



**Desarrollo de un modelo analítico para la
segmentación de asociados en una cooperativa de
ahorros y créditos**

William Alberto Jovel Tamayo

Universidad Nacional de Colombia

Facultad de Minas, Área Curricular de Sistemas e Informática

Medellín, Colombia

2020



UNIVERSIDAD
NACIONAL
DE COLOMBIA

**Development of an analytical model for the
segmentation of affiliates in a savings and loan
cooperative**

William Alberto Jovel Tamayo

Universidad Nacional de Colombia

Faculty of Mines, Systems and Informatics Curriculum Area

Medellín, Colombia

2020

Desarrollo de un modelo analítico para la segmentación de asociados en una cooperativa de ahorros y créditos

William Alberto Jovel Tamayo

Trabajo de grado presentado como requisito parcial para optar al título de:

Magister en Ingeniería - Analítica

Director:

Juan David Velásquez Henao, MSc, PhD.

Universidad Nacional de Colombia

Facultad de Minas, Área Curricular de Sistemas e Informática

Medellín, Colombia

2020

Agradecimientos

A mi esposa y mi hija por su comprensión, paciencia y apoyo permanente para realizar este proyecto.

A todos los profesores de la Maestría en Ingeniería – Analítica, y especialmente al Doctor Juan David Velázquez, mi director de tesis, por su orientación y colaboración con mi formación académica y el desarrollo de este proyecto.

Resumen

Este trabajo presenta el desarrollo de un modelo analítico realizado para una cooperativa de ahorros y créditos de Antioquia con el fin de obtener un mejor conocimiento de sus asociados y que este pueda ser utilizado para fortalecer los modelos de gestión comercial, de servicios y de riesgos financieros. Se aplicaron varias técnicas de análisis de agrupamiento a los asociados activos, usando información histórica transaccional; el saldo en productos de ahorro y crédito, el número de productos adquiridos en la cooperativa, el interés pagado en productos de crédito y la fecha de la última transacción realizada. Finalmente se presenta un análisis descriptivo de los grupos de asociados activos de la entidad, según los agrupamientos realizados del modelo seleccionado. El documento contiene la descripción de los datos, los detalles de la metodología aplicada, basada en *CRISP-DM*, los modelos de agrupamiento usados, la evaluación de los modelos y finalmente un producto de datos.

Palabras clave: RFM, minería de datos, cooperativa de ahorro y crédito, *CRISP-DM*, segmentación de clientes.

Abstract

This paper presents the development of an analytical model carried out for a savings and loan cooperative in Antioquia in order to obtain a better knowledge of its associates and It can be used to strengthen the commercial, service and financial risk management models. Various clustering analysis techniques were applied to active associates, using transactional historical information; the balance in savings and credit products, the number of products purchased in the cooperative, the interest paid on credit products and the date of the last transaction made. Finally, a descriptive analysis of the groups of active associates of the entity is presented, according to the groupings made of the selected model. The document contains the description of the data, the details of the applied methodology, based on CRISP-DM, the grouping models used, the evaluation of the models and finally a data product.

Keywords: RFM, Data Mining, savings and loan cooperative, CRISP-DM, Customer segmentation.

Contenido

Resumen.....	V
Lista de figuras.....	IX
Lista de tablas.....	X
Introducción	1
1. Minería de datos.....	3
1.1 Aplicación de la minería de datos en servicios financieros.....	3
1.2 Segmentación de clientes	4
2. Entendimiento del negocio.....	9
2.1 Descripción de la situación.....	9
2.2 Solución planteada	10
2.3 Criterios de éxito del negocio.....	11
2.4 Objetivos de la minería de datos.....	11
2.5 Objetivos específicos.....	12
2.6 Criterios de éxito de la minería de datos.....	12
2.7 Evaluación inicial de herramientas y técnicas	12
2.8 Conclusión	13
3 Entendimiento de los datos.....	14
3.1 Recolección de datos	14
3.2 Descripción del conjunto de datos	15
3.3 Limpieza de datos.....	17
3.4 Verificación de la calidad de los datos	18
3.5 Exploración de datos	19
3.5.1 Análisis exploratorio de asociados con ahorros activos	19
3.5.2 Análisis exploratorio de asociados con créditos activos	31
3.6 Conclusiones	43
4 Modelamiento	44
4.1 Propuestas de segmentación de asociados	44
4.2 Construcción de los modelos.....	45
4.3 Segmentación de asociados con ahorros activos.....	47
4.3.1 Test de agrupamiento.....	47
4.3.2 Selección del número óptimo de grupos.....	47
4.3.3 Resultados del agrupamiento.....	48
4.3.4 Evaluación de los modelos	51
4.3.5 Descripción de los agrupamientos y asociados típicos obtenidos	53

4.4	Segmentación de asociados con créditos activos.....	55
4.4.1	Test de agrupamiento	55
4.4.2	Selección del número óptimo de grupos.....	55
4.4.3	Resultados del agrupamiento.....	57
4.4.4	Evaluación de los modelos	58
4.4.5	Variables que más influyeron en la segmentación.....	60
4.4.6	Descripción de los agrupamientos y asociado típicos obtenidos	61
4.5	Segmentación de asociados con ahorros y créditos activos.	63
4.5.1	Test de agrupamiento	63
4.5.2	Selección del número óptimo de grupos.....	63
4.5.3	Resultados del agrupamiento.....	65
4.5.4	Evaluación de los modelos	68
4.5.5	Variables que más influyeron en la segmentación.....	69
4.5.6	Descripción de los agrupamientos y asociados típicos obtenidos	71
4.6	Conclusiones	77
5.	Evaluación	78
6.	Producto de datos.....	80
7.	Conclusiones y recomendaciones.....	84
7.1	Conclusiones.....	84
7.2	Recomendaciones.....	84
	Bibliografía.....	86

Lista de figuras

Figura 1: Análisis de correlación por pares entre variables numéricas	17
Figura 2: Evaluación de calidad de los datos	18
Figura 3 Método gráfico del codo para la selección del número óptimo de clúster para asociados ahorradores.	48
Figura 4 Método gráfico de Silhouette para la selección del número óptimo de clúster para asociados ahorradores	48
Figura 5 Árbol de decisión de la segmentación de asociados ahorradores	55
Figura 6 Método gráfico del codo para la selección del número óptimo de clúster para asociados con créditos activos.	56
Figura 7 Método gráfico de Silhouette para la selección del número óptimo de clúster para asociados con créditos activos.	56
Figura 8 Árbol de decisión de la segmentación de asociados con créditos activos	60
Figura 9 Método gráfico del codo para la selección del número óptimo de clúster para asociados con ahorros y créditos activos.	64
Figura 10 Método gráfico de Silhouette para la selección del número óptimo de clúster para asociados con ahorros y créditos activos.	65
Figura 11 Árbol de decisión de la segmentación de asociados con ahorros y créditos activos	70
Figura 12 Dashboard desarrollado en Google Data Studio	81
Figura 13 Producto de datos desarrollado en R shiny	81
Figura 14 Dashboard de clúster de asociados con créditos desarrollado en R shiny	83

Lista de tablas

Tabla 1: Descripción de la fuente de datos	15
Tabla 2: Descripción atributos numéricos del conjunto de datos	15
Tabla 3: Descripción atributos categóricos del conjunto de datos	16
Tabla 4: Descripción atributos categóricos ordinales del conjunto de datos	16
Tabla 5: Descripción atributos numéricos de los asociados ahorradores	20
Tabla 6: Quintil de saldos en ahorros	21
Tabla 7: Cuartil de cantidad de ahorros	21
Tabla 8: Cuartil de recencias en ahorros	22
Tabla 9: Descripción de los atributos categóricos de los asociados ahorradores	22
Tabla 10: Distribución de saldos ahorrados por género	23
Tabla 11: Distribución de saldos ahorrados por rango de edad	24
Tabla 12: Distribución de saldos ahorrados por nivel educativo	24
Tabla 13: Distribución de saldos ahorrados por estado civil	25
Tabla 14: Distribución de saldos ahorrados por estrato socioeconómico	26
Tabla 15: Distribución de saldos ahorrados por tipo de vivienda	27
Tabla 16: Distribución de saldos ahorrados por tipo de contrato	27
Tabla 17: Distribución de saldos ahorrados por rango de ingresos	28
Tabla 18: Valores promedio del asociado agrupados por género, y estrato socioeconómico	29
Tabla 19: Moda de variables categóricas del asociado agrupados por género, y estrato socioeconómico	30
Tabla 20: Descripción atributos numéricos de asociados con créditos activos	31
Tabla 21: Distribución en quintiles de saldos en créditos	32
Tabla 22: Distribución en quintiles de interés pagados por créditos	33
Tabla 23: Rango de cantidad de créditos adquiridos	33
Tabla 24: Rango de recencia de créditos	34
Tabla 25: Descripción de atributos categóricos de asociados con créditos activos	34
Tabla 26: Distribución de créditos activos por género	35
Tabla 27: Distribución de créditos activos por rango de edad	35
Tabla 28: Distribución de créditos activos por nivel educativo	36
Tabla 29: Distribución de créditos activos por estado civil	37
Tabla 30: Distribución de créditos activos por tipo de vivienda	37
Tabla 31: Distribución de créditos activos por estrato socioeconómico	38
Tabla 32: Distribución de créditos activos por tipo de contrato	39
Tabla 33: Distribución de créditos activos por rango de ingresos	40
Tabla 34: Valores promedio de asociados con créditos activos agrupados por género, y estrato socioeconómico	40
Tabla 35: Valores promedio de asociados con créditos activos agrupados por género, y estrato socioeconómico	41
Tabla 36: Métricas de número óptimo de clusters para asociados ahorradores	47
Tabla 37: Resultados de los Clusters de asociados ahorradores por K-means	49
Tabla 38: Resultados complementarios de los Clusters de asociados ahorradores por K-Means	49
Tabla 39: Resultados complementarios de los Clusters de asociados ahorradores por K-Means	49
Tabla 40: Resultados de los Clusters de asociados ahorradores por DbSCAN	50
Tabla 41: Resultados complementarios de los Clusters de asociados ahorradores por DbSCAN	50
Tabla 42: Resultados complementarios de los clusters de asociados ahorradores por DbSCAN	51
Tabla 43: Métricas de evaluación de los métodos aplicados para la segmentación de asociados con ahorros activos.	52

Tabla 44: Adjusted Rand Index de los métodos aplicados para la segmentación de asociados con ahorros activos.	53
Tabla 45: Métricas de número óptimo de clusters para asociados con créditos activos	56
Tabla 46: Resultado del Clúster de asociados con créditos activos por K-means	57
Tabla 47: Resultados complementarios del Clúster de asociados con créditos activos por K-means	57
Tabla 48: Resultados complementarios del Clúster de asociados con créditos activos por K-means	58
Tabla 49: Resultados complementarios del Clúster de asociados con créditos activos por K-means	58
Tabla 50: Métricas de evaluación de los métodos aplicados para la segmentación de asociados con créditos activos.	59
Tabla 51: Adjusted Rand Index de los métodos aplicados para la segmentación de asociados con créditos activos.	59
Tabla 52: Caracterización transaccional de los asociados con créditos activo representantes de cada grupo obtenido por K-means	61
Tabla 53: Caracterización socio-demográfica de los asociados con créditos activo representantes de cada grupo obtenido por K-means.	61
Tabla 54: Caracterización transaccional de los asociados con créditos activo representantes de cada grupo obtenido por PAM.	62
Tabla 55: Métricas de número óptimo de clusters para asociados con ahorros y créditos activos	64
Tabla 56: Datos del comportamiento transaccional en ahorros de los asociados con ahorros y créditos activos obtenidos por K-means.	65
Tabla 57: Datos complementarios del comportamiento transaccional en ahorros de los asociados con ahorros y créditos activos obtenidos por K-means.	66
Tabla 58: Datos de Comportamiento transaccional en Créditos	66
Tabla 59: Datos complementarios de Comportamiento transaccional en Créditos	67
Tabla 60: Datos de pagos de Interés por créditos	67
Tabla 61: Datos complementarios	68
Tabla 62: Evaluación de los agrupamientos de Asociados con ahorros y créditos activos	68
Tabla 63: Adjusted Rand Index de los métodos aplicados para la segmentación de asociados con ahorros y créditos activos.	69
Tabla 64: Caracterización transaccional de los asociados con ahorros y créditos activos representantes de la segmentación de ahorradores obtenidos por K-means.	71
Tabla 65: Caracterización sociodemográfica de los asociados con ahorros y créditos activos representantes de la segmentación de ahorradores por K-means	71
Tabla 66: Caracterización transaccional de los asociados con ahorros y créditos activos representantes de la segmentación de ahorradores obtenidos por PAM.	72
Tabla 67: Caracterización sociodemográfica de los asociados con ahorros y créditos activos representantes de la segmentación de ahorradores por PAM.	73

Introducción

El contexto actual de negocios exige que las organizaciones de todo tipo realicen una mejor gestión de sus datos, para que les permita un mayor soporte para la toma de decisiones estratégicas, tácticas y operativas que ayuden al desarrollo y sostenimiento de una propuesta de valor diferenciada, oportuna y consistente a su público de interés. Este trabajo aplica algunas técnicas de minería de datos con el fin de desarrollar un modelo de segmentación que permita a una cooperativa de ahorro y crédito obtener un mayor y mejor conocimiento de los asociados y que este pueda a corto plazo ser utilizado en fortalecer los modelos de gestión comercial, de servicios y de riesgos financieros. El modelo de segmentación se realizará utilizando las variables netamente transaccionales de RFM (Recencia, Frecuencia, Monto) en productos de ahorro y productos de crédito, y se hace una modificación al modelo tradicional de RFM, al incluir el monto de interés pagados por los asociados en productos de créditos.

Para el desarrollo del proyecto se utilizó la metodología *CRISP-DM* de IBM Corporation®, que son las siglas de *Cross-Industry Standard Process for Data Mining*, el cual es un método probado para orientar proyectos de minería de datos [1]. *CRISP-DM* incluye las descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. Como modelo de proceso, *CRISP-DM* ofrece un resumen del ciclo vital de minería de datos, la cual establece que los pasos a seguir en el proceso de la minería de datos son: 1) Comprensión del negocio; 2) Comprensión de los datos; 3) Preparación de los datos; 4) Modelado; 5) Evaluación; 6) Implantación [1]. A continuación, se describen las fases desarrolladas:

- a. **Comprensión del negocio.** En esta fase se realizó la identificación, delimitación y conocimientos de las necesidades y expectativas de la organización con el fin de plantear el proyecto en términos de un problema de analítica.
- b. **Comprensión de los datos:** Se llevó a cabo la creación de una bodega de datos con la información requerida para desarrollar los modelos de segmentación, posteriormente se utilizaron diferentes técnicas de análisis descriptivo sobre los datos.
- c. **Preparación de los datos:** En esta fase las siguientes actividades fueron realizadas: 1) Selección de las variables a monitorear en términos de los atributos o características a analizar; 2) Limpieza de los datos, para gestionar los datos ausentes, los errores o las incoherencias en estos; 3) Creación de nuevas variables a partir de los datos existentes.
- d. **Modelado:** Se desarrollaron tres propuestas de modelo de segmentación aplicando varias técnicas de clustering utilizando el análisis clásico de RFM, en el cual solo se considera variables de

comportamiento transaccional de los asociados en la entidad.

e. **Evaluación:** Se analizaron los resultados de la segmentación proporcionados por los modelos desarrollados. En esta etapa se realizaron análisis de criterios de rendimiento desde el punto de vista analítico y de expertos del negocio, de tal forma que se pudieran dar interpretación y valor a los resultados obtenidos en el modelado.

f. **Desarrollo de un producto de datos:** Finalmente, se desarrolló un producto de datos, el cual permite facilidades de visualización dinámica e interactiva de los datos de los asociados, la creación de agrupamientos por diferentes métodos, y la descripción de los grupos resultantes.

El nombre de la entidad y los datos de asociados que se utilizaron en este proyecto no son de carácter público, y, por tanto, no serán publicados por exigencias de confidencialidad por parte de la cooperativa. Tampoco se realizará la presentación de recomendaciones de estrategias de tipo comercial y/o de gestión de asociados para la entidad.

El documento se organiza de la siguiente forma: en el capítulo 1 se presenta una introducción al concepto de la minería de datos y su aplicación en el contexto de empresas de prestación de servicios financieros, y la segmentación de clientes. Los capítulos 2, 3, 4, 5 y 6 están estrechamente relacionados con los objetivos específicos propuestos en el trabajo y la metodología *CRISP-DM*. En el capítulo 2 se realiza la fase de comprensión del negocio; en el capítulo 3 se realiza el entendimiento de los datos y preparación de los datos; en el capítulo 4 se realiza la presentación del modelo de segmentación; el capítulo 5 trata sobre la evaluación de los resultados obtenidos desde el punto de vista de los objetivos de la entidad; en el capítulo 6 se presenta un producto de datos que permite la visualización dinámica e interactiva de los grupos de asociados obtenidos; finalmente, en el capítulo 7 se presentan las conclusiones y recomendaciones.

1. Minería de datos

En este capítulo se describe el concepto de minería de datos y su aplicación en el contexto de empresas de servicios financieros.

Para SAS Institute [2], los datos fluyen a velocidades y volúmenes nunca antes vistos, y de todas partes, sin embargo, tomar decisiones basadas en hechos no depende de la cantidad de datos que se tenga; sino que, dependerá de lo rápido que las organizaciones puedan descubrir conocimientos en todos esos datos y utilizarlos para llevar a cabo mejores acciones que les permita el cumplimiento de sus objetivos.

Existen múltiples definiciones de minería de datos en [3], “la minería de datos es el proceso de recopilar, limpiar, procesar, analizar y obtener información útil de los datos para generar conocimiento”. En [4], los autores definen la minería de datos como “el proceso de descubrir patrones interesantes y conocimiento a partir de grandes cantidades de datos. Las fuentes de datos pueden incluir bases de datos, bodegas de datos, la Web, u otros repositorios de datos”. En [2], “La minería de datos es el proceso de hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados”.

1.1 Aplicación de la minería de datos en servicios financieros

La minería de datos es un apoyo fundamental para la administración de los sistemas de gestión comercial y de riesgos financieros, los cuales tienen como base primordial un conocimiento de los clientes y/o asociados, tanto en los aspectos sociodemográficos, como de comportamiento transaccional. La minería de datos ofrece una serie de tecnologías que ayudan a las organizaciones a anticipar resultados futuros, descubrir nuevas oportunidades y mejorar el desempeño del negocio. Esta puede aplicarse en entidades de servicios financieros para la segmentación de clientes, la detección de fraudes, la calificación de riesgo de crédito, identificar problemas de seguridad cibernética, aumentar las tasas de respuesta de las campañas de marketing, generar las próximas mejores ofertas, reducir el abandono de los clientes [2].

En Colombia, en los últimos años las empresas de servicios financieros han realizado inversiones en la incorporación de tecnologías con el fin de brindar un mejor servicio a sus clientes, se destaca

el crecimiento de las Fintech¹, Colombia es el país en Latinoamérica que lidera la implementación de servicios Fintech, a septiembre de 2020 habían más de 100 compañías en este segmento abarcando ese espacio de consumidores de productos financieros que no tienen acceso al sistema tradicional de financiación y una mejor oferta de servicios digitales transaccionales lo que ha propiciado una nueva dinámica comercial al permitir que más personas tengan un mayor acceso a servicios financieros desde canales digitales y/o no presenciales [5]. La UIAF [6], desarrolla publicaciones y sistemas de información de inteligencia financiera, dirigidos a las organizaciones financieras con presencia en Colombia, para que hagan una adecuada gestión de los datos de sus clientes y asociados con el fin de prevenir operaciones de lavado de activos y de apoyo al narcotráfico usando minería de datos. Así mismo, La UIAF utiliza la minería de datos para identificar y analizar patrones sobre los comportamientos típicos de los individuos, los cuales permiten caracterizar operaciones extrañas que merecen una investigación de inteligencia financiera más detallada, potencializando sus esfuerzos en la lucha contra el lavado de activos y la financiación del terrorismo en Colombia [7].

1.2 Segmentación de clientes

Un uso común de la minería de datos y de las técnicas de aprendizajes de máquinas es la segmentación automática de los clientes por comportamiento transaccional, variables demográficas o actitudes, para entender mejor las necesidades de grupos específicos y atenderlos de una manera más dirigida. Esta segmentación analítica, ayuda a identificar a grupos de clientes que son similares y que podrían reaccionar a ciertas ofertas o actividades de manera parecida. Utilizando estos segmentos, se pueden crear modelos para cada grupo para predecir la siguiente mejor oferta o actividad a la cual probablemente responderán. Puede complementarse el modelo de adquisición con un modelo de calificación de riesgos para averiguar quién tiene un bajo riesgo de crédito y que vale la inversión para adquirirlo o retenerlo. [2].

Es posible realizar muchos modelos de segmentación de clientes, dependiendo del objetivo deseado por la entidad, para el caso de conocer cuáles son los clientes más “leales” o cuales representan un mayor valor financiero para la entidad, se encontró en la revisión de literatura que la mayoría de los casos de estudio utilizan modelos de RFM (siglas de Recencia, Frecuencia, Montos) y CLV (*Customer Lifetime Value*). De acuerdo con [8], el análisis RFM consiste en clasificar a los clientes

¹ Fintech es la utilización de tecnología para ofrecer productos o servicios financieros innovadores.

por su valor en función de tres variables: 1) Recencia (R): tiempo transcurrido desde la última compra o transacción financiera en la entidad; 2) Frecuencia (F): número de compras por período de tiempo. En el caso de una entidad financiera será el número de productos adquiridos de ahorros y créditos; 3) Montos (M): Valor de las compras totales realizadas por el cliente en el tiempo de análisis. Para el caso de entidades financieras serían los saldos en productos de ahorro y /o de crédito también puede considerarse como el interés pagado por un crédito.

El CLV se define como el valor presente de todas las utilidades generadas y las esperadas por parte del cliente a lo largo de su relación con la compañía [8]. Dentro de los beneficios que brindan estas técnicas a las organizaciones se encuentran:

- Formular e implementar estrategias para la gestión de clientes con el fin de aumentar las ganancias que se obtienen a lo largo del ciclo de vida de la relación cliente-empresa, y buscando a la vez aumentar el largo de esta.
- Saber cuánto se puede invertir en conservar al cliente con el fin de obtener retorno sobre la inversión.
- Identificar cuáles son los clientes más y menos rentables para la empresa segmentándolos por características similares de hábitos de compra.
- Conocer qué tipo de clientes tiene la empresa, qué tipo de clientes han dejado de comprar, identificar aquellos clientes que gastan más y menos dinero.
- Diferenciar a los clientes más rentables para poder enfocarse en ellos de mejor forma.

Para realizar procesos de segmentación de clientes se pueden usar modelos de aprendizaje no supervisado, usando técnicas de clustering. A continuación, se presenta una breve descripción de las referencias analizadas:

En [8], los autores presentan un modelo de segmentación de clientes de una empresa de salud y belleza. Se utilizan dos enfoques: primero, el modelo clásico de RFM y posteriormente una propuesta extendida de RFM con un parámetro adicional, llamado “elemento de conteo”. La comparación de los resultados de estos, muestra que agregar un elemento de recuento como un nuevo parámetro al método RFM no influye en el resultado de la agrupación. El CLV se calcula con el método de RFM ponderado para cada segmento.

