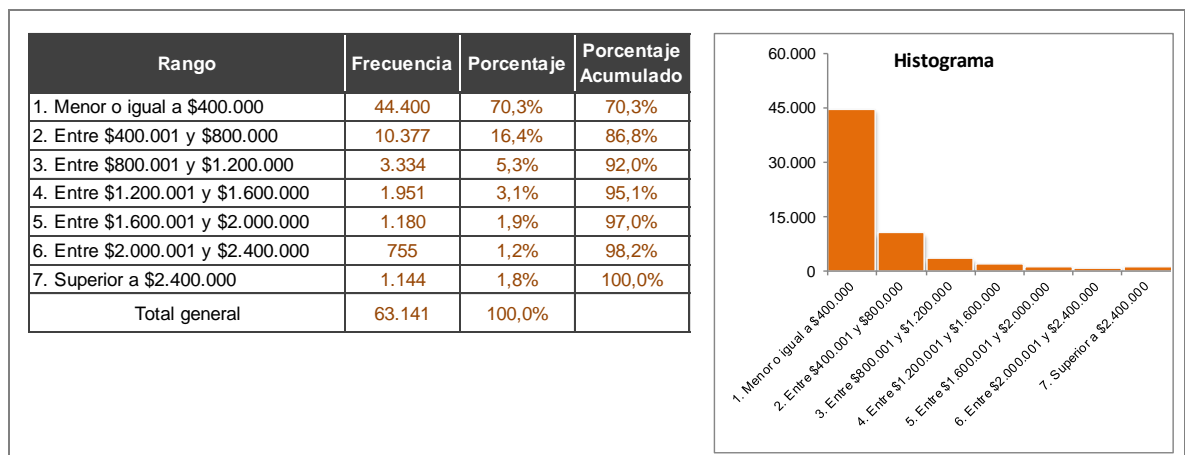
Figura 18. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo PAP.



Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

Este es un producto de ahorro programado, es decir, el asociado determina el plazo del ahorro y el monto mensual a consignar. La media es de \$412.000, con una desviación estándar de \$594.000, es decir tiene un coeficiente de variación igual al 1,4. La moda es igual a \$10.000, valor con el cual normalmente los asociados abren los planes de ahorro.

Figura 19. Tabla de frecuencias e histograma. Variable Saldo PAP.

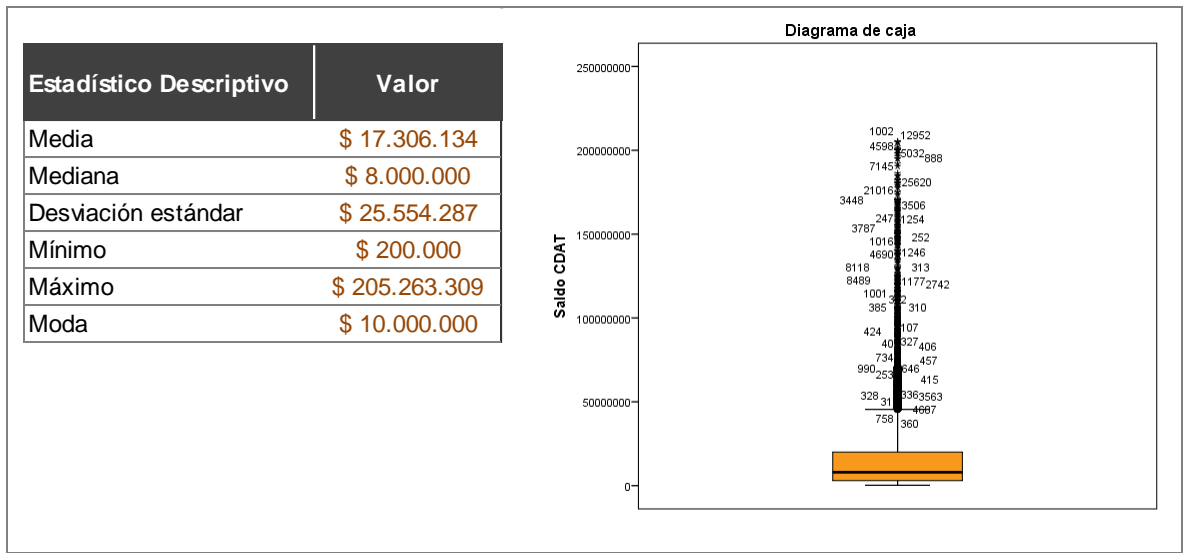


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 51% de los asociados de la base de clientes analizada tienen este producto, de los cuales, el 70% manejan un saldo menor o igual a \$400.000. Solo el 3% de los asociados poseen más de \$2.000.000 en su saldo de PAP.

2.3.1.8 Saldo CDAT

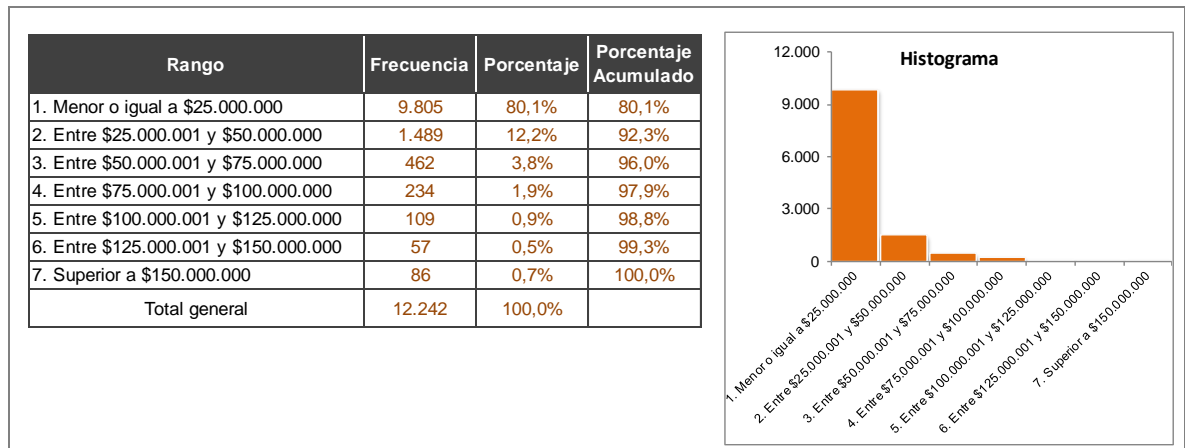
Figura 20. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo CDAT.



Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

Dadas las características de este producto (tasa de interés más alta), es apetecido por los asociados que son inversionistas; de los cuales, algunos son muy representativos convirtiéndose en casos outliers (poco comunes), razón por la cual se realizó depuración en esta variable. La media en el saldo del producto es de \$17.300.000, con una desviación estándar de \$25.554.000. El mínimo es de \$200.000, valor que corresponde al mínimo con el cual se puede abrir el producto.

Figura 21. Tabla de frecuencias e histograma. Variable Saldo CDAT.

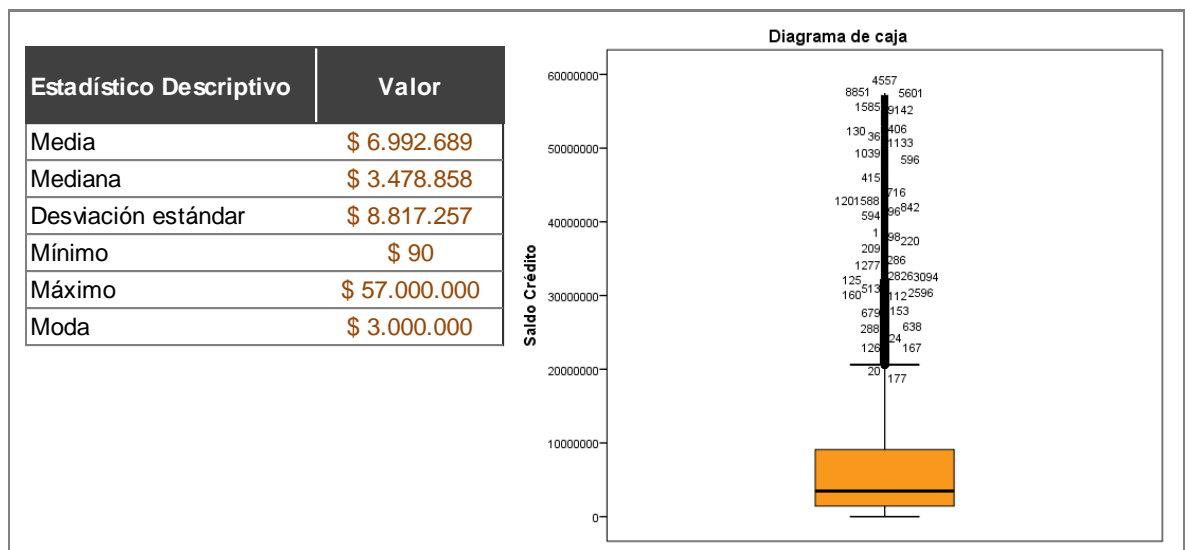


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

Aproximadamente el 10% de los asociados poseen este producto (12.242 de 124.955); de los cuales el 80% tienen un saldo menor o igual a \$25.000.000 y el 2% un saldo superior a \$100.00.000.

2.3.1.9 Saldo Crédito

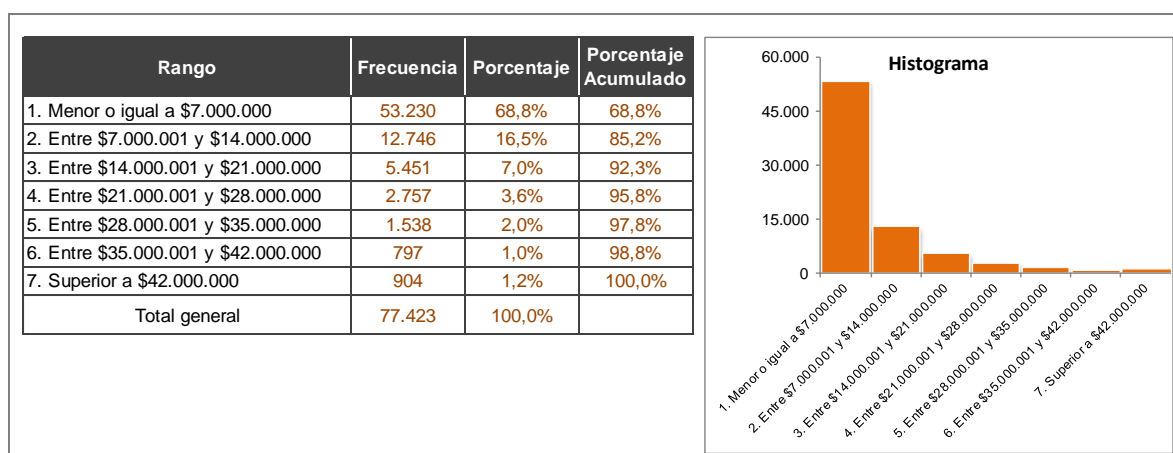
Figura 22. Tabla estadísticos descriptivos y diagrama de caja. Variable Saldo Crédito.



Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

En el saldo de crédito la media es de aproximadamente \$6.993.000 con una mediana y moda cercana a los \$3.000.000. Lo anterior se debe, entre otros, a un enfoque hacia el microcrédito, producto que maneja saldos inferiores a otras líneas de crédito. La desviación estándar es de \$8.817.000, con un coeficiente de variación del 1,3 (valor inferior comparado con otras variables de saldo).

Figura 23. Tabla de frecuencias e histograma. Variable Saldo Crédito.



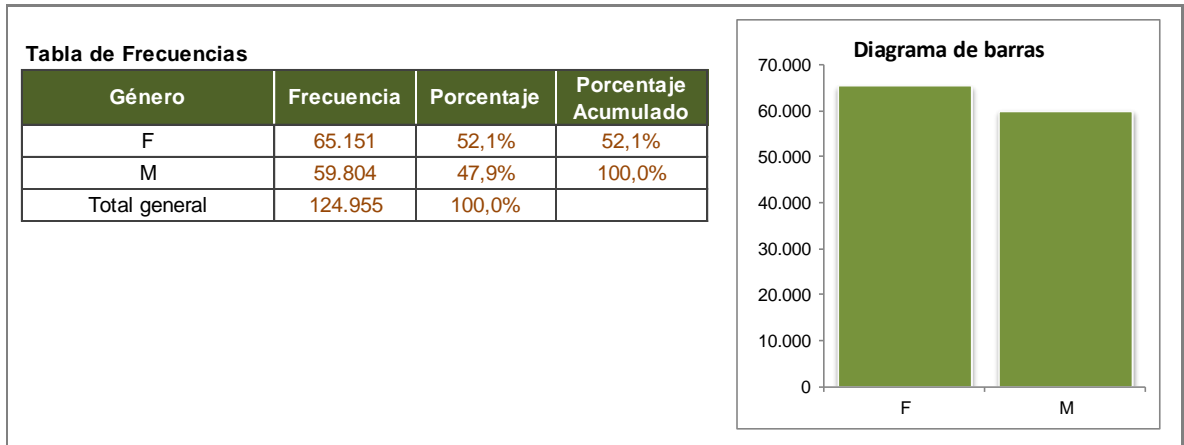
Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 62% de los asociados en este análisis poseen el producto; de los cuales el 85% tienen un saldo inferior a \$21.000.000 en su crédito, y solo el 2% superan los \$35.000.000 en el saldo del producto.

2.3.2 Variables cualitativas

2.3.2.1 Género

Figura 24. Tabla de frecuencias y diagrama de barras. Variable Género.

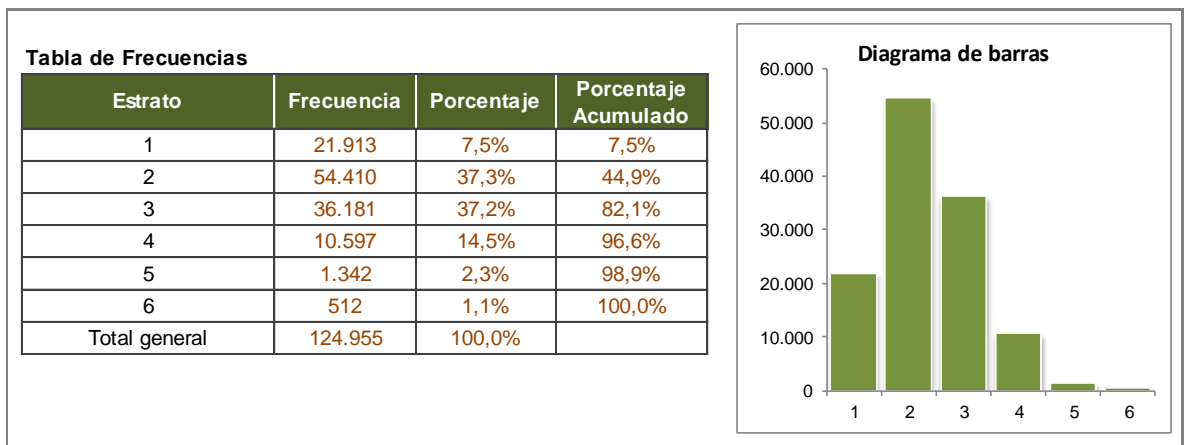


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

Dado que La Cooperativa no tiene un enfoque específico de su portafolio hacia un género en particular, se observa que su participación en el total de asociados es similar, donde el género femenino participa con el 52% y el restante 48% pertenece al género masculino.

2.3.2.2 Estrato

Figura 25. Tabla de frecuencias y diagrama de barras. Variable Estrato.

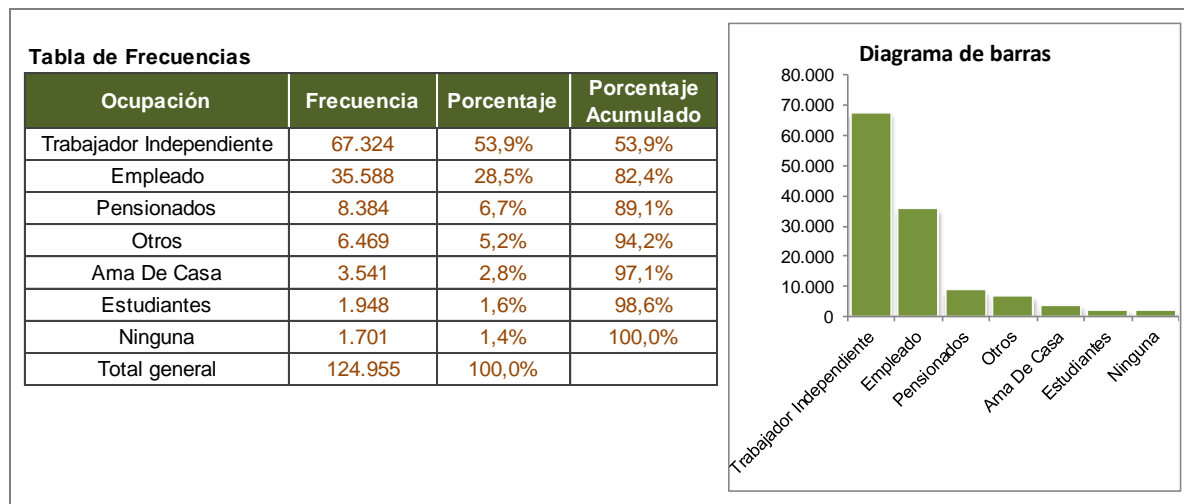


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

Como ya se mencionó, el mercado objetivo de La Cooperativa son las personas naturales, mayores de edad, de estratos 2, 3 y 4, donde precisamente la participación de estos tres estratos suman más del 89% de los asociados.

2.3.2.3 Ocupación

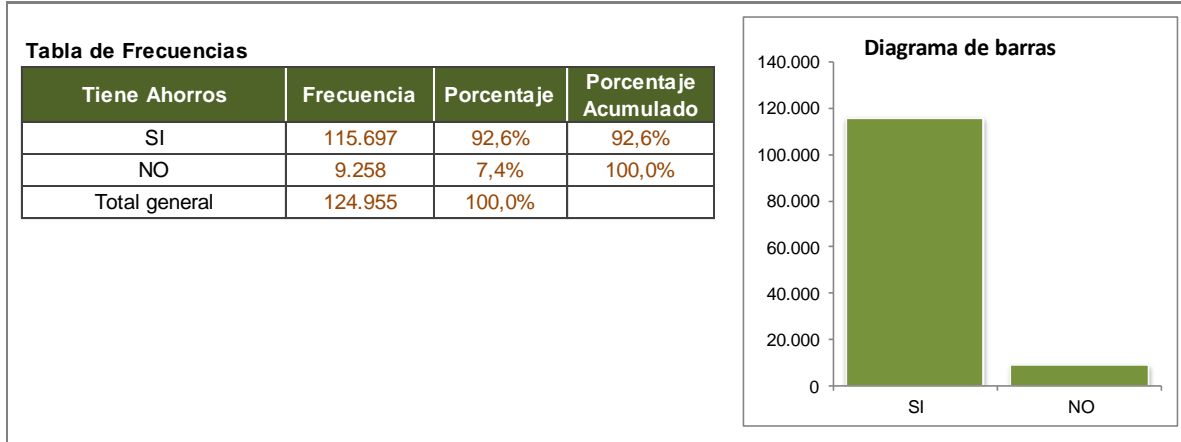
Figura 26. Tabla de frecuencias y diagrama de barras. Variable Ocupación.



El 54% de los asociados personas naturales son trabajadores independientes, principalmente empresarios (de la micro o pequeña empresa). Esto se debe en parte al trabajo realizado por el área de Microfinanzas y de fuerzas comerciales que posee La Cooperativa. Por otra parte, el 35% de los asociados son empleados activos o ya retirados (pensionados). Llama la atención la participación de ocupaciones como “Ama de Casa” o “Estudiante”; aunque no son participaciones altas, representan el 4,4% de los asociados analizados.

2.3.2.4 Tiene Ahorros

Figura 27. Tabla de frecuencias y diagrama de barras. Variable Tiene Ahorros.

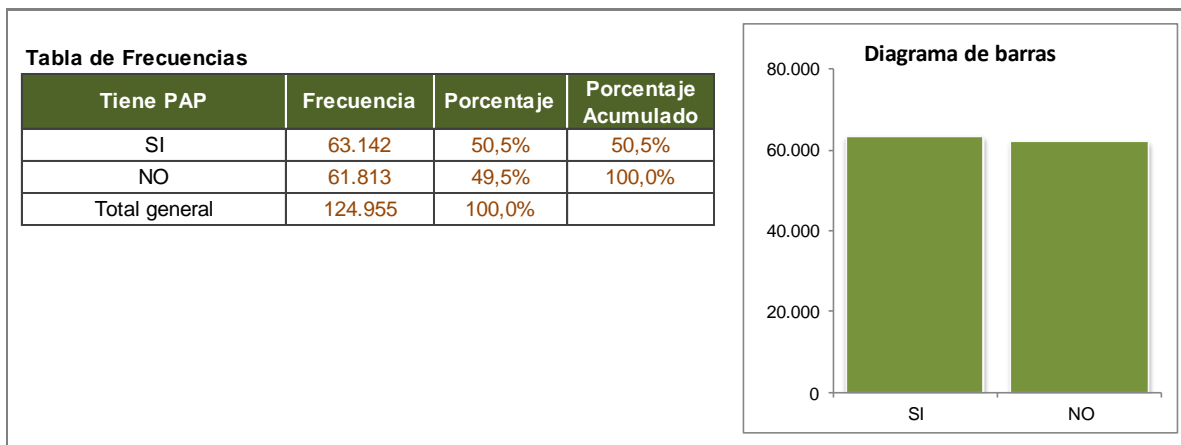


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 93% de los asociados poseen por lo menos una cuenta de ahorros; sin embargo, como se observó en el análisis descriptivo del saldo, la mayoría de cuentas de ahorro poseen saldos muy bajos.

2.3.2.5 Tiene PAP

Figura 28. Tabla de frecuencias y diagrama de barras. Variable Tiene PAP.

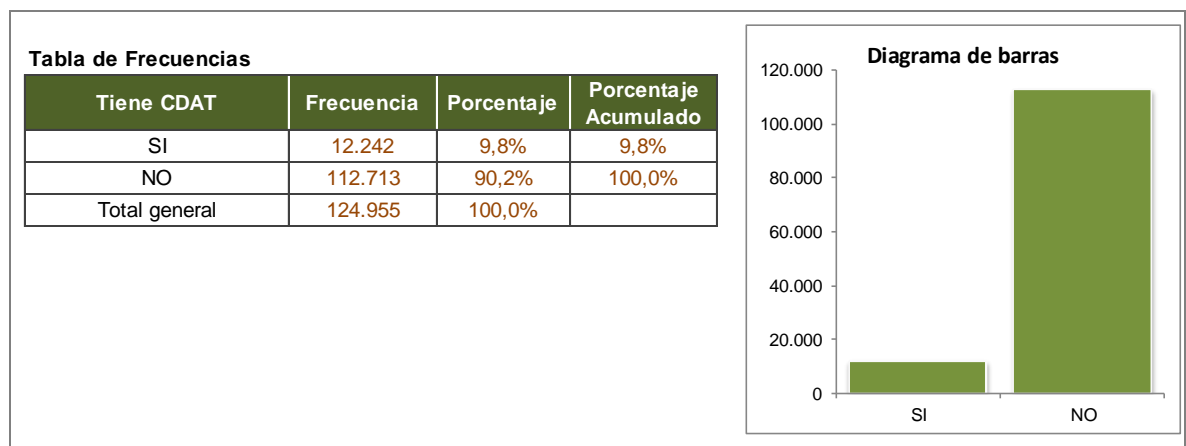


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

La participación entre los asociados que tienen PAP y los que no es muy similar; donde el 50,5% de los asociados tienen el producto y el restante 40,5% no lo tienen.

2.3.2.6 Tiene CDAT

Figura 29. Tabla de frecuencias y diagrama de barras. Variable Tiene CDAT.

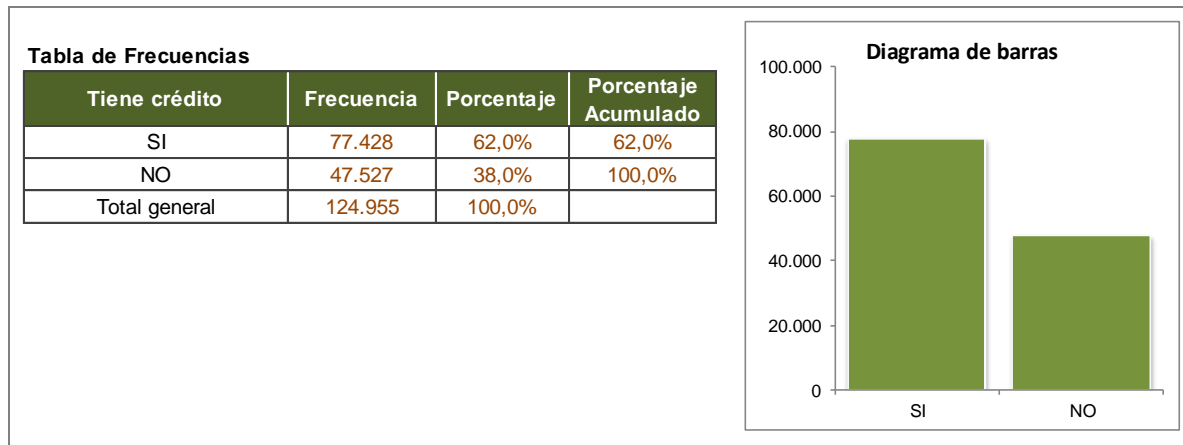


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

Cerca del 10% de los asociados tienen por lo menos un CDAT. De los productos de captación que ofrece La Cooperativa, éste es el producto de menor participación en número pero de mayor participación en saldo.

2.3.2.7 Tiene Crédito

Figura 30. Tabla de frecuencias y diagrama de barras. Variable Tiene Crédito.

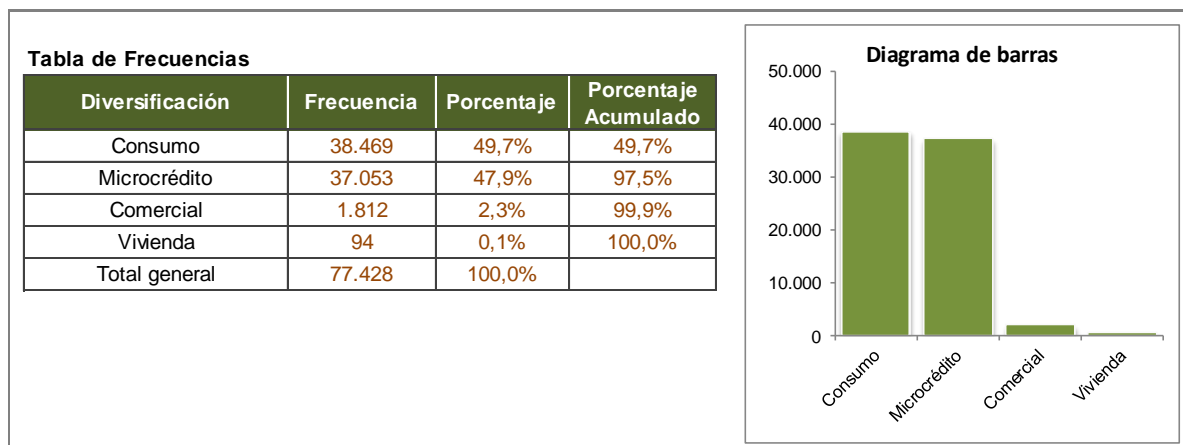


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 62% de los asociados analizados poseen crédito vigente con La Cooperativa. Con la implementación en los últimos años de fuerzas comerciales, éste porcentaje ha venido en incremento.

2.3.2.8 Clasificación Crédito

Figura 31. Tabla de frecuencias y diagrama de barras. Variable Clasificación Crédito.

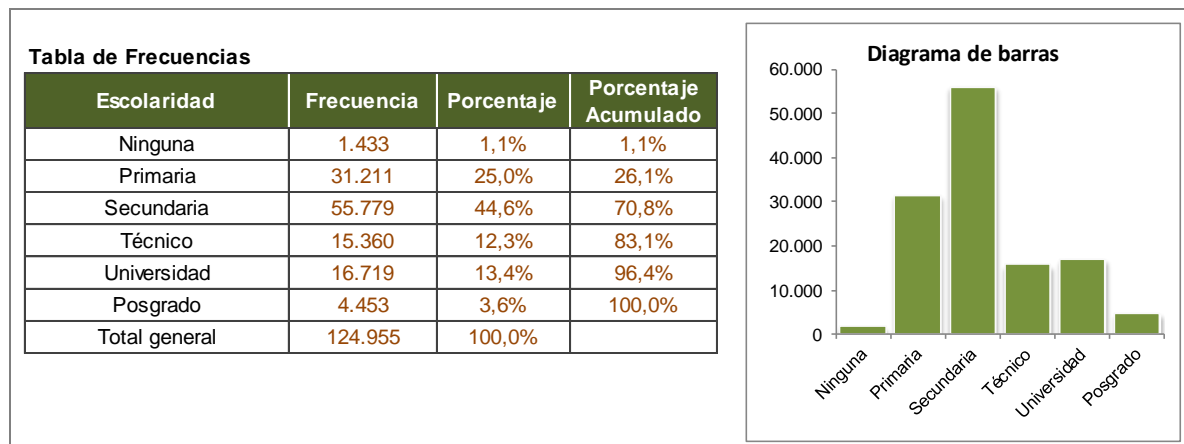


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

En la modalidad “Consumo” se encuentran aproximadamente el 50% de los asociados que tienen crédito. El 48% están clasificados en la modalidad de “Microcrédito”. Pocos asociados tienen crédito “Comercial” (2,3%) y crédito de “Vivienda” (0,1%).

2.3.2.9 Escolaridad.

Figura 32. Tabla de frecuencias y diagrama de barras. Variable Escolaridad.

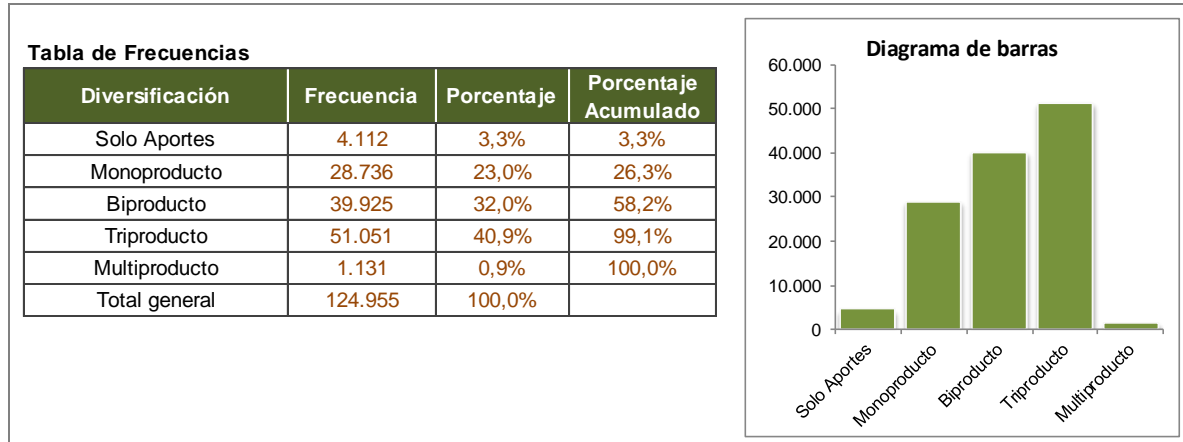


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 70% de los asociados tienen estudio hasta Primaria o Secundaria; el 29% tienen estudios superiores (técnicos, universitario o posgrado). Solo el 1,1% de los asociados manifiesta no tener escolaridad.

2.3.2.10 Diversificación.

Figura 33. Tabla de frecuencias y diagrama de barras. Variable Diversificación.

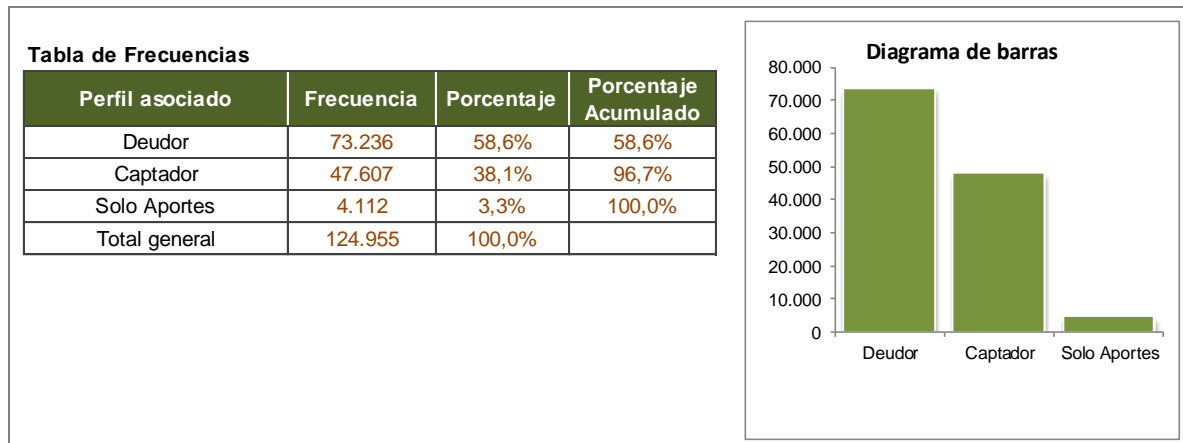


Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 3,3% de los asociados solo tienen el producto de Aportes. El 73% tienen entre 2 y 3 productos (adicional a los Aportes). Menos del 1% manejan más de 3 productos con La Cooperativa.

2.3.2.11 Perfil Asociado

Figura 34. Tabla de frecuencias y diagrama de barras. Variable Perfil Asociado.



Fuente: La Cooperativa, base de datos con corte a marzo de 2016.

El 69% de los asociados tienen un perfil “Deudor”; es decir, su saldo de crédito es mayor al saldo de las captaciones. El 38% de los asociados tienen un perfil “Captador”; es decir, su saldo de captaciones es mayor al saldo del crédito.

2.4 CATEGORIZACIÓN DE VARIABLES

2.4.1 Categorización de variables cuantitativas. Con el objetivo de facilitar el análisis y la posterior interpretación de resultados, se hace necesario categorizar las variables cuantitativas que se emplearán para el análisis de conglomerados.

2.4.1.1 Antigüedad.

Tabla 1. Categorías variable Antigüedad.

Antigüedad	Frecuencia	Porcentaje	Porcentaje Acumulado
1. Menos de 2 años	49.454	39,6%	39,6%
2. Entre 2 y 5 años	33.852	27,1%	66,7%
3. Entre 6 y 15 años	27.706	22,2%	88,8%
4. Más de 15 años	13.943	11,2%	100,0%
Total general	124.955	100%	

Fuente: La Cooperativa, base de datos con corte a marzo de 2016

La antigüedad de los asociados en La Cooperativa se dividió en 4 categorías. En la tabla 1.1 se observa que cerca del 40% de los asociados tienen menos de dos años de antigüedad en La Cooperativa. Las posibles razones fueron expuestas en la sección 2.3. Por otra parte, el 11% tiene más de 15 años de antigüedad. Para esta categorización se tuvo en cuenta los estadísticos descriptivos de las figuras 8 y 9.

2.4.1.2 Total Ingresos.

Tabla 2. Categorías variable Total Ingresos.

Total Ingresos	Frecuencia	Porcentaje	Porcentaje Acumulado
1. Menor a \$700.000	11.789	9,4%	9,4%
2. Entre \$700.000 y \$1.400.000	29.716	23,8%	33,2%
3. Entre \$1.400.001 y \$3.500.000	48.356	38,7%	71,9%
4. Entre \$3.500.001 y \$10.000.000	25.796	20,6%	92,6%
5. Superior a \$10.000.000	9.298	7,4%	100,0%
Total general	124.955	100%	

Fuente: La Cooperativa, base de datos con corte a marzo de 2016

La variable Total Ingresos se dividió en 5 categorías. Se toma como base una aproximación del salario mínimo legal mensual en Colombia para el año 2016 (\$689.454 que se aproxima a \$700.000). En la tabla 1.2 se muestra que solo el 9% de los asociados tienen ingresos menores a \$700.000 (menos de un salario mínimo mensual). El 62% de los asociados registran ingresos totales entre \$700.001 y \$3.500.000.

Solo se categorizan estas dos variables porque son las únicas variables continuas que se utilizarán en el análisis de conglomerados y el perfilamiento final (ver tabla 1.3).

2.5 RESUMEN VARIABLES SELECCIONADAS

A continuación se relacionan las variables inicialmente seleccionadas para el análisis descriptivo (capítulo 2), las variables seleccionadas para el análisis de conglomerados (capítulo 3) y las variables seleccionadas para el perfilamiento y descripción de los conglomerados (capítulos 3 y 4).

Las variables inicialmente seleccionadas fueron las principales encontradas en la base de clientes que posee La Cooperativa (y con la información más completa). Las variables usadas en el análisis de conglomerados, se seleccionaron después de realizar análisis con diferentes algoritmos de clasificación, donde se encontró que podían tener alguna influencia en la definición de conglomerados. Finalmente, para el perfilamiento (descripción) de los conglomerados se utilizaron las mismas variables del análisis cluster y adicional otras variables que apoyan este perfilamiento (por esta misma razón en esta etapa no se tuvieron en cuenta variables iniciales como Género y Edad).

Tabla 3. Resumen variables seleccionadas

Total Variables Analizadas	Variables Usadas en el Análisis de Conglomerados	Variables Usadas en el Perfilamiento
Tiene Ahorros Tiene PAP Tiene CDAT Tiene Crédito Antigüedad Estrato Total Ingresos Ocupación Escolaridad Saldo Aportes Saldo Ahorros Saldo PAP Saldo CDAT Saldo Crédito Clasificación Crédito Diversificación Perfil Asociado Género Edad Total Activos	Tiene Ahorros Tiene PAP Tiene CDAT Tiene Crédito Antigüedad Estrato Total Ingresos Ocupación	Tiene Ahorros Tiene PAP Tiene CDAT Tiene Crédito Antigüedad Estrato Total Ingresos Ocupación Escolaridad Clasificación Crédito Diversificación Perfil Asociado

Fuente: Elaboración propia.

CAPÍTULO 3. ANÁLISIS DE CONGLOMERADOS

Una vez seleccionadas, depuradas y definidas las variables, se procede a definir las medidas de distancia y similitud, analizando la posibilidad de estandarización de variables y posteriormente la selección del algoritmo de clasificación.

3.1 SELECCIÓN DEL ALGORITMO DE CLASIFICACIÓN

Para computar las distancias, todas las variables del conglomerado deben tener valores válidos. Los algoritmos de clasificación jerárquicos y el procedimiento bietápico requieren de datos completos, mientras el algoritmo de k-medias tiene la opción que permite datos perdidos.

Entre las variables inicialmente seleccionadas, se encuentran cinco que hacen referencia al saldo de los productos. Dado que no todos los asociados tienen todos los productos, estas variables carecen de datos en muchos casos (las filas que representan los asociados), razón por la cual los algoritmos los definen como “valores perdidos” y no se considerarían los datos de estos casos dentro del análisis. Al respecto se tienen dos opciones a considerar:

Opción 1. Usar método de K-medias. Este método se empleó con variables que involucran saldos, en algunas ocasiones estandarizando y en otras no; sin embargo, los resultados obtenidos no muestran conglomerados claramente diferenciados, con lo cual no se cumpliría con el objetivo del estudio.

Opción 2. No usar las variables con altas concentraciones de datos perdidos. Se acogió esta opción, razón por la cual en los algoritmos de clasificación no se emplearán variables de saldos de productos, pero sí las otras variables conductuales que hacen referencia a la tenencia de productos. Las variables de saldo se utilizarán más adelante para la descripción de los conglomerados.

Por otra parte, los algoritmos de clasificación jerárquicos no soportan grandes volúmenes de casos. Al ingresar la base de datos seleccionada de La Cooperativa (con 124.955 casos) no se obtiene una respuesta dado que el número de casillas excede la capacidad de la herramienta.

Aquí nuevamente se presentan dos opciones:

Opción 1. Tomar una muestra aleatoria y representativa de la base de datos.

Opción 2. Emplear el procedimiento bietápico, que funciona eficientemente con grandes volúmenes de casos.

Inicialmente se tomó la opción de una muestra aleatoria (por medio de un software que genera números aleatorios, asignándolos a cada uno de los casos de la base de clientes, posteriormente se ordenaron y seleccionaron aproximadamente el 10% de la base total). Se ejecutó un análisis de cluster jerárquico, utilizando diversos métodos de agrupación (vecino más cercano, vecino más alejado, centroides, inter-grupos, Ward, entre otros). Los resultados, aunque eran satisfactorios, dificultaban la descripción de los conglomerados y no permitían el uso de la totalidad de casos (124.955 asociados de la base objetivo).

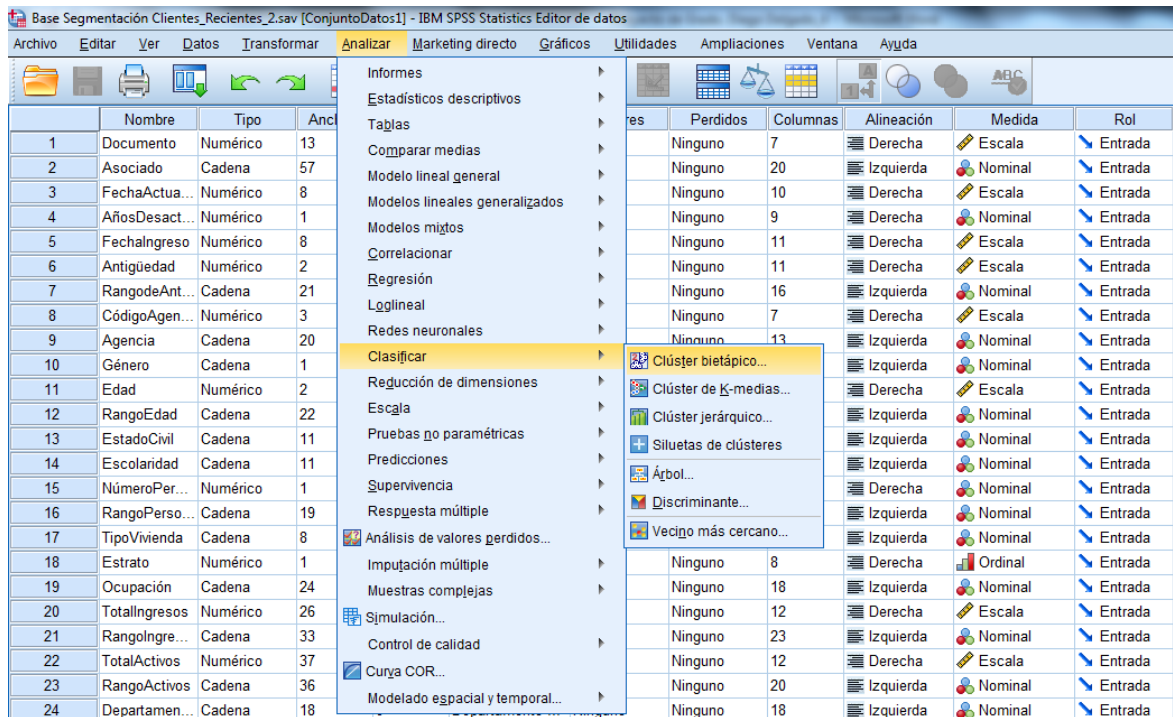
Teniendo en cuenta las ventajas del método bietápico (descritas en la sección 1.5.5), se escogió por la opción 2, el cual entre otros, permite la utilización de la base total inicialmente seleccionada.

3.2 PROCEDIMIENTO PARA EL ANÁLISIS DE CONGLOMERADOS UTILIZANDO LA HERRAMIENTA SPSS

A continuación se describe paso a paso el procedimiento empleado para el análisis de conglomerados mediante el método bietápico, utilizando SPSS como herramienta de análisis.

1. Una vez consolidada la base de datos con las variables seleccionadas y depuradas, se procede a abrir el archivo en SPSS.
2. En la barra de herramientas se busca el tipo de cluster a realizar (en este caso cluster bietápico), con la siguiente ruta: Analizar/Clasificar/Clúster Bietápico (ver figura 35).

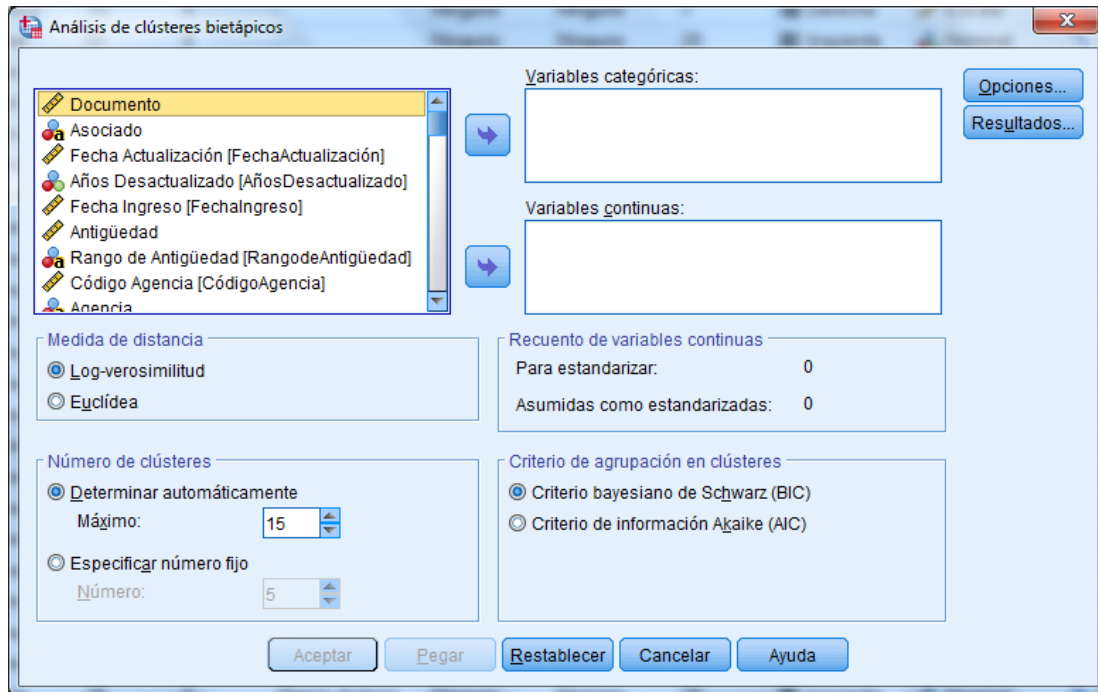
Figura 35. Ruta para realizar un cluster Bietápico.



Fuente: *Elaboración propia*

3. Se seleccionan las variables con las cuales se busca hacer el análisis de conglomerados, diferenciando las variables categóricas de las continuas. Se elige entre las diferentes opciones la medida de distancia, el número de clusters y el criterio de agrupación (ver figura 36).

Figura 36. Cuadro de diálogo de análisis de cluster Bietápico.



Fuente: Elaboración propia

El área “Medida de distancia” permite seleccionar entre la distancia *Log-verosimilitud* y la distancia *Euclídea*. Si se utilizan variables categóricas, se desactiva la opción de distancia *Euclídea*.

El área “Recuento de variables continuas” resume cuántas variables continuas van a estandarizarse y cuántas no. Por defecto, todas las variables continuas son estandarizadas para que las variables con mayor varianza no tengan mayor influencia en el análisis cuando se selecciona la medida de distancia euclídea; las soluciones estandarizadas y no estandarizadas son idénticas cuando se selecciona el criterio log-verosimilitud. Puede quitarse la opción de estandarización para las variables seleccionadas en el cuadro de diálogo “Opciones”. Esto podría hacerse si todas las variables continuas estuvieran en la misma escala y tuvieran una variación similar, o si está usando la distancia euclídea y se

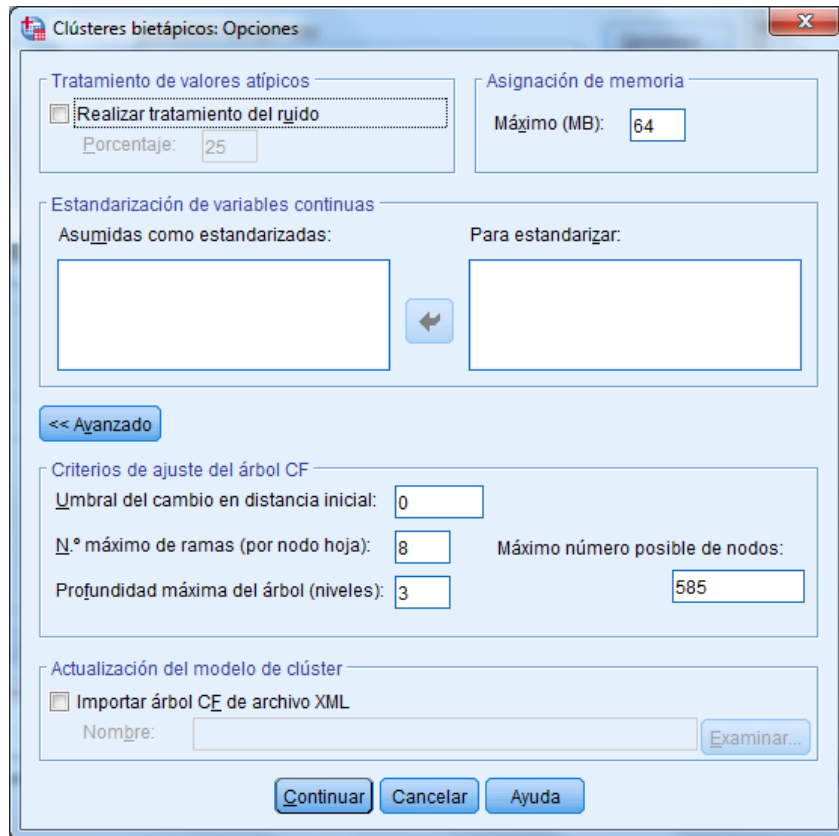
desean variables con mayor varianza para tener más influencia en la solución.

En el área “Número de conglomerados” hay opciones para que el método bietápico determine el número de conglomerados automáticamente (de manera predeterminada) o se especifique el número de conglomerados. Si el número de conglomerados se determina automáticamente, entonces puede establecerse el número máximo de conglomerados a ser considerados por el algoritmo. Por ejemplo, si sólo se está interesado en un pequeño número de conglomerados, entonces el valor máximo podría reducirse.

También ofrece dos “criterios de agrupación en clústeres”: el *Criterio bayesiano de Schwarz (BIC)* y el *Criterio de información Akaike (AIC)*. Generalmente, los dos se usan para evaluar la bondad de ajuste del modelo. Debido a su forma, no suelen interpretarse de manera directa, pero se usan para comparar los modelos alternativos (los valores más pequeños son mejores).

Al hacer clic en el botón “Opciones” se muestra el *Tratamiento de valores atípicos* y la *Estandarización de variables continuas*, indicando qué variables continuas (todas en la configuración predeterminada) serán estandarizadas previo al análisis (permite modificar la designación). De igual forma, en este cuadro se muestra el botón “Avanzado” con Criterios de ajuste al árbol CF (árbol de Características del Conglomerado). Ver figura 37. Si el árbol CF se completa durante el paso del pre-conglomerado, pre-conglomerados con valores atípicos serán colocados en un pre-conglomerado único. Esto tendría a reducir la ocurrencia de conglomerados pequeños debidos a los puntos de datos atípicos.

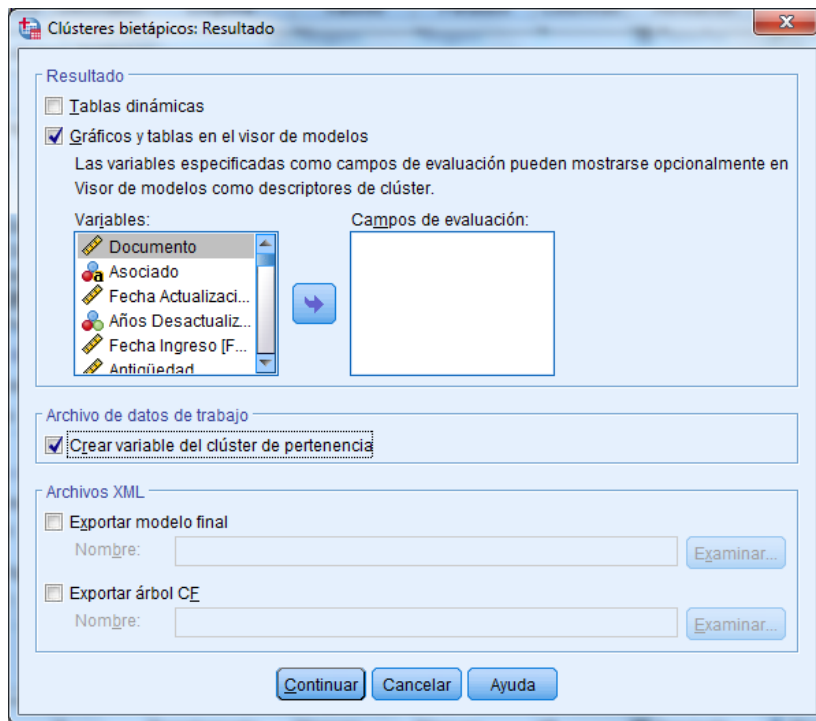
Figura 37. Cuadro de diálogo “Opciones” en conglomerado bietápico.



Fuente: Elaboración propia

Al hacer clic en “Resultados” se presenta la figura 38 con opciones para generar *Tablas dinámicas* y *Gráficos y tablas en el visor de modelos*. Aquí es importante en el campo “Archivo de datos de trabajo” seleccionar la opción *Crear variables de clúster de pertenencia*, con las cuales posteriormente se realizarán análisis de los conglomerados. También hay opciones disponibles para exportar el modelo final y el árbol CF en archivos XML.

Figura 38. Cuadro de diálogo “Resultados” en conglomerado Bietápico.



Fuente: *Elaboración propia*

4. Se da clic en “Aceptar” en el cuadro de diálogo de conglomerado bietápico (figura 36) y se procede con la interpretación y análisis de la información.
5. Para la interpretación y análisis, SPSS en el visor de resultados presenta entre otros cuadros:
 - *Distribución de clúster.* Cantidad y porcentaje de participación en el total de casos para cada uno de los clústeres.
 - *Perfiles de clúster.* Para las variables continuas presenta la media y desviación estándar en cada uno de los clústeres.

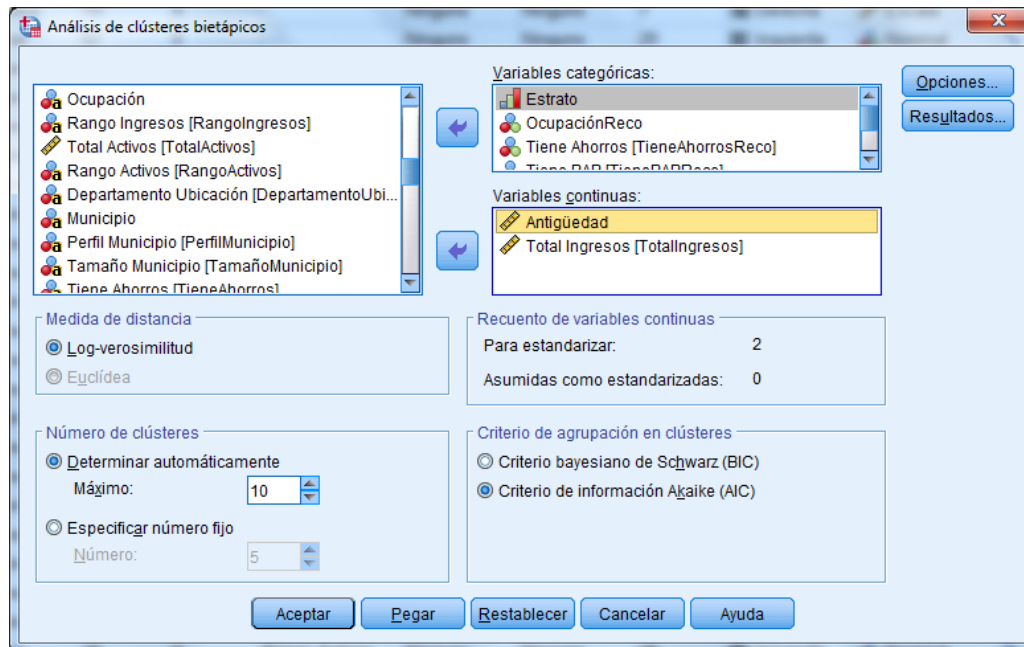
- *Frecuencias.* Para cada una de las variables categóricas presenta una tabla de frecuencias (con cantidad y porcentaje de participación) en cada uno de los clústeres.
- Una figura con el “Resumen del modelo” y la “Calidad de clúster”.

6. Para enriquecer el análisis de recomendación realizar “cruces” de información (por medio de tablas) con otras variables no utilizadas en el análisis cluster pero de gran importancia al momento de describir los conglomerados.

Concretamente para la segmentación de clientes de La Cooperativa, en el procedimiento Bietápico se seleccionaron seis (6) variables categóricas: Estrato, Ocupación, Tiene Ahorros, Tiene PAP, Tiene CDAT y Tiene Crédito; y dos (2) variables continuas: Antigüedad y Total Ingresos. Ver figura 39.

Por otra parte, se empleó la medida de distancia *Log-verosimilitud* (predeterminada por el tipo de variables usadas) y se estandarizaron las variables continuas. Se seleccionó el criterio de agrupación Akaike (AIC). Como se busca un número menor de conglomerados, se determinó cinco (5) como el número máximo de clusters.

Figura 39. Cuadro de diálogo análisis de cluster bietápico con las variables y opciones seleccionadas.

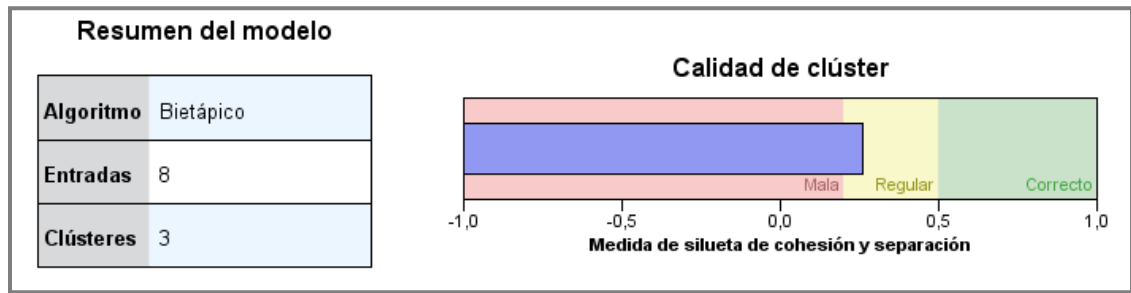


Fuente: Elaboración propia

3.3 RESULTADOS DEL ANÁLISIS DE CONGLOMERADOS

Al aplicar el procedimiento descrito en el punto anterior para análisis de conglomerados utilizando SPSS mediante cluster bietápico, se obtuvieron los siguientes resultados:

Figura 40. Resumen del modelo y calidad del cluster.



Fuente: Elaboración propia

En el resumen del modelo (figura 40) se muestra la utilización del algoritmo bietápico, con ocho (8) variables de entrada y la salida de tres (3) conglomerados. En cuanto a la calidad del conglomerado, ésta se clasifica como “Regular”. En esta etapa se intentó con diferentes variables (categóricas y continuas) y otras opciones de criterio y estandarización; sin embargo, no se obtuvo un mejor resultado.

En la tabla 4 se puede observar que del total de casos (124.955), el 31% pertenecen al conglomerado 1, el 29% al conglomerado 2 y el restante 40% al conglomerado 3. Una distribución relativamente uniforme con la cual se realizará el análisis de conglomerados.

Tabla 4. Frecuencias y porcentajes de participación por conglomerado

Conglomerado	Frecuencia	Porcentaje	Porcentaje acumulado
1	38.221	30,6%	30,6%
2	36.427	29,2%	59,7%
3	50.307	40,3%	100,0%
Total	124.955	100,0%	

Fuente: Elaboración propia.

A continuación se realiza un análisis detallado por variable, tanto las variables utilizadas en el análisis de conglomerados como las demás variables inicialmente seleccionadas. En cada tabla se resalta en color naranja las participaciones más representativas entre conglomerados (análisis horizontal) y en color verde las participaciones más representativas dentro de los conglomerados (análisis vertical).

Tabla 5. Perfil de conglomerados. Variable Tiene Ahorros

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Tiene	NO	25	0,1%	0,3%	360	1,0%	3,9%	8.873	17,6%	95,8%	9.258
Ahorros	SI	38.196	99,9%	33,0%	36.067	99,0%	31,2%	41.434	82,4%	35,8%	115.697
	Total Columna	38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

El 96% de los asociados que No tienen Ahorros se encuentran en el conglomerado 3. Dentro de los conglomerados 1 y 2 mínimo el 99% de los asociados tienen ahorros en La Cooperativa. Ver tabla 5.

Tabla 6. Perfil de conglomerados. Variable Tiene PAP

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Tiene	NO	9.023	23,6%	14,6%	16.279	44,7%	26,3%	36.511	72,6%	59,1%	61.813
PAP	SI	29.198	76,4%	46,2%	20.148	55,3%	31,9%	13.796	27,4%	21,8%	63.142
	Total Columna	38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

El 59% de los asociados que No tienen PAP se encuentran en el conglomerado 3. Dentro del conglomerado 1 el 77% tienen PAP. Dentro del conglomerado 2 el 55% de los asociados tienen PAP. Ver tabla 6.

Tabla 7. Perfil de conglomerados. Variable Tiene CDAT

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Tiene	NO	37.740	98,7%	33,5%	35.578	97,7%	31,6%	39.395	78,3%	35,0%	112.713
CDAT	SI	481	1,3%	3,9%	849	2,3%	6,9%	10.912	21,7%	89,1%	12.242
Total Columna		38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

El 89% de los asociados que Tienen CDAT se encuentran en el conglomerado 3. Dentro de los conglomerados 1 y 2 por lo menos el 98% de los asociados no tienen el producto CDAT. Ver tabla 7.

Tabla 8. Perfil de conglomerados. Variable Tiene Crédito.

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Tiene	NO	147	0,4%	0,3%	87	0,2%	0,2%	47.293	94,0%	99,5%	47.527
Crédito	SI	38.074	99,6%	49,2%	36.340	99,8%	46,9%	3.014	6,0%	3,9%	77.428
Total Columna		38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

El 99,5% de los asociados que No tienen Crédito se encuentran en el conglomerado 3. Dentro de los conglomerados 1 y 2 por lo menos el 99% de los asociados tienen crédito. Ver tabla 8.

Tabla 9. Perfil de conglomerados. Clasificación de Crédito.

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Clasificación de Crédito	No tiene Crédito	147	0,4%	0,3%	87	0,2%	0,2%	47.293	94,0%	99,5%	47.527
	Comercial	1.526	4,0%	84,2%	267	0,7%	14,7%	19	0,0%	1,0%	1.812
	Consumo	6.113	16,0%	15,9%	29.633	81,3%	77,0%	2.723	5,4%	7,1%	38.469
	Microcrédito	30.391	79,5%	82,0%	6.394	17,6%	17,3%	268	0,5%	0,7%	37.053
	Vivienda	44	0,1%	46,8%	46	0,1%	48,9%	4	0,0%	4,3%	94
Total Columna		38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

Por lo menos el 82% de los asociados que tienen crédito “comercial” o “microcrédito” se encuentran en el conglomerado 1. El 77% de los asociados que tienen crédito de “consumo” y el 49% que tiene crédito de “vivienda” se encuentran en el conglomerado 2. Ver tabla 9.

Tabla 10. Perfil de conglomerados. Variable Diversificación.

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Diversificación	Solo Aportes	0	0,0%	0,0%	0	0,0%	0,0%	4.112	6,2%	100,0%	4.112
	Monoproducto	75	0,2%	0,3%	342	0,9%	1,2%	27.950	56,8%	98,5%	28.367
	Biproducto	9.006	23,6%	22,6%	15.613	42,9%	39,1%	15.298	31,1%	38,3%	39.917
	TripProducto	28.698	75,1%	55,8%	20.052	55,0%	39,0%	2.672	5,4%	5,2%	51.422
	Multiproducto	442	1,2%	38,8%	420	1,2%	36,9%	276	0,6%	24,2%	1.138
	Total Columna	38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

El 100% de los asociados que solo tienen Aportes y el 99% de los asociados que tienen un producto (adicional a Aportes) se encuentran en el conglomerado 3. Por lo menos el 39% de los asociados que tienen dos o más productos se encuentran en el conglomerado 2. El 56% de los asociados que tienen tres productos se encuentran en el conglomerado 1. Ver tabla 10.

Tabla 11. Perfil de conglomerados. Variable Perfil Asociado.

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Perfil Asociado	Aportes	4	0,0%	0,1%	0	0,0%	0,0%	4.108	8,2%	99,9%	4.112
	Captador	1.451	3,8%	3,0%	1.429	3,9%	3,0%	44.727	88,9%	94,0%	47.607
	Deudor	36.766	96,2%	50,2%	34.998	96,1%	47,8%	1.472	2,9%	2,0%	73.236
		Total Columna	38.221	100%		36.427	100%		50.307	100%	

Fuente: Elaboración propia

Aproximadamente el 100% de los asociados que solo tienen Aportes y el 94% que son “Captadores” se encuentran en el conglomerado 3. Por otra parte, el 98% de los asociados que son “Deudores” se encuentran en los conglomerados 1 y 2. Ver tabla 11.

Tabla 12. Perfil de conglomerados. Variable Ocupación.

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Ocupación	Trabajador Independiente	38.018	99,5%	56,5%	4.990	13,7%	7,4%	24.316	48,3%	36,1%	67.324
	Empleado	42	0,1%	0,1%	22.423	61,6%	63,0%	13.123	26,1%	36,9%	35.588
	Pensionados	11	0,0%	0,1%	4.673	12,8%	55,7%	3.700	7,4%	44,1%	8.384
	Ama de casa	0	0,0%	0,0%	8	0,0%	0,2%	3.533	7,0%	99,8%	3.541
	Estudiante	0	0,0%	0,0%	0	0,0%	0,0%	1.948	3,9%	100,0%	1.948
	Ninguna	2	0,0%	0,1%	552	1,5%	32,5%	1.147	2,3%	67,4%	1.701
	Otros	148	0,4%	2,3%	3.781	10,4%	58,4%	2.540	5,0%	39,3%	6.469
Total Columna		38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

En el conglomerado 3 se encuentra el 100% de los asociados con ocupación “Estudiante” y el 99,8% de los asociados “Ama de casa” y el 67% de los que manifiestan no tener ocupación. En el conglomerado 2 se encuentra el 63% de los asociados “Empleados” y el 56% de los asociados “Pensionados”. En el conglomerado 1 se encuentra el 57% de los “Trabajadores Independientes”. Ver tabla 12.

Tabla 13. Perfil de conglomerados. Variable Antigüedad (categorizada).

		Conglomerado									
		1			2			3			
Variable	Descripción	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Total Fila
Antigüedad	1. Menos de 2 años	19.935	52,2%	40,3%	14.837	40,7%	30,0%	14.682	29,2%	29,7%	49.454
	2. Entre 2 y 5 años	9.347	24,5%	27,6%	10.363	28,4%	30,6%	14.142	28,1%	41,8%	33.852
	3. Entre 6 y 15 años	6.653	17,4%	24,0%	7.491	20,6%	27,0%	13.562	27,0%	48,9%	27.706
	4. Más de 15 años	2.286	6,0%	16,4%	3.736	10,3%	26,8%	7.921	15,7%	56,8%	13.943
Total Columna		38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

En el conglomerado 3 se ubican el 57% de los asociados que tienen más de 15 años de antigüedad y el 49% de los asociados que tienen entre 6 y 15 años de antigüedad con La Cooperativa. En el conglomerado 1 se ubican el 40% los asociados que tienen menos de dos años de antigüedad. Ver tabla 13.

Tabla 14. Perfil de conglomerados. Variable Estrato.

Variable	Descripción	Conglomerado									Total Fila
		1			2			3			
		Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	
Estrato	1	9.554	25,0%	43,6%	3.889	10,7%	17,7%	8.470	16,8%	38,7%	21.913
	2	21.796	57,0%	40,1%	12.806	35,2%	23,5%	19.808	39,4%	36,4%	54.410
	3	4.582	12,0%	12,7%	16.172	44,4%	44,7%	15.427	30,7%	42,6%	36.181
	4	1.969	5,2%	18,6%	3.105	8,5%	29,3%	5.523	11,0%	52,1%	10.597
	5	235	0,6%	17,5%	346	0,9%	25,8%	761	1,5%	56,7%	1.342
	6	85	0,2%	16,6%	109	0,3%	21,3%	318	0,6%	62,1%	512
	Total Columna	38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

En la tabla 14, se observar que por lo menos el 52% de los asociados que son de estratos 4, 5 y 6 se ubican en el conglomerado 3. El 45% de los asociados de estrato 3 se ubican en el conglomerado 2. Mínimo el 40% de los asociados que están en los estratos 1 y 2 se ubican en el conglomerado 1.

Tabla 15. Perfil de conglomerados. Variable Total Ingresos (categorizada).

Variable	Descripción	Conglomerado									Total Fila
		1			2			3			
		Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	
Total Ingresos	1. Menor a \$700.000	193	0,5%	1,6%	617	1,7%	5,2%	10.979	21,8%	93,1%	11.789
	2. Entre \$700.000 y \$1.400.000	3.912	10,2%	13,2%	11.467	31,5%	38,6%	14.337	28,5%	48,2%	29.716
	3. Entre \$1.400.000 y \$3.500.000	15.193	39,8%	31,4%	17.271	47,4%	35,7%	15.892	31,6%	32,9%	48.356
	4. Entre \$3.500.000 y \$10.000.000	12.415	32,5%	48,1%	6.370	17,5%	24,7%	7.011	13,9%	27,2%	25.796
	5. Superior a \$10.000.000	6.508	17,0%	70,0%	702	1,9%	7,6%	2.088	4,2%	22,5%	9.298
	Total Columna	38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

El 93% de los asociados con ingresos menores a \$700.000 y el 48% de los asociados con ingresos entre \$700.000 y \$ 1.400.000, es decir, asociados con bajos ingresos, se encuentran en el conglomerado 3. El 48% de los asociados con ingresos entre \$3.500.00 y \$10.000.000 y el 70% de los asociados con ingresos superiores a \$10.000.000, es decir, asociados con ingresos altos se encuentran en el conglomerado 1. Ver tabla 15.

Tabla 16. Perfil de conglomerados. Variable Escolaridad.

Variable	Descripción	Conglomerado									Total Fila
		1			2			3			
		Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	Cantidad	% Participa Columna	% Participa Fila	
Escolaridad	Ninguna	404	1,1%	28,2%	121	0,3%	8,4%	908	1,8%	63,4%	1.433
	Primaria	12.616	33,0%	40,4%	4.668	12,8%	15,0%	13.927	27,7%	44,6%	31.211
	Secundaria	19.821	51,9%	35,5%	14.548	39,9%	26,1%	21.410	42,6%	38,4%	55.779
	Técnico	3.005	7,9%	19,6%	7.267	19,9%	47,3%	5.088	10,1%	33,1%	15.360
	Universidad	2.135	5,6%	12,8%	7.264	19,9%	43,4%	7.320	14,6%	43,8%	16.719
	Posgrado	240	0,6%	5,4%	2.559	7,0%	57,5%	1.654	3,3%	37,1%	4.453
	Total Columna	38.221	100%		36.427	100%		50.307	100%		124.955

Fuente: Elaboración propia.

Según la tabla 16. El 63% de los asociados que manifiestan no tener “Escolaridad” se encuentran en el conglomerado 3, de igual forma el 45% de los asociados con

nivel “Primaria”. En el conglomerado 2 se encuentran los asociados con niveles educativos superiores (técnico, universitario o posgrado).

Tabla 17. Resumen de los resultados del análisis.

Variable	CONGLOMERADO		
	1	2	3
Tiene Ahorros	Sí	Sí	No
Tiene PAP	%	%	No
Tiene CDAT	No	No	Sí
Tiene Crédito	Sí	Sí	No
Clasificación Crédito	Microcrédito o Comercial	Consumo	N.A.
Perfil Asociado	Deudor	Deudor	Solo Aportes o Captador
Diversificación (Productos)	Tres productos	Dos o Tres productos	Solo Aportes o CDAT
Ocupación	Trabajador Independiente	Empleado o Pensionado	Ama de Casa, Estudiante, Pensionado, Otros
Escolaridad	Básica (Primaria)	Superior (Técnica, Universitaria o Posgrado)	%
Antigüedad	Reciente (menos de dos años)	%	Antiguo (más de dos años)
Estrato	Principalmente 1 y 2	Principalmente 3	Principalmente 4, 5 o 6
Ingresos	Superiores (Mayores a 5 smmlv)	Intermedios (Entre 1 y 5 smmlv)	Bajos (Menos de 2 smmlv)

Fuente: Elaboración propia.

CAPÍTULO 4. DESCRIPCIÓN DE LOS CONGLOMERADOS

4.1 DESCRIPCIÓN CONGLOMERADO 1.

Este conglomerado está compuesto por 38.221 asociados (30,6% de la base total).

“EI DEUDOR INDEPENDIENTE”

Trabajador independiente. Con perfil “Deudor”; es decir con crédito, en las modalidades Microcrédito y Comercial. Maneja principalmente 3 productos (aparte de Aportes). Asociado reciente a La Cooperativa (menos de 2 años). Con nivel de estudios en primaria y secundaria. Principalmente de estratos 1 y 2, aunque con ingresos superiores a 5 salarios mínimos mensuales (mayor a \$3.500.000).

4.2 DESCRIPCIÓN CONGLOMERADO 2.

Este conglomerado está compuesto por 36.427 asociados (29,2% de la base total).

“EI DEUDOR ASALARIADO”

Empleado o pensionado. Con perfil “Deudor”; es decir con crédito, en la modalidad Consumo. Maneja entre 2 y 3 productos con La Cooperativa (aparte de Aportes). Con nivel de estudios superiores (es decir, Técnico, Universitario o Posgrado). Principalmente de estrato 3, con ingresos intermedios entre 1 a 5 salarios mínimos mensuales (entre \$700.000 y \$3.500.000).

4.3 DESCRIPCIÓN CONGLOMERADO 3.

Este conglomerado está compuesto por 50.307 asociados (40,3% de la base total).

“EL INVERSIONISTA DE ESTRATO”

Con La Cooperativa solo maneja Aportes y/o CDAT's, es decir, tiene un perfil “Captador”. Presenta diversas ocupaciones como “Pensionado”, “Ama de Casa”, “Estudiante”, entre otros. En su mayoría ya tienen antigüedad con La Cooperativa (más de dos años). Pertenecen principalmente a estratos 4, 5 o 6, aunque con ingresos inferiores a 2 salarios mínimos mensuales (menor a \$1.400.000 mensual).

CONCLUSIONES

Para realizar segmentación de clientes, los métodos multivariados presentan diferentes alternativas. La decisión de emplear una u otra depende del objetivo del estudio y de la información que se tenga disponible o se pueda conseguir.

Para el proceso de segmentación de clientes, objetivo de este análisis, se empleó la técnica descriptiva llamada Análisis de Conglomerados.

Dadas las ventajas que ofrece el procedimiento bietápico, entre ellas, el permitir la utilización de variables continuas y categóricas y la base completa de asociados, se escogió este método para el análisis de conglomerados.

De aproximadamente 50 variables con las que se puede contar en la base de datos de clientes, se seleccionaron para el análisis descriptivo 20 (entre variables sociodemográficas y conductuales). De éstas, solo ocho (8) fueron escogidas para el análisis cluster. Posteriormente, para la descripción de los conglomerados, adicional a las ocho (8) se emplearon cuatro (4) variables más.

Finalmente, con estas 12 variables: Tiene Ahorros, Tiene PAP, Tiene CDAT, Tiene Crédito, Clasificación de Crédito, Diversificación, Perfil Asociado, Ocupación, Escolaridad, Antigüedad, Estrato y Total Ingresos, se ejecutó del procedimiento Bietápico, hallando tres (3) clusters o conglomerados: “El deudor independiente”, “El deudor asalariado” y “El inversionista de estrato”, los cuales son descritos en el capítulo 4.

Teniendo en cuenta los conglomerados hallados, las áreas Comercial y de Mercadeo de La Cooperativa, podrían emplear estas características en la redefinición de sus estrategias de venta. Por ejemplo, según conglomerado 3 “El

inversionista de estrato”, si se desea incrementar el número de asociados con el producto de CDAT, se podría buscar asociaciones o agrupaciones de pensionados o amas de casa ubicados en estratos altos y/o emplear imágenes de este tipo de personas en su publicidad. Basados en el conglomerado 1 “El deudor independiente”, si se desea mejorar los resultados de las líneas de microcrédito o comercial, se podría buscar estos perfiles en entidades que agrupen o apoyen a trabajadores independientes (Cámaras de Comercio, Gobernaciones, Alcaldías, etc) cruzando información en sus bases de datos de las personas que desarrollen estas actividades económicas, de tal manera que allí se enfoquen los esfuerzos de los asesores comerciales de La Cooperativa.

Es necesario que La Cooperativa enriquezca la base de datos de clientes con otro tipo de variables como las sicográficas (con gustos y preferencias de sus asociados), las cuales no son fáciles de conseguir, pero se pueden utilizar herramientas como las encuestas. De igual forma, construir y mantener “variables calculadas” como “Rentabilidad por cliente” mostrando qué tan rentable es el cliente para La Cooperativa o variables como “Profundización” que es resultado del cruce entre la tenencia y cantidad de productos con el saldo de los mismos. Estas variables seguramente permitirían análisis más completos y con mayor aporte a las estrategias y objetivos de La Cooperativa.

En cuanto a la calidad del conglomerado, en el resumen del modelo ésta se clasifica como “Regular”. Es importante que para nuevos análisis se incorporen por ejemplo las variables mencionadas en el punto anterior, y se realicen trabajos de actualización de datos de los asociados, de tal forma que en la depuración de los datos sean menos los casos que se eliminen por este criterio, contando con una base más amplia.

BIBLIOGRAFÍA

BEERLY, A y GARCÍA, J.M. La segmentación del mercado y su aplicación al mercado bancario minorista. En: Revista europea de dirección y economía de la empresa. 1992, vol. 1 N°1, p. 79-89.

BELLO, Daniel. La segmentación de mercados en el sistema financiero español. Tesis Master oficial en banca y finanzas. España: Universidad Da Coruña. Facultad de Ciencias Económicas y Empresariales, 2012. 77p.

KOTLER, Philip y ARMSTRONG, Gary. Fundamentos de marketing (11^a edición). Pearson Educación. México, 2013. 648p. ISBN 9786073217224.

PEÑA SÁNCHEZ, Daniel. Análisis de datos multivariantes. McGraw-Hill Interamericana de España S.L., 2002. 529p. ISBN 9788448136109.