



**Modelo de segmentación conductual y por beneficios buscados aplicado al  
segmento de persona natural en la banca colombiana**

Fabián Ricardo Bautista Rueda

Universidad EAN

Facultad de administración, finanzas y ciencias económicas

Maestría en administración de empresas - MBA

Bogotá, Colombia

12/Noviembre/2021

**Modelo de segmentación conductual y por beneficios buscados aplicado al  
segmento de persona natural en la banca colombiana**

**Fabián Ricardo Bautista Rueda**

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

**Magíster en Administración de Empresas - MBA**

Director:

Jonathan David Franco Ospina

Modalidad:

**Monografía**

Universidad EAN

Facultad de administración, finanzas y ciencias económicas

Maestría en administración de empresas - MBA

Bogotá, Colombia

12/Noviembre/2021

Nota de aceptación:


---

Firma del jurado

---

Firma del jurado

---

Firma del director del trabajo de grado

Ciudad, día/mes/año

## Resumen

La segmentación de clientes consiste en identificar grupos de clientes con características similares permitiendo adaptar la estrategia organizacional a las características de cada grupo facilitando el éxito de las organizaciones. Estas características se relacionan con la demografía, geografía, el uso de productos o servicios, estilos de vida y con los beneficios buscados por los clientes. El sector bancario representa un pilar fundamental del crecimiento económico y este sector ha presentado un gran cambio en los últimos años debido a los avances tecnológicos derivados de la Industria 4.0, en particular en la aplicación del Big Data y la minería de datos, herramientas recomendadas para realizar la segmentación de clientes. Las organizaciones de este sector han utilizado la segmentación conductual (uso de canales digitales y tenencia de cuentas de ahorro y créditos), por beneficios buscados y según una variable de interés para la organización pero deben utilizar diferentes fuentes de información de los clientes en el proceso de segmentación.

Es por esta última razón que esta investigación se propuso identificar qué tipos de segmentos de clientes persona natural se obtienen al usar información conductual (tenencia y saldo de los productos cuentas de ahorro, CDTs, tarjetas de crédito, créditos y uso de canales) y por beneficios buscados (calificación de los costos, de los servicios y recomendación del banco) de los clientes persona natural de la banca colombiana utilizando el procedimiento estadístico K-medias con información simulada de los clientes. Se encontraron cinco segmentos con características diferenciadas según la información analizada, permitiendo establecer una estrategia de oferta de productos y servicios, enfocada en los canales que usan los clientes de cada uno de los segmentos y teniendo en cuenta la percepción que tienen con respecto a los costos, a los servicios y su recomendación del banco. El haber utilizado distintos tipos de información permitió caracterizar los segmentos de manera más completa que al haber utilizado un solo tipo de información, logrando tener una visión más completa de los clientes persona natural y posibilitando tomar decisiones más precisas. Esto repercutirá en la fidelización, en la retención y en la satisfacción de los clientes persona natural.

**Palabras clave:** Segmentación, mercadeo, sector bancario, Industria 4.0, minería de datos, Big Data.

### **Abstract**

Customer segmentation is the process of identifying groups of customers with similar characteristics allowing organizations to adapt the strategy to the characteristics of each group, facilitating the success of organizations. These characteristics are related to demographics, geography, the use of products or services, lifestyles and the benefits sought by customers. The banking sector represents a fundamental pillar of economic growth and this sector has undergone a great change in recent years due to technological advances derived from Industry 4.0, in particular in the application of Big Data and Data Mining, recommended tools to perform customer segmentation. Organizations in this sector have used behavioral segmentation (use of digital channels and holding of savings and credit accounts), segmentation by benefits sought by customers and according to a variable of interest to the organization, but they must use different sources of customer information in the segmentation process.

It is for this last reason that this research aims at identifying what types of individual customer segments are obtained using behavioral information (holding and balance of saving accounts, deposits, credit cards, credits and use of channels) and the benefits sought (rating of costs, services and bank recommendation) by individual customers from a Colombian banking institution using the K-means statistical procedure with simulated customer information. Five individual customer segments were found with differentiated characteristics according to the information analyzed, allowing to establish a strategy for the offering of products and services, focused on the channels used by the individual customers of each of the segments and taking into account their perception of the costs, the services and their recommendation of the bank. Having used different types of information allowed one to characterize the individual customer segments in a more complete way than having used a single type of information, achieving a more complete view of the individual customers and making it possible to make more precise decisions. This will have an impact on loyalty, retention and satisfaction of the individual customers.

**Keywords:** Segmentation, marketing, banking sector, Industry 4.0, Data Mining, Big Data.

## Contenido

	<b>Pág.</b>
<b>Lista de Figuras .....</b>	<b>8</b>
<b>Lista de Tablas .....</b>	<b>9</b>
<b>Introducción .....</b>	<b>10</b>
<b>Objetivos .....</b>	<b>13</b>
<i>Objetivo general.....</i>	<i>13</i>
<i>Objetivos específicos .....</i>	<i>13</i>
<b>Justificación.....</b>	<b>14</b>
<b>Marco Teórico .....</b>	<b>16</b>
<i>Industria 4.0 .....</i>	<i>17</i>
<i>Sector bancario .....</i>	<i>18</i>
<i>Segmentación de clientes.....</i>	<i>19</i>
<i>Segmentación de clientes en el sector bancario .....</i>	<i>22</i>
<i>Recomendación del uso de la minería de datos y el Big Data para la segmentación de clientes en el sector bancario .....</i>	<i>23</i>
<i>Segmentación de clientes en el sector bancario según una variable de interés .....</i>	<i>25</i>
<i>Segmentación de clientes en el sector bancario según el uso de canales.....</i>	<i>26</i>
<i>Segmentación de clientes en el sector bancario según el uso de productos.....</i>	<i>27</i>
<i>Segmentación de clientes en el sector bancario según los beneficios buscados.....</i>	<i>28</i>
<b>Hipótesis .....</b>	<b>30</b>

<b>Metodología</b> .....	<b>31</b>
<i>Enfoque y alcance de la investigación</i> .....	31
<i>Población y muestra</i> .....	32
<i>Instrumentos</i> .....	32
<i>Técnicas para el análisis de la información</i> .....	32
<b>Variables</b> .....	<b>32</b>
<b>Trabajo de Campo</b> .....	<b>38</b>
<i>Procesamiento de los datos</i> .....	38
<i>Análisis de resultados</i> .....	41
<i>Propuesta de solución a la problemática</i> .....	46
<b>Discusión</b> .....	<b>48</b>
<b>Conclusiones y Trabajo Futuro</b> .....	<b>50</b>
<i>Conclusiones</i> .....	50
<i>Trabajo futuro</i> .....	51
<b>Referencias</b> .....	<b>53</b>
<b>A. Anexo. Resultados estadísticos por segmento</b> .....	<b>57</b>
<b>B. Anexo. Código en R para obtener la segmentación</b> .....	<b>58</b>

## Lista de Figuras

	<b>Pág.</b>
Figura 1. Árbol de la ciencia. ....	17
Figura 2. Muestra base de datos. ....	39
Figura 3. Selección número de segmentos. ....	41
Figura 4. Tamaño de los segmentos. ....	42
Figura 5. Tenencia productos por segmento. ....	43
Figura 6. Promedio saldo productos por segmento. ....	43
Figura 7. Uso canales por segmento. ....	44
Figura 8. Promedio calificaciones por segmento. ....	45

### Lista de Tablas

	<b>Pág.</b>
Tabla 1. Tenencia de productos .....	39
Tabla 2. Saldo de los productos .....	40
Tabla 3. Uso de canales .....	40
Tabla 4. Calificaciones costos, servicios y recomendación .....	41

## Introducción

La cuarta revolución industrial o la Industria 4.0 se caracteriza por la aplicación de nuevas tecnologías como el Internet de las cosas, la computación en la nube y el uso de dispositivos inteligentes en las organizaciones y busca la digitalización de los ecosistemas industriales. Esta industria utiliza sensores inteligentes, la inteligencia artificial, la ciencia de datos y la analítica para analizar en tiempo real la información de los sistemas y los procesos de las organizaciones permitiendo la automatización y la optimización de muchas tareas (Xu, Xu y Li, 2018). En la actualidad y en la cuarta revolución industrial, la tecnología ha permeado todas las áreas de la sociedad, la producción, las finanzas, los servicios, el transporte y las comunicaciones (Mehdiabadi et al, 2020).

El sistema bancario representa un pilar fundamental del crecimiento económico y la estabilidad macroeconómica, especialmente en el contexto de la globalización. El establecimiento de un sistema bancario sólido y eficiente es un prerrequisito importante para un crecimiento económico sostenible. El sector bancario es uno de los sectores que ha presentado un gran cambio debido a los avances tecnológicos derivados de la Industria 4.0, en particular logrando ventajas como el mejoramiento de la productividad, la creación de productos y servicios innovadores, transacciones y transferencias rápidas, sistemas de información en tiempo real y el manejo eficiente del riesgo (Mehdiabadi et al, 2020).

En la mayoría de los mercados donde operan las organizaciones, las necesidades, los deseos y los intereses de sus clientes son muy variados, así como sus hábitos de compra y uso de los productos o servicios, lo que hace casi imposible satisfacer a todos los clientes con una oferta única. Las organizaciones consideran esencial la segmentación de mercados donde se logran identificar grupos de clientes con hábitos de compra y requisitos similares (Lambin, Gallucci y Sicurrello, 2009). Identificar estos grupos con características similares permite a los gerentes adaptar productos y servicios acorde a las necesidades de cada grupo facilitando el éxito de las organizaciones (Ferrel y Hartline, 2018).