En [9], se presenta un modelo para el cálculo del CLV en un banco en Alemania, basado en una combinación de modelado de cadenas de Markov de primer orden y árboles de clasificación y

regresión. Además, se basaron en el análisis de grupos homogéneos en lugar de clientes individuales. El modelo fue validado utilizando una muestra de aproximadamente 6.2 millones de datos. El conjunto de datos de los clientes usados para este proyecto corresponde a: edad, variables sociodemográficas, estilos de vida, productos activos e historial transaccional.

En [10], realizan una investigación para identificar las técnicas de minería de datos usadas en el sector bancario en diferentes lugares del mundo, con el fin de ayudar a una mejor identificación de los clientes y diseñar estrategias de mercadeo más eficientes. La literatura analizada en este estudio por los autores cubre los años 2001-2017. Los resultados de estos casos son comparados en términos de los conjuntos de datos utilizados, el modelo usado y la precisión de predicción. Si bien se analizan varias técnicas usadas, los autores concluyen que la mayoría de las entidades bancarias analizadas, para segmentar clientes aplicaron algoritmos de clustering: K-means utilizando modelos de CLV y de RFM. Adicionalmente concluyen que las redes neuronales fueron el mejor modelo para predecir el comportamiento de los clientes.

En [11], los autores describen la aplicación de técnicas de minería de datos para segmentación de clientes, sobre datos reales de una empresa dedicada a la comercialización de productos tecnológicos de la región sur del Ecuador. Se aplicó la metodología CRISP-DM para el proceso de minería de datos, usaron el modelo de RFM, aplicaron algoritmos de clustering: K-means y mapas auto-organizados (SOM). Para validar el resultado de los algoritmos de clustering y seleccionar el que proporcionara grupos de mejor calidad, se usó la técnica de evaluación en cascada; para ello se aplicó un algoritmo de clasificación, colocando como etiqueta de clase a los grupos encontrados por los algoritmos de clustering y luego se midió la precisión de predicción con respecto a estos grupos. El algoritmo SOM fue el que proporcionó los mejores resultados.

En [12], se presentan los comportamientos de compra de los clientes de una empresa que comercializa solo por internet. Para ello, se aplicó la segmentación a partir de análisis de RFM, que calcularon mediante dos metodologías que son el método clásico de RFM y el método de las 2-Tuplas. En el primero, el autor realizó una clasificación numérica mediante quintiles que se enumeran de 1 a 5 tanto para la Recencia, la Frecuencia y el Valor Monetario. En el segundo, realizó otra clasificación de los clientes más precisa, más detallada y con la ventaja que ofrece un valor lingüístico para poder entender mejor a que clúster pertenece cada cliente. Finalmente, el autor presenta unos análisis de clúster con el método de K-means con diferentes segmentos. Todo esto con el fin de dar respuesta a la empresa, sobre cómo es el comportamiento de compra de cada cliente, cuáles son los más importantes, cuáles son los menos importantes, cuántos han dejado de

comprar, etc.

En [13], describen la aplicación del modelo de RFM a las cuentas de los socios de la COAC Jardín Azuayo, el proyecto se desarrolló dada la necesidad de la institución de contar con herramientas eficientes y eficaces para la toma de decisiones y conocimiento del socio. Primero, determinaron la importancia de construir una herramienta de inteligencia de negocios dentro de Jardín Azuayo que permitiera obtener información clara y concisa en tiempo real para la toma de decisiones. Segundo, continuaron con el desarrollo de metodologías para la gestión del valor del socio a través del conocimiento de sus necesidades analizando la información histórica de su última transacción realizada, la frecuencia con la que acude para acceder a los servicios que ofrece la cooperativa y el monto promedio por transacción.

En [14], se presenta una metodología y los resultados obtenidos en la segmentación de clientes utilizando un modelo de RFM extendido con nuevas variables, las cuales fueron seleccionadas mediante análisis multivariado. La propuesta fue aplicada en una empresa en la que se usaron variables, como; ganancia, porcentaje de ganancia y días de vencimiento de facturación, y se logró establecer una segmentación de clientes más detallada que el RFM clásico.

En [15], los autores describen una propuesta metodológica que combina dos técnicas: un algoritmo de fuzzy c-means y algoritmos genéticos para la segmentación de clientes en la industria de acero. Los clientes fueron divididos en dos grupos usando las variables de LRFM (Antigüedad, Recencia, Frecuencia, Monto de compras). Se encontró que el algoritmo combinado en este estudio tuvo un menor error medio cuadrado comparado con el obtenido cuando solo se utilizó fuzzy c-means.

En [16], se desarrolló un modelo con el objetivo de identificar el nivel de fidelidad de un cliente de comercio electrónico a partir del CLV, calculado a partir de clustering de clientes utilizando el algoritmo K-means y el análisis RFM. Se utilizaron 3 conjuntos de datos de transacciones de 3 diferentes marcas locales de comercio electrónico de moda en Indonesia. Este estudio encontró cinco grupos de clientes que fueron etiquetados basados en las clasificaciones de CLV, así: 1) el mejor cliente, 2) valiosos, 3) potencialmente valiosos, 4) promedio y 5) potencialmente invaluable. El CLV para cada segmento se calculó como un RFM ponderado. Los pesos de las variables de RFM, se obtuvieron a través de juicio de expertos y los valores originales de RFM de cada clúster fueron normalizados usando el método min-max.

En [17], realizan una segmentación de clientes usando una variación al método RFM, al cual incluyen el costo por cliente (C) y lo llaman RFM-C, y, desarrollan dos modelos, uno con el

algoritmo C4.5 y otro con Random Forest. Este proyecto fue realizado en una compañía de empaques plásticos en Indonesia, con el fin de clasificar los clientes y predecir compras futuras basado en los datos históricos de ventas. Después de procesar, analizar y evaluar los dos modelos, encontraron que con el algoritmo C4.5 se tuvo mayor precisión y mejor valor del índice de kappa comparado con el algoritmo de Random Forest, finalmente concluyen que con el algoritmo C 4.5 la compañía puede realizar una mejor segmentación de clientes y predicción de compras.

En [18], los autores proponen un procedimiento para la segmentación de clientes usando el modelo RFM, y clustering con K-means. La determinación del número óptimo de k grupos en el método K-means se evaluó utilizando ocho índices de validez, a saber, el método Elbow, el índice Silhouette, el índice de Calinski-Harabasz, el índice Davies-Bouldin, el índice Ratkowski, el índice Hubert, el índice Ball-Hall y el índice de Krzanowski-Lai, con el fin de mejorar la objetividad y la precisión en la segmentación, en comparación con el uso de un solo método. El conjunto de datos utilizado para este estudio contenía registros de ventas realizadas entre enero y diciembre de 2015 de una farmacia en Palembang (Indonesia). Los resultados de la evaluación mostraron que el número óptimo de grupos para el método K-means aplicado en el análisis RFM era tres, con un valor de varianza de 0,191.

2. Entendimiento del negocio

En este capítulo se describe el contexto donde se aplicará el proyecto, dado que está es una actividad que se requiere dentro de la metodología *CRISP-DM* donde se entiende el proyecto desde una perspectiva de negocio.

2.1 Descripción de la situación

La cooperativa es una entidad perteneciente al sector de servicios financieros con más de 58 años de operación, que tiene un cubrimiento a nivel nacional; actualmente, tiene seis oficinas ubicadas en los departamentos de Antioquia y Chocó; su departamento comercial tiene 20 asesores financieros. La cooperativa tiene una base social de aproximadamente 14.500 asociados activos. Su portafolio financiero consta de dos categorías: productos de ahorros (a la vista, contractuales y a término fijo), y productos de crédito (comercial, microcrédito, vivienda y de consumo).

La entidad analizada es consciente que en Colombia se ha venido ampliando y mejorando la oferta de servicios financieros, y cada vez es más complicado competir con las entidades bancarias, por lo que ella ha venido trabajando en varios proyectos que le permitan el diseño e implementación de estrategias comerciales y de servicio apoyadas en tecnologías de la información y comunicación para atraer, satisfacer, fidelizar y retener a sus asociados.

Dicha cooperativa requiere mejorar sus indicadores de desempeño comercial, a nivel de base social (nuevas afiliaciones), créditos, ahorros, y morosidad, tal que pueda mejorar su sostenibilidad, su permanencia en el sector financiero, y la generación de valor para sus asociados. Desde el punto de vista comercial ha logrado, en los últimos años, una mejor promoción y comercialización de su portafolio, a través de un mayor número de actividades publicitarias, aumento de pautas comunicacionales en múltiples medios, la contratación de un servicio externo de telemercadeo, la incorporación de más asesores comerciales para mayores acciones comerciales. Sin embargo, a pesar de las mejoras que se han hecho, la entidad no es eficiente en la gestión comercial de sus asociados; los principales problemas identificados son:

- No hay un conocimiento profundo de sus asociados por parte de la cooperativa.
- No desarrollan propuestas comerciales personalizadas a sus asociados.
- Para el cumplimiento de las metas comerciales anuales se realizan diferentes campañas, pero no se generan listados de asociados con mayores posibilidades de compra,

y, actualmente, se debe llamar a toda la base social para ofrecer los productos, o envían un correo electrónico masivo genérico o también envían mensajes de texto masivos.

- No tienen un conocimiento de cuáles son los asociados más valiosos.
- No realizan estrategias de servicio diferenciadas, que sean de acuerdo con el valor de los asociados.

2.2 Solución planteada

Con el fin de dar solución a las necesidades de negocios planteadas, y alineado con su plan estratégico, la cooperativa ha propuesto el desarrollo las siguientes estrategias a futuro:

- Aumentar la rentabilidad de las campañas comerciales, al optimizar los canales de contacto comercial con los asociados, específicamente en: la cantidad de llamadas telefónicas realizadas, la cantidad de mensajes de voz enviados, el número de correos electrónicos enviados, el número de mensajes de texto a celulares.
- Aumentar la venta de productos y servicios financieros en su base actual de asociados.
- Disminuir los indicadores de deserción de asociados, al poder tener un mejor conocimiento el perfil de los asociados y así desarrollar productos, servicios y estrategias de fidelización que aumenten el tiempo de permanencia promedio de los asociados en la entidad.
- Desarrollar campañas personalizadas priorizando la promoción en aquellos asociados con mayores probabilidades de compra.
- Aumentar el indicador de productividad comercial de los asesores comerciales (relación entre el número de asesorías realizadas y ventas obtenidas).
- Identificar cliente con mayores niveles de riesgo de crédito, con el fin de desarrollar acciones que minimicen el incumplimiento de los asociados y el deterioro de la cartera de créditos.

Se espera que todas estas acciones contribuyan a mejorar el estado de resultados de la cooperativa, sin embargo, en la actualidad estas no pueden ser implementadas porque la entidad no tienen los datos segmentados de los asociados, ni un perfil de los asociados lo suficientemente estructurado que les permita entre otros: 1) Identificar en cuales asociados se deben priorizar los esfuerzos y/o recursos; 3) a cuales se les deben dirigir las campañas comerciales, y tener mayor probabilidad de éxito; 4) la creación de un plan de fidelización y de recompensas para los “mejores asociados” desde el punto de vista transaccional, con el fin de que los ahorradores mantengan y aumenten sus

ahorros en la cooperativa y los deudores paguen oportunamente sus créditos; 5) Ofrecer productos y/o servicios que se ajusten a su comportamiento transaccional y estilo de vida; 6) Identificar asociados con mayor tendencia a retirarse de la entidad.

2.3 Criterios de éxito del negocio

La entidad considera como criterio de éxito empresarial si se logra los siguientes resultados a mediano plazo:

- Incrementar las ventas de productos financieros en al menos un 10%, en los clientes actuales.
- Disminuir los costos de la operación comercial y en la ejecución de las campañas comerciales en al menos un 15%, aumentando la rentabilidad de las campañas, al poder planear y ejecutar campañas focalizadas. Esta disminución en los costos principalmente será resultado por la optimización del número de llamadas a realizar por parte del centro externo de llamadas, los correos electrónicos comerciales, los costos de los mensajes de texto a celulares, los tiempos de asesoría del personal del área comercial.
- Disminuir el indicador de deserción anual de asociados en un 5% con respecto al valor actual.
- Aumentar la productividad comercial promedio anual de los asesores comerciales en un 15% con respecto al valor actual.
- Llevar el indicador de cartera de créditos morosos a un valor inferior al 5%.

2.4 Objetivos de la minería de datos

El objetivo de este proyecto es desarrollar un modelo analítico para la segmentación de clientes, que le permita a la cooperativa obtener un mejor conocimiento del comportamiento de sus asociados, para ser usado en el diseño de estrategias de comercialización y de fidelización, que le permita el cumplimiento de sus presupuestos comerciales anuales, mejorar los indicadores de riesgo y retención de asociados, y finalmente que las estrategias y acciones realizadas aporten a su competitividad y sostenibilidad en el sector de servicios financieros en Colombia.

2.5 Objetivos específicos

- Desarrollar un modelo de segmentación de asociados a partir de datos del comportamiento histórico transaccional en la entidad usando técnicas de clustering.
- Emplear un modelo predictivo de clasificación de árboles de decisión a los resultados de la segmentación, con el fin de conocer el conjunto de reglas de decisión que definen la composición de cada segmento.
- Realizar un análisis descriptivo de cada uno de los segmentos obtenidos, con el fin de tener una mayor comprensión de los tipos de asociados atendidos, generar un perfil de los asociados y que la entidad pueda usar este conocimiento para la personalización de los productos y/o servicios, el modelo de atención y la gestión del riesgo financiero.

2.6 Criterios de éxito de la minería de datos

Se determina como factor de éxito del modelo analítico, cuando los resultados o agrupamientos hallados permitan:

- Desarrollar una propuesta de segmentación de los asociados que permita identificar cada tipo de asociado (a que clúster pertenece) a partir de su comportamiento histórico transaccional en productos de ahorro y productos de crédito.
- Obtener un análisis descriptivo de los grupos y el “asociado típico” de cada grupo, para que sirva de base para la estrategia de segmentación del mercado y tener una mejor base para la construcción de un perfil del cliente deseado.
- Implementar un tablero de mando comercial donde se pueda visualizar el detalle de los agrupamientos y el perfil de los asociados que componen cada uno de los segmentos obtenidos.

2.7 Evaluación inicial de herramientas y técnicas

Como herramienta de trabajo en el proceso de extracción de conocimiento se utilizaron varias aplicaciones: Júpiter®, R software®, Rapid Miner® y Tableau®, para la visualización se usaron Google Data Studio® y R-Shiny®.

2.8 Conclusión

En este capítulo se presentó la etapa de comprensión del contexto de negocio, según la metodología *CRISP-DM*, donde el trabajo abordado se contempla desde un análisis desde la perspectiva de negocio y posteriormente desde una perspectiva de la minería de datos, se exponen los criterios de éxito del trabajo para satisfacer los objetivos propuestos.

3 Entendimiento de los datos

En este capítulo se realiza la descripción de los datos con el fin de realizar una exploración preliminar de los registros y verificar la calidad de estos. Se deben considerar la detección de datos faltantes y de datos atípicos que puedan presentarse para obtener confiabilidad en el conjunto de datos a trabajar.

3.1 Recolección de datos

Como parte de este proyecto se creó una bodega de datos en el sistema de gestión de bases de datos de la entidad (Ms SQL Server ®); sin embargo, se decidió no trabajar con una conexión directa al motor de bases de datos, sino que el conjunto de datos fue exportado a un archivo de Ms Excel ®. Las operaciones de filtrado de datos, validación de la integridad y una primera tarea de limpieza de datos fueron previamente gestionados en la actividad de creación de la bodega de datos.

Para el presente proyecto se consideran irrelevantes los siguientes datos de los asociados:

- Tipo de documento de identificación.
- Número de documento de identificación.
- Nombre completo del asociado.
- Dirección de residencia.
- Número de teléfono.
- Número de celular.
- Correo electrónico.

Para identificar cada registro se mapeó cada cédula a un nuevo valor entero autonumérico en un campo llamado Id.

El conjunto de datos a trabajar consta de 14.114 registros de asociados activos con información personal y transaccional registrados desde mayo de 1967 hasta julio 30 de 2020. A nivel de los asociados, se tiene datos sociodemográficos, historial de aportes, créditos aperturados, saldos actuales en productos de ahorros, aportes y créditos, así como comportamiento histórico de cartera. Los datos de los asociados fueron exportados a en un archivo en formato xlsx; las características del archivo con la información se encuentran en la Tabla 1.

Tabla 1: Descripción de la fuente de datos

Archivo:	asociados.xlsx
Número de instancias:	14.114
Número de atributos:	17

3.2 Descripción del conjunto de datos

La descripción de datos corresponde a los atributos contenidos en el archivo a utilizar para la tarea de minería. La dimensionalidad del conjunto de datos es multivariada, en lo que se encontró datos de tipo categóricos y numéricos. El conjunto completo de datos contiene 14.114 instancias con 17 atributos, de los cuales 8 son de tipo numéricos y 9 categóricos. Estos se describen en la Tabla 2, Tabla 3 y Tabla 4 respectivamente.

Tabla 2: Descripción atributos numéricos del conjunto de datos

Atributo	Promedio	Desviación estándar	Nulos	Únicos	Mínimo	Máximo
edad (años)	43	13,8	0	75	19	94
ingresos (\$)	\$ 2.128.303	\$ 3.224.608	0	3.349	\$ 0	\$ 184.189.118
antigüedad (meses)	57,142	59,67	0	305	0	638
saldoahorros (\$)	\$ 2.813.323	\$ 11.705.976	0	13.018	\$ 2.396	\$ 919.866.567
cantidadahorros	3	3	0	49	1	99
cantidadcreditoss	2,	12	0	112	0	380
recenciacreditos (meses)	8,	21,5	7,5	143	0	282
saldocreditos (\$)	\$ 2.866.732	\$ 7.416.639	0	4.516	\$ 0	\$ 175.690.124
cuotamensualcreditos (\$)	\$ 129.485	\$ 282.081	0	3.163	\$ 0	\$ 9.172.866
interespagadocreditos (\$)	\$ 1.068.308	\$ 3.718.032	0	6.461	\$ -	\$ 111.571.528
recenciaahorros (meses)	17,6	30,5	0	125	0	125

Tabla 3: Descripción atributos categóricos del conjunto de datos

Atributo	Cantidad	Nulos	Únicos	Listado de únicos
estadocivil	14.108	6	5	Soltero(a), Casado(a), Unión Libre, Separado(a), Viudo(a)
genero	14.109	5	2	Femenino, Masculino
niveleducativo	14.112	2	7	Sin estudio Formal, Primaria, Secundaria, Tecnológico, Técnico, Universitario Posgrado,
Rangoantigüedad (meses)	14.114	0	5	[0.0, 18.0), [18.0, 34.0), [34.0, 52.0), [52.0, 81.0), [81.0, 638.1)
Rangoedad (años)	14.114	0	5	18-30, 31-40, 41-50, 51-60, 60-100
Rangoingresos (Salarios mínimos)	14.114	0	5	<1, 1-2, 2-4, 4-6, > 6
tipodecontrato	14.111	3	14	Asalariado, contratista, Dependiente de tercero, Empleado oficial, Independiente, Inferior a Un año, Obra o Labor, Pensionado, Prestación de Servicios, Provisional, Temporales, Término Fijo, Término Indefinido, Otro,
tipodevivienda	14.070	44	3	Arrendada, Familiar, Propia

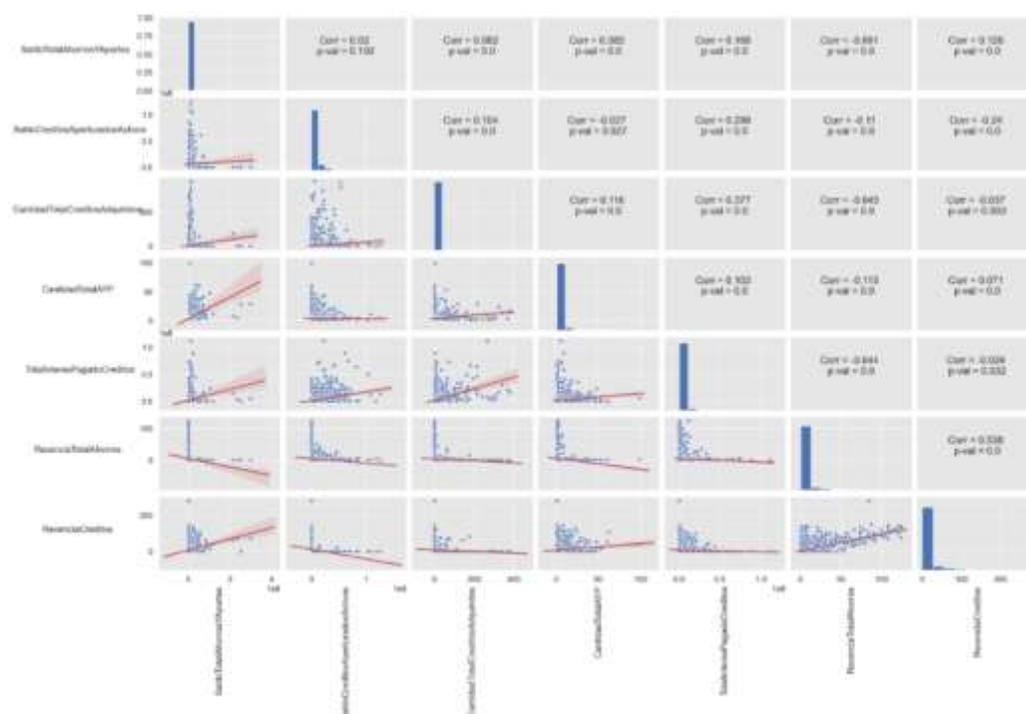
Tabla 4: Descripción atributos categóricos ordinales del conjunto de datos

Atributo	Mínimo	Máximo	Nulos	Únicos	Lista de únicos
estrato	0	6	0	7	0, 1, 2, 3, 4, 5, 6
rangoordinalantigüedad	1	5	0	5	1, 2, 3, 4, 5
rangoordinaledad	1	5	0	5	1, 2, 3, 4, 5
rangoordinalingresos	1	5	0	5	1, 2, 3, 4, 5

Para analizar las variables, y posibles correlaciones entre estas se presenta un gráfico de correlación

en la Figura 1. En este no se identifica una correlación entre las variables numéricas transaccionales de RFM y no se considera necesario realizar reducción de dimensionalidad.

Figura 1: Análisis de correlación por pares entre variables numéricas



3.3 Limpieza de datos

En la etapa de modelamiento se requieren actividades de limpieza de datos con el fin de ajustar los datos con los requerimientos del algoritmo a aplicar. Los valores faltantes en general son pocos. Se presentan datos faltantes en las variables:

- Género
- Estado civil.
- Tipo de contrato y
- Tipo de vivienda.

En el caso de la variable género y estado civil existen seis datos faltantes, en la variable género hay cinco datos faltantes, en la variable estado civil hay tres datos faltantes. El campo con más datos faltantes es el tipo de vivienda (44). Con el fin de continuar con el proceso se decidió eliminar todos los registros con datos faltantes.

3.4 Verificación de la calidad de los datos

Los datos almacenados de las transacciones financieras de los asociados son confiables, ya que por ser una entidad financiera es vigilada por varios entes de control externos, que le exigen el cumplimiento de varias normativas, y por tanto técnicamente la entidad tiene controles implementados en sus diferentes sistemas de información que garantizan las reglas de integridad y exactitud de las operaciones financieras.

Para la evaluación de la calidad de los datos se usó Rapidminer software, el cual analiza el conjunto de datos en términos de unicidad, estabilidad y cantidad de datos faltantes. Los resultados de la evaluación se presentan en la Figura 2. De acuerdo con los resultados de las métricas de dicha figura se determina que la calidad de los datos es apropiada para el desarrollo de los modelos de segmentación.

Figura 2: Evaluación de calidad de los datos

Row No. ↑	Attribute	ID-ness	Stability	Missing	Text-ness
1	Genero	0.000	0.515	0	0.038
2	NivelEducativo	0.002	0.347	0	0.069
3	EstadoCivil	0.001	0.386	0	0.136
4	Estrato	0.002	0.379	0	0.005
5	TipodeContrato	0.003	0.442	0	0.283
6	TipodeVivienda	0.001	0.472	0	0.033
7	SaldoAhorrosYAportes	0.986	0.000	0	0
8	CantidadAhorrosYAportesAdquiridos	0.008	0.456	0	0
9	CantidadCreditosAdquiridos	0.022	0.381	0	0
10	RecenciaCreditos	0.006	0.559	0	0
11	SaldoCreditosAperturadosActivos	0.967	0.002	0	0
12	TotalInteresPagadoCreditos	0.978	0.014	0	0
13	RecenciaAhorros	0.009	0.600	0	0
14	RangoOrdinalEdad	0.001	0.266	0	0.005
15	RangoOrdinalIngresos	0.001	0.396	0	0.005
16	OrdinalNivelEducativo	0.002	0.347	0	0.005
17	RangoOrdinalAntiguedad	0.001	0.246	0	0.005

3.5 Exploración de datos

El análisis exploratorio es una actividad que permite analizar en detalles algunas variables e identificar características de los asociados y formular algunas preguntas de interés con relación a los objetivos de minería definidos anteriormente. La exploración de los datos se puede realizar mediante gráficos, tablas y otras herramientas de visualización. En la exploración de datos se pretende dar respuesta a los siguientes interrogantes:

1. ¿Se puede identificar una relación entre el género de los asociados y su comportamiento transaccional en la entidad?
2. ¿Se puede identificar una relación entre el nivel educativo de los asociados y su comportamiento transaccional en la entidad?
3. ¿Se puede identificar una relación entre el estrato socioeconómico de los asociados y su comportamiento transaccional en la entidad?
4. ¿Se puede identificar una relación entre el estado civil de los asociados y su comportamiento transaccional en la entidad??
5. ¿Se puede identificar una relación entre el tipo de contrato de los asociados y su comportamiento transaccional en la entidad?
6. ¿Se puede identificar una relación entre el rango mensual de ingresos de los asociados y su comportamiento transaccional en la entidad?

El poder dar respuesta a los anteriores interrogantes permitirán establecer hipótesis de posibles criterios de variables a utilizar para el agrupamiento de los asociados y ayudar a la construcción de los perfiles.

El análisis exploratorio se realizará de manera independiente para el conjunto de datos de asociados con ahorros activos y el conjunto de datos de asociados con créditos activos.

3.5.1 Análisis exploratorio de asociados con ahorros activos

Con el fin de tener un mayor contexto de los asociados con ahorros activos en la cooperativa, se realizó un análisis exploratorio y descriptivo utilizando las variables transaccionales de RFM junto con las principales variables sociodemográficas. Con este análisis se espera identificar posibles relaciones de las variables sociodemográficas con el comportamiento transaccional de ahorros en la

entidad y tener datos para la construcción de los perfiles de los asociados ahorradores. El conjunto de datos de ahorradores activos consta de 14.062 registros de asociados con 15 atributos, de los cuales 7 son numéricos y 8 categóricos, los cuales se presentan en la Tabla 5.

Tabla 5: Descripción atributos numéricos de los asociados ahorradores

Atributo	Promedio	Valores nulos	Valores únicos	Mínimo	Máximo	Desviación estándar
Edad (años)	43	0	75	19	94	13,8
Ingresos (\$)	\$ 2.128.303	0	3.349	\$	18.189.118	\$ 3.224.608
Egresos (\$)	\$ 708.529	0	1.256	\$ 0	15.499.845	\$ 1.709.090
Antigüedad (meses)	57	0	305	0	638	59,7
SaldoAhorros (\$)	\$ 2.813.323	0	13.018	2.396	919.866.567	11.705.976
CantidadAhorros	3	0	49	1	99	3,6
RecenciaAhorros (meses)	18	0	125	0	125	30,6

De la Tabla 5, se encuentra que el rango de saldos de ahorros va desde \$ 2.396 hasta \$ 919.866.567. En cantidad de ahorros adquiridos hay asociados con un solo producto de ahorros hasta asociados con 99 productos adquiridos; a nivel de recencia de ahorros, los valores oscilan entre 0 y 125 meses; el promedio de saldos ahorrados es de \$ 2.813.323, de esta misma tabla se observa que el rango de edad de los asociados ahorradores oscila entre 19 y 94 años; la antigüedad en la cooperativa oscila entre 0 y 638 meses y finalmente existen en la entidad asociados con ingresos mensuales desde \$ 0 pesos hasta un máximo de \$ 18.189.118.

Para analizar la distribución porcentual de los saldos de ahorros, se realizó la división de los asociados en quintiles (cinco grupos con igual cantidad de asociado), los resultados son presentados en la Tabla 6. Se observa que el 80% de los asociados correspondientes a los quintiles 1, 2, 3 y 4 tienen saldos ahorrados menores a \$ 3.065.480 pesos, el último quintil de los asociados tienen saldos ahorrados entre \$ 3.065.866 y \$ 919.866.567. Este grupo corresponde a un 20% de los asociados ahorradores, los cuales representan el 74,2 % del total de saldos ahorrados en la entidad.

Tabla 6: Quintil de saldos en ahorros

Quintil	Saldo Ahorros \$	Porcentaje de saldos en ahorros	Mínimo	Máximo	Cantidad de asociados	Porcentaje de asociados	Porcentaje acumulado de asociados
Q1	\$ 179.872.275	0,46 %	\$ 2.396	\$ 133.000	2813	20 %	20 %
Q2	\$ 982.485.539	2,49 %	\$ 133.094	\$ 643.315	2812	20 %	40 %
Q3	\$ 2.981528.086	7,55 %	\$ 643.600	\$ 1.503.006	2812	20 %	60 %
Q4	\$ 6.049.359.287	15,31 %	\$ 1.503.077	\$ 3.065.480	2812	20 %	80 %
Q5	\$ 29.312.205.374	74,2 %	\$ 3.065.866	\$ 919.866.567	2813	20 %	100 %

Para analizar la distribución porcentual de cantidad de ahorros, se realiza una división de los asociados en cuatro grupos, en la Tabla 7, se presentan los resultados.

Tabla 7: Cuartil de cantidad de ahorros

Cuartil	Cantidad de ahorros	Porcentaje de ahorros	Mínimo	Máximo	Cantidad de asociados	Porcentaje de asociados	Porcentaje acumulado de asociados
C1	1.491	3,26 %	1	1	1.491	10,6 %	10,6 %
C2	13.890	30,4 %	2	2	6.945	49,4 %	60,0 %
C3	7.362	16,1 %	3	3	2.454	17,5 %	77,4 %
C4	22.960	50,2 %	4	99	3.172	22,6 %	100 %

De la Tabla 7, se identifica que el 60 % de los asociados con ahorros activos tienen dos o menos productos de ahorros, el cuartil 4 el cual tiene el 22,6 % de los asociados con ahorros activos tienen entre 4 y 99 productos de ahorro y además este mismo grupo tiene el 50,2 % de la cantidad total de cuentas de ahorros aperturadas.

Se realiza una división de los asociados en cuatro grupos para analizar la recencia en ahorros, los resultados de esta distribución se presentan en la Tabla 8. De esta se puede deducir los siguientes aspectos, el 39,9% de los asociados realizaron una operación de ahorros en el último mes, el 58,2 % de los asociados tienen una recencia en ahorros promedio menor a 4 meses. Hay un grupo compuesto

por el 20,3 % de los asociados con una recencia entre 32 y 125 meses.

Tabla 8: Cuartil de recencias en ahorros

Grupo	Recencia mínima (meses)	Recencia máxima (meses)	Cantidad de asociados	Porcentaje de asociados	Porcentaje acumulado de asociados
G1	0	0	5.611	39,9 %	39,9 %
G2	1	4	2.566	18,3 %	58,2 %
G3	5	31	3.028	21,5 %	79,7 %
G4	32	125	2.857	20,3 %	100 %

En la Tabla 9, se presenta la descripción de las variables categóricas de los asociados ahorradores.

Tabla 9: Descripción de los atributos categóricos de los asociados ahorradores

Atributo	Cantidad	Nulos	Únicos	Listado de únicos
estadocivil	14.062	0	5	Soltero(a), Casado(a), Unión Libre, Separado(a), Viudo(a)
genero	14.062	0	2	Femenino, Masculino
niveleducativo	14.062	0	7	Sin estudio Formal, Primaria, Secundaria, Tecnológico, Técnico, Universitario Posgrado,
rangoantiguedad	14.062	0	5	[0.0, 18.0), [18.0, 34.0), [34.0, 52.0), [52.0, 81.0), [81.0, 638)
rangoedad	14.062	0	5	18-30, 31-40, 41-50, 51-60, 60-100
rangoingresos	14.062	0	5	<1, 1-2, 2-4, 4-6, > 6

tipodecontrato	14.062	3	14	Asalariado, contratista, Dependiente de tercero, Empleado oficial, Independiente, Inferior a Un año, Obra o Labor, Pensionado, Prestación de Servicios, Provisional, Temporales, Término Fijo, Término Indefinido, Otro,
tipodevivienda	14.062	44	3	Arrendada, Familiar, Propia

La Distribución de saldos ahorrados por género se presenta en la Tabla 10, se encuentra que las mujeres constituyen el 56,8% de los ahorradores, igualmente presentan un mayor porcentaje de saldos ahorrados (60,8 %) y de cantidad de productos de ahorros adquiridos (62,5 %) con respecto a los hombres. En términos de recencia las mujeres presentan un menor promedio (16 meses).

Tabla 10: Distribución de saldos ahorrados por género

Género	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldos ahorrados	Porcentaje de productos de ahorros	Recencia de ahorros promedio (meses)
femenino	7.994	56,8%	60,8 %	62,5 %	16
masculino	6.068	43,2%	39,2 %	37,5 %	20

En la Tabla 11, se presentan los datos comportamiento en ahorros por rango de edad. De esta, se puede identificar que el grupo de asociados mayores de 60 años, los cuales constituyen el 13% del total de ahorradores, tienen el mayor porcentaje de saldos ahorrados (31,4 %), mientras que el grupo de asociados con rango de edad entre 18 y 30 años, que representan un 19% del total de ahorradores, son los que tienen el menor porcentaje ahorrado (8,2 %). La mayor cantidad de porcentaje de ahorros adquiridos (26,4 %) se encuentran en el grupo de rango de edad entre 31 y 40 años. En términos de recencia en ahorros, los asociados en el rango de edad entre 18 y 30 años tienen el menor promedio, y los asociados entre 41 y 50 años el mayor.

Tabla 11: Distribución de saldos ahorrados por rango de edad

Rango edad (años)	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldos ahorrados	porcentaje de productos de ahorros	Recencia de ahorros promedio (meses)
18-30	2.608	19%	8,2 %	14,8 %	14
31-40	4.083	29%	20,6 %	26,4 %	19
41-50	3.076	22%	19,6 %	21,1 %	20
51-60	2.451	17%	20,2 %	18,8 %	18
60-100	1.844	13%	31,4 %	18,8 %	15

En la Tabla 12, se presenta los resultados del comportamiento en ahorros con respecto al nivel educativo. Se encuentra los siguientes aspectos, los asociados con nivel educativo *Universitario* son un grupo con el 26% de los ahorradores, tienen el mayor porcentaje de saldos ahorrados (29,7 %) y el segundo en mayor porcentaje en productos de ahorro (25,8 %), los asociados con nivel educativo *Secundaria* constituyen el 31,6 % de los ahorradores, son el grupo de mayor cantidad de productos de ahorros adquiridos (29,4 %) y el segundo grupo con mayor porcentaje de saldos ahorrados (22,0 %). Los asociados de nivel educativo *Posgrado* (1,02 %) tienen el menor porcentaje de saldos ahorrados (2,2 %). A nivel de recencia los de menor promedio (13) son los de nivel educativo *Primaria* y los de mayor promedio son el grupo de los asociados *Sin estudios formales* (43 meses).

Tabla 12: Distribución de saldos ahorrados por nivel educativo

Nivel educativo	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldos ahorrados	Porcentaje de productos de ahorros	Recencia de ahorros promedio (meses)
Sin estudio formal	1.118	7,95 %	15,8 %	10,3 %	43
Primaria	2.001	14,23 %	14,1 %	14,3 %	13
Secundaria	4.443	31,6 %	22,0 %	29,4 %	16

Técnico	1.536	10,92 %	8,7 %	10,6 %	17
Tecnológico	1.148	8,16 %	7,6 %	8,2 %	14
Universitario	3.673	26,12 %	29,7 %	25,8 %	16
Posgrado	143	1,02 %	2,2 %	1,4 %	15

En la Tabla 13, se presenta la distribución de asociados ahorradores con respecto al estado civil. De acuerdo con lo presentado en esta, el mayor porcentaje de asociados ahorradores tienen estado civil Soltero(a) los cuales representan el 41,84 %, luego siguen los asociados con estado civil Unión libre con un 27,36 %. Los asociados con mayor porcentaje de saldos ahorrados (39,9 %) y de cantidad de ahorros adquiridos (39,2 %) tienen estado civil Soltero(a). Los asociados de estado civil Viudo(a) son los que menos porcentaje tienen en cantidad de ahorros adquiridos (3,2 %) y de saldos ahorrados (3,6 %). En términos de recencia de ahorros, los Viudos(a)s son los de menor valor promedio en recencia (8 meses) y los Separados(a)s son los de mayor valor promedio (24 meses).

Tabla 13: Distribución de saldos ahorrados por estado civil

Estado civil	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldos ahorrados	Porcentaje de productos de ahorro	Recencia promedio de ahorros (meses)
Soltero(a)	5.884	41,84 %	39,9 %	39,2 %	21
Casado(a)	3.493	24,84 %	32,5 %	28,8 %	14
Unión Libre	3.848	27,36 %	19,3 %	24,9 %	17
Separado(a)	533	3,79 %	4,7 %	4,0 %	24
Viudo(a)	304	2,16 %	3,6 %	3,2 %	8

En la Tabla 14, se presenta la distribución de asociados ahorradores con respecto al estrato socioeconómico.

Tabla 14: Distribución de saldos ahorrados por estrato socioeconómico

Estrato	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldos ahorrados	Porcentaje de productos de ahorros	Recencia de ahorros promedio (meses)
0	338	2,4 %	0,8 %	1,6 %	3
1	6.450	45,87 %	37,4 %	42,3 %	23
2	4.135	29,41 %	25,7 %	29,0 %	15
3	2.451	17,43 %	20,7 %	19,6 %	14
4	426	3,03 %	6,9 %	4,7 %	10
5	212	1,51 %	7,7 %	2,2 %	15
6	50	0,36 %	0,8 %	0,6 %	13

De acuerdo lo presentado en la Tabla 14, el mayor porcentaje de asociados ahorradores se encuentran ubicado en los estratos 1 y 2, los cuales en conjunto constituyen el 75,28 % del total de asociados y los cuales aportan el 63,1 % del total de los saldos en ahorros. El grupo de los asociados de estrato socioeconómico 1 son los de mayor porcentaje en saldos ahorrados (37,4 %), y en porcentaje de cantidad de productos activos (42,3 %), mientras que los asociados del estrato 0 son los de menor porcentaje en saldos ahorrados (0,8 %) y en cantidad de productos de ahorros activos (1,6 %). En términos de recencia de ahorros los asociados de estrato 0 tiene un menor valor promedio y los de estrato 2 y 5 tienen un mayor valor promedio.

En la Tabla 15, se presenta la distribución de asociados ahorradores con respecto al tipo de vivienda. De acuerdo con los datos de esta, los asociados con vivienda propia son los de mayor porcentaje tanto en cantidad de productos adquiridos (43,39 %) y en saldos ahorrados (55,6 %); mientras que los asociados con vivienda arrendada representan el menor porcentaje en las mismas variables mencionadas. En términos de recencia de ahorro, los asociados con vivienda familiar tienen un menor valor promedio (14 meses) y los de vivienda propia un mayor valor (21 meses).

Tabla 15: Distribución de saldos ahorrados por tipo de vivienda

Tipo de vivienda	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldos ahorrados	Porcentaje de productos de ahorro	Recencia de ahorros promedio (meses)
Arrendada	2.673	19,01 %	11,8 %	16,63 %	19
Familiar	5.938	42,23 %	32,7 %	40,09 %	14
Propia	5.451	38,76 %	55,6 %	43,29 %	21

En la Tabla 16, se presenta la distribución de asociados ahorradores con respecto al tipo de contrato, se cuantifica que el 81,24 % de los asociados tienen contrato laboral *a Término indefinido*, o son *Independientes* o tienen un *contrato a Término fijo*. El grupo de los asociados con contrato *a Término Indefinido* (40,78 %) son los de mayor porcentaje tanto en cantidad de productos de ahorros activos (39,3 %), como de saldos ahorrados (34,1 %). Los asociados con contratos de tipo *Temporal* (0,07 %), y *Contratos inferiores a un año* (0,2 %) son los de menor saldo ahorrado. Los *Empleados Oficiales* (0,26 %) son los de mayor recencia de ahorros (35 meses) y los de menor recencia los *Dependientes de tercero* (1 mes).

Tabla 16: Distribución de saldos ahorrados por tipo de contrato

Tipo de contrato	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldo de ahorros	Porcentaje de productos de ahorros	Recencia en ahorros promedio (meses)
Asalariado	42	0,3 %	0,2 %	0,3 %	6
Contratista	136	0,97 %	1,1 %	1,2 %	30
Dependiente de tercero	17	0,12 %	0,1 %	0,1 %	1
Empleado oficial	37	0,26 %	0,4 %	0,3 %	35
Independiente	3.748	26,65 %	24,4 %	26,3 %	15
Inferior a un año	28	0,2 %	0,2 %	0,2 %	33

Obra o labor	294	2,09 %	0,9 %	1,6 %	19
Otro	387	2,75 %	3,7 %	3,2 %	19
Pensionado	594	4,22 %	11,8 %	7,0 %	6
Prestación de servicios	932	6,63 %	5,7 %	6,2 %	18
Provisional	161	1,14 %	1,1 %	1,0 %	13
Temporales	10	0,07 %	0,1 %	0,1 %	3
Término fijo	1.942	13,81 %	16,3 %	13,1 %	31
Término indefinido	5.734	40,78 %	34,1 %	39,3 %	16

En la Tabla 17, se presenta la distribución de asociados ahorradores con respecto al rango de ingresos. De acuerdo con los datos de esta, el 37,42 % de los asociados ahorradores tienen un rango de ingresos mensuales entre 2 y 4 salarios mínimos, Luego sobresale un grupo del 27,42 % de los asociados con un rango de ingresos de 4 a 6 salarios mínimos. El grupo de asociados con rango de ingresos mensuales entre 1 y 2 salarios mínimos (4,95 %), son los de mayor porcentaje en cantidad de productos adquiridos (35,2 %), el mayor porcentaje de saldos ahorrados (30,8 %) se encuentra el grupo de asociados con rango de ingresos mensuales entre 2 y 4 salarios mínimos. En términos de recencia de ahorro, los de menor promedio (10 meses) son los asociados con rango de ingresos mayores a 6 salarios mínimos, y los mayores promedios (27 meses) son los asociados con rango de ingresos mensuales menores a 1 salario mínimo.

Tabla 17: Distribución de saldos ahorrados por rango de ingresos

Rango ingresos (salarios mínimos)	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldo de ahorros	Porcentaje de productos de ahorros	Recencia en ahorros promedio (meses)
< 1	3.189	22,68 %	17,3 %	21,9 %	27
[1-2)	696	4,95 %	27,7 %	35,2 %	17
[2-4)	5.262	37,42 %	30,8 %	28,7 %	14

[4-6)	3.856	27,42 %	16,4 %	8,9 %	11
> 6	1.059	7,53 %	7,9 %	5,4 %	10

Con las estadísticas descriptivas anteriores y haciendo un análisis de los valores promedio de las variables de comportamiento transaccional de los asociados agrupados por género, y estrato socioeconómico que se presentan en la Tabla 18 y a Tabla 19, se podrá tener más elementos para una comprensión y descripción del perfil de los asociados de la cooperativa. En la Tabla 18 se presenta los valores promedio de variables numéricas de asociados con ahorros activos agrupados por género y estrato socioeconómico, en la Tabla 19, se presentan las modas de las variables categóricas.

Tabla 18: Valores promedio del asociado agrupados por género, y estrato socioeconómico

Genero	Estrato	Edad (años) (Promedio)	Ingresos \$ (Promedio)	Antigüedad (meses) (Promedio)	Saldo ahorros \$ (Promedio)	Cantidad ahorros (Promedio)	Recencia ahorros (meses) (Promedio)
Femenino	0	38	\$ 1.784.241	10	\$ 977.995	2	4
	1	43	\$ 1.766.160	61	\$ 2.510.250	3	20
	2	43	\$ 1.689.900	50	\$ 2.739.917	4	14
	3	46	\$ 2.284.668	60	\$ 3.692.737	4	12
	4	53	\$ 2.683.702	93	\$ 5.303.982	5	7
	5	53	\$ 2.911.934	93	\$ 11.018.875	5	13
	6	59	\$ 3.243.729	161	\$ 8.052.034	7	14
Masculino	0	39	\$ 2.004.159	9	\$ 803.260	2	3
	1	42	\$ 2.005.864	60	\$ 1.955.771	3	26
	2	41	\$ 2.290.738	46	\$ 2.148.735	3	16
	3	45	\$ 3.065.668	58	\$ 2.904.050	3	15
	4	53	\$ 4.145.392	95	\$ 8.262.100	5	14
	5	54	\$ 3.929.818	100	\$ 20.646.294	4	18
	6	49	\$ 5.368.918	79	\$ 2.712.880	4	12

Tabla 19: Moda de variables categóricas de los asociados agrupados por género, y estrato socioeconómico

Genero	Estrato	Tipo de Vivienda (moda)	Tipo de Contrato (moda)	Nivel Educativo (moda)	Estado Civil (moda)
Femenino	0	Familiar	Término Indefinido	Universitario	Soltero(a)
	1	Propia	Término Indefinido	Universitario	Soltero(a)
	2	Familiar	Término Indefinido	Secundaria	Soltero(a)
	3	Familiar	Término Indefinido	Secundaria	Soltero(a)
	4	Propia	Término Indefinido	Universitario	Soltero(a)
	5	Propia	Término Indefinido	Universitario	Soltero(a)
	6	Propia	['Independiente' 'Pensionado']	Universitario	Soltero(a)
Masculino	0	Familiar	Término Indefinido	Secundaria	Soltero(a)
	1	Propia	Término Indefinido	Secundaria	Unión Libre
	2	Familiar	Término Indefinido	Secundaria	Soltero(a)
	3	Familiar	Término Indefinido	Secundaria	Soltero(a)
	4	Propia	Término Indefinido	Universitario	Casado(a)
	5	Propia	Término Indefinido	Universitario	Casado(a)
	6	['Familiar' 'Propia']	Pensionado	Universitario	Soltero(a)

Al analizar los resultados de la Tabla 18 y la Tabla 19, se obtuvieron los siguientes hallazgos estadísticos:

1. A nivel de mujeres, una descripción de un representante típico de los asociados ahorradores, a partir de los valores promedio de las variables numéricas y de las modas de las variables categóricas sería: *“Mujer con una edad promedio mayor de 38 años, estado civil soltera, con nivel educativo universitario, vivienda propia, estrato socioeconómico inferior a 3, vinculación laboral por un contrato a término indefinido”*.
2. Las mujeres de mayores saldos en ahorros promedio pertenecen al estrato socioeconómico cinco.
3. A nivel de hombres una descripción de un representante típico a partir de los valores promedio de las variables numéricas y de las modas de las variables categóricas sería: *“Hombre mayor de 39 años, estado civil soltero, con nivel educativo secundaria, con vivienda propia o familiar, vinculación laboral por un contrato a término indefinido”*.

4. Los hombres que mejor promedio en saldos en ahorros son de estrato socioeconómico cinco.
5. No se puede concluir que con los dos perfiles descriptos anteriormente se puede caracterizar el asociado típico de la entidad, son aproximaciones que desde un punto de vista de negocio pueden servir, sin embargo, es necesario realizar análisis estadísticos más completos los cuales no son el objetivo de este proyecto. Para un mejor y mayor análisis por parte de los funcionarios de la entidad, en la tabla anterior se pueden tener aproximaciones de 14 perfiles de asociados representantes o “típicos”, esto es, el perfil femenino y el perfil masculino de un asociado por cada estrato socioeconómico. Por ejemplo, la descripción de una mujer con ahorros activos típico del estrato socioeconómico 2 sería: *“Mujer asociada con una edad promedio de 43 años, estado civil soltera, con nivel educativo universitario, tiene vivienda propia, vinculación laboral por un contrato a término indefinido, ingresos mensuales promedio de \$ 1.689.900, una antigüedad en la cooperativa promedio de 50 meses, con cuatro productos de ahorros, un saldo ahorrado promedio de \$ 2.739.917 y una recencia promedio de 14 meses”.*

3.5.2 Análisis exploratorio de asociados con créditos activos

De manera análoga a lo realizado con los asociados ahorradores con el fin de tener un mayor contexto de los asociados con créditos activos en la cooperativa, se presenta un detalle del análisis exploratorio de los datos, acompañado de un análisis descriptivo utilizando las variables transaccionales usadas para la segmentación RFM. Con este análisis se espera identificar posibles relaciones sociodemográficas con el comportamiento transaccional en créditos en la entidad. El conjunto de datos final de asociados deudores contiene 4.645 instancias, las variables numéricas se presentan en la Tabla 20.

Tabla 20: Descripción atributos numéricos de asociados con créditos activos

Atributo	Promedio	Desviación estándar	Nulos	Únicos	Mínimo	Máximo
Edad (años)	44	14	0	71	19	93
Ingresos (\$)	\$ 2.636.622	\$ 3.328.552	0	1.750	\$ 0	\$ 98.000.000
Antigüedad (meses)	56	60	0	242	0	580
SaldoCreditosActivos (\$)	\$ 8.638.872	\$ 10.512.164	0	4.639	\$ 4	\$ 126.147.319

CantidadCreditosAdquiridos	5	20	0	103	1	380
RecenciaCreditos (meses)	1	2	0	26	0	37
InteresPagadoCreditos (\$)	\$ 2.307.680	\$ 5.620.260	0	4.545	\$0	\$ 111.571.528

Analizando los datos de la Tabla 20 se encuentra que el rango de saldos de crédito adeudados va de \$ 4 hasta \$126.147.319; En cantidad de créditos adquiridos hay asociados con un solo crédito hasta asociados con 380 créditos adquiridos; los valores de recencia de créditos se encuentran entre 0 y 37 meses; el rango de interés pagado por créditos esta entre cero (créditos recientemente otorgados) y \$ 111.571.528. De esta misma tabla se observa que el rango de edad de los asociados se encuentra entre 19 y 93 años; la antigüedad en la cooperativa oscila entre 0 y 580 meses; a nivel de ingresos mensuales existen valores desde \$ 0 hasta un máximo de \$ 98.000.000.

Para analizar la distribución porcentual de los saldos de créditos, se realizó la división de los asociados en quintiles (cinco grupos con igual cantidad de asociados) y se obtuvieron los resultados presentados en la Tabla 21. De esta, se encuentra que el 20 % de los asociados tienen saldos adeudados entre \$ 11.904.987 y \$ 126.147.319, el 80% restantes tiene saldos adeudados menores a \$ 11.890.995.

Tabla 21: Distribución en quintiles de saldos en créditos

Quintil	Saldo créditos	Porcentaje saldos de créditos	Saldo de créditos mínimo	Saldo de créditos máximo	Cantidad de asociados	Porcentaje de asociados	Porcentaje acumulado de asociados
Q1	\$ 995.582.380	2,48 %	\$ 4	\$ 2.051.251	929	20 %	20 %
Q2	\$ 2.885.400.015	7,19 %	\$ 2.057.652	\$ 4.153.274	929	20 %	40 %
Q3	\$5.066.612.303	12,63 %	\$ 4.156.695	\$ 6.960.190	929	20 %	60 %
Q4	\$.418.420.638	20,98 %	\$ 6.964.907	\$ 11.890.995	929	20 %	80%
Q5	\$ 22.7615.44.687	56,72 %	\$ 11.904.987	\$ 126.147.319	929	20 %	100%

Para analizar la distribución porcentual de interés pagado, se realiza una división de los asociados en cuatro grupos, en la Tabla 22 se presentan los resultados. Se identifica que el 20 % de los asociados han pagado por interés valores entre \$ 2.991.720 y \$ 111.571.528, el 80 % restante han pagado un interés menor a \$ 2.979.169.

Tabla 22 Distribución en quintiles de interés pagados por créditos

Quintil	Interés pagado por créditos \$	Porcentaje de interés pagado por créditos	Interés mínimo \$	Interés máximo \$	Cantidad de asociados	Porcentaje de asociados	Porcentaje acumulado de asociados
Q1	16.329.919	0,15 %	0	\$ 43.511	929	20 %	20 %
Q2	87.507.737	0,82 %	\$ 43.512	\$ 182.350	929	20 %	40 %
Q3	442.448.656	4,13 %	\$ 182.425	\$ 904.074	929	20 %	60 %
Q4	1.643.043.641	15,33 %	\$ 904.254	\$ 2.979.169	929	20 %	80 %
Q5	8.529.845.327	79,58 %	\$ 2.991.720	\$ 111.571.528	929	20 %	100 %

Para analizar la distribución porcentual de cantidad de productos de crédito adquiridos, se realizó una división de los asociados en cuatro grupos, la cual se presenta en la Tabla 23. Analizando los resultados se obtuvo que el 38,08 % de los asociados con créditos activos han adquirido un solo producto de crédito, mientras que un 25,58 % de los asociados igualmente con créditos activos han adquirido más de 4 créditos.

Tabla 23: Rango de cantidad de créditos adquiridos

Cantidad de créditos adquiridos	Cantidad de créditos	Porcentaje de créditos	mínimo	máximo	Cantidad de asociados	Porcentaje de asociados	Porcentaje acumulado de asociados
1	1.769	7,4 %	1	1	1.769	38,08 %	38,08 %
2	4.037	16,8 %	2	3	1.688	36,34 %	74,42 %
3	18.280	75,9 %	4	380	1.188	25,58 %	100 %

Se realiza una división de los asociados en cuatro grupos para analizar la recencia en créditos, los resultados de esta distribución se presentan en la Tabla 24. Al analizarla, se puede deducir los siguientes aspectos: El 38,88 % de los asociados con créditos activos realizaron una operación de créditos en el último mes. El restante tiene recencias que van de 1 mes hasta 37 meses.

Tabla 24: Rango de recencia de créditos

Recencia créditos (meses)	Mínimo	Máximo	Frecuencia	Porcentaje de asociados	Porcentaje acumulado de asociados
1	0	0	1.806	38,88 %	38,88 %
2	1	37	2.839	61,12 %	100 %

En la Tabla 25 se presenta la descripción de las variables categóricas de los asociados con créditos activos.

Tabla 25: Descripción de atributos categóricos de asociados con créditos activos

Atributo	Cantidad	Nulos	Únicos	Listado de únicos
Estadocivil	4.645	0	5	Soltero(a), Casado(a), Unión Libre, Separado(a), Viudo(a)
Genero	4.645	0	2	Femenino, Masculino
Niveleducativo	4.645	0	7	Sin estudio Formal, Primaria, Secundaria, Tecnológico, Técnico, Universitario Posgrado,
Rangoantigüedad	4.645	0	5	[0.0, 18.0), [18.0, 34.0), [34.0, 52.0), [52.0, 81.0), [81.0, 638)
Rangoedad	4.645	0	5	18-30, 31-40, 41-50, 51-60, 60-100
Rangoingresos	4.645	0	5	<1, 1-2, 2-4, 4-6, > 6
Tipodecontrato	4.645	3	14	Asalariado, contratista, Dependiente de tercero, empleado oficial, Independiente, Inferior a Un año, Obra o Labor, Pensionado, Prestación de Servicios, Provisional, Temporales, Término Fijo, Término Indefinido, Otro,
Tipodevivienda	4.645	44	3	Arrendada, Familiar, Propia

A continuación, se realizará una descripción más amplia de las variables categóricas presentadas en la Tabla anterior.

La Distribución de saldos de créditos por genero se presenta en la Tabla 26, se encuentra que las

mujeres constituyen el 51,4% de los asociados deudores. Los asociados con género masculino presentan un leve mayor porcentaje de saldos adeudados en créditos con respecto a las mujeres, sin embargo, las mujeres son las que mayor porcentaje del total de interés han pagado. En términos de recencia de crédito promedio ambos géneros tienen el mismo valor.

Tabla 26: Distribución de créditos activos por genero

Genero	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldos de créditos adeudados	Porcentaje de interés pagado por créditos	Porcentaje de créditos adquiridos	Recencia de créditos promedio (meses)
Femenino	2.392	51,4 %	49,6 %	52,4 %	50,6 %	0,9
Masculino	2.253	48,5 %	50,4 %	47,6 %	49,4 %	0,9

Con respecto al comportamiento en créditos por rango de edad, en la Tabla 27 se presentan los datos.

Tabla 27: Distribución de créditos activos por rango de edad

Rango Edad (años)	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldo de créditos	Porcentaje interés pagado por créditos	Porcentaje de Créditos adquiridos	Recencia en créditos promedio (meses)
18-30	873	19 %	15,4 %	6,2 %	8,1 %	1,0
31-40	1.235	27 %	27,5%	18,4 %	24,4 %	0,9
41-50	1.025	22 %	23,7 %	23,8 %	20,4 %	0,9
51-60	882	19 %	19,5 %	23,2 %	20,1 %	0,8
60-100	630	14 %	14,0 %	28,3 %	27,0 %	0,7

De la Tabla 27 se puede identificar los siguientes aspectos: El rango de edad donde existe un mayor porcentaje de asociados deudores (27,5 %) se encuentra entre 31 y 40 años seguido del rango entre 41 y 50 años (23,7 %). El mayor porcentaje de ingresos (28,3 %) por interés de créditos proviene de asociados mayores de 60 años, seguido de asociados en el rango de edad entre 41 y 50 (23,8 %). Los asociados mayores de 60 años son los que menor promedio de recencia de crédito tienen, y los

de mayor promedio en recencia son los de asociados entre 18 y 30 años de edad.

Con respecto al nivel educativo, en la Tabla 28 se presentan los resultados. Analizando se encuentra los siguientes aspectos Los mayores deudores de la entidad tienen nivel académico *Secundaria* con un 33,2 % de participación en los saldos, seguido de *Universitario* con un 24,6 %, y los de menor porcentaje tienen nivel educativo *Posgrado* con un valor de 1,7 %. A nivel de recencia los de menor promedio son los de nivel educativo *Posgrado* y los de mayor promedio los de nivel de estudio *Tecnológico*.

Tabla 28: Distribución de créditos activos por nivel educativo

Nivel educativo	Cantidad de asociados	Porcentaje de asociados	Porcentaje de Saldo de Créditos	Porcentaje de interés pagado por créditos	Porcentaje de créditos adquiridos	Recencia de créditos promedio (meses)
Sin Estudio Formal	296	6,37 %	7,1 %	22,4 %	21,7 %	0,8
Primaria	746	16,06 %	14,6 %	10,6 %	8,0 %	0,9
Secundaria	1.612	34,7 %	33,2 %	23,8 %	18,9 %	0,8
Técnico	522	11,24 %	10,1 %	9,7 %	11,7 %	0,9
Tecnológico	423	9,11 %	8,7 %	7,1 %	8,1 %	1,0
Universitario	993	21,38 %	24,6 %	24,9 %	29,7 %	0,8
Posgrado	53	1,14 %	1,7 %	1,5 %	2,0 %	0,7

En la Tabla 29 se presenta la distribución de asociados deudores con respecto al estado civil. Los asociados con mayor porcentaje de saldos adeudados en créditos y de cantidad de créditos adquiridos tienen estado civil *Soltero(a)*, los que más han pagado interés son de estado civil *Casado(a)*. Los asociados de estado civil *Separado(a)* son los que menos porcentaje tiene en cantidad de créditos adquiridos y de saldos adeudados. En términos de recencia de crédito no hay una diferencia promedia notoria entre asociados de diferentes estados civil, sin embargo, los de estado civil *Casado(a)* son los de menor valor promedio en recencia y los *Separados(a)s* son los de mayor valor promedio.

Tabla 29: Distribución de créditos activos por estado civil

Estado civil	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldo de crédito	Porcentaje de interés pagado por créditos	Porcentaje de créditos adquiridos	Recencia de créditos promedio (meses)
Soltero(a)	1.794	38,62 %	35,7 %	32,8 %	39,1 %	0,9
Casado(a)	1.305	28,09 %	32,6 %	37,7 %	29,9 %	0,8
Unión libre	1.271	27,36 %	26,3 %	22,2 %	22,8 %	0,9
Separado(a)	178	3,83 %	3,3 %	4,9 %	5,3 %	1,0
Viudo(a)	97	2,09 %	2,2 %	2,4 %	3,0 %	0,6

En la Tabla 30 se presenta la distribución de asociados con créditos activos con respecto al tipo de vivienda. De acuerdo con los datos de la Tabla 29, Los asociados con vivienda propia son los de mayor porcentaje tanto en cantidad de productos adquiridos (50,4 %), saldos de créditos adeudados (44,8 %) e interés pagado (59,9 %), mientras que los asociados con vivienda arrendada representan el menor porcentaje en las mismas variables mencionadas. En términos de recencia de crédito los asociados de vivienda propia tienen un menor valor promedio y los de vivienda arrendada un mayor valor.

Tabla 30: Distribución de créditos activos por tipo de vivienda

Tipo de vivienda	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldo de crédito	Porcentaje de interés pagado por créditos	Porcentaje de créditos Adquiridos	Recencia de créditos promedio (meses)
Arrendada	716	15,41 %	12,4 %	9,3 %	10,0 %	1,0
Familiar	2193	47,21 %	42,8 %	30,8 %	39,6 %	0,9
Propia	1.736	37,37 %	44,8 %	59,9 %	50,4 %	0,8

En la Tabla 31 se presenta la distribución de asociados con créditos con respecto al estrato socioeconómico. El grupo de los asociados de estrato socioeconómico 2 son los de mayor porcentaje en saldos de créditos adeudado (37,1 %), mientras que los asociados del estrato 1 son los de mayor porcentaje en cantidad de productos adquiridos (34,6 %) e interés pagado (38,4 %). Los asociados de estrato 6 y 5 respectivamente son los de menor saldo en créditos adeudados. En términos de recencia de crédito los asociados de estrato 6 tiene un menor valor promedio y lo de estrato 0 y 3 tienen un mayor valor promedio.

Tabla 31: Distribución de créditos activos por estrato socioeconómico

Estrato	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldo de créditos	Porcentaje de interés pagado por créditos	Porcentaje de créditos adquiridos	Recencia de créditos promedio (meses)
0	83	1,79 %	1,3 %	0,1 %	0,5 %	1,0
1	1.583	34,08 %	32,0 %	38,4 %	34,6 %	0,8
2	1.760	37,89 %	37,1 %	27,1 %	28,7 %	0,9
3	1.032	22,22 %	24,8 %	26,1 %	24,5 %	1,0
4	129	2,78 %	3,8 %	6,0 %	7,6 %	0,8
5	45	0,97 %	0,8 %	0,9 %	1,1 %	0,7
6	13	0,28 %	0,2 %	1,4 %	2,9 %	0,6

En la Tabla 32 se presenta la distribución de asociados con créditos activos con respecto al tipo de contrato. El grupo de los asociados con contrato a Término indefinido son los de mayor porcentaje tanto en cantidad de productos adquiridos (40,5 %), saldos de créditos adeudados (42,1 %) e interés pagado (39,0 %), seguidos de los asociados con actividad económica Independiente. Los asociados con contratos de tipo temporales, y contratos inferiores a un año son los de menor saldo en créditos adeudados. En términos de recencia de crédito los asociados con contrato de tipo Obra o Labor son los de mayor recencia de crédito y lo Dependientes de Tercero son los de menor recencia.

Tabla 32: Distribución de créditos activos por tipo de contrato

Tipo de Contrato	Cantidad de asociados	Porcentaje de asociados	Porcentaje de saldo de créditos	Porcentaje de interés pagado por créditos	Porcentaje de créditos adquiridos	Recencia de créditos promedio (meses)
Asalariado	14	0,3 %	0,3 %	0,1 %	0,1 %	1,1
Contratista	32	0,69 %	0,7 %	1,7 %	3,0 %	0,8
Dependiente de tercero	2	0,04 %	0,0 %	0,0 %	0,0 %	0,5
Empleado oficial	15	0,32 %	0,3 %	0,6 %	0,3 %	0,6
Independiente	1.314	28,29 %	34,4 %	27,3 %	23,5 %	0,8
Inferior a Un año	6	0,13 %	0,1 %	0,1 %	0,5 %	0,7
Obra O Labor	73	1,57 %	1,25 %	1,0 %	1,1 %	1,3
Pensionado	55	1,18 %	4,4 %	10,4 %	13,9 %	0,7
Prestación de Servicios	234	5,04 %	5,1 %	3,7 %	4,3 %	1,2
Provisional	244	5,25 %	0,8 %	1,0 %	0,6 %	0,7
Temporales	38	0,82 %	0,1 %	0,1 %	0,1 %	1,0
Término Fijo	5	0,11 5	9,2 %	13,2 %	10,4 %	0,8
Término Indefinido	559	12,03 %	42,1 %	39,0 %	40,5 %	0,9
Otro	2.054	44,22 %	1,2 %	1,7 %	1,7 %	1,4

En la Tabla 33 se presenta la distribución de asociados con créditos activos con respecto al rango de ingresos. De acuerdo con los datos presentados, el grupo de asociados con rango de ingresos mensuales entre 2 y 4 salarios mínimos son los de mayor porcentaje tanto en cantidad de productos adquiridos (38,8 %), saldos de créditos adeudados (35,2 %) e interés pagado (36,1 %), mientras que

los asociados con ingresos menores a un salario mínimo representan el menor porcentaje de en las mismas variables mencionadas. En términos de recencia de crédito no hay una diferencia promedia notoria entre asociados de diferentes rangos de ingresos.

Tabla 33: Distribución de créditos activos por rango de ingresos

Rango de ingresos (Salarios mínimos)	Cantidad de asociados	Porcentaje de asociados	Porcentaje en saldo de créditos	Porcentaje de interés pagado por créditos	Porcentaje de créditos adquiridos	Recencia en créditos promedio (meses)
< 1	1.543	33,22 %	3,3 %	3,1 %	3,9 %	1,0
1-2	388	8,35 %	24,0 %	19,9 %	24,0 %	0,9
2-4	401	8,63 %	35,2 %	36,1 %	38,8 %	0,9
4-6	1.840	39,61 %	16,4 %	19,8 %	19,7 %	0,9
> 6	401	8,63 %	21,0 %	21,2 %	13,6 %	0,8

Con las estadísticas descriptivas anteriores y haciendo un análisis de los valores promedio de las variables de comportamiento transaccional de los asociados agrupados por género, y estrato socioeconómico que se presentan en la Tabla 34 se podrá tener más elementos para una comprensión y descripción del perfil de los asociados deudores de la cooperativa.

Tabla 34: Valores promedio de asociados con créditos activos agrupados por género, y estrato socioeconómico

Genero	Estrato	Promedio	Promedio	Promedio	Promedio	Promedio	Promedio	Promedio
		Edad (años)	Ingresos \$	Antigüedad (meses)	Saldo créditos \$	Interés pagado por créditos \$	Cantidad de créditos adquiridos	Recencia créditos (meses)
Femenino	0	36	\$ 1.874.214	11	\$ 5.483.924	\$ 123.173	2	1,2

	1	46	\$ 2.356.164	68	\$ 8.292.880	\$ 2.806.841	6	0,8
	2	43	\$ 2.088.062	49	\$ 7.656.108	\$ 1.553.150	4	0,9
	3	46	\$ 2.693.667	57	\$ 9.415.791	\$ 2.643.228	6	0,9
	4	52	\$ 3.056.348	117	\$ 9.919.429	\$ 3.710.852	10	0,8
	5	49	\$ 3.555.505	68	\$ 6.794.589	\$ 1.747.754	7	0,8
	6	58	\$ 4.036.541	250	\$ 8.257.884	\$ 8.800.519	10	0,5
Masculino	0	37	\$ 2.308.878	10	\$ 6.984.856	\$ 92.255	1	0,9
	1	44	\$ 2.417.576	59	\$ 7.862.764	\$ 2.306.384	5	0,9
	2	41	\$ 2.821.038	42	\$ 9.079.928	\$ 1.728.220	4	0,9
	3	44	\$ 3.551.915	57	\$ 9.877.265	\$ 2.788.955	6	1,0
	4	53	\$ 4.363.596	113	\$ 14.505.131	\$ 6.788.589	21	0,7
	5	54	\$ 5.203.067	95	\$ 7.575.482	\$ 2.735.566	4	0,6
	6	52	\$ 3.687.835	98	\$ 6.853.985	\$ 13.759.992	92	0,7

Tabla 35: Valores promedio de asociados con créditos activos agrupados por género, y estrato socioeconómico

		Edad (años)	Tipo de Vivienda	Tipo de Contrato	Nivel Educativo	Estado Civil
Genero	Estrato	Promedio	Moda	Moda	Moda	Moda
Femenino	0	36	Familiar	Término Indefinido	Secundaria	Soltero(a)
	1	46	Propia	Término Indefinido	Universitario	Soltero(a)

	2	43	Familiar	Término Indefinido	Secundaria	Soltero(a)
	3	46	Familiar	Término Indefinido	Secundaria	Soltero(a)
	4	52	Propia	Término Indefinido	Universitario	Soltero(a)
	5	49	Propia	Término Indefinido	Universitario	Soltero(a)
	6	58	Propia	['Pensionado' 'Prestación de Servicios']	Sin Estudio Formal	Casado(a)
Masculino	0	37	Familiar	Término Indefinido	Secundaria	Soltero(a)
	1	44	Propia	Término Indefinido	Secundaria	Unión Libre
	2	41	Familiar	Término Indefinido	Secundaria	Soltero(a)
	3	44	Familiar	Término Indefinido	Secundaria	Soltero(a)
	4	53	Propia	Término Indefinido	Universitario	Casado(a)
	5	54	Propia	Término Indefinido	Universitario	Casado(a)
	6	52	['Familiar' 'Propia']	Pensionado	Universitario	Soltero(a)

Analizando los resultados de las Tablas 34 y 35 se puede concluir los siguientes hallazgos estadísticos con respecto los asociados con créditos activos:

1. A nivel de mujeres, la descripción de una asociada típica, a partir de los valores promedio de las variables numéricas y de las modas de las variables categóricas sería:

“Mujer con una edad promedio mayor de 38 años, estado civil soltera, con nivel educativo universitario, tiene vivienda propia, vinculación laboral por un contrato a término indefinido”.

2. Las mujeres con mayor promedio en saldos en créditos pertenecen al estrato socioeconómico cuatro.

3. A nivel de hombres una descripción de un asociado “típico”, a partir de los valores promedio de las variables numéricas y de las modas de las variables categóricas sería: *“Hombre con una edad promedio mayor de 39 años, estado civil soltero, nivel educativo secundario, con vivienda propia o familiar, vinculación laboral por un contrato a término indefinido”*

4. Los hombres con mayor promedio en saldos en créditos son de estrato socioeconómico cuatro.

5. No se puede concluir que con los dos perfiles descriptos anteriormente se puede caracterizar el asociado con créditos activos de la entidad, son aproximaciones que desde un punto de vista de negocio pueden servir, sin embargo, es necesario realizar análisis estadísticos más completos los cuales no son el objetivo de este proyecto. Para un mejor y mayor análisis por parte de los funcionarios de la entidad, en la tabla anterior se pueden tener aproximaciones de 14 perfiles de asociados representantes o “prototipo”, esto es, el perfil femenino y el perfil masculino de un asociado por cada estrato socioeconómico. Por ejemplo, la descripción de una mujer con créditos activos, típico o representante del estrato socioeconómico 2 sería:

“Mujer con una edad promedio de 43 años, estado civil soltera, con nivel educativo secundaria, tiene vivienda familiar, vinculación laboral por un contrato a término indefinido, ingresos mensuales promedio de \$ 2.088.062, una antigüedad en la cooperativa promedio de 49 meses, con un historial de cuatro productos de créditos adquiridos, un saldo en créditos promedio de \$ \$ 7.656.108, una recencia promedio de 1 mes y ha pagado \$ 1.553.150 por interés en créditos”.

3.6 Conclusiones

En este capítulo se expuso la etapa de entendimiento de los datos y se cumple en parte con el primer objetivo que comprende los pasos de recolección de datos, exploración inicial y verificación de calidad de datos. Cabe resaltar que esta etapa es crucial, de mayor esfuerzo y la que determina el avance y validez de las demás etapas.

La actividad de verificación de los datos permitió conocer el conjunto de datos, identificando los atributos con valores perdidos o faltantes, los cuales fueron eliminados. Finalmente se seleccionaron los datos para el procesamiento y la construcción del modelo que serán analizados en profundidad en la siguiente fase de este proyecto.

El proceso de comprensión de los datos permite tener un panorama más amplio del contexto de la cooperativa. Con el análisis estadístico desarrollado, se obtuvo un mayor conocimiento del comportamiento de los asociados con ahorros activos y con créditos activos. Esta información combinada con otros tipos de análisis le permitirá a la entidad una mejor caracterización de sus asociados, y realizar un mejor perfilamiento de las estrategias comerciales, de riesgos y de servicio de acuerdo a los datos de comportamiento transaccional y los estilos de vida de los asociados.

4 Modelamiento

En este capítulo se trata la fase de modelamiento, la cual consiste en seleccionar una metodología de minería de datos que pueda aplicarse a los datos disponibles y obtener el resultado esperado; en esta fase puede ser necesario volver a la fase de preparación de datos para ajustarlos de acuerdo a los requerimientos del algoritmo seleccionado.

En la etapa de comprensión del negocio se definió que el alcance de este proyecto es obtener una segmentación que permita a la cooperativa obtener un mayor y mejor conocimiento de los asociados y que este pueda a corto plazo ser utilizado en el diseño, promoción, venta de productos y servicios financieros, así como en fortalecer el modelo de gestión comercial y de servicio. La justificación de usar las variables de RFM es porque son las variables recomendadas y más ampliamente usadas de acuerdo con la bibliografía analizada para proyectos de segmentación de clientes, donde hay un interés en diferenciar los mejores clientes desde el punto de vista de valor financiero para la entidad, y también para creación de modelos de fidelización de clientes.

Para el desarrollo de este proyecto se aplicaron técnicas no supervisadas que realizan agrupamiento de datos por su similitud en variables de comportamiento transaccional, con el fin de encontrar segmentos homogéneos. Al finalizar esta fase de modelamiento se espera dar respuestas a las siguientes preguntas:

1. ¿Cómo podrían segmentarse los asociados de la entidad en comportamiento de ahorros?
2. ¿Cómo podrían segmentarse los asociados de la entidad en comportamiento de créditos?
3. ¿Cuáles son las variables que más influyen en la segmentación obtenida de los asociados?
4. ¿Cómo puede describirse el perfil de los asociados de acuerdo a las agrupaciones obtenidas?

4.1 Propuestas de segmentación de asociados

El portafolio financiero de la cooperativa donde se realizó este proyecto se divide en dos categorías: productos de ahorro y productos de crédito, los cuáles desde el punto de vista comercial y de riesgos financieros requieren tratamientos diferentes. Es por esto que en este proyecto se realizaron tres propuestas de segmentación, una primera para diferenciar los asociados ahorradores entre sí, una segunda donde solo se consideró los asociados con créditos activos y, finalmente una segmentación donde se analizaron los asociados que tienen productos activos de ahorro y de crédito. En este proyecto se desarrollarán tres propuestas, así:

Segmentación de ahorradores: En este conjunto de datos solo se tomaron los asociados con saldo en productos de ahorro y aportes voluntarios mayor a cero. Las variables a usar para esta segmentación son: 1) saldo en productos de ahorro y aportes voluntarios; 2) cantidad de productos de ahorro y aportes voluntarios adquiridos; 3) recencia en productos de ahorro y aportes voluntarios.

Segmentación de deudores: En este conjunto de datos solo se tomaron los registros de asociados con créditos activos, es decir con saldo actual en productos de crédito mayor a cero. Las variables a usar para la segmentación son: 1) saldo en productos de crédito; 2) cantidad de productos de crédito adquiridos; 3) recencia en productos de crédito.

Segmentación de asociados con portafolio mixto: Las variables a usar para esta segmentación son: 1) saldo en productos de crédito; 2) cantidad de productos de crédito adquiridos; 3) recencia en productos de crédito; 4) saldo en productos de ahorro y aportes voluntarios; 5) cantidad de productos de ahorro y aportes voluntarios adquirido; 6) recencia en productos de ahorro y aportes voluntarios; 7) monto de Interés pagado en créditos. En este conjunto de datos se tomaron los registros de asociados con saldo en productos de ahorro y aportes voluntarios mayor a cero, saldos en productos de créditos mayor a cero e intereses pagados en créditos mayor a cero.

4.2 Construcción de los modelos

El proceso de modelamiento consiste de las siguientes actividades, las cuales se ejecutarán para las tres propuestas de segmentación:

Test de clustering: La primera evaluación realizada al conjunto de datos es la de tendencia de agrupamiento, la cual consiste en examinar si los datos son agrupables, es decir, si los datos contienen cualquier estructura de agrupación inherente. Existen varias técnicas para esto, en este proyecto se usó el estadístico de Hopkins.

Selección de la técnica de agrupamiento: Teniendo en cuenta que las variables a usar para la segmentación en los tres casos propuestos son del tipo numéricas, además que la cantidad de registros a procesar y de acuerdo a la revisión bibliográfica, en este proyecto se usó una técnica de agrupamiento Jerárquico, tres técnicas de agrupamiento basado en particiones: K-means, PAM (Partition Around Medoids), Fuzzy C-Means, y una técnica basada en modelos de densidad: Dbscan.

Selección del número óptimo de grupos (K): La mayoría de los métodos a usar de agrupamiento requieren que se especifique el número de grupos deseados, excepto con Dbscan. La siguiente actividad realizada fue obtener el valor óptimo del número de grupos para aquellos métodos que lo requieren. Para esto se usaron las siguientes técnicas: el método del codo, y el método del coeficiente de Silhouethe.

Aplicación de la técnica de agrupamiento: Una vez se tuvo el conjunto de datos con el procesado requerido, definido las técnicas a aplicar, y el número óptimo de grupos a obtener, se realizaron los procesos de clustering usando R software y Jupyter Notebook.

Evaluación de los modelos: Para definir cuál fue el mejor método se realizó un análisis comparativo de los índices de cada agrupamiento obtenido. El rendimiento de los modelos fue evaluado usando las siguientes métricas: Dunn, Davies Bouldin, Silhouette, Calinski Harabasz, PBM, Ray Turi Y Wemmert Gancarski. Adicionalmente, para la evaluación de la similitud en las agrupaciones realizadas por los diferentes métodos, para cada una de las propuestas de segmentación se utilizó el Adjusted Rand Index, el cual presenta un valor de 1 cuando los agrupamientos son iguales.

Identificación de variables que más influenciaron en la segmentación: Luego de seleccionado el modelo y el número de clúster definitivos para cada propuesta de segmentación, se procedió a identificar las variables que más influyeron en el agrupamiento. Para lograr esto, se usó un árbol de clasificación, donde se especificó como variable dependiente el clúster de pertenencia (una variable de tipo categórica) y como variables independientes las variables transaccionales de RFM.

Descripción de los grupos obtenidos y de los asociados típicos: Se realizó una descripción a cada uno de los grupos obtenidos y del asociado “típico” o “representante” de cada agrupamiento usando las variables sociodemográficas y transaccionales disponibles. Para la descripción del asociado típico para los agrupamientos por K-means se usaron los centroides hallados por el algoritmo, para las variables continuas estos coinciden con la media de las variables transaccionales, y posteriormente se obtuvieron las modas para las variables categóricas. Igualmente, para la descripción del asociado típico a partir de los resultados obtenido por el algoritmo de PAM se usaron los medoides, los cuales son un asociado existente en cada uno de los agrupamientos. Para la descripción del asociado típico por el algoritmo de Dbscan, dado que no se obtiene directamente por el modelo, se construyó a partir de los grupos obtenidos y etiquetados, para las variables continuas se tomó la media, y la moda para las variables categóricas.

4.3 Segmentación de asociados con ahorros activos.

4.3.1 Test de agrupamiento

Estadístico de Hopkins: Se obtuvo un valor de 0,012 el cual es menor a 0,5 lo que indica que no hay una distribución uniforme de los datos y si es posible realizar agrupamiento.

4.3.2 Selección del número óptimo de grupos

En la Tabla 36, se presentan los resultados de las métricas obtenidas del número óptimo de grupos recomendados para la segmentación de asociados con ahorros activos. Con estas métricas, complementadas con las Figuras 3 y 4 para determinar de manera gráfica el número óptimo de agrupamientos de asociados con ahorros activos, se eligió como número óptimo de agrupamiento $k=2$.

Tabla 36: Métricas de número óptimo de clusters para asociados ahorradores

Método	K-means	PAM	Fuzzy C-means	Dbscan
Métrica / Número óptimo de Clúster	2	2	2	6
Dunn	0,54	0,54	0,54	0,13
Silhouette	0,74	0,74	0,74	0,32

Figura 3 Método gráfico del codo para la selección del número óptimo de clúster para asociados ahorradores.

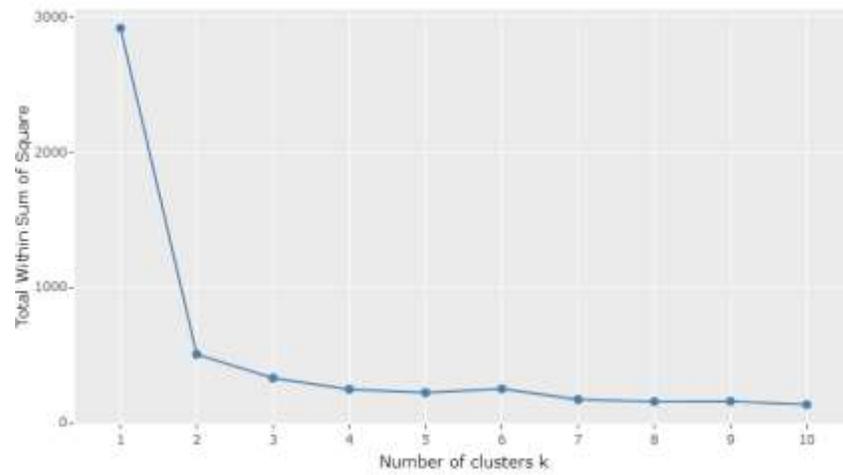
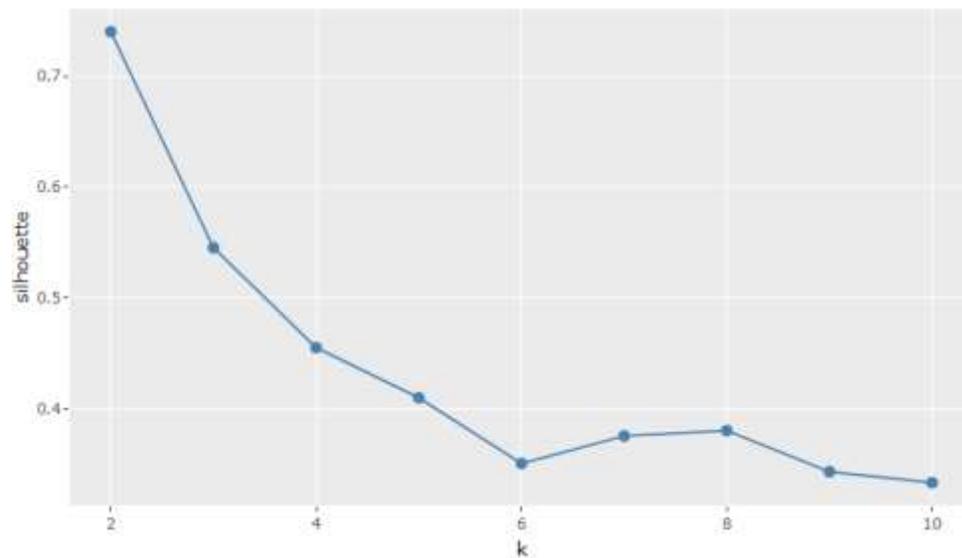


Figura 4 Método gráfico de Silhouette para la selección del número óptimo de clúster para asociados ahorradores



4.3.3 Resultados del agrupamiento

Resultados del clúster de asociados ahorradores por K-means: Por este modelo se recomienda dos grupos. En las Tablas 37, 38 y 39 se presentan los resultados obtenidos por K-means.

Tabla 37: Resultados de los Clusters de asociados ahorradores por K-means

Clúster	número de asociados	Porcentaje de asociados	Saldo total en ahorros \$	Porcentaje en Saldo de ahorros	Saldo en ahorros (media)	Recencia de ahorros (media)	Cantidad de ahorros (media)	Cantidad de ahorros (suma)	Porcentaje de cuentas de ahorros
1	5.611	39,9 %	\$ 26.547.100.774	67,2 %	\$ 4.731.260	0	4	23.242	50,9 %
2	8.451	60,1 %	\$ 12.958.349.788	32,8 %	\$ 1.533.351	29	3	22.461	49,1 %
Total	14.062	100%	\$ 39.505.450.563	100 %				45.703	100%

Tabla 38: Resultados complementarios de los Clusters de asociados ahorradores por K-Means

Clúster	Saldos en ahorros (mínimo)	Saldos en ahorros (máximo)	Saldos en ahorros (desviación estándar)	Recencia de ahorros (mínimo)	Recencia de ahorros (máximo)	Recencia de ahorros (moda)	Recencia de ahorros (desviación estándar)	Cantidad de ahorros (mínimo)	Cantidad de ahorros (máximo)	Cantidad de ahorros (moda)	Cantidad de ahorros (desviación estándar)
1	\$ 26.767	\$ 919.866.567	\$ 16.908.565	0	0	0	0	1	63	2	5
2	\$ 2.396	\$ 292.059.951	\$ 5.744.673	1	125	1	35	1	99	2	3

Tabla 39: Resultados complementarios de los Clusters de asociados ahorradores por K-Means

Clúster	Saldos en ahorros (Mediana)	Recencia de ahorros (Mediana)	Cantidad de ahorros (Mediana)
1	\$ 2.061.081	0	3
2	\$ 397.217	12	2

Resultados de los Clusters de asociados ahorradores por Dbscan: Por este modelo se recomienda seis grupos, pero la cantidad de registros por grupo y sus métricas presentan diferencias a las obtenidas por los métodos presentados anteriormente. En las Tablas 40, 41 y 42 se presentan los resultados obtenidos por el método de Dbscan.

Tabla 40: Resultados de los Clusters de asociados ahorradores por Dbscan

Clúster	número de asociados	Porcentaje de asociados	Saldo total Ahorros \$s	Porcentaje Saldos ahorros	Saldos ahorros \$ (media)	Recencia ahorros (media)	Cantidad de ahorros (media)	Cantidad de ahorros (suma)	Porcentaje de cuentas de ahorros
0	173	1,2 %	\$ 7.037.760.429	17,2 %	\$ 406.806.962	9	18	3.036	6,6 %
1	4.695	33,4 %	\$ 3.833.627.165	9,7 %	\$ 816.534	29	2	9.390	20,5 %
2	2.247	16,0 %	\$ 5.345.569.114	13,5 %	\$ 2.378.980	0	2	4.494	9,8 %
3	2.483	17,7 %	\$ 5.854.187.219	14,8 %	\$ 2.357.707	14	4	10.325	22,6 %
4	1.153	8,2 %	\$ 486.452.463	1,2 %	\$ 421.902	65	1	1.153	2,5 %
5	3.004	21,4 %	\$ 16.128.510.351	40,8 %	\$ 5.369.011	0	6	16.998	37,2 %
6	307	2,2 %	\$ 819.343.822	2,1 %	\$ 2.668.872	0	1	307	0,7 %
Total	14.062	100 %	\$ 39.505.450.563	100%				45.703	100 %

Tabla 41: Resultados complementarios de los Clusters de asociados ahorradores por Dbscan

Clúster	Saldo ahorros (mínimo)	Saldo Ahorros (máximo)	Saldo Ahorros (desviación estándar)	Recencia Ahorros (Mínimo)	Recencia Ahorros Máxima	Recencia ahorros (moda)	Recencia Ahorros (desviación estándar)	Cantidad de Ahorros (mínimo)	Cantidad de Ahorros (máximo)	Cantidad de ahorros (moda)	Cantidad de Ahorros (desviación estándar)
0	\$ 2.396	\$ 919.866.567	\$ 87.601.588	0	80	0	18	1	99	1	17
1	\$ 13.989	\$ 21.956.949	\$ 1.455.009	1	125	1	31	2	2	2	0

2	\$ 26.767	\$ 50.245.700	\$ 3.180.494	0	0	0	0	2	2	2	0
3	\$ 24.000	\$ 35.303.520	\$ 3.649.243	1	124	1	21	3	17	3	2
4	\$ 24.000	\$ 7.249.252	\$ 898.078	1	125	124	47	1	1	1	0
5	\$ 85.100	\$ 110.608.700	\$ 8.801.478	0	0	0	0	3	34	3	4
6	\$ 70.200	\$ 33.088.399	\$ 3.654.803	0	0	0	0	1	1	1	0

Tabla 42: Resultados complementarios de los clusters de asociados ahorradores por Dbscan

Clúster	Saldo ahorros (mediana)	Recencia ahorros (mediana)	Cantidad de ahorros (mediana)
0	\$ 13.248.832	1	12
1	\$ 278.092	18	2
2	\$ 1.469.641	0	2
3	\$ 1.272.794	5	3
4	\$ 64.051	65	1
5	\$ 2.664.722	0	4
6	\$ 1.446.254	0	1

4.3.4 Evaluación de los modelos

Para las técnicas de segmentación se obtuvieron las métricas presentadas en la Tabla 43. De acuerdo a los resultados, los métodos de K-means, PAM, Fuzzy C-means coincidieron en el número de grupos recomendados, el número de registros por grupo y por tanto en las métricas de agrupamiento.

Tabla 43: Métricas de evaluación de los métodos aplicados para la segmentación de asociados con ahorros activos.

Método	K-means	PAM	Fuzzy C-means	Dbscan
Métrica / número de clusters	2	2	2	6
Dunn	0,54	0,54	0,54	0,13
Davies Bouldin	0,38	0,38	0,38	1,22
Silhouette	0,74	0,74	0,74	0,32
Calinski Harabasz	67103,02	67103,02	67103,02	25902,60
Pbm	1,25	1,25	1,25	0,36
Ray turi	0,05	0,05	0,05	0,87
Wemmert gancarski	0,81	0,81	0,81	0,34

Adjusted Rand Index: En la Tabla 44 se presentan los resultados del *Adjusted Rand Index* obtenidos para los diferentes modelos usados para la segmentación. Considerando especialmente los índices de Silhouette, Dunn y Calinski- Harabasz presentados en las tablas 38 y 39, se puede concluir que el modelo óptimo de agrupamiento es el obtenido por K-means o por PAM o por Fuzzy C-means. Cualquiera de estos es igualmente válido.

Analizando las métricas de la Tabla 43, la Tabla 44 y por temas de interpretabilidad se decidió seleccionar el modelo de agrupamiento obtenido por K-means de 2 grupos.

En este trabajo sólo se presentará el análisis de las variables que influyeron en la segmentación, y la interpretación de los agrupamientos de acuerdo a los resultados de K-means. Adicionalmente La descripción de los asociados típicos se realizará teniendo en cuenta los resultados obtenidos por K-Means y PAM.

Tabla 44: Adjusted Rand Index de los métodos aplicados para la segmentación de asociados con ahorros activos.

	K-means	PAM	Fuzzy C-Means	DBSCAN
K-means	1	1	1	0,415
PAM	1	1	1	0,415
Fuzzy C-Means	1	1	1	0,415
HClust	1	1	1	0,415
DBSCAN	0,415	0,415	0,415	1,000

4.3.5 Descripción de los agrupamientos y asociados típicos obtenidos

Descripción de los grupos obtenidos por k-means

- **Clúster 1: Asociados de alta cultura en ahorros.** Representa a los asociados con mayor antigüedad en la cooperativa, mayor cantidad de productos, saldos más altos y baja recencia. Está constituido por el 39,9 % de los asociados, los cuales aportan el 67,2 % de los saldos en ahorros. Todos los asociados de este grupo han realizado una operación de ahorros en el último mes.
- **Clúster 2: Asociados de baja cultura en ahorros.** Representa a los asociados con menor antigüedad, menor cantidad de productos y menores saldos en ahorros, mayor recencia. Está constituido por el 60,1 % de los asociados los cuales aportan el 33,8 % de los saldos disponibles en ahorros. Todos los asociados de este grupo no han realizado una operación de ahorros en el último mes.

A continuación, se presenta la descripción de los asociados típicos representantes, los cuales se hicieron tomando los centroides obtenidos por K-means y los medoides obtenidos por PAM.

Descripción de asociados típicos obtenidos por *K-means*

- **Asociado típico Clúster 1:** Mujer de 41 años, con nivel educativo secundario, estado civil soltero, vive en estrato 1, su tipo de vivienda es familiar, su contrato es a término indefinido, tiene un ingreso mensual menor de 2 salarios mínimos. Es un asociado con cuatro productos de ahorros, un saldo ahorrado de \$ 4.731.260 y ha realizado abonos en el último mes.
- **Asociado típico Clúster 2:** Mujer de 35 años, con nivel educativo secundario, estado civil soltero, vive en estrato 1, su tipo de vivienda es familiar, su contrato es a término indefinido, tiene un ingreso mensual menor de 2 salarios mínimos. Es un asociado con antigüedad promedio en la cooperativa de 54 meses, tiene tres productos de ahorros, un saldo ahorrado promedio de \$ 1.533.352 y hace 29 meses que no realiza un ahorro.

Descripción de asociados típicos a partir del agrupamiento PAM

- **Asociado típico Clúster 1:** Hombre de 28 años, con nivel educativo secundaria, estado civil unión libre, vive en estrato 1, su tipo de vivienda es familiar, su ocupación es independiente, y su rango de ingresos mensuales es menor de 2 salarios mínimos. Es un asociado con antigüedad en la Cooperativa de 32 meses, tiene tres productos de ahorros, un saldo ahorrado de \$1.942.623 y ha realizado abonos en ahorros en el último mes.
- **Asociado típico Clúster 2:** Mujer de 30 años, con nivel educativo secundaria, estado civil Unión libre, vive en estrato 1, su tipo de vivienda es familiar, su ocupación es independiente, y su rango de ingresos mensuales esta entre 2 y 4 salarios mínimos legales vigente. Es un asociado con una antigüedad de 30 meses en la entidad, tiene dos productos de ahorros, un saldo ahorrado de \$ 337.104, hace trece meses que no realiza un abono en alguna de sus cuentas de ahorros.

Variables que influyeron en la segmentación

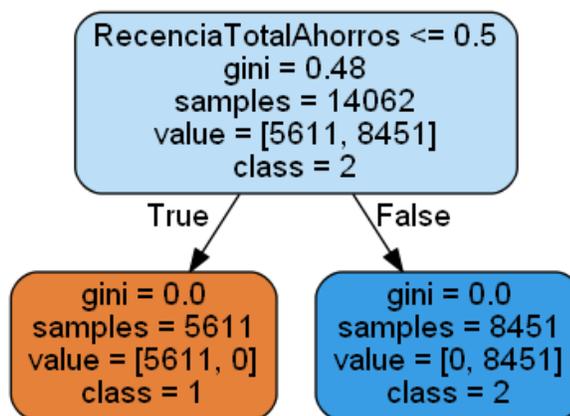
Con el fin de identificar las reglas que más influyeron y las reglas que definieron el modelo de segmentación de asociados, se usó un modelo de árboles de decisión con los datos de asociado etiquetados con la pertenencia a un clúster.

Segmentación de asociados ahorradores: Para este modelo de segmentación por K-means se obtiene el árbol de decisión presentado en la Figura 5.

De la Figura 5 se puede identificar que las reglas de clasificación para este modelo de clasificación fueron:

- **Regla 1:** Si $\text{RecenciaAhorros} \leq 0,5$ Entonces Asociado pertenece a Grupo 1.
- **Reglas 2:** Si $\text{RecenciaAhorros} > 0,5$ Entonces Asociado pertenece a Grupo 2.

Figura 5 Árbol de decisión de la segmentación de asociados ahorradores



Se observa que 5.611 asociados son clasificados con la regla 1, y 8.451 con la regla 2.

Del análisis de las reglas de decisión, se puede concluir que la variable que influyó en la creación de los agrupamientos para asociados con ahorros activos fue la recencia en ahorros.

4.4 Segmentación de asociados con créditos activos.

4.4.1 Test de agrupamiento

Estadístico de Hopkins: Se obtuvo un valor de 0,01 el cual es menor a 0,5, lo que indica que no hay una distribución uniforme de los datos y que si es posible realizar agrupamiento.

4.4.2 Selección del número óptimo de grupos

En la Tabla 45 se presentan los resultados de las métricas obtenidas del número óptimo de grupos para la segmentación de asociados con créditos activos:

Tabla 45: Métricas de número óptimo de clusters para asociados con créditos activos

	K-means	PAM	Fuzzy C-means	Dbscan
Métrica / Número de Clúster	2	2	2	2
Dunn	0,52	0,52	0,52	0,67
Silhouette	0,75	0,75	0,75	0,75

La Tabla 45 se complementó con las Figuras 6 y 7 de indicadores para determinar el número óptimo de agrupamientos de asociados con ahorros activos. Se eligió como número óptimo de agrupamientos $k=2$.

Figura 6 Método gráfico del codo para la selección del número óptimo de clúster para asociados con créditos activos.

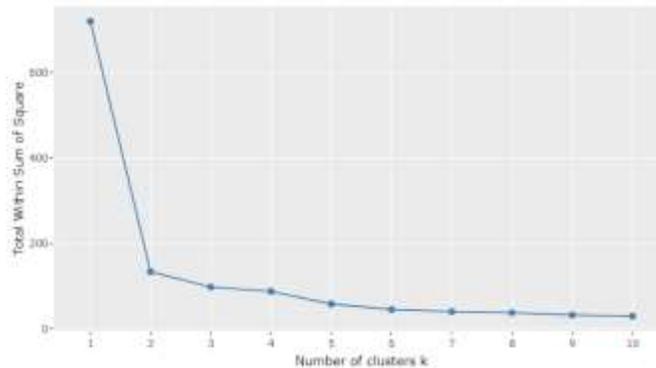
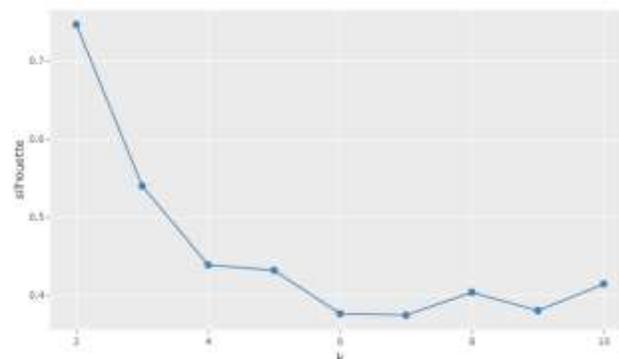


Figura 7 Método gráfico de Silhouette para la selección del número óptimo de clúster para asociados con créditos activos.



4.4.3 Resultados del agrupamiento

Se presentan a continuación los resultados obtenidos por el método de K-Means.

Resultado del Clúster de asociados con créditos activos por K-means: Por este modelo se recomienda dos grupos. En las Tablas 46, 47, 48 y 49 se presentan los resultados obtenidos por K-means.

Tabla 46: Resultado del Clúster de asociados con créditos activos por K-means

Clúster	número de asociados	Porcentaje de asociados	Saldo total créditos \$	Porcentaje Saldo Créditos	Saldo Créditos \$ (media)	Recencia Créditos (Media)	Cantidad de créditos (Media)	Cantidad de créditos (suma)	Porcentaje de Cantidad créditos
1	2.839	61,1 %	\$ 23.091.309.180	57,5 %	\$ 8.133.607	1	4	11.846	49,2 %
2	1.806	38,9 %	\$ 17.036.250.844	42,5 %	\$ 9.433.140	0	7	12.240	50,8 %
Total	4.645	100%	\$ 40.127.560.024	100 %				24.086	100%

Tabla 47: Resultados complementarios del Clúster de asociados con créditos activos por K-means

Clúster	Saldo créditos \$ (Mínimo)	Saldo créditos \$ (máximo)	Saldo créditos \$ (Desviación estándar.)	Recencia créditos (meses) (Mínimo)	Recencia créditos (meses) (Máximo)	Recencia (meses) (Moda)	Recencia créditos (meses) (Desviación estándar.)	Cantidad de créditos (mínimo)	Cantidad de créditos (máximo)	Cantidad de créditos (Desviación estándar.)	Cantidad de créditos (moda)
1	\$ 198	\$ 108.348.867	\$ 9.743.438	1	37	1	2	1	380	16	1
2	\$ 4	\$ 126.147.319	\$ 11.576.542	0	0	0	0	1	370	26	1

Tabla 48: Resultados complementarios del Clúster de asociados con créditos activos por K-means

Clúster	Saldo créditos \$ (Mediana)	Recencia Créditos (meses) (Mediana)	Cantidad de créditos (Mediana)
1	\$ 5.896.363	1	2
2	\$ 5.124.303	0	2

Tabla 49: Resultados complementarios del Clúster de asociados con créditos activos por K-means

Clúster	Interés pagado (\$ (suma))	Porcentaje interés pagado	Interés pagado \$ (Mínimo)	Interés pagado \$ (Máximo)	Interés pagado \$ (Desviación estándar)	Interés pagado \$ (media)	Interés pagado \$ (mediana)
1	\$ 6.070.104.475	56,6 %	\$ 0	\$111.571.528	\$ 4.995.00	\$ 2.138.114	\$ 451.193
2	\$ 4.649.070.805	43,4 %	\$0	\$ 88.959.639	\$ 6.474.895	\$ 2.574.236	\$ 416.633

4.4.4 Evaluación de los modelos

Para este modelo de segmentación se obtuvieron las métricas presentadas en la Tabla 50 para cada uno de los métodos aplicados. De acuerdo a éstas, especialmente las de Silhouette, Dunn y Calinski Harabasz se encuentra que los modelos de K-means, PAM, y Fuzzy c-means obtuvieron los mismos grupos de clúster con los mismos registros y por tanto los mismos valores en los indicadores, se puede afirmar que estos modelos son igualmente válidos para esta segmentación.

Tabla 50: Métricas de evaluación de los métodos aplicados para la segmentación de asociados con créditos activos.

	K-means	Pam	Fuzzy c-means	Dbscan	Hclust
Métrica / número de clúster	2	2	2	2	2
Dunn	0,52	0,52	0,52	0,67	0,27
Davies bouldin	0,38	0,38	0,38	0,37	0,38
Silhouette	0,75	0,75	0,75	0,75	0,75
Calinski harabasz	20305,29	20305,29	20305,29	22139,49	20210,93
Pbm	1,00	1,00	1,00	1,04	1,00
Ray turi	0,05	0,05	0,05	0,05	0,05
Wemmert gancarski	0,82	0,82	0,82	0,82	0,82

Adjusted Rand Index: En la Tabla 51 se presentan los resultados del Adjusted Rand Index obtenidos para los diferentes modelos usados para la segmentación de asociados con créditos activos.

Tabla 51: Adjusted Rand Index de los métodos aplicados para la segmentación de asociados con créditos activos.

	K-means	Pam	Fuzzy c-means	Hclust	Dbscan
K-means	1	1	1	0,999	0,987
Pam	1	1	1	0,999	0,987
Fuzzy c-means	1	1	1	0,999	0,987
Hclust	0,999	0,999	0,999	1,000	0,987
Dbscan	0,987	0,987	0,987	0,987	1,000

Analizando las métricas presentados en las Tablas 50 y 51 se seleccionar el modelo de agrupamiento

por K-means de 2 grupos. En este informe sólo se presentará el análisis de las variables que influyeron en la segmentación, y la interpretación de los agrupamientos de acuerdo a los resultados de K-means. La descripción de los asociados típicos o representantes se realizará teniendo en cuenta los resultados obtenidos por K-Means y PAM.

4.4.5 Variables que más influyeron en la segmentación

Con el fin de identificar las reglas que más influyeron y las reglas que definieron el modelo de segmentación de asociados con créditos activos, se usó un modelo de árboles de decisión con los datos de los asociados etiquetados con la pertenencia a un clúster. El resultado del mismo para cada uno de los modelos desarrollados se presenta a continuación.

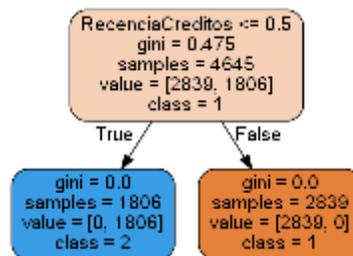
Segmentación de asociados deudores: Para este modelo de segmentación por K-means se obtuvo el árbol de decisión presentado en la Figura 8.

De la Figura 8 se puede identificar que las reglas de clasificación para este modelo de clasificación fueron:

- **Regla 1:** Si $\text{RecenciaCreditos} \leq 0,5$ Entonces Asociado pertenece a Grupo 2
- **Regla 2:** Si $\text{RecenciaCreditos} > 0,5$ Entonces Asociado pertenece a Grupo 1

Se observa que 1.806 asociados son clasificados con la regla 1, y 2.839 con la regla 2.

Figura 8 Árbol de decisión de la segmentación de asociados con créditos activos



La única variable que influyó en la creación de los agrupamientos para asociados con créditos activos fue la recencia en créditos.

4.4.6 Descripción de los agrupamientos y asociado típicos obtenidos

Caracterización transaccional y socio-demográfica de los asociados con créditos activos típicos o representantes de cada grupo obtenido por K-means.

Tabla 52: Caracterización transaccional de los asociados con créditos activos representantes de cada grupo obtenido por K-means

Clúster	Recencia créditos (meses) (media)	Cantidad créditos (media)	Saldo créditos \$ (media)	Interés pagado \$ (media)
1	1	4	\$ 8.133.607	\$ 2.138.114
2	0	7	\$ 9.433.140	\$ 2.574.236

Tabla 53: Caracterización socio-demográfica de los asociados con créditos activos representantes de cada grupo obtenido por K-means.

Clúster	Genero (moda)	Edad (años) (media)	Edad (años) (moda)	Nivel educativo (moda)	Estado Civil (moda)	Estrato (moda)	Tipo de Vivienda (moda)	Tipo de Contrato (moda)	Rango de Ingresos (SM) (moda)	Antigüedad en la cooperativa (meses) (media)	Rango antigüedad (meses) (moda)
1	Femenino	44	38	Secundaria	Soltero(a)	2	Familiar	Término Indefinido	1-2	56	[18, 34]
2	Femenino	44	31	Secundaria	Soltero(a)	2	Familiar	Término Indefinido	1-2	57	[34, 52]

Caracterización transaccional y socio-demográfica de los asociados representantes de cada grupo obtenido por PAM.

Tabla 54: Caracterización transaccional de los asociados con créditos activos representantes de cada grupo obtenido por PAM.

Clúster	Id	Recencia créditos (meses)	Cantidad créditos	Saldo créditos \$	Interés pagado \$
1	35188	1	2	\$ 4.933.215	\$ 3.501.624
2	34883	0	2	\$ 5.552.575	\$ 2.449.286

Descripción de los grupos obtenidos por k-means

- **Clúster 1: Asociados con riesgo de crédito Medio.** El 61,1 % de los asociados aportan el 57,5 % de la cartera de saldos en créditos. Representa a los asociados con menor promedio en saldos créditos, menor cantidad de historial de productos de crédito, asociados con mayor recencia de créditos. Analizando la recencia de créditos, estos asociados se pueden calificar como en un riesgo de crédito medio, ya que, a pesar de tener créditos activos, en promedio en el último mes no han realizado operaciones de crédito, se recomienda hacerles un plan de seguimiento prioritario con el fin de identificar asociados que pueden tener dificultades en el pago de sus obligaciones y diseñar estrategias de cobro preventivo o de reliquidación de créditos.
- **Clúster 2: Asociados con riesgo de crédito Bajo.** El 38,9 % de los Asociados aportan el 42,5% de la cartera de saldos en créditos. Representa a los asociados con mayor promedio en saldos créditos, mayor cantidad de historial de productos de crédito, asociados con menor recencia de créditos. Todos los asociados en este grupo han realizado una operación de crédito en el último mes. Analizando la recencia de créditos, estos asociados se pueden calificar como en un riesgo de crédito bajo.

Descripción de asociados típicos obtenidos por K-means.

- **Asociado típico Clúster 1:** Mujer de 44 años en promedio, con nivel educativo secundaria, estado civil soltero, vive en estrato 2, su tipo de vivienda es familiar tiene contrato a término indefinido y con un ingreso mensual menor a 2 salarios mínimos y una antigüedad en la entidad entre 18 y 34 meses. Asociado con historial de cuatro productos de créditos, un saldo pendiente

de \$ 8.133.607, hace un mes que no realiza un abono o una operación de crédito.

- **Asociado típico Clúster 2:** Mujer de 44 años en promedio, con nivel educativo Secundaria, estado civil Soltero, vive en estrato 2, su tipo de vivienda es Familiar tiene contrato a Término indefinido, tiene un ingreso mensual menor a 2 salarios mínimos y una antigüedad en la entidad entre 34 y 52 meses. Asociado con historial de siete productos de crédito, un saldo pendiente de \$ 9.433.140, ha realizado un abono o una operación de crédito en el último mes.

Descripción de asociados típicos obtenidos por PAM

- **Asociado típico Clúster 1:** Mujer de 52 años en promedio, con nivel educativo Secundaria, estado civil Unión libre, vive en estrato 2, su tipo de vivienda es Familiar tiene contrato a Término indefinido, tiene un ingreso mensual de 1 salario mínimo y una antigüedad en la entidad menor a 18 meses. Asociado con historial de dos productos de créditos, un saldo pendiente de \$ 4.933.215, ha pagado \$ 3.501.624 en interés, hace un mes que no realiza un abono o una operación de crédito.
- **Asociado típico Clúster 2:** Mujer de 56 años en promedio, con nivel educativo Técnico, estado civil Unión libre, vive en estrato 1, su tipo de vivienda es Familiar tiene contrato a Término indefinido, tiene un ingreso mensual de 1 salario mínimo y una antigüedad en la entidad entre 58 y 81 meses. Asociado con historial de dos productos de créditos, un saldo pendiente de \$ 5.552.575, ha pagado \$ 2.449.286 en interés, ha realizado un abono o una operación de crédito en el último mes.

4.5 Segmentación de asociados con ahorros y créditos activos.

4.5.1 Test de agrupamiento

Estadístico de Hopkins: Se obtuvo un valor de 0,062, el cual es menor a 0,5, lo que indica que no hay una distribución uniforme de los datos y que si es posible realizar agrupamiento.

4.5.2 Selección del número óptimo de grupos

En la Tabla 55 se presenta los resultados de las métricas obtenidas del número óptimo de grupos para la segmentación de asociados con ahorros y créditos activos:

Tabla 55: Métricas de número óptimo de clusters para asociados con ahorros y créditos activos

	K-means	PAM	Fuzzy C-means	Dbscan
Métrica / Número de Clusters	4	4	4	4
Dunn	0,420	0,420	0,420	0,581
Davies Bouldin	0,711	0,711	0,711	0,667
silhouette	0,562	0,562	0,562	0,584
Calinski Harabasz	5.148,50	5.148,50	5.148,50	5.921,96

La Tabla 55 se puede complementar con las siguientes gráficas de indicadores para determinar el número óptimo de agrupamientos de asociados con ahorros y créditos activos.

Figura 9 Método gráfico del codo para la selección del número óptimo de clúster para asociados con ahorros y créditos activos.

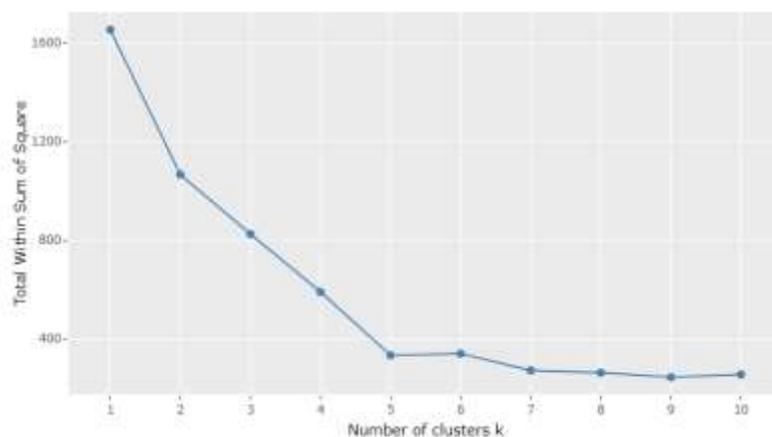
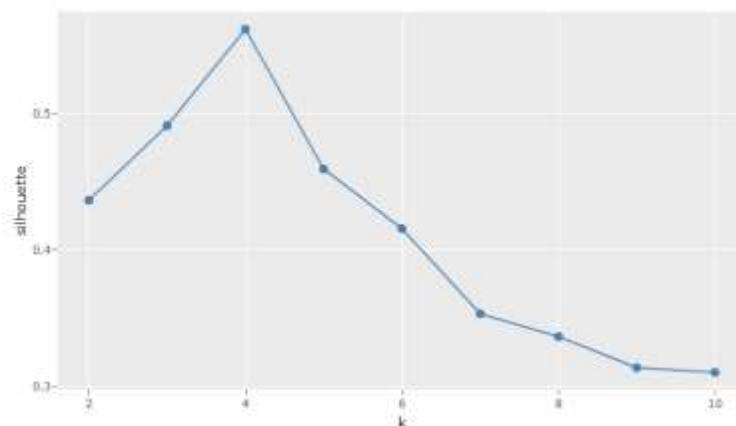


Figura 10 Método gráfico de Silhouette para la selección del número óptimo de clúster para asociados con ahorros y créditos activos.



Del análisis de las métricas de la Tabla 55 y las Figuras 9 y 10 presentadas se tomó como número óptimo de agrupamiento $k=4$.

4.5.3 Resultados del agrupamiento

Segmentación por K-means y PAM: Los resultados del agrupamiento al utilizar los métodos de K-means, PAM, Fuzzy C-means, fueron equivalente tanto en número de grupos como en el número de registros por grupo, por tanto, en las métricas de agrupamiento. Por lo anterior solo se presentarán los resultados por K-means en las Tablas: 56, 57, 58, 59, 60 y 61.

Datos de Comportamiento transaccional en Ahorros

Tabla 56: Datos del comportamiento transaccional en ahorros de los asociados con ahorros y créditos activos obtenidos por K-means.

Clúster	número de asociados	Porcentaje de asociados	Saldo total Ahorros	Porcentaje Saldo Ahorros	Saldo Ahorros (media)	Recencia Ahorros (media)	Cantidad de Ahorros (suma)	Porcentaje de ahorros	Cantidad de Ahorros (media)
1	1.294	28,3 %	\$ 2.656.956.617	19,5 %	\$ 2.053.290	5	3.981	24,2	3

2	1.209	26,4 %	\$ 4.283.294.719	31,5 %	\$ 3.542.841	0	4.784	29,1	4
3	1.544	33,7 %	\$ 5.585.109.405	41,0 %	\$ 3.617.299	0	6.145	37,4	4
4	531	11,6 %	\$ 1.081.276.570	7,9 %	\$ 2.036.302	6	1.513	9,2	3

Tabla 57: Datos complementarios del comportamiento transaccional en ahorros de los asociados con ahorros y créditos activos obtenidos por K-means.

Clúster	Saldo Ahorros \$ (Mínimo)	Saldo Ahorros \$ (máximo)	Saldo Ahorros \$ (Desviación estándar)	Recencia Ahorros (meses) (Mínimo)	Recencia Ahorros (meses) (Máximo)	Recencia Ahorros (meses) (Desviación estándar)	Recencia Ahorros (meses) (moda)	Cantidad de Ahorros (mínimo)	Cantidad de Ahorros (máximo)	Cantidad de Ahorros (Desviación estándar)	Cantidad de Ahorros (moda)
1	\$ 3.167	\$ 61.390.193	\$ 2.830.399	1	44	5	1	1	38	3	2
2	\$ 173.230	\$ 67.658.909	\$ 4.884.751	0	0	0	0	1	44	4	2
3	\$ 170.341	\$ 253.603.300	\$ 8.094.074	0	0	0	0	1	46	4	2
4	\$ 120.491	\$ 25.034.933	\$ 2.392.277	1	52	6	4	1	23	2	2

Tabla 58: Datos de Comportamiento transaccional en Créditos

Clúster	número de asociados	Porcentaje de asociados	Saldo total de créditos \$	Porcentaje Saldo Créditos	Saldo Créditos \$ (media)	Recencia Créditos (meses) (media)	Cantidad de créditos (suma)	Porcentaje de créditos	Cantidad de créditos (media)
1	1294	28,3 %	\$ 11.876.804.242	29.2 %	\$ 9.178.365	2	4172	17.4 %	3
2	1209	26,4 %	\$ 10.474.879.479	26.4 %	\$ 8.664.086	0	9934	41.4 %	8
3	1544	33,7 %	\$ 11.214.417.063	28.3 %	\$ 7.263.223	1	7673	31.9 %	5

4	531	11,6 %	\$	15,4 %	\$	0	2238	9,3 %	4
			6.116.204.544		11.518.276				

Tabla 59: Datos complementarios de Comportamiento transaccional en Créditos

Clúster	Saldo créditos \$ (Mínimo)	Saldo créditos \$ (máximo)	Saldo Créditos \$ (Desviación estándar)	Recencia créditos (meses) (Mínimo)	Recencia créditos (meses) Máxima	Recencia créditos (meses) (Desviación estándar)	Recencia créditos (meses) (moda)	Cantidad de créditos (mínimo)	Cantidad de créditos (máximo)	Cantidad de créditos (Desviación estándar)	Cantidad de créditos (moda)
1	\$ 3.097	\$ 108.348.867	\$ 10.810.971	1	37	3	1	1	242	10	1
2	\$ 4.202	\$ 126.147.319	\$ 10.699.591	0	0	0	0	1	37	30	1
3	\$ 198	\$ 84.233.319	\$ 8.657.651	1	8	0	1	1	380	19	1
4	\$ 103.445	\$ 118.015.933	\$ 13.502.629	0	0	0	0	1	286	16	1

Tabla 60: Datos de pagos de Interés por créditos

Clúster	Interés Pagado Por Créditos \$ (suma)	Interés Pagado Por Créditos (Porcentaje)	Interés Pagado Por Créditos \$ (Media)	Interés Pagado Por Créditos \$ (Mínimo)	Interés Pagado Por Créditos \$ (Máximo)	Interés Pagado Por Créditos \$ (Desviación estándar)
1	\$ 2.446.252.030	22,8 %	\$ 1.890.458	\$ 126	\$ 74.418.980	\$ 4.741.014
2	\$ 3.569.814.778	33,3 %	\$ 2.952.700	\$ 24	\$ 88.959.639	\$ 7.180.387
3	\$ 3.623.852.445	33,8 %	\$ 2.347.055	\$ 21	\$ 111.571.528	\$ 5.191.966
4	\$ 1.079.256.027	10,1 %	\$ 2.032.497	\$ 546	\$ 64.002.284	\$ 4.880.635

Tabla 61: Datos complementarios

Clúster	Número de Asociados	Saldo ahorros \$ (mediana)	Recencia Ahorros (meses) (mediana)	Cantidad de ahorros (mediana)	Saldo créditos \$ (mediana)	Recencia Créditos (meses) (mediana)	Cantidad de créditos (mediana)	Interés Pagado Por Créditos \$ (mediana)
1	1.294	\$ 1.466.540	4	2	\$ 5.733.359	1	2	\$ 403.261
2	1.209	\$ 2.064.179	0	3	\$ 5.473.121	0	2	\$ 563.072
3	1.544	\$ 2.172.189	0	3	\$ 4.737.845	1	2	\$ 504.600
4	531	\$ 1.459.548	4	2	\$ 7.455.153	0	2	\$ 371.962

4.5.4 Evaluación de los modelos

En la Tabla 62 se presentan las métricas de evaluación para cada uno de los métodos aplicados. De acuerdo a estos datos, usando los modelos de K-means, PAM y Fuzzy c-means se obtuvieron los mismos grupos de clúster con los mismos registros y los mismos valores en los indicadores. Todos los modelos recomiendan 4 grupos para este modelo de segmentación.

Tabla 62: Evaluación de los agrupamientos de Asociados con ahorros y créditos activos

	K-means	PAM	Fuzzy C-means	Dbscan	HClust
Métrica / Número de Clusters	4	4	4	4	4
Dunn	0,420	0,420	0,420	0,581	0,420
Davies Bouldin	0,711	0,711	0,711	0,667	0,711
silhouette	0,562	0,562	0,562	0,584	0,562
Calinski Harabasz	5.148,499	5148,499	5148,499	5.921,952	5.148,499

PBM	0,380	0,380	0,380	0,413	0,380
Ray Turi	0,160	0,160	0,160	0,133	0,160
Wemmert Gancarski	0,671	0,671	0,671	0,685	0,671

Adjusted Rand Index: En la Tabla 63 se presenta los resultados del *Adjusted Rand Index* para los diferentes modelos usados para la segmentación. Se observa que las agrupaciones realizadas por los modelos de K-means, PAM y Fuzzy C-means son los mismos, confirmando la razón de la igualdad en las métricas de desempeño anteriormente presentadas

Tabla 63: Adjusted Rand Index de los métodos aplicados para la segmentación de asociados con ahorros y créditos activos.

	K-means	PAM	FUZZY	HClust	DBSCAN
K-means	1	1	1	1	0,948
PAM	1	1	1	1	0,948
FUZZY	1	1	1	1	0,948
HClust	1	1	1	1	0,948
DBSCAN	0,948	0,948	0,948	0,948	1

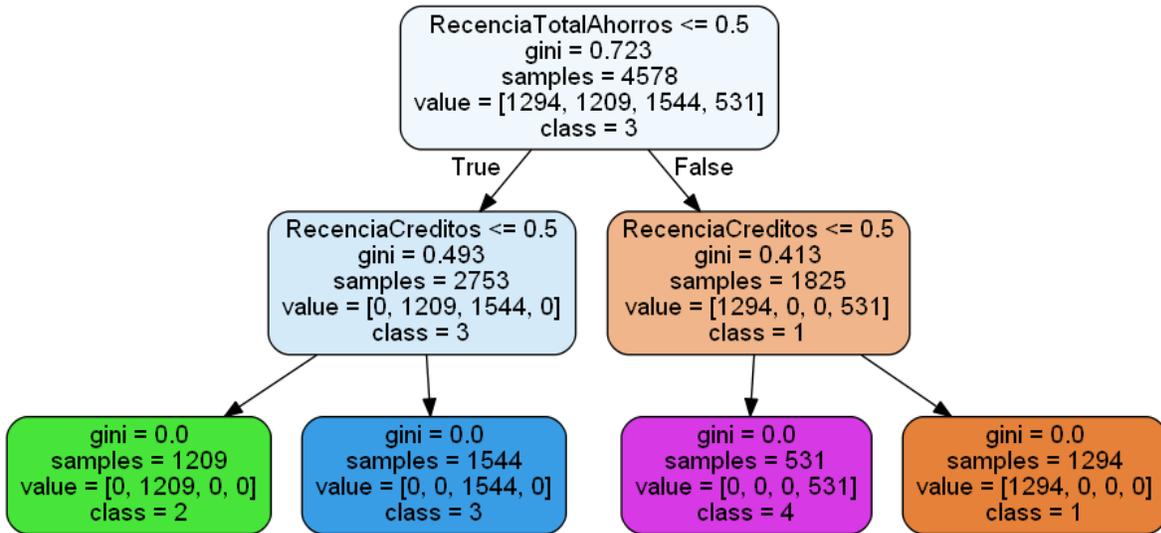
Analizando las métricas y por temas de interpretabilidad se eligió como óptimo el modelo de agrupamiento obtenido por K-means de 4 grupos.

4.5.5 Variables que más influyeron en la segmentación

Con el fin de identificar las reglas que más influyeron y las reglas que definieron el modelo de segmentación de asociados, se usó un modelo de árboles de decisión con los datos de los asociados etiquetados con la pertenencia a un clúster. El resultado del mismo se presenta a continuación.

Segmentación de asociados con ahorros y créditos activos: Para este modelo de segmentación por K-means se obtuvo el árbol de decisión presentado en la Figura 11.

Figura 11 Árbol de decisión de la segmentación de asociados con ahorros y créditos activos



De la Figura 11 se puede identificar que las reglas de clasificación para este modelo de clasificación fueron:

- **Regla 1:** Si $RecenciaAhorros \leq 0,5$ y $RecenciaCreditos \leq 0,5$, entonces el asociado pertenece al Grupo 2.
- **Regla 2:** Si $RecenciaAhorros \leq 0,5$ y $RecenciaCreditos > 0,5$, entonces el asociado pertenece al Grupo 3.
- **Regla 3:** Si $RecenciaAhorros > 0,5$ y $RecenciaCreditos \leq 0,5$, entonces el asociado pertenece al Grupo 4.
- **Regla 4:** Si $RecenciaAhorros > 0,5$ y $RecenciaCreditos > 0,5$, entonces el asociado pertenece al Grupo 1.

Se observa que 1.209 asociados son clasificados con la regla 1, 1.544 con la regla 2, 531 asociados son clasificados con la regla 3 y 1.294 asociados son clasificados con la regla 4.

Igualmente, de la figura 9 y del conjunto de reglas de clasificación se puede concluir que las variables que más influenciaron en la creación de los agrupamientos para asociados con ahorros y créditos activos fueron la recencia en créditos y la recencia en ahorros respectivamente.

4.5.6 Descripción de los agrupamientos y asociados típicos obtenidos

Caracterización transaccional y socio-demográfica de los asociados con ahorros y créditos activos típicos de los grupos obtenido por K-means

A continuación, se presenta las características transaccionales y sociodemográficas del asociado típico o representante de cada agrupamiento.

Caracterización transaccional

Tabla 64: Caracterización transaccional de los asociados con ahorros y créditos activos representantes de la segmentación de ahorradores obtenidos por K-means.

Clúster	Número de asociados	Recencia ahorros (meses) (media)	Cantidad Ahorros (media)	Saldo Ahorros \$ (media)	Recencia créditos (meses) (media)	Cantidad de créditos (media)	Saldo de Créditos \$ (media)	Interés pagado \$ (media)
1	1.294	5	3	\$ 2.053.289	2	3	\$ 9.178.364	\$ 1.890.457
2	1.209	0	4	\$ 3.542.840	0	8	\$ 8.664.085	\$ 2.952.700
3	1.544	0	4	\$ 3.617.298	1	5	\$ 7.263.223	\$ 2.347.054
4	531	6	3	\$ 2.036.302	0	4	\$ 11.518.275	\$ 2.032.497

Caracterización sociodemográfica: A continuación, se presentan los datos sociodemográficos obtenidos a partir de análisis de las modas de las variables categóricas de cada grupo.

Tabla 65: Caracterización sociodemográfica de los asociados con ahorros y créditos activos típico de la segmentación de ahorradores por K-means

Clúster	Genero (moda)	Rango Edad (moda)	Nivel Educativo (moda)	Estado Civil (moda)	Estrato (moda)	Tipo de Vivienda (moda)	Tipo de Contrato (moda)	Rango de Ingresos (SM) (moda)	Rango antigüedad (moda)	Antigüedad (moda)
1	Masculino	31- 40	Secundaria	Soltero(a)	2	Familiar	Término Indefinido	1-2	[18 - 34]	24, 31
2	Femenino	31- 40	Secundaria	Soltero(a)	2	Familiar	Término	1-2	[3	

							Indefinido		4 - 52)	
3	Femenino	31- 40	Secundaria	Soltero(a)	1	Familiar	Término Indefinido	1-2	[81 - 638)	22
4	Masculino	31- 40	Secundaria	Soltero(a)	2	Familiar	Independient e	2-4	[18 - 34)	20

Caracterización transaccional y socio-demográfica de los asociados típicos de cada grupo obtenido por PAM

En la Tabla 66 se presenta las características transaccionales y sociodemográficas de los asociados típicos o representantes de cada agrupamiento.

Tabla 66: Caracterización transaccional de los asociados con ahorros y créditos activos típicos de la segmentación de ahorradores obtenidos por PAM.

Clúster	Número de asociados	ID	Recencia ahorros (meses) (media)	Cantidad ahorros (media)	Saldo ahorros \$ (media)	Recencia créditos (media)	Cantidad créditos (media)	Saldo créditos \$ (media)	Interés pagado \$ (media)
1	1.209	25280	0	3	\$ 1.574.361	0	2	\$ 5.512.990	\$ 426.817
2	1.294	28744	3	3	\$ 1.047.081	1	2	\$ 7.374.510	\$ 488.910
3	531	42940	3	2	\$ 1.059.581	0	2	\$ 6.068.706	\$ 296.968
4	1.544	48589	0	3	\$ 1.530.508	1	2	\$ 4.139.087	\$ 365.949

Caracterización sociodemográfica: En la Tabla 67, se presentan los datos sociodemográficos obtenidos a partir de análisis de las modas de las variables categóricas de cada grupo.

Tabla 67: Caracterización sociodemográfica de los asociados con ahorros y créditos activos típicos de la segmentación de ahorradores por PAM.

Clúster	ID	Genero (moda)	Rango Edad (años) (moda)	Nivel Educativo (moda)	Estado Civil (moda)	Estrato (moda)	Tipo de Vivienda (moda)	Tipo de contrato (moda)	Rango de Ingresos (SM) (moda)	Rango antigüedad (meses) (moda)
1	2528 0	Masculino	18- 30	Técnico	Soltero(a)	2	Familiar	Término Indefinido	1-2	[34 - 52)
2	2874 4	Femenino	18- 30	Sin Estudio Formal	Unión Libre	1	Propia	Término Indefinido	1-2	[18 - 34)
3	4294 0	Masculino	51- 60	Técnico	Soltero(a)	1	Propia	Término Indefinido	2-4	[34 - 52)
4	4858 9	Femenino	41- 50	Primaria	Unión Libre	3	Propia	Otro	1-2	[34 - 52)

Descripción de los agrupamientos y asociados típicos con ahorros y créditos activos

Descripción de los grupos obtenidos por K-MEANS

A continuación, se presenta una descripción de las características de los grupos de asociados obtenidos

- Clúster 1: Asociados con baja cultura de ahorros y riesgo de crédito Alto:** Los asociados en este clúster representan un 28,3 % del total de asociados a la cooperativa. Tienen una participación del 19,5 % en el total de ahorros y el 29,2 % en el total de la cartera de créditos. Son el tercer grupo con mayor participación en ahorros y el tercero en la cartera de créditos, tienen una recencia promedio de 5 meses en ahorros y de 2 meses de recencia en créditos. Es el grupo con mayor recencia promedio en créditos, hay una alta varianza en la recencia de ahorros, se encuentran en este grupo asociados con valores de recencia que van desde 1 hasta 44 meses, así mismo es donde se encuentran los asociados con los menores saldos en ahorros. Son el tercer grupo con el mayor porcentaje de interés pagado (22,8 %). Analizando la recencia de créditos posiblemente son asociados que no están al día, o tienen periodo de pagos mayores a sesenta 60 días, estos asociados se pueden calificar como en un riesgo de crédito alto, ya que, a pesar de tener créditos activos, en promedio en los últimos 60 días no han realizado operaciones de

crédito, se recomienda hacerles un plan de seguimiento prioritario con el fin de identificar asociados que pueden tener dificultades en el pago de sus obligaciones y diseñar estrategias de cobro preventivo o de reliquidación de créditos. Desde el punto de vista del comportamiento en ahorros son asociados con un bajo hábito en ahorros.

- **Clúster 2: Asociados con buena cultura de ahorros y riesgo de crédito Bajo:** Los asociados en este clúster representan un 26,4 % del total de asociados a la cooperativa. Tienen una participación del 31,5 % en el total de ahorros y el 26,4 % en el total de la cartera de créditos. Son el segundo grupo con mayor participación en los saldos totales de ahorros y el primero en la cartera de créditos, todos los asociados de este grupo tienen recencia cero tanto en ahorros como en créditos, lo que indica que han realizado operaciones de ahorro y de crédito en el último mes. Son el segundo grupo con el mayor porcentaje de interés pagado (33,3 %). Son asociados con hábito de uso de crédito frecuente, y se mantienen al día, podrían calificarse como el grupo de los mejores asociados deudores analizando su recencia y por tanto se les puede asignar una calificación de riesgo de crédito baja.

- **Clúster 3: Asociados con buena cultura de ahorros y riesgo de crédito medio:** Los asociados en este clúster representan un 33,7 % del total de asociados a la cooperativa. Tienen una participación del 41 % en el total de ahorros y el 28,3 % en el total de la cartera de créditos. Son el grupo con mayor porcentaje de participación en los saldos totales de ahorros al igual que en número de productos de ahorros adquiridos, es el segundo grupo con mayor participación en la cartera de saldos de créditos, todos los asociados de este grupo tienen recencia en ahorros de 0 lo que indica que han realizado operaciones de ahorro en el último mes. A nivel de créditos todos los asociados de este grupo no han realizado operaciones de créditos en el último mes, se observa valores de recencia que van de 1 a 8 meses. Son el grupo con el mayor porcentaje de interés pagado (33,8 %). Analizando la recencia de créditos posiblemente son asociados que no están al día, o tienen periodo de pagos mayores a sesenta 60 días estos asociados se pueden calificar como en un riesgo de crédito medio.

- **Clúster 4: Asociados con Baja cultura de ahorros y riesgo de crédito medio:** Los asociados en este clúster representan un 11,6 % del total de asociados a la cooperativa. Tienen una participación del 7,9 % en el total de ahorros y el 15,4 % en el total de la cartera de créditos. Son el grupo con el menor porcentaje de participación en los saldos totales de ahorros al igual que en número de productos de ahorros adquiridos, es el grupo con menor participación en la cartera de saldos de créditos, todos los asociados de este grupo tienen recencia en ahorros mayor a 0 lo que indica que no han realizado operaciones de ahorro en el último mes. A nivel de créditos todos los asociados de este grupo han realizado operaciones de créditos en el último

mes. Son el grupo con el menor porcentaje de interés pagado (10,1 %). En aspectos de crédito se mantienen al día o han realizado operaciones de crédito en el último mes, esto se evidencia en la recencia igual a cero, en promedio hace más de 6 meses que no han hecho un ahorro.

Descripción de asociados típicos obtenidos por K-MEANS

- **Asociado típico Clúster 1:** Hombre en rango de edad entre 31y 40 años, con nivel de estudios secundaria, estado civil soltero, vive en estrato 2, su tipo de vivienda es Familiar, tiene contrato a Término indefinido, sus ingresos mensuales son menores a 2 salarios mínimos. Asociado con una antigüedad en la entidad entre 18 y 34 meses. A nivel de comportamiento en ahorros, se identifica que tiene 3 productos de ahorro, un saldo promedio de \$ 2.053.289 y una recencia de 5 meses. Con respecto a su comportamiento en créditos, ha adquirido tres productos, tiene un saldo de \$ 9.178.364, ha pagado \$ 1.890.457 en interés y tiene una recencia de 2 meses. Hace dos meses que no realiza una operación de créditos, hace cinco que meses que no realiza una operación de ahorros.
- **Asociado típico Clúster 2:** Mujer en rango de edad entre 31y 40 años, con nivel de estudios Secundaria, estado civil Soltera, vive en estrato 2, su tipo de vivienda es Familiar, tiene contrato a Término indefinido, sus ingresos mensuales son menores a 2 salarios mínimos. Asociado con una antigüedad en la entidad entre 34 y 52 meses. A nivel de comportamiento en ahorros, se identifica que tiene 4 productos de ahorro, un saldo promedio de \$ 3.542.840 y una recencia de 0. Con respecto a su comportamiento en créditos, ha adquirido 8 productos, tiene un saldo de \$ 8.664.085, ha pagado \$ 2.952.700 en interés y tiene una recencia de 0. Ha realizado operaciones de ahorro y de crédito en el último mes. Tiene buen hábito de pago en créditos y buena cultura de ahorros, es un buen cliente de crédito y ha pagado un buen monto en interés por los créditos adquiridos. Se pueden calificar como en un riesgo de crédito bajo.
- **Asociado típico Clúster 3:** Mujer en rango de edad entre 31y 40 años, con nivel de estudios Secundaria, estado civil Soltera, vive en estrato 1, su tipo de vivienda es Familiar, tiene contrato a Término indefinido, sus ingresos mensuales son menores a 2 salarios mínimos. Asociado con una antigüedad en la entidad entre 81 y 638 meses. A nivel de comportamiento en ahorros, se identifica que tiene 4 productos de ahorro, un saldo ahorrado promedio de \$ 3.617.298 y una recencia de 0. Con respecto a su comportamiento en créditos, ha adquirido cinco productos, tiene un saldo de \$ 7.263.223, ha pagado \$ 2.347.054 en interés y tiene una recencia de 1 mes.
- **Asociado típico Clúster 4:** Hombre en rango de edad entre 31 y 40 años, con nivel de estudios Secundaria, estado civil Soltero, vive en estrato 2, su tipo de vivienda es Familiar, su actividad económica es Independiente, sus ingresos mensuales son menores a 2 salarios

mínimos. Asociado con una antigüedad en la entidad entre 18 y 34 meses. A nivel de comportamiento en ahorros, se identifica que tiene 3 productos de ahorro, un saldo promedio en ahorros de \$ 2.036.302 y una recencia de 6 meses, con respecto a su comportamiento en créditos, ha adquirido 4 productos, tiene un saldo promedio de \$ 11.518.275, ha pagado \$ 2.032.497 en interés y tiene una recencia de 0. Hace seis meses que no realiza una operación de ahorros, ha realizado operaciones de créditos en el último mes.

Descripción de asociados típicos obtenidos por PAM

- **Asociado típico Clúster 1:** Hombre en rango de edad entre 18 y 30 años, con nivel de estudios Técnico, estado civil Soltero, vive en estrato 2, su tipo de vivienda es Familiar, tiene contrato a término indefinido, sus ingresos mensuales son menores a 2 salarios mínimos. Asociado con una antigüedad en la entidad entre 34 y 52 meses. A nivel de comportamiento en ahorros, se identifica que tiene 3 productos de ahorro, un saldo promedio de \$ 1.574.361 y una recencia de 0. Con respecto a su comportamiento en créditos, ha adquirido dos productos, tiene un saldo de \$ 5.512.990, ha pagado \$ 426.817 en interés y tiene una recencia de 0 meses. Ha realizado operaciones de ahorro y de créditos en el último mes.
- **Asociado típico Clúster 2:** Mujer en rango de edad entre 18 y 30 años, sin estudios formales, estado civil Unión libre, vive en estrato 1, su tipo de vivienda es Propia, tiene contrato a Término indefinido, sus ingresos mensuales son menores a 2 salarios mínimos. Asociado con una antigüedad en la entidad entre 18 y 34 meses. A nivel de comportamiento en ahorros, se identifica que tiene 3 productos de ahorro, un saldo promedio de \$ 1.047.081 y una recencia de 3 meses. Con respecto a su comportamiento en créditos, ha adquirido dos productos, tiene un saldo de \$ 7.374.510, ha pagado \$ 488.910 en interés y tiene una recencia de 1 mes.
- **Asociado típico Clúster 3:** Hombre en rango de edad entre 51 y 60 años, con nivel educativo técnico, estado civil soltero, vive en estrato 1, su tipo de vivienda es Propia, tiene contrato a Término indefinido, sus ingresos mensuales están entre 2 y 4 salarios mínimos. Asociado con una antigüedad en la entidad entre 34 y 52 meses. A nivel de comportamiento en ahorros, se identifica que tiene 2 productos de ahorro, un saldo promedio de \$ 1.059.581 y una recencia de 3 meses. Con respecto a su comportamiento en créditos, ha adquirido dos productos, tiene un saldo de \$ 6.068.706, ha pagado \$ 296.968 en interés y tiene una recencia de 0.
- **Asociado típico Clúster 4:** Mujer en rango de edad entre 41 y 50 años, con nivel educativo Primaria, estado civil Unión libre, vive en estrato 3, su tipo de vivienda es Propia, no tiene especificado su ocupación, sus ingresos mensuales son menores a 2 salarios mínimos. Asociado

con una antigüedad en la entidad entre 34 y 52 meses. A nivel de comportamiento en ahorros, se identifica que tiene 3 productos de ahorro, un saldo promedio de \$ 1.530.508 y una recencia de 0. Con respecto a su comportamiento en créditos, ha adquirido dos productos, tiene un saldo de \$ 4.139.087, ha pagado \$ 488.910 en interés y tiene una recencia de 1 mes.

4.6 Conclusiones

En este capítulo se presentó una propuesta de tres modelos de segmentación, uno para cada perfil de asociados de acuerdo al producto financiero activo. Para cada uno de estos se usaron varias técnicas de clustering (Jerárquico aglomerativo, K-means, PAM, Fuzzy C-means, Dbscan), y luego se eligió el mejor de acuerdo a un análisis comparativo de los resultados de las métricas de agrupamiento: Silhoutthe, Dunn, Davies Bouldin, Calinski Harabasz. Solo se presentaron los resultados de los agrupamientos obtenidos por K-means y se complementó con los medoides obtenidos por el método de PAM.

Desde el análisis de las métricas evaluadas todos los métodos de agrupamiento basados en particionamiento coincidieron en que las propuestas de clustering para la entidad eran óptimas con 2 segmentos de asociados ahorradores, 2 segmentos de asociados con créditos y 4 segmentos para los asociados con portafolio mixto.

Se identificaron las variables que más influyeron en la creación de los agrupamientos usando árboles de clasificación, se presentó posteriormente una descripción de los segmentos obtenidos. Adicionalmente a partir de la información de los centroides de K-means y de los medoides por el método de PAM se pudo presentar 2 propuestas de asociados “típicos” de cada uno de los clústeres obtenidos, lo cual permitirá que la cooperativa pueda usarlos como modelos para el diseño y desarrollo de productos y servicios financieros y estrategias de relacionamiento con sus grupos de asociados.

5. Evaluación

En este capítulo se evalúan los resultados de los modelos desarrollado desde una perspectiva de los objetivos de negocio, está es una actividad requerida dentro de la metodología CRISP-DM. Una vez realizada esta evaluación, se debe decidir si los objetivos han sido cumplidos y de ser así se puede avanzar a la fase de implantación, de lo contrario se tendría que identificar cualquier factor que se haya podido pasar por alto y hacer una revisión del proceso.

En el capítulo anterior se hizo la presentación de los resultados de la evaluación técnica de las tres propuestas de segmentación desarrolladas utilizando varias métricas analíticas, en el cual no se tiene en cuenta que los resultados de los grupos conformados si tengan sentido o no para el negocio. Desde el punto de vista de negocios y en conjunto con la evaluación por parte de un grupo de experto de la entidad luego de hacer la interpretación a los mismos tanto como grupo y a nivel del representante o asociado “típico” para cada uno de los agrupamientos permiten evidenciar que los modelos presentados entregan información útil, interpretable y aprovechable para un mejor conocimiento y gestión de los asociados y específicamente en el desarrollo de propuestas de gestión comercial, de servicio y de riesgos financieros, cumpliendo con los objetivos propuestos.

Las propuestas de segmentación presentadas en este proyecto son un insumo importante y complementario para enriquecer las gestiones de los procesos comercial, de riesgo de crédito, liquidez, la tesorería y de análisis de provisiones por posibles pérdidas por incumplimiento en pagos de créditos. Son muchos los análisis que permitirán los modelos de segmentación presentados en este proyecto que por sí mismos o complementados por datos de otras fuentes ayudarán a la entidad entre otros a:

- Identificar cuáles son los mejores asociados desde su comportamiento en ahorros y créditos, desde el punto de vista de su contribución en la cartera de saldos ahorrados o de interés pagado por el uso de productos de crédito.
- Conocer a los asociados más leales, considerando aquellos como los que tengan mayores saldos en ahorros, mayor cantidad de productos adquiridos y menor recencia de ahorros, y/o los asociados con créditos que tengan recencia 0, mayor cantidad de interés pagado, mayor cantidad de créditos adquiridos.
- Agrupar a los asociados con mayor posibilidad de caer en riesgo de crédito, por ejemplo, aquellos asociados con créditos activos con alto saldo en crédito y alta recencia.
- A sí mismo, permiten conocer en cuales asociados están concentrados la mayor parte del

total de ahorros en la entidad, ya que son de alto impacto en el caso de su retiro de la entidad, y por eso es necesario conocerlos para desarrollar propuestas de retención ajustadas al perfil transaccional, sociodemográfico y de estilo de vida.

Los indicadores de agrupamiento mostraron mejores valores para los agrupamientos obtenidos por modelos basados en particiones como K-means, PAM, fuzzy c-means comparados con DbSCAN.

Desde el punto de vista de negocios y en conjunto con la evaluación por parte de un grupo de expertos de la entidad se concluye que las tres propuestas de segmentación obtenidas por el modelo de K-means cumplen con los criterios de éxito de negocio. Posteriormente la entidad desarrollará un plan para que estos sean implementados en la entidad. Este no hace parte del alcance de este proyecto.

6. Producto de datos

En este capítulo se presenta un producto de datos desarrollado, el cual dentro de la metodología de CRIPSP-DM hace parte de la fase de Implementación.

En la sección de análisis descriptivo de los resultados obtenidos, se presentó un resumen de los grupos y de los asociados típicos para las tres propuestas de segmentación. Para un mejor análisis de los asociados que incluso permita realizar “microsegmentaciones”, se recomienda hacer un análisis estadístico de datos para cada uno de los clústeres obtenidos con las variables transaccionales y sociodemográficas. Sin embargo, este no se realizará como parte del alcance de este proyecto.

Es fundamental la presentación a los directivos y de los líderes de procesos de la cooperativa de los resultados obtenidos, en un formato que facilite una interpretación desde el punto de vista del negocio, y que para operarlo y generar los informes requeridos no requiera conocimientos técnicos. Para cumplir con lo anterior se desarrolló un producto de datos usando inicialmente Google Data Studio software y posteriormente se realizó una versión más completa usando R-Shiny software.

El producto desarrollado a pesar de que es funcional no tiene como propósito principal ser implementado en producción, es un producto mínimo viable que permitirá ser usado como un prototipo para una etapa posterior de evaluación e implementación de una solución que se adapte a las necesidades de la entidad.

El producto de datos desarrollado tiene varias funcionalidades, así: un módulo de análisis exploratorio básico de los datos, un módulo de visualización de los de agrupamientos para los datos de la entidad usando K-means, PAM, Fuzzy C-means, Dbscan y Optics, un dashboard donde se presentan los resultados de los agrupamientos obtenidos de manera tal que sean interpretables para los funcionarios de la entidad y finalmente la posibilidad de exportar los asociados de un grupo en particular en un archivo de texto plano.

El aplicativo desarrollado con Google Data Studio se encuentra disponible en la siguiente dirección de internet: <https://datastudio.google.com/s/17m9LFtTXHU>. Una imagen de uno de los tableros del aplicativo se presenta en la imagen 15, donde se puede apreciar el análisis del comportamiento en transacciones de ahorros de acuerdo a varias variables sociodemográficas de los asociados.

Figura 12 Dashboard desarrollado en Google Data Studio



El producto final de datos en R- Shiny permiten de manera interactiva visualizar los detalles de las tres propuestas de segmentación, es posible ver los detalles transaccionales y sociodemográficos de manera global o por cada uno de los agrupamientos y exportar el listado de asociados que componen cada uno de los segmentos. El aplicativo se encuentra disponible en la siguiente dirección de internet: <https://william-jovel.shinyapps.io/sica/>

Figura 13 Producto de datos desarrollado en R shiny



El aplicativo permite de manera gráfica e interactiva dar respuesta a los siguientes aspectos de interés:

- ¿Cómo están distribuidos los asociados ahorradores y asociados con crédito en los diferentes estratos socioeconómicos de manera global o para un clúster en particular?
- ¿Cómo están distribuidos los asociados ahorradores y asociados con crédito de acuerdo al rango de edad de manera global o para un clúster en particular?
- ¿Cómo están distribuidos los asociados ahorradores y asociados con crédito de acuerdo al estado civil de manera global o para un clúster en particular?
- ¿Cómo están distribuidos los asociados ahorradores y asociados con crédito de acuerdo al tipo de vivienda de manera global o para un clúster en particular?
- ¿Cómo están distribuidos los asociados ahorradores y asociados con crédito de acuerdo al nivel académico de manera global o para un clúster en particular?
- ¿Cómo están distribuidos los asociados ahorradores y asociados con crédito de acuerdo al tipo de contrato de manera global o para un clúster en particular?

En términos generales el producto de datos permite dar respuesta a los objetivos de este proyecto y se convierte en la base para el desarrollo de un sistema de información analítico para la gestión de asociados en una entidad del sector solidario.

Una imagen del aplicativo se presenta a continuación, donde se muestra un tablero con la distribución de créditos en la entidad analizada desde múltiples variables sociodemográficas de los asociados

Figura 14 Dashboard de clúster de asociados con créditos desarrollado en R shiny



7. Conclusiones y recomendaciones

7.1 Conclusiones

Este proyecto cumplió con los objetivos planteados: Se desarrollaron tres modelos de segmentación de asociados a partir de variables de tipo transaccional (RFM) aplicando cuatro técnicas de clustering diferentes. Se identificaron las variables que más influyeron en la creación de los agrupamientos usando árboles de clasificación, se presentó una descripción de los segmentos obtenidos y de los asociados representantes de cada clúster y finalmente se desarrolló un producto de datos usando R Shiny que permite la visualización de los agrupamientos obtenidos. Desde el punto de vista metodológico este proyecto se realizó siguiendo las recomendaciones propuestas por la metodología CRISDP-DM, permitiendo cumplir con los objetivos propuestos.

El desarrollo del proyecto de segmentación de asociados para la cooperativa es de alta importancia ya que permitió identificar a partir de las variables de RFM quienes son sus mejores clientes. El producto de datos desarrollado le permitirá a la Cooperativa un mejor análisis de sus asociados, y de esta manera, podrán usar esta información para desarrollar estrategias que ayuden a una mejor gestión de sus asociados.

Para este proyecto a pesar que se obtuvieron los mismos grupos por los diferentes métodos aplicados, los asociados típicos de cada grupo obtenido por PAM, difieren de los obtenidos por K-means, ya que el primero se basa en el medoide, mientras que el segundo en el centroide. Por esto resulta complejo definir un perfil único del tipo de asociado de la Cooperativa por grupo. Si bien los centroides y los medoides pueden dar una idea de las características sociodemográficas y transaccionales de un asociado típico de cada clúster, se recomienda hacer un mayor análisis de los asociados para establecer si es necesario la creación por cada clúster de varios perfiles de asociados, para esto se podría usar el producto de datos desarrollado.

7.2 Recomendaciones

Se recomienda para proyectos posteriores desarrollar y evaluar modelos de segmentación de asociados que incluyan otras variables como por ejemplo rentabilidad por asociado, ingresos mensuales del asociado, antigüedad del asociado, edad del asociado, que permitan complementar las propuestas presentadas en este trabajo.

Se recomienda desarrollar una propuesta de modelos de ranking de asociados, en el cual se asignan

calificaciones a cada asociado a partir de las variables de RFM. En este proyecto no se llegó a ese punto. Como beneficio de esos modelos es que permiten diferenciar de manera cuantitativa un asociado de otro, a pesar de que pueden estar en un mismo clúster. Para hacerlo se puede hacer a través de las variables de RFM, donde la calificación se obtiene a través de un proceso matemático con los valores de estas para cada asociado.

Otra posibilidad para proyectos futuros es modificar el modelo de agrupamiento asignando pesos a las variables de RFM, en las lecturas realizadas este tipo de modelo se trabajan como RFM ponderado o RFM Scoring, y lo que buscan es diferenciar los clientes, dándole mayor importancia a unas variables de otras.

En este proyecto se aplicaron las técnicas de agrupamiento solo para asociados activos, esto es, tienen al menos una cuenta de ahorros activa o un crédito activo. Se recomienda usar este tipo de técnicas para el análisis de asociados inactivos, para identificar quienes pueden ser potenciales para reactivarlos a través de campañas dirigidas, por ejemplo, asociados que alguna vez han tenido crédito y en el momento no tienen, o campañas para asociados que en algún momento han tenido ahorros importantes y en el momento no son ahorradores.

Se recomienda realizar un nuevo modelo de agrupamiento de asociados, pero haciendo filtros adicionales a la base de datos, tales como filtrar para incluir solo asociados con una antigüedad mayor a un año, asociados con un monto mínimo en saldos en ahorros, asociados con un monto mínimo en saldos de créditos adeudados, asociados con un monto mínimo en interés pagados por créditos. Esto permitiría incluir asociados con historial en la entidad y que en el análisis de la variable de recencia en ahorros se incluya solo asociados que han realizado operaciones adicionales al aporte voluntario en el proceso de afiliación, esto también permitirá reducir la cantidad de registros atípicos.

Este tipo de actividades de obtención de agrupamientos se recomienda realizarlo al menos una vez al año, y compararlos con los realizados en periodos anteriores, de esta forma podría ser utilizado para analizar si las estrategias implementadas desde el punto de vista comercial y de servicio se reflejan en cambios visibles en los agrupamientos de asociados, sin embargo, dado el tipo de producto que se comercializa en una cooperativa, los cuales tiene unas condiciones particulares de adquisición y recompra, es posible que los resultados de las acciones estratégicas no se visualicen de forma inmediata, y están supeditados a la capacidad financiera y del nivel de endeudamiento de un asociado, así que si bien estos modelos pueden ayudar a caracterizar clientes, se deben integrar con modelos de análisis de riesgo financiero para un mejor diseño de las propuestas comerciales.

Bibliografía

- [1] Manual CRISP-DM de IBM SPSS Modeler, IBM Corporation, 2012. [En línea]. Disponible: <ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>.
- [2] La Minería de Datos de la A a la Z: Cómo descubrir conocimientos y crear mejores oportunidades. SAS Institute Inc. 2015. White paper. [En línea]. Disponible en <https://www.sas.com/content/dam/SAS/documents/marketing-whitepapers-ebooks/sas-whitepapers/es/data-mining-104937.pdf>.
- [3] C. Aggarwal, Data Mining: The Textbook, DOI 10.1007/978-3-319-14142-8 1, ISBN 978-3-319-14142-8 (eBook). Springer International Publishing Switzerland 2015. [En línea]. Disponible en https://doc.lagout.org/Others/Data%20Mining/Data%20Mining_%20The%20Textbook%20%5BAggarwal%202015-04-14%5D.pdf.
- [4] H. Jiawei, K. Micheline y P. Jian. Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems) Data Mining. Third Edition. Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems). ISBN-13: 978-9380931913.
- [5] Colombia Fintech. Las Fintech revolucionan los servicios financieros. [En línea]. Disponible en <https://www.colombiafintech.co/novedades/las-fintech-revolucionan-los-servicios-financieros>. [Accedido: 8-sep-2020].
- [6] Técnicas de minería de datos para la detección y prevención del lavado de activos y la financiación del terrorismo (LA/FT). Unidad de Información y Análisis Financiero (UIAF), Unidad Administrativa Especial adscrita al Ministerio de Hacienda y Crédito Público. Primera edición. Colombia, 2014. ISBN: 978-958-58578-5-8. [En línea]. Disponible en https://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Tecnicas-de-mineria-de-datos-para-la-prevencion-de.pdf.
- [7] Aplicabilidad de la minería de datos y el análisis de redes sociales en la inteligencia financiera. Unidad de Información y Análisis Financiero (UIAF), Unidad Administrativa Especial adscrita al Ministerio de Hacienda y Crédito Público. Primera edición. Colombia, 2014. ISBN: 978-958-58578-3-4. [En línea]. Disponible en [https://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Aplicabilidad-de-la-mineria-de-datos-\(1\).pdf](https://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Aplicabilidad-de-la-mineria-de-datos-(1).pdf).

- [8] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh. "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study, 2011". [En línea]. Disponible: <https://www.sciencedirect.com/science/article/pii/S1877050910003868>.
- [9] M. Haenlein, A. Kaplan, A. Beeser. "A Model to Determine Customer Lifetime Value in a Retail Banking Context". *European Management Journal* Vol. 25, No. 3, pp. 221–234, 2007. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0263237307000163>.
- [10] S. Hossein, M. Khorrami. "Data mining techniques for analyzing bank customers: A survey. January 2018", *Intelligent Decision Technologies* 12(7):1-19. DOI: 10.3233/IDT-180335. [En línea]. Disponible: https://www.researchgate.net/publication/322611443_Data_mining_techniques_for_analyzing_bank_customers_A_survey.
- [11] R. Figueroa-Diaz, S. Chamba, R. Guaman-Quinche1 y M. Cueva-Hurtado. "Mapas Autoorganizados aplicados a la segmentación de clientes en entornos empresariales", Universidad Nacional de Loja. *Revista Tecnológica ESPOL – RTE*, Vol. 29, N. 1, 130-143 (Julio 2016).
- [12] M. Mas Rosa. "Análisis del modelo RFM según el método convencional y el método de las 2-tuplas". master en minería de datos e inteligencia de negocios universidad complutense de Madrid, 2016. [En línea]. Disponible: https://eprints.ucm.es/38546/1/Trabajo%20Fin%20de%20Master_Rosa%20Maria%20Mas.pdf.
- [13] S. V. Chuva Bueno. "Inteligencia de negocios aplicando la metodología RFM a las cuentas de los Socios de la COAC Jardín Azuayo," B.S. Thesis. Cuenca, 2016-03-07. [En línea]. Disponible: <http://dspace.ucuenca.edu.ec/handle/123456789/25323>.
- [14] A. Cuadros, C. Gonzales, C. Jiménez. "Análisis multivariado para segmentación de clientes basada en RFM", Universidad del Valle. Noviembre 2017. [En línea]. Disponible: https://www.researchgate.net/publication/326548750_Analisis_multivariado_para_segmentacion_de_clientes_basada_en_RFM.
- [15] A. Ansari, A. Riasi. "Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms". *International Journal of Business and Management*; Published by Canadian Center of Science and Education. Vol. 11, No. 7; 2016. En línea]. Disponible: <https://pdfs.semanticscholar.org/483b/93e535e40d30301d98005cd38ddb718c5156.pdf>.

[16] M. Dachyar, F. M. Esperanca, y R. Nurcahyo, "Loyalty Improvement of Indonesian Local Brand Fashion Customer Based on Customer Lifetime Value ({CLV}) Segmentation", {IOP} Conf. Ser. Mater. Sci. Eng., vol. 598, p. 12116, sep. 2019.

[17] W. Gata, Iskandar, H. Basri, D. A. Puspitawati, S. Hidayat, y Walim, "Implementation of Decision Tree Algorithm in Customer Recency, Frequency, Monetary, and Cost Profiling: a Case Study of Plastic Packing Industry". {IOP} Conf. Ser. Mater. Sci. Eng., vol. 662, p. 22032, nov. 2019.

[18] R. Gustriansyah, N. Suhandi, F. Antony. "Clustering optimization in RFM analysis based on k-means". Indonesian Journal of Electrical Engineering and Computer Science, Vol. 18, No. 1, April 2020, pp. 470~477. ISSN: 2502-4752, DOI: 10.11591/ijeecs.v18.i1. pp 470-477.