El sector bancario colombiano se enfrenta a nuevos retos como la llegada al país de empresas disruptivas como las fintech, la adopción de nuevas tecnologías como el Big Data, el Blockchain, las nuevas tendencias en analítica de datos, la crisis derivada por la

pandemia del COVID-19 y en general la situación propia del país. Según Xu et al (2018), uno de los ejes de la Industria 4.0 es la analítica, la que permite la toma de mejores decisiones de negocio en las organizaciones y de acuerdo con Berger (2017), la Industria 4.0 provee flexibilidad a los procesos de producción y así ayuda a crear productos y servicios específicos para un segmento objetivo, satisfaciendo necesidades personalizadas a un costo marginal. La segmentación de clientes es un ejemplo de lo que las organizaciones del sector bancario colombiano pueden lograr con la analítica y como mencionan Villanueva y De-Toro (2017), segmentar a los clientes de manera creativa puede proporcionar ventajas competitivas a las organizaciones al permitir adaptar, para cada segmento, la mejor oferta de producto o servicio, la política de precios, los canales de distribución y las demás variables comerciales.

La definición de las organizaciones y su manera de administrarlas ha evolucionado con el paso del tiempo, empezando con la teoría administrativa científica que hace énfasis en las tareas, hasta llegar al nuevo paradigma de las organizaciones responsables y sostenibles que responden a los problemas del mundo actual y a un consumidor más informado y exigente. Las organizaciones responsables y sostenibles deben ser abiertas y sensibles hacia el entorno, manteniendo un diálogo auténtico con los diferentes grupos de interés y deben considerar el largo plazo, construyendo relaciones basadas en la confianza y permitiendo el beneficio mutuo mientras dure esta relación (Olcese, Rodríguez y Alfaro, 2008).

Las organizaciones responsables y sostenibles son aquellas que se preocupan por sus clientes, sus proveedores, sus empleados, por la comunidad donde operan y por el medio ambiente, y buscan que sus relaciones con cada uno de estos grupos de interés sean cada vez mejores, duraderas y confiables (Rojas, Ramírez y Vélez, 2013). Los clientes son uno de los grupos de interés más importantes para las organizaciones dado que estas se deben a ellos y sin ellos no existirían y teniendo en cuenta la definición del mercadeo de la Asociación Americana de Mercadeo como “la actividad y los procesos para crear, comunicar, entregar e intercambiar productos y servicios que tengan valor para los clientes”, esto nos remite a la segmentación como una herramienta para ofrecer productos y servicios adecuados a cada grupo de clientes. Adicionalmente, las organizaciones deben empezar o continuar su transformación digital, con la inclusión de nuevas tecnologías como la analítica de datos.

El problema de investigación surge como una duda profesional y académica. Como se mencionó anteriormente, el sector bancario es importante para el desarrollo económico sostenible de las comunidades humanas, es fundamental que las organizaciones estén atentas al ambiente en el que se desarrollan y deben ser conscientes de la importancia de utilizar las nuevas tecnologías de la Industria 4.0, y en particular las que les permitan relacionarse de la mejor manera con un grupo de interés vital para las organizaciones como lo son los clientes.

Por lo anterior, es necesario determinar las decisiones que pueden tomar las organizaciones del sector bancario colombiano al segmentar a sus clientes del segmento persona natural usando información conductual y por beneficios buscados, identificar la cantidad de segmentos de clientes obtenidos al usar esta información y las características de estos segmentos. Esto permite plantear la siguiente pregunta de investigación: ¿Qué tipos de segmentos se pueden obtener usando información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana?.

El documento está organizado de la siguiente manera, primero se plantea el objetivo general y los objetivos específicos del trabajo de grado, posteriormente se presenta la justificación y se desarrolla el marco teórico, después se plantea la hipótesis de trabajo, la metodología de la investigación y las variables analizadas para finalizar con el trabajo de campo y la discusión de las conclusiones obtenidas.

## **Objetivos**

A continuación se presenta el objetivo general y los objetivos específicos del trabajo de grado:

### **Objetivo general**

Identificar qué tipos de segmentos de clientes persona natural se obtienen usando información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana.

### **Objetivos específicos**

- Evaluar la cantidad de segmentos de clientes obtenidos al usar información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana.
- Conocer las características de los segmentos de clientes persona natural de la banca colombiana usando información conductual y por beneficios buscados.
- Determinar las decisiones que pueden tomar las organizaciones del sector bancario colombiano al segmentar a sus clientes persona natural usando información conductual y por beneficios buscados.

### **Justificación**

Según Bhuvana, Thirumagal y Vasantha (2016), la falta de integración de información o los silos de información, son una barrera para que las organizaciones puedan analizar grandes volúmenes de información. Las organizaciones del sector bancario necesitan tener una visión completa de sus clientes para poder ofrecerles el producto o servicio correcto y adaptarse a las necesidades de sus clientes. El desarrollo de las capacidades para manejar y analizar información les permitirá a las organizaciones del sector bancario colombiano superar la falta de integración de la información lo que limita tener una visión unificada de sus clientes. Es por esto que el presente trabajo de grado quiere identificar la cantidad de segmentos, sus características y las decisiones que pueden tomar las organizaciones del sector bancario colombiano al segmentar a sus clientes persona natural usando distintas fuentes de información, como la tenencia y el uso de productos, el uso de canales y los beneficios buscados por los clientes persona natural en la relación con su banco.

En el sector bancario, diseñar productos y servicios según las preferencias de cada uno de los clientes no es posible debido al costo y al tiempo. Por tanto, algún grado de estandarización es necesario, lo que se logra con la segmentación de clientes. Cómo se va a observar en el marco teórico, las organizaciones del sector bancario han segmentado sus clientes según una variable de interés para la organización (Ogwueleka et al y Smeureaunu, Ruxanda y Badea), según el uso de sus canales (Aryuni, Didik Madyatmadja y Miranda y Chawla y Joshi), según la tenencia y el uso de ciertos productos y servicios (Mihova y Pavlov y Sundjaja) o por los beneficios buscados por los clientes en la relación con su banco (Dimitriadis, Kaur, Sharma y Mahajan, Alfansi y Sargeant y Machauer y Morgner). Sin embargo, ninguna de las investigaciones ha analizado la situación en países Latinoamericanos ni usado una combinación de distintas fuentes de información de sus clientes, por tanto, existe una oportunidad de conocer las características de los segmentos de clientes persona natural que pueden obtener las organizaciones del sector bancario colombiano usando información conductual y por beneficios buscados.

La definición de las organizaciones y su manera de administrarlas ha evolucionado con el paso del tiempo, empezando con la teoría administrativa científica que hace énfasis en las tareas, hasta llegar al nuevo paradigma de las organizaciones responsables y

sostenibles que responden a los problemas del mundo actual y a un consumidor más informado y exigente. Las organizaciones responsables y sostenibles deben ser abiertas y sensibles hacia el entorno, manteniendo un diálogo auténtico con los diferentes grupos de interés y deben considerar el largo plazo, construyendo relaciones basadas en la confianza y permitiendo el beneficio mutuo mientras dure esta relación (Olcese et al, 2008).

Los clientes son uno de los grupos de interés más importantes para las organizaciones dado que estas se deben a ellos y sin ellos no existirían y teniendo en cuenta la definición del mercadeo de la Asociación Americana de Mercadeo como “la actividad y los procesos para crear, comunicar, entregar e intercambiar productos y servicios que tengan valor para los clientes”, esto nos remite a la segmentación como una herramienta que permite adaptar, para cada segmento, la mejor oferta de producto o servicio, la política de precios, los canales de distribución y las demás variables comerciales, mejorando la relación con sus clientes y así conseguir una ventaja competitiva.

### **Marco Teórico**

Para la organización del marco teórico, se usó un árbol de la ciencia que permite identificar los documentos más citados y que proporcionan los conceptos teóricos base de la investigación (raíz), los documentos que proveen la evolución y la especificidad al tema de investigación (tronco) y los documentos que muestran las tendencias del tema de investigación (hojas) (Gil, Aguilera y Suárez, 2020). Para la presente investigación, los documentos que se encuentran en la raíz corresponden a los autores que han dado los conceptos base de la segmentación de clientes, sus categorías generales y sus ventajas y desventajas, los documentos del tronco, corresponden a las investigaciones que se han realizado sobre la segmentación de clientes en el sector bancario y en las hojas se muestran las investigaciones que marcan las tendencias actuales en el tema de investigación y corresponden a la recomendación de que las organizaciones del sector bancario empleen herramientas de la industria 4.0, como la minería de datos y el Big Data para segmentar a sus clientes.

El marco teórico de la investigación esta compuesto por la definición de la Industria 4.0, la identificación de la importancia del sector bancario, la definición de la segmentación de clientes y las investigaciones que se han realizado con respecto a esta en el sector bancario, el cual se desarrolla a continuación.

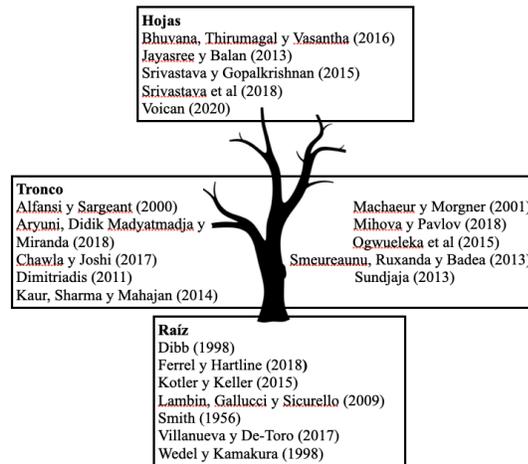


Figura 1. Árbol de la ciencia. Fuente: elaboración propia basado en artículos de Scopus.

## Industria 4.0

La producción industrial ha pasado por cuatro momentos importantes dependiendo de las fuentes de energía o tecnologías utilizadas. La primera revolución industrial comenzó al final del siglo dieciocho y se caracterizó por la introducción de sistemas de producción mecánicos que utilizaban el poder del vapor. Por otro lado, la segunda revolución industrial comenzó al final del siglo diecinueve y se distinguió por la producción en masa por medio del uso de la energía eléctrica. La tercera revolución industrial comenzó en la mitad del siglo veinte e introdujo la automatización y la microelectrónica a la producción. Por último, la cuarta revolución industrial o la Industria 4.0 se caracteriza por el aplicación de nuevas tecnologías como el Internet de las cosas, la computación en la nube y el uso de dispositivos inteligentes en las organizaciones y busca la digitalización de los ecosistemas industriales (Xu, Xu y Li, 2018).

La Industria 4.0 utiliza sensores inteligentes, la inteligencia artificial, la ciencia de datos y la analítica para analizar en tiempo real la información de los sistemas y los procesos de las organizaciones permitiendo la automatización y la optimización de muchas tareas (Xu et al, 2018).

En la actualidad y en la cuarta revolución industrial, la tecnología ha permeado todas las áreas de la sociedad, la producción, las finanzas, los servicios, el transporte y las comunicaciones. Estos desarrollos tecnológicos se basan en dispositivos y procesos capaces de transmitir y procesar grandes volúmenes de información (integración digital) y la disponibilidad de máquinas capaces de realizar tareas de complejidad media o alta (automatización) (Mehdiabadi et al, 2020).

La omnipresencia del Internet y de los teléfonos inteligentes junto con el ascenso de tecnologías como el Internet de las cosas, la biometría, el Big Data, la analítica avanzada, la inteligencia artificial y el Blockchain ha permitido a las organizaciones diseñar y desarrollar productos y servicios personalizados para sus clientes (Mehdiabadi et al, 2020).

### **Sector bancario**

El sector bancario representa un pilar fundamental del crecimiento económico y la estabilidad macroeconómica, especialmente en el contexto de la globalización. Sin embargo, la evolución del sector bancario en cada país es continuamente afectado por las dinámicas cambiantes del sistema financiero internacional. El establecimiento de un sector bancario sólido y eficiente es un prerrequisito importante para un crecimiento económico sostenible (Mehdiabadi et al, 2020).

El sector bancario es uno de los sectores que ha presentado un gran cambio debido a los avances tecnológicos derivados de la Industria 4.0. Algunas de las ventajas derivadas del uso de estos avances tecnológicos que pueden lograr las organizaciones del sector bancario son el mejoramiento de la productividad, la creación de productos y servicios innovadores, transacciones y transferencias rápidas, sistemas de información en tiempo real y el manejo eficiente del riesgo (Mehdiabadi et al, 2020).

Actualmente, una porción significativa de los clientes bancarios son jóvenes y jóvenes adultos que tienen expectativas y preferencias diferentes a la generación anterior. Cumplir estas expectativas y preferencias no es posible con los modelos de banco actuales y será posible únicamente con el uso de tecnologías, herramientas y procesos de la Industria 4.0 (Mehdiabadi et al, 2020).

De acuerdo con Berger (2017), la Industria 4.0 provee flexibilidad a los procesos de producción y así ayuda a crear productos y servicios específicos para un segmento objetivo satisfaciendo necesidades personalizadas a un costo marginal. Esto nos permite deducir qué el uso de las tecnologías de la Industria 4.0 les permitirá a las organizaciones del sector bancario segmentar de mejor manera a sus clientes.

### **Segmentación de clientes**

Los avances tecnológicos y en específico, los informáticos, que permiten el almacenamiento de la información de los clientes de las organizaciones han permitido comprender las necesidades de los clientes, sus deseos y las similitudes o diferencias entre grupos de clientes. Identificar estos grupos de clientes con características similares permite adaptar productos y servicios acorde a las necesidades de cada grupo facilitando el éxito de las organizaciones. La segmentación de clientes es la responsable de la increíble variedad de productos y servicios que disfrutamos actualmente e incluso los clientes esperan que las organizaciones profundicen en sus necesidades y en sus deseos, ajustando sus productos y servicios a estos. Esta variedad de ofertas era impensable hace 50 años, donde se ofrecía a los clientes una única oferta ya fuera en sabor o en estilo (Ferrel y Hartline, 2018).

Para Ferrel y Hartline (2018), la estrategia de segmentación de clientes utilizada por las organizaciones va a depender de su capacidad para identificar las características de sus clientes y de la información disponible sobre estos. Estas características pueden estar relacionadas con el uso del producto o servicio, la demografía de los clientes, sus estilos de vida o su ubicación geográfica y definen las cuatro categorías generales de segmentación para estos autores: conductual, demográfica, psicográfica y geográfica.

La segmentación conductual utiliza el comportamiento del cliente o el uso del producto o servicio para hacer distinciones entre segmentos. Estas distinciones están relacionadas con las razones por las que los clientes compran y usan los productos o servicios. La segmentación demográfica divide los clientes usando características como el género, la edad, el ingreso, la educación, la ocupación o el estado civil. La segmentación psicográfica aborda los motivos, las actitudes, las opiniones, los valores, los estilos de vida, los intereses y la personalidad de los clientes y por último, la segmentación geográfica

utiliza la ubicación del cliente. Vale la pena resaltar que estas cuatro categorías no son excluyentes y que las organizaciones pueden usar una combinación de las cuatro para segmentar a sus clientes (Ferrel y Hartline, 2018).

Para Kotler y Keller (2015), un segmento consiste de un grupo de clientes con necesidades y gustos similares y es tarea del administrador identificar el número y la naturaleza de estos segmentos. Estos autores concuerdan en las cuatro categorías generales de segmentación que puede usar una organización, conductual, demográfica, psicográfica y geográfica, e independientemente del tipo de segmentación utilizado, su recomendación es ajustar la estrategia para cada segmento. Adicionalmente, estos autores mencionan los posibles niveles de segmentación que puede usar una organización, pasando por no segmentar a los clientes, usar uno o varios segmentos, hasta tratar a cada cliente como un segmento, llegando a una personalización total de la oferta de sus productos o servicios.

En la investigación de Goyat (2011), se determina el origen del concepto de segmentación de clientes cerca del año 1950 en los estudios del profesor de mercadeo estadounidense Wendell R. Smith. Este autor concuerda en las cuatro categorías generales de segmentación que puede usar una organización, conductual, demográfica, psicográfica y geográfica y la elección de una categoría dependerá del tipo de producto, la naturaleza de la demanda, el método de distribución del producto, los medios disponibles para comunicar la oferta y las motivaciones de los clientes. Adicionalmente, para obtener una segmentación adecuada, los clientes del mismo segmento deben ser similares entre sí y distintos a los de los otros segmentos y deben responder de manera similar a las ofertas.

Para Lambin, Gallucci y Sicurello (2009), las organizaciones pueden implementar la segmentación de clientes de cuatro maneras distintas: descriptiva (características sociodemográficas y geográficas de los clientes), por beneficios buscados (considera explícitamente la categoría del producto y el sistema de valores del cliente), conductual (comportamiento de compra de los clientes) o por estilo de vida (características socioculturales de los clientes).

La segmentación descriptiva supone que los clientes que tienen distintos perfiles sociodemográficos también tienen diferentes necesidades y expectativas respecto de los productos y servicios. En este caso, las variables sociodemográficas permiten un acercamiento al conocimiento de las necesidades de los clientes. Esta segmentación se usa actualmente para conocer a los clientes, para seleccionar los medios que tienen mejor probabilidad de alcanzar a los clientes y para encontrar clientes potenciales de un nuevo producto o servicio. Las ventajas de esta segmentación son su bajo costo y su facilidad de aplicación mientras que entre sus desventajas se encuentra que no permite conocer las motivaciones de los clientes y que ha perdido importancia en países desarrollados, donde las personas han adoptado patrones de consumo similares sin importar su nivel adquisitivo (Lambin et al, 2009).

En la segmentación por beneficios buscados, el objetivo es identificar los beneficios que buscan los clientes al comprar un producto o servicio. Al aplicar esta segmentación se supone que los clientes con el mismo perfil sociodemográfico pueden tener sistemas de valores distintos y que los clientes pueden valorar de muy distinta manera los productos o servicios que usan. Para implementar este tipo de segmentación es necesario que la organización cuente con la información de los atributos o beneficios asociados con la categoría del producto o servicio y la importancia que le dan los clientes a cada beneficio. Las desventajas de esta segmentación son la dificultad de identificar los beneficios buscados por los clientes, el costo de recolectar información primaria para identificar estos beneficios, el recurrir a información sociodemográfica para poder interpretar mejor los resultados y que se debe usar alguna técnica estadística para agrupar a los clientes con valores similares (Lambin et al, 2009).

La segmentación conductual tiene como objetivo agrupar a los clientes de acuerdo a su comportamiento de compra de los productos o los servicios y los criterios más utilizados son el uso del producto o servicio, el volumen comprado o la frecuencia de uso y el nivel de lealtad. Para realizar este tipo de segmentación, la organización debe tener esta información en sus bases de datos (Lambin et al, 2009).

Por último, la segmentación por estilo de vida, incorpora elementos como actividades, actitudes, intereses, opiniones, percepciones y preferencias de los clientes, permitiendo un acercamiento a la motivación y la personalidad de estos. El estilo de vida

se refiere a la manera cómo los clientes viven, gastan su tiempo y dinero y puede medirse desde los rasgos de personalidad hasta en los productos y servicios que compran y en el modo en que los consumen o utilizan (Lambin et al, 2009).

Por otra parte, para Villanueva y De-Toro (2017) sólo existen dos categorías generales de segmentación que puede usar una organización, la primera relacionada con variables que describen a los clientes cómo individuos (edad, ingresos, gustos, etc.) y la segunda categoría, relacionada con variables que describen el uso histórico del producto o servicio por parte de los clientes o la manera en que lo han adquirido. Generalmente, las organizaciones deben emplear una combinación de las dos categorías de segmentación y la elección de las variables depende de la situación particular de cada organización y de la información disponible o qué se pueda obtener de los clientes. Para estos autores, segmentar a los clientes de manera creativa puede proporcionar ventajas competitivas a las organizaciones al permitir adaptar, para cada segmento, la mejor oferta de producto o servicio, la política de precios, los canales de distribución y las demás variables comerciales. Por último, recomiendan repetir el análisis de segmentación con cierta periodicidad dado que los grupos de clientes pueden variar con el tiempo.

Para finalizar con aspectos sobre la aplicación de la segmentación de clientes, en la investigación de Dibb (1998) se abordan los problemas con los que se encuentran las organizaciones al usar la segmentación de clientes y cómo se pueden evitar, encontrando que el principal problema es la falta de entendimiento del concepto de segmentación y que el compromiso de la alta gerencia es vital en la implementación de la segmentación. Los autores proponen que se debe seguir investigando sobre los factores que hacen que una segmentación sea exitosa y la manera de cuantificar este éxito.

### **Segmentación de clientes en el sector bancario**

Al revisar la literatura sobre la segmentación de clientes en las organizaciones del sector bancario, se encontró que las investigaciones se han centrado en cinco temas: la recomendación del uso de herramientas como la minería de datos y el Big Data para la segmentación de clientes (Jayasree y Balan, Voican, Srivastava y Gopalkrishnan, Srivastava et al y Bhuvana, Thirumagal y Vasantha), segmentar a los clientes según una variable de interés para la organización (Ogwueleka et al y Smeureaunu, Ruxanda y

Badea), segmentar a los clientes usando información del uso de sus canales (Aryuni, Didik Madyatmadja y Miranda y Chawla y Joshi), segmentar usando la tenencia y el uso de ciertos productos y servicios (Mihova y Pavlov y Sundjaja) y por último, segmentar según las percepciones que tienen los clientes con respecto a la relación con su banco (Dimitriadis, Kaur, Sharma y Mahajan, Alfansi y Sargeant y Machauer y Morgner). A continuación se resumen las investigaciones en cada uno de los temas encontrados.

### **Recomendación del uso de la minería de datos y el Big Data para la segmentación de clientes en el sector bancario**

La minería de datos es el proceso de extraer información oculta, desconocida, válida y accionable de grandes bases de datos que puede ser usada en la toma de decisiones de negocio. Esta extracción se logra por medio de la aplicación de técnicas estadísticas y de aprendizaje automático (Jayasree y Balan, 2013).

Según la investigación de Jayasree y Balan (2013), la información se puede convertir en el activo más valioso de las organizaciones del sector bancario si estas logran extraer conocimiento de sus datos, lo cual pueden lograr con el uso de la minería de datos. Según los autores, las organizaciones del sector bancario han utilizado la minería de datos especialmente para segmentar a sus clientes, realizar predicciones del valor de sus inversiones, detectar transacciones fraudulentas, manejar el riesgo y para predecir los clientes que van a dejar de pagar sus deudas. En referencia a la segmentación, la minería de datos ha permitido a las organizaciones del sector bancario encontrar grupos de clientes dependiendo de sus necesidades y preferencias, les ha permitido detectar diferencias entre los clientes interesados sólo en los servicios digitales y los que prefieren ir a las oficinas y les ha permitido encontrar grupos de clientes que sólo tienen un producto a los cuales se les puede ofrecer uno complementario.

Por su parte, según la investigación de Voican (2020), los bancos han utilizado la minería de datos para predecir fraude en sus tarjetas de crédito y para segmentar a sus clientes para identificar grupos de clientes más propensos a aceptar ofertas de productos y servicios permitiendo la retención de los clientes más rentables. El autor indica que las organizaciones del sector bancario fueron de las primeras en utilizar la minería de datos,

principalmente para conocer qué motiva a sus clientes, analizar cómo pueden ser retenidos, mejorando así la rentabilidad, en la determinación del otorgamiento o no de préstamos a sus clientes y en el mercadeo de productos y servicios. Dependiendo de la información de la que disponga la organización de sus clientes y utilizando la minería de datos, las organizaciones pueden segmentar a sus clientes por su valor para la organización, por la tenencia y uso de productos y servicios, por su probabilidad de adquirir productos o servicios, por su lealtad, por sus características sociodemográficas y por sus necesidades y actitudes. Esta investigación recalca la importancia de la aplicación de la minería de datos en el sector bancario pero no realizó una segmentación de clientes de las organizaciones de este sector.

El Big Data es definido como una herramienta del siglo veintiuno que permite a las organizaciones crear, manipular, manejar y almacenar grandes volúmenes de información que no podrían ser manejadas con las herramientas y técnicas tradicionales. Esta información se caracteriza por su variedad, volumen y velocidad (Srivastava y Gopalkrishnan, 2015).

El incremento de la población mundial ha aumentado la cantidad de clientes y de transacciones de las organizaciones del sector bancario y la digitalización del mundo ha creado grandes volúmenes de información haciendo que los bancos de Estados Unidos y de otros países recurran a la analítica del Big Data para manejar esta situación y esto les ha permitido encontrar patrones en la información y aumentar la rentabilidad de sus organizaciones (Srivastava et al, 2018).

Según Srivastava y Gopalkrishnan (2015), las organizaciones del sector bancario internacional están utilizando el Big Data de manera satisfactoria para conocer los patrones de consumo de sus clientes, conocer los canales por los que usan sus productos y servicios, conocer los sentimientos de sus clientes, segmentarlos, generar ofertas y combatir el fraude. En su investigación, por medio del análisis de la información de una muestra de clientes de un banco de la India, donde la información del banco se mantuvo anónima para preservar la confidencialidad, los autores realizaron ejemplos de análisis que los bancos pueden obtener utilizando el Big Data. Esta investigación recomienda extender el estudio para cuantificar los beneficios financieros y no financieros obtenidos por el banco

analizado al usar Big Data y cubrir las diferentes técnicas de minería de datos que los bancos pueden usar para mejorar la calidad de los análisis.

La investigación de Srivastava et al (2018) encuentra que los bancos de la India no han aprovechado los grandes volúmenes de información que tienen de sus clientes debido a problemas como la conectividad. Los autores revisan la conveniencia de aplicar la analítica del Big Data en el sector bancario de la India dado que les permitirá a las organizaciones de este sector extraer conocimiento que se convertirá en acciones para la toma de decisiones estratégicas y estar por encima de la competencia con el uso de esta tecnología. Dentro de los análisis que los bancos pueden realizar con la información, los autores mencionan la segmentación de clientes, lo que permite dividir la base de clientes en grupos de personas que son similares en ciertas características que son importantes para las áreas de negocio y de mercadeo como la edad, el género, las condiciones financieras, los intereses y los hábitos de consumo, lo que permitirá a los bancos entregar a sus clientes exactamente lo que sus clientes están buscando. Adicionalmente, mencionan otros análisis de utilidad para los bancos como la detección del fraude y el manejo del riesgo.

Por último, según Bhuvana, Thirumagal y Vasantha (2016), las organizaciones del sector bancario utilizan el Big Data tanto para el beneficio del banco como el de sus clientes. Esta investigación analiza la manera en que los bancos de la India han utilizado esta tecnología, los factores que tienen mayor impacto para los bancos al manejar grandes volúmenes de información y como la analítica crea valor para estas organizaciones. El estudio encontró que los factores que tiene mayor impacto para la adopción del Big Data en los bancos de ese país son el manejo del riesgo, la detección de fraude y la segmentación de clientes y que los bancos de la India deben actualizar su software y adoptar nuevas tecnologías para poder manejar grandes volúmenes de información.

### **Segmentación de clientes en el sector bancario según una variable de interés**

En la investigación de Ogwueleka et al (2015) se recalca la importancia de la segmentación de clientes para la toma de decisiones y como una herramienta para satisfacer los deseos de los clientes y aumentar su satisfacción. En esta investigación, los autores primero segmentaron a los clientes del producto tarjeta de crédito del

Intercontinental Bank Plc con sede en Nigeria usando la frecuencia de sus compras, su monto y la variación del monto en un periodo de tiempo y después obtuvieron un modelo estadístico que clasifica a los clientes como leales ó no leales utilizando redes neuronales con información demográfica, transaccional y monetaria de los clientes. Esta investigación advierte que sus resultados no pueden ser extendidos a clientes de otros productos o servicios del banco analizado ni a canales transaccionales como el Internet dado que sólo analizaron las compras realizadas de manera presencial y sugieren seguir realizando estudios que clasifiquen a los clientes de las organizaciones del sector bancario.

Por su parte, la investigación de Smeureaunu, Ruxanda y Badea (2013) recalca la importancia de la segmentación de clientes para el sector bancario privado para lograr el desarrollo de negocios rentables permitiendo a las organizaciones de este sector dirigir sus productos y servicios a grupos de clientes similares y utiliza dos técnicas populares de aprendizaje automático, redes neuronales y maquinas de vector soporte. Según los autores, una cuidadosa selección de las variables y la técnica de segmentación estadística usada son requisitos para que el proceso de segmentación de clientes traiga beneficios a la organización. La investigación analizó una muestra de clientes con tarjeta de crédito de un banco de Rumania y por medio del análisis transaccional de los últimos seis meses y del domicilio del cliente diferenciaron a los clientes en dos grupos, los clientes con mayor capacidad adquisitiva y los de menor capacidad adquisitiva. Esta investigación recomienda extender el estudio comparando los resultados obtenidos con los que se pueden obtener usando otras técnicas estadísticas.

### **Segmentación de clientes en el sector bancario según el uso de canales**

La investigación de Aryuni, Didik Madyatmadja y Miranda (2018) utiliza técnicas de segmentación estadística como K-medias y K-mediodes para encontrar grupos de clientes basados en la información del uso de la banca digital (Internet) analizada previamente con un modelo de recencia, frecuencia y monto (RFM) de un año de información de los clientes de un banco de Indonesia, donde la información del banco se mantuvo anónima para preservar la confidencialidad. Los autores explican que con los resultados de la segmentación, el banco puede aplicar estrategias de relacionamiento y mercadeo para cada grupo. Esta investigación recomienda extender el estudio, no sólo utilizando

información de la banca digital, sino también información sociodemográfica de los clientes y tenencia de productos y servicios.

Por su parte, en la investigación de Chawla y Joshi (2017) se analizaron factores como la facilidad de uso, la conveniencia, la eficiencia, la confianza y el estilo de vida que influyen la adopción de la banca digital por parte de los clientes de bancos privados que viven en grandes ciudades de la India, con un nivel educativo alto, jóvenes, empleados y estos factores se usaron para segmentar los clientes usando las técnicas de segmentación estadística K-medias y segmentación jerárquica. Cada uno de los grupos de clientes encontrados fue posteriormente caracterizado demográficamente, por los servicios de la banca móvil usados y por las actitudes e intenciones hacia estos servicios. Esta investigación recomienda extender el estudio a clientes con un perfil distinto al analizado, incluir clientes de bancos públicos, analizar clientes de otros países tanto en desarrollo como desarrollados y comparar los factores asociados a la satisfacción y a la confianza de los clientes en los diferentes canales, no sólo la banca móvil.

### **Segmentación de clientes en el sector bancario según el uso de productos**

Según Mihova y Pavlov (2018) una de las mayores prioridades de las organizaciones del sector bancario es mejorar su segmentación de clientes y tenerla en cuenta en el diseño y distribución de nuevos productos y servicios. Una estrategia común para mejorar la competitividad es el diseño de productos y servicios específicos dirigidos a clientes leales o la oferta de descuentos a estos clientes. Esto se conoce como el programa de lealtad e incluye la entrega de cierto tipo de tarjeta a estos clientes. Esta investigación analizó el uso de préstamos bancarios y en particular variables como el monto del préstamo, la antigüedad en el banco y el número de pagos incumplidos en los últimos 12 meses, en una muestra de clientes de un banco y utilizó la técnica de segmentación estadística K-medias. Esta investigación advierte que la segmentación de clientes no es una tarea de una sola vez sino que debe ser una tarea continua para poder detectar los cambios que se presenten en el mercado y en las necesidades y en los comportamientos de los clientes.

La investigación de Sundjaja (2013) discute el uso de la segmentación demográfica, basada en la profesión de los clientes con cuentas de ahorro de un banco de Indonesia,

donde la información del banco se mantuvo anónima para preservar la confidencialidad. Cada segmento fue caracterizado según el saldo en sus cuentas de ahorro, su transacción más realizada y el canal preferido para realizar transacciones. Esta investigación advierte que existen otras técnicas de segmentación más avanzadas que la segmentación demográfica y sugiere analizar los clientes de otros productos y comparar los resultados que se puedan obtener con los clientes de otros bancos de Indonesia.

### **Segmentación de clientes en el sector bancario según los beneficios buscados**

La investigación de Dimitriadis (2011) usó la técnica de segmentación estadística K-medias en una muestra de clientes de los cuatro principales bancos de Grecia analizando variables relacionadas con el costo y el beneficio percibidos por los clientes en su relación con su banco y estableciendo la relevancia de usar estas variables para la segmentación de clientes. El autor demostró que los grupos encontrados difieren por los beneficios y los costos analizados además de las características demográficas, psicográficas y de comportamiento de los clientes (antigüedad en el banco, número de productos con el banco y frecuencia de interacción con el banco). Esta investigación advierte que los resultados no se pueden extender a otros bancos o países dado que la percepción de los costos y los beneficios es subjetivo por parte de los clientes.

La investigación de Kaur, Sharma, y Mahajan (2014) intenta predecir los clientes que van a dejar su banco actual en la ciudad de Jammu en la India por medio del análisis de variables como la calidad, el costo, la satisfacción, la confianza, el servicio, la lealtad, los costos y las barreras de salida. Se utilizó la técnica de segmentación estadística K-medias a las respuestas a una encuesta telefónica realizada al azar a clientes de los bancos en la ciudad mencionada. Los autores afirman que encontrar las características de los clientes que pueden dejar el banco facilita a los gerentes formular diferentes estrategias de retención, buscando retener a los clientes más leales. Esta investigación advierte que los resultados de este estudio no se pueden generalizar a clientes fuera de la ciudad de Jammu de la India.

Por su parte, Alfansi y Sargeant (2000) exploran la utilización de la segmentación por beneficios buscados en las organizaciones bancarias de Indonesia para retener a sus clientes. En esta investigación, a una muestra de mil clientes bancarios de la ciudad de

Bengkulu en Indonesia se le aplicó la técnica de segmentación estadística K-medias utilizando información de los beneficios que buscaban en sus bancos. Esta investigación concluyó que las ventajas para las organizaciones bancarias de Indonesia al usar la segmentación por beneficios buscados son limitadas dado que no se encontró una relación entre los segmentos y variables demográficas, posiblemente debido al tamaño de la muestra usado, por lo que los autores proponen realizar ejercicios similares con tamaños de muestra mayores.

Por último, en la investigación de Machauer y Morgner (2001), se entrevistaron a doscientos ochenta y cinco estudiantes de la Universidad de Mannheim en Alemania para conocer sus actitudes sobre la banca digital y los beneficios esperados en la relación con su banco y posteriormente se utilizó la técnica estadística de segmentación jerárquica para encontrar los grupos de clientes dependiendo de sus respuestas a la entrevista. Como conclusión, se encontró que la segmentación por beneficios buscados ofreció mayor información que las variables demográficas, sin embargo, el tamaño de la muestra y el perfil de los entrevistados no permiten que los resultados puedan ser extendidos a otros países o perfiles de clientes.

### **Hipótesis**

La pregunta de investigación planteada, ¿qué tipos de segmentos se pueden obtener usando información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana?, permitirá a las organizaciones del sector bancario colombiano tener una visión más completa de los clientes persona natural, posibilitándoles tomar mejores decisiones enfocadas al perfilamiento del cliente, a la oferta de productos y servicios, a la definición de precios y canales y otras variables comerciales, consiguiendo fidelizar y retener al cliente e incluso disminuir el riesgo.

Al analizar las iniciativas de segmentación de clientes en otros países e instituciones del sector bancario se encontró que estas se han centrado en la recomendación del uso de herramientas como la minería de datos y el Big Data para la segmentación de clientes, en la segmentación de clientes según una variable de interés para la organización, según el uso de sus canales, según la tenencia y el uso de ciertos productos y servicios y según los beneficios buscados por los clientes en la relación con su banco y se encontró que al utilizar herramientas de minería de datos y Big Data les ha permitido a las organizaciones del sector bancario estar a la vanguardia tecnológica, diferenciarse de la competencia, innovar, mejorar la competitividad y la rentabilidad al igual que lograr una mejor relación con sus clientes.

## **Metodología**

A continuación se especifica el enfoque y el alcance de la investigación, la población y la muestra de interés, los instrumentos de recolección de la información y las técnicas para el análisis de la información para llevar a cabo la investigación propuesta en el trabajo de grado.

### **Enfoque y alcance de la investigación**

Según Hernández-Sampieri y Mendoza (2018, p.4), la investigación se define como un “conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de un fenómeno o problema con el resultado (o el objetivo) de ampliar su conocimiento”, existiendo tres enfoques para investigar, el cuantitativo, el cualitativo y el mixto.

El enfoque cuantitativo se debe usar cuando se quiere estimar la magnitud o la ocurrencia de un fenómeno, el cualitativo cuando se examina el fenómeno y se revisan los estudios previos de manera simultánea con el objetivo de generar una teoría que sea consistente con lo que se esta observando y el mixto cuando se usan los enfoques anteriores en su forma pura o una adaptación, alteración o síntesis de ambos (Hernández-Sampieri y Mendoza, 2018).

La pregunta de investigación planteada, ¿qué tipos de segmentos se pueden obtener usando información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana? se enmarca dentro del enfoque cuantitativo dado que se quiere comparar distintos grupos de clientes usando variables conductuales y de beneficios buscados de los clientes persona natural.

El diseño de la investigación planteada es descriptiva dado que se quieren describir las características de los segmentos de clientes persona natural que pueden obtener las organizaciones del sector bancario colombiano al usar información conductual y por beneficios buscados de sus clientes y no se busca establecer relaciones o medir la relación entre variables en un contexto, características de la investigación correlacional. Adicionalmente, no se manipularán variables como en la realización de un experimento y

la recolección de la información se realizará de manera transversal, es decir en un solo momento.

### **Población y muestra**

La población objeto de estudio son los clientes persona natural de una organización del sector bancario colombiano. Se simularán variables relacionadas con información conductual y por beneficios buscados de los clientes persona natural para una muestra de quinientos clientes. Este tamaño de muestra se basa en los tamaños de muestra usados en las investigaciones analizadas en el marco teórico y en el tamaño de muestra sugerido para estimar proporciones en poblaciones de gran tamaño con un nivel de confianza del 95% y un margen de error del 5%.

### **Instrumentos**

Para la recolección de la información se simularán variables relacionadas con información conductual y por beneficios buscados de clientes persona natural para una muestra de quinientos clientes persona natural de una organización del sector bancario colombiano. Cada una de las variables será simulada usando una distribución probabilística teniendo en cuenta los valores que puede tomar cada una de las variables. La información propuesta debe ser simulada dada la sensibilidad de estas organizaciones con respecto a la información de sus clientes.

### **Técnicas para el análisis de la información**

La información simulada de la muestra de clientes persona natural de la banca colombiana se segmentará usando un análisis de conglomerados por medio el procedimiento estadístico K-medias. La información se estandarizará para que las variables sean analizadas en la misma escala. El análisis de conglomerados es un análisis de estadística multivariada que busca agrupar objetos basados en información que describe estos objetos. El objetivo de este análisis es que los objetos dentro de un grupo sean similares entre sí y distintos a los objetos de los demás grupos. Entre mayor sea la similitud dentro de un grupo y entre mayor sea la diferencia entre los grupos, los resultados

del análisis serán mejores. La técnica estadística K-medias es una de las técnicas más importantes y más usadas dentro del análisis de conglomerados y busca encontrar un número específico de grupos ( $k$ ) representados por el centroide o promedio de los objetos dentro de cada grupo (Tan, Steinbach y Kumar, 2005).

La técnica estadística K-medias consiste en particionar un conjunto de  $n$  objetos en  $k$  grupos. Siendo  $X_{i,j}$  el valor del  $i$ -ésimo objeto en la  $j$ -ésima variable, con  $i = 1, \dots, n$  y  $j = 1, \dots, p$ . El promedio de la  $j$ -ésima variable en el  $g$ -ésimo grupo se denotará por  $\bar{X}_{(g)j}$ ,  $g = 1, \dots, k$  y  $n_{(g)}$  será el número de objetos en el  $g$ -ésimo grupo. La distancia de un objeto a un grupo será

$$D_{(i,g)} = \left( \sum_{j=1}^p (X_{i,j} - \bar{X}_{(g)j})^2 \right)^{1/2}.$$

Se define el componente de error de la partición como

$$\varepsilon\{\rho(n, K)\} = \sum_{i=1}^n [D(i, g(i))]^2,$$

Donde  $g(i)$  es el grupo que contiene al  $i$ -ésimo objeto, y  $D(i, g(i))$  es la distancia euclidiana entre el objeto  $i$  y el centroide o promedio del grupo que contiene al individuo. La técnica consiste en encontrar la partición con el error  $\varepsilon$  más pequeño, moviendo los objetos de un grupo a otro hasta que se establezca la reducción de  $\varepsilon$ . En resumen, se trata de reubicar los objetos, de manera que se logren grupos con la menor variabilidad posible.

En otras palabras, el algoritmo funciona de la siguiente manera:

1. Seleccionar aleatoriamente  $k$  objetos como los centroides o promedios iniciales. Siendo  $k$  el número de grupos deseados.
2. Cada objeto es asignado al centroide más cercano. Los objetos asignados a un centroide conforman un grupo.
3. El centroide de cada grupo es actualizado basado en los objetos asignados al grupo.
4. Se repiten los pasos 2 y 3 hasta que ningún objeto cambie de grupo o equivalentemente, hasta que los centroides no cambien.

Cuando el algoritmo termina, dado que no ocurren más cambios en la asignación de los objetos a los grupos, los centroides han identificado los segmentos naturales de la información analizada. Para asignar un objeto al centroide más cercano, se necesita de una medida de proximidad que cuantifique el concepto de cercanía. Generalmente se usa la distancia euclídea pero se pueden usar otras medidas de distancia.

Para a la selección de número de segmentos,  $k$ , se calculará la suma de cuadrados del error ( $SSE = \sum_{g=1}^k \sum_{x \in G_i} (g_i - x)^2$  donde  $G_i$  es el  $i$ -ésimo segmento,  $x$  es un objeto en  $G_i$  y  $g_i$  es el promedio del  $i$ -ésimo segmento) usando un número de posibles segmentos de dos a quince y se seleccionará el número de segmentos en el que exista una disminución significativa de esta medida (Tan, Steinbach y Kumar, 2005).

Para el análisis de la información se utilizará el programa estadístico R versión 4.1.0, programa en el que se puede usar el procedimiento estadístico K-medias.

## Variables

A continuación se definen las variables, tanto conceptual como operacionalmente, que se tendrán en cuenta para segmentar a los clientes persona natural de la banca colombiana usando información conductual y por beneficios buscados. Cabe la pena mencionar que todas las variables se clasifican como independientes para esta investigación y que su selección se basó en las variables utilizadas en las investigaciones analizadas en el marco teórico y en la experiencia del investigador. Adicionalmente, la selección de los parámetros de las distribuciones se basó en la experiencia laboral del investigador en el sector bancario colombiano y guardan relación con lo que se puede encontrar en una organización de este sector y país. Por último, todas las variables propuestas deben ser analizadas como variables cuantitativas y así poder ser analizadas por medio de la técnica estadística K-medias.

Las variables de tenencia de productos fueron simuladas usando una distribución binomial, cuya función de densidad esta dada por  $p(x) = \binom{n}{x} p^x (1-p)^{(n-x)}$  siendo  $n$  el tamaño de la muestra,  $p$  la probabilidad de tener el producto y  $x = 0, \dots, n$  y las variables de saldo fueron simuladas usando una distribución log-normal cuya función de densidad esta dada por  $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$  donde  $\mu$  y  $\sigma$  son la media y la desviación estándar del logaritmo. Para obtener una distribución con una media deseada  $\mu_X$  y varianza  $\sigma_X^2$ , se utiliza  $\mu = \ln\left(\frac{\mu_X^2}{\sqrt{\mu_X^2 + \sigma_X^2}}\right)$  y  $\sigma^2 = \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)$  (Blanco, 2004).

Las variables relacionadas con la calificación de los costos, de los servicios y la recomendación del banco fueron identificadas en investigaciones como las de Dimitriadis (2011) y Kaur et al (2014) que tratan de la segmentación de clientes en el sector bancario según los beneficios buscados por los clientes en la relación con su banco. Estas variables fueron simuladas usando una distribución uniforme cuya función de densidad es  $f(x) = \frac{1}{(max-min)}$  con  $min \leq x \leq max$  (Blanco, 2004).

- Tenencia de cuenta de ahorros. Posibles valores 1 (tiene) o 0 (no tiene). Simulada con una distribución binomial con una probabilidad de 90%. Indica si el cliente tiene o no cuenta de ahorros en el banco.
- Saldo cuenta de ahorros. Posibles valores entre \$0 y \$7.000.000. Simulada con una distribución log-normal con media \$2.000.000 y desviación estándar \$1.000.000. Indica el saldo del cliente en su cuenta de ahorros.
- Tenencia de CDT. Posibles valores 1 (tiene) o 0 (no tiene). Simulada con una distribución binomial con una probabilidad de 30%. Indica si el cliente tiene o no un CDT en el banco.
- Saldo CDT. Posibles valores entre \$0 y \$16.000.000. Simulada con una distribución log-normal con media \$5.000.000 y desviación estándar \$3.000.000. Indica el saldo del cliente en su CDT.
- Tenencia de tarjeta de crédito. Posibles valores 1 (tiene) o 0 (no tiene). Simulada con una distribución binomial con una probabilidad de 70%. Indica si el cliente tiene o no tarjeta de crédito en el banco.
- Saldo tarjeta de crédito. Posibles valores entre \$0 y \$47.000.000. Simulada con una distribución log-normal con media \$10.000.000 y desviación estándar \$6.000.000. Indica el saldo del cliente en su tarjeta de crédito.
- Tenencia de crédito. Posibles valores 1 (tiene) o 0 (no tiene). Simulada con una distribución binomial con una probabilidad de 50%. Indica si el cliente tiene o no un crédito en el banco.
- Saldo crédito. Posibles valores entre \$0 y \$203.000.000. Simulada con una distribución log-normal con media \$50.000.000 y desviación estándar \$30.000.000. Indica el saldo del cliente en su crédito.
- Cliente App. Posibles valores 1 (usuario) o 0 (no usuario). Simulada con una distribución binomial con una probabilidad de 80%. Indica si el cliente es usuario de la aplicación móvil del banco.
- Cliente Web. Posibles valores 1 (usuario) o 0 (no usuario). Simulada con una distribución binomial con una probabilidad de 90%. Indica si el cliente es usuario de la página Web del banco.
- Cliente oficina. Posibles valores 1 (usuario) o 0 (no usuario). Simulada con una distribución binomial con una probabilidad de 50%. Indica si el cliente es usuario de las oficinas del banco.

- Calificación de los costos. Posibles valores entre 1 (muy bajo) y 10 (muy alto). Simulada con una distribución uniforme con valores entre 1 y 10. Indica la percepción del cliente en relación a los costos por los servicios ofrecidos por el banco.
- Calificación de los servicios. Posibles valores entre 1 (muy insatisfecho) y 10 (muy satisfecho). Simulada con una distribución uniforme con valores entre 1 y 10. Indica la percepción del cliente en relación con la satisfacción con los servicios ofrecidos por el banco.
- Recomendación del banco. Posibles valores entre 1 (no lo recomendaría en lo absoluto) y 10 (lo recomendaría totalmente). Simulada con una distribución uniforme con valores entre 1 y 10. Indica el nivel de recomendación del cliente del banco a otros usuarios.

## **Trabajo de Campo**

El trabajo de campo consistió inicialmente en la simulación, según las distribuciones definidas en el capítulo anterior, de las catorce variables relacionadas con la tenencia de productos, el saldo de estos productos, el uso de canales y las calificaciones de los costos, de los servicios y la recomendación del banco para una muestra de quinientos clientes persona natural de una organización del sector bancario colombiano. Posteriormente, se procesó estadísticamente la información simulada por medio de estadísticas descriptivas de cada una de las variables, a continuación se procedió con la selección del número de segmentos según el procedimiento estadístico K-medias, por medio de la obtención de la suma de cuadrados del error, para un número de posibles segmentos entre dos y quince y por último, la descripción de los segmentos obtenidos con el número de grupos definido en el análisis anterior. Estos resultados estadísticos se explican a continuación.

### **Procesamiento de los datos**

El primer paso consistió en la simulación de la base de datos de quinientos clientes persona natural de una organización del sector bancario colombiano con la tenencia de los productos bancarios cuenta de ahorros, CDT, tarjeta de crédito, crédito, los saldos de cada uno de estos productos, el uso de los canales aplicación móvil, página web y oficinas de servicios, la calificación de los costos, de los servicios y la recomendación del banco según las distribuciones definidas previamente. Una muestra de los diez primeros registros de la base de datos se presenta en la figura 2.

	tenencia_cuenta_ahorros	saldo_cuenta_ahorros	tenencia_cdt	saldo_cdt	tenencia_tarjeta_credito	saldo_tarjeta_credito	tenencia_credito	saldo_credito	cliente_app	cliente_web	cliente_oficina
1	1	1873854	0	0	1	7147036	0	0	1	1	0
2	1	1987355	0	0	1	3529790	0	0	1	1	0
3	1	2087057	0	0	1	6317767	1	59032021	1	1	1
4	1	523502	0	0	0	0	0	0	1	1	0
5	0	0	0	0	1	5649642	1	39057092	1	1	1
6	1	2358856	0	0	1	8259010	1	59014736	1	1	0
7	1	3011257	1	8981271	1	6876250	1	169940954	1	1	1
8	1	3745598	1	2179696	1	5863543	1	42116525	0	1	0
9	1	2707137	0	0	1	17820916	1	27368185	1	1	0
10	1	2888866	0	0	1	3068058	0	0	0	1	0
11	0	0	1	2723291	0	0	0	0	1	1	1
12	1	1381709	0	0	1	7799123	1	41845309	1	0	0
13	1	1188829	1	8039120	1	18987213	1	25522836	1	0	0
14	1	3870532	0	0	1	8297769	1	54584918	1	1	1
15	1	680615	1	4854775	1	15218245	0	0	1	0	0
16	1	2019502	0	0	0	0	0	0	1	1	0
17	1	1822961	0	0	1	5393235	0	0	1	1	1
18	1	2206170	0	0	1	10586092	0	0	0	1	1
19	1	2661132	0	0	0	0	1	70285689	1	1	0
20	0	0	0	0	1	4439426	0	0	1	1	1
21	1	2609242	1	3382028	1	16809147	1	18055595	1	1	1
22	1	2162837	0	0	1	9641293	0	0	0	1	1
23	1	1735131	0	0	1	3561485	0	0	1	1	1
24	0	0	1	7007267	1	6107028	0	0	1	1	1
25	1	2152401	0	0	1	12015596	0	0	1	1	0

Figura 2. Muestra base de datos. Fuente: elaboración propia en R Studio.

Posteriormente, cada una de las variables fue analizada estadísticamente. Las variables relacionadas con la tenencia de productos fueron analizadas por medio del promedio o media, estadística que en este caso representa la proporción de tenencia de cada producto en los clientes analizados, donde el 90% de los clientes tiene cuenta de ahorros, el 30% cuenta con CDT, el 72% tiene tarjeta de crédito y el 50% de los clientes tiene un crédito. En la tabla 1 se muestran estas proporciones junto con la cantidad de clientes que tiene cada producto.

Tabla 1. Tenencia de productos

Producto	Clientes	% Clientes
Cuenta de ahorros	452	90%
CDT	148	30%
Tarjeta de crédito	358	72%
Crédito	245	50%

Fuente: elaboración propia.

Las variables relacionadas con el saldo de cada producto fueron analizadas por medio del saldo mínimo, la mediana del saldo, el promedio o media del saldo, el saldo máximo y la desviación estándar del saldo. Estos resultados estadísticos se muestran en la tabla 2.

Tabla 2. Saldo de los productos

Producto	Mínimo	Mediana	Media	Máximo	Desviación estándar
Cuenta de ahorros	\$0	\$1.750.996	\$1.848.210	\$6.937.840	\$1.126.410
CDT	\$0	\$0	\$1.397.905	\$15.682.172	\$2.639.052
Tarjeta de crédito	\$0	\$6.342.904	\$6.886.990	\$46.149.780	\$6.451.324
Crédito	\$0	\$0	\$25.271.651	\$202.778.730	\$34.178.189

Fuente: elaboración propia.

Las variables relacionadas con el uso de los canales también fueron analizadas por medio del promedio o media, estadística que en este caso representa la proporción de uso de cada canal por parte de los clientes analizados, donde el 80% de los clientes usa la aplicación móvil del banco, el 89% usa la página Web del banco y el 48% usa las oficinas de servicio del banco. En la tabla 3 se muestran estas proporciones junto con la cantidad de clientes que usa cada canal.

Tabla 3. Uso de canales

Canal	Clientes	% Clientes
App	396	80%
Web	443	89%
Oficina	241	48%

Fuente: elaboración propia.

Por último, las variables relacionadas con la calificación de los costos, de los servicios y la recomendación del banco también fueron analizadas por medio de la calificación mínima, la mediana de las calificaciones, el promedio o media de las calificaciones, la calificación máxima y la desviación estándar de las calificaciones. Estos resultados estadísticos se muestran en la tabla 4.

Tabla 4. Calificaciones costos, servicios y recomendación

Calificaciones	Mínimo	Mediana	Media	Máximo	Desviación estándar
Calificación de los costos	1,0	6,0	5,6	10,0	2,62
Calificación de los servicios	1,0	5,5	5,6	10,0	2,61
Recomendación del banco	1,0	5,0	5,3	10,0	2,65

Fuente: elaboración propia.

### Análisis de resultados

Después de analizar estadísticamente cada una de las variables de la base de datos, se procedió con la selección del número de segmentos según el procedimiento estadístico K-medias por medio de la obtención de la suma de cuadrados del error para un número de posibles segmentos entre dos y quince y seleccionado el número de segmentos en el que exista una disminución significativa de esta medida (Tan et al, 2005). En la figura 3 se muestra la suma de cuadrados del error en función del número de segmentos. Se observa que en cinco segmentos la suma de cuadrados del error presenta la mayor disminución de la suma de cuadrados del error y con números mayores de segmentos la disminución de esta medida no es tan significativa.

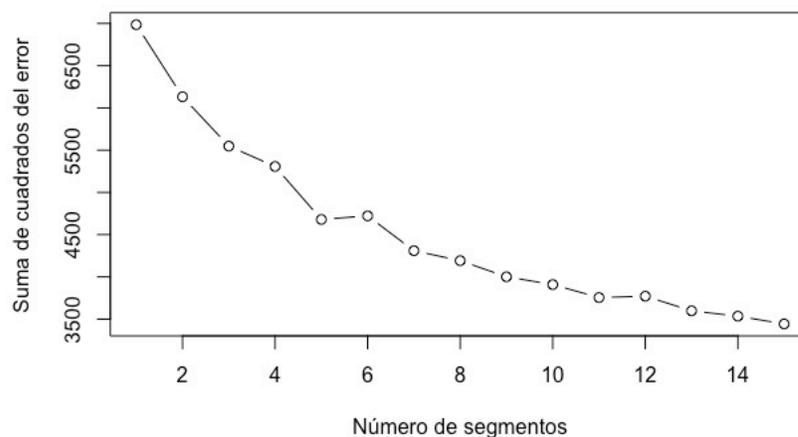


Figura 3. Selección número de segmentos.

Fuente: elaboración propia en R Studio.

Este resultado permite dar respuesta al objetivo de evaluar la cantidad de segmentos de clientes persona natural obtenidos al usar información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana, que en el caso de la información analizada se puede concluir que se obtienen cinco segmentos.

Con el número de segmentos identificado, se procedió a ejecutar el procedimiento estadístico K-medias con esta cantidad de segmentos y así poder caracterizar cada uno de los cinco segmentos según el promedio de cada una de las variables tenidas en cuenta y dar respuesta al objetivo de conocer las características de los segmentos de clientes persona natural de la banca colombiana usando información conductual y por beneficios buscados. El detalle de la caracterización de cada uno de los segmentos en las variables analizadas se encuentra en el anexo A y de manera gráfica en las figuras 4 a 8. El análisis de estos resultados se presenta a continuación.

En la figura 4 se muestra la proporción de clientes en cada uno de los segmentos obtenidos. El segmento más grande es el 1 con el 33% de los clientes, seguido por los segmentos 2 y 5 con el 21% de los clientes, el segmento 4 con 14% de los clientes y el segmento 3 con el 11% de los clientes.

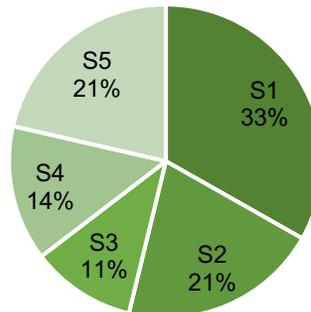


Figura 4. Tamaño de los segmentos.

Fuente: elaboración propia.

En la figura 5 se muestra la proporción de tenencia de cada uno de los productos para los segmentos obtenidos. Se resalta que la tenencia de CDT en segmento 1 es 0% mientras que el segmento 2 es 100%, la tenencia de tarjeta de crédito es de 0% en el

segmento 4 mientras que en el segmento 5 es 100% y la tenencia de crédito es del 0% en el segmento 1, del 100% en los segmentos 4 y 5 y menor en los demás segmentos.

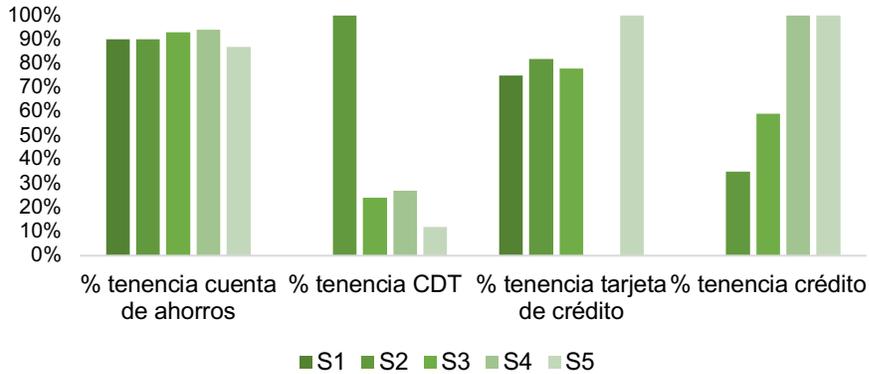


Figura 5. Tenencia productos por segmento.

Fuente: elaboración propia.

En la figura 6 se analiza el promedio del saldo de cada producto por segmento, donde se resalta el promedio del saldo del CDT en el segmento 2 que es de \$5.309.118 y menor en los demás segmentos, el promedio del saldo de la tarjeta de crédito es mayor en el segmento 5 con un valor de \$10.329.432 y el promedio del saldo del crédito tiende a aumentar a medida que se incrementa el número del segmento.

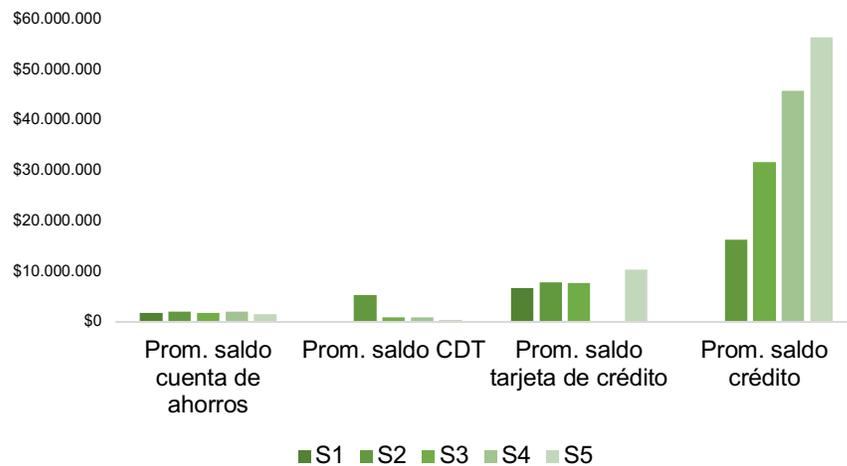


Figura 6. Promedio saldo productos por segmento.

Fuente: elaboración propia.

En la figura 7 se muestra el uso de los canales por segmento, donde a se presentan algunas pequeñas diferencias por segmento y se resalta que ninguno de los clientes del segmento 3 usa la página Web del banco.

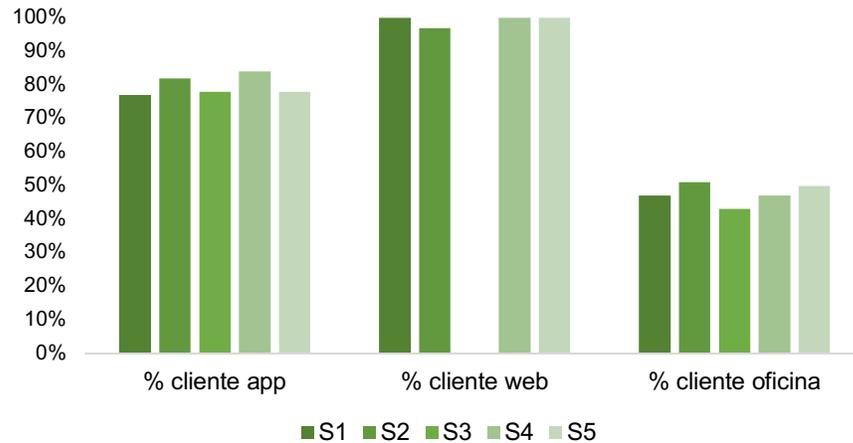


Figura 7. Uso canales por segmento.

Fuente: elaboración propia.

En la figura 8 se analizan las calificaciones de los costos, de los servicios y la recomendación del banco por segmento. Se resalta que el promedio de calificación de los costos es mayor en los segmentos 4 y 5, con un valor de 5,9, el promedio de calificación de los servicios es mayor en el segmento 3 con un valor de 6,1 seguido del segmento 5 con un valor de 5,9 y el promedio de recomendación del banco es mayor en el segmento 2 con un valor de 6,1.

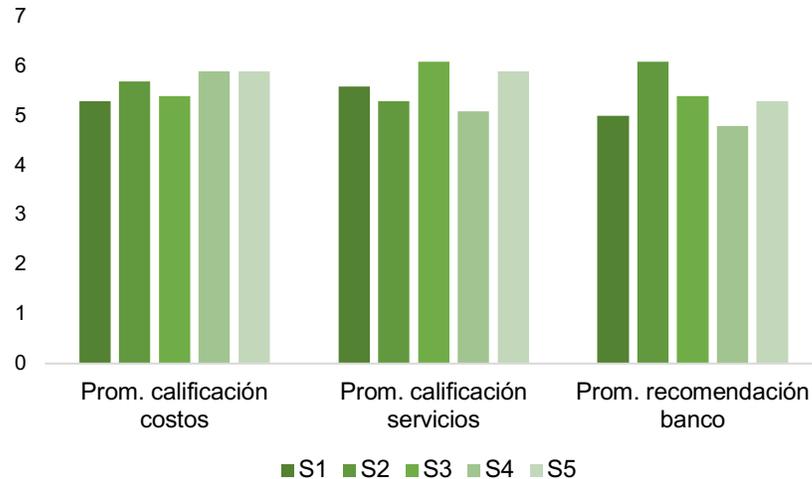


Figura 8. Promedio calificaciones por segmento.

Fuente: elaboración propia.

El segmento 1 se puede denominar el segmento de clientes oportunidad para crédito y CDT dado que ninguno de los clientes de este segmento tienen estos productos. Adicionalmente, el 75% de los clientes de este segmento tiene tarjeta de crédito pero su saldo promedio es el segundo más bajo de todos los segmentos (\$6.727.885), el 77% usan la aplicación móvil del banco, el porcentaje más bajo de todos los segmentos y su promedio de calificación de los costos es el más bajo de todos los segmentos (5,3).

El segmento 2 se puede denominar el segmento de clientes CDT dado que todos los clientes de este segmento tienen este producto y su saldo promedio en este producto es el más alto de todos los segmentos (\$5.309.118). Adicionalmente, los clientes de este segmento tienen el saldo promedio más alto en su cuenta de ahorros (\$2.095.274) y el promedio de recomendación del banco más alto de todos los segmentos (6,1). Sin embargo, la tenencia de crédito de los clientes de este segmento es el segundo más bajo de todos los segmentos (35%) y el saldo en este producto también es el segundo más bajo de todos los segmentos (\$16.334.068).

El segmento 3 se puede denominar el segmento de clientes móviles dado que ninguno de los clientes de este segmento utilizan la página Web del banco y su proporción de uso de las oficinas de servicio del banco es el más bajo de todos los segmentos (43%). Adicionalmente, su promedio de calificación de los servicios es el más alto de todos los

segmentos (6,1) y la proporción de tenencia de productos de los clientes de este segmento no es la más baja ni la más alta comparada con los demás segmentos.

El segmento 4 se puede denominar el segmento de oportunidad para tarjeta de crédito dado que ninguno de los clientes de este segmento tiene este producto pero todos los clientes de este segmento tienen crédito, con el segundo mejor saldo promedio de todos los segmentos (\$45.793.890). Adicionalmente, este segmento presenta la mayor tenencia de cuentas de ahorro de todos los segmentos (94%), el segundo mejor saldo promedio en este producto (\$2.067.295), la mayor proporción de uso de la aplicación móvil del banco de todos los segmentos (84%), el promedio más alto de calificación de los costos de todos los segmentos (5,9), el promedio más bajo de calificación de los servicios de todos los segmentos (5,1) y el promedio más bajo de recomendación del banco (4,8).

El segmento 5 se puede denominar el segmento de clientes del activo dado que todos los clientes de este segmento tienen créditos y tarjetas de crédito y sus saldos promedio en estos productos son los más altos de todos los segmentos (\$56.417.435 en crédito y \$10.329.432 en tarjeta de crédito). Adicionalmente, este segmento presenta la menor tenencia de cuentas de ahorro de todos los segmentos (87%), el menor saldo promedio en este producto (\$1.585.103) y su promedio de calificación de los costos es el más alto de todos los segmentos (5,9).

### **Propuesta de solución a la problemática**

En el análisis de los resultados se respondieron dos de los objetivos específicos del trabajo de grado, evaluar la cantidad de segmentos de clientes persona natural obtenidos al usar información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana y conocer las características de los cinco segmentos de clientes obtenidos. Solo restaría determinar las decisiones que pueden tomar las organizaciones del sector bancario colombiano al segmentar a sus clientes persona natural usando información conductual y por beneficios buscados, lo cual se hará basándose en las características de cada uno de los cinco segmentos obtenidos.

Los clientes del segmento 1 pueden ser profundizados ofreciéndoles los productos que no tienen, crédito y CDT, el costo de estos nuevos productos no debe ser muy alto

para no afectar la percepción positiva que tienen los clientes de este segmento con respecto al costo de los servicios ofrecidos por el banco y la oferta se podría hacer por el canal Web, dado que todos los clientes de este segmento utilizan este canal. Por último, el uso de la aplicación móvil por parte de estos clientes puede ser incentivada para alcanzar las proporciones de los demás segmentos y así evitar que los clientes de este segmento dependan de una sola manera de usar los servicios del banco.

Los clientes del segmento 2 pueden tener un trato diferencial y pueden ser asesorados personalmente en productos de inversión al ser los clientes con mayor saldo promedio en sus cuentas de ahorro y al tener todos un CDT. Adicionalmente, estos clientes podrían convertirse en promotores del banco y sus servicios dado que el promedio de recomendación del banco es el más alto de todos los segmentos.

Los clientes del segmento 3 deben tener una oferta de productos móviles amplia dado que ninguno de los clientes de este segmento utilizan la página Web del banco y su proporción de uso de las oficinas de servicio del banco es el más bajo de todos los segmentos. Estos clientes deben ser retenidos ante la llegada de nuevos competidores como las Fintech y pueden ser analizados como un caso de éxito dado que la calificación promedio de los servicios es el más alto de todos los segmentos.

Los clientes del segmento 4 pueden ser profundizados ofreciéndoles tarjeta de crédito dado que ninguno de los clientes de este segmento tienen este producto. Dado que este segmento presenta la mayor tenencia de cuentas de ahorro de todos los segmentos con el segundo mejor saldo promedio en este producto, estos clientes podrían pagar anticipadamente sus créditos. Los clientes de este segmento pueden ser objeto de campañas de comunicación por medio de la aplicación móvil del banco para mejorar su percepción de los costos, de los servicios y su recomendación del banco.

Por último, los clientes del segmento 5 pueden ser objeto de una campaña de educación financiera dado que todos los clientes de este segmento tienen créditos y tarjetas de crédito, sus saldos promedio en estos productos son los más altos de todos los segmentos y tienen el menor saldo promedio en sus cuentas de ahorro. Cualquier afectación económica de los clientes de este segmento puede afectar el cobro de estos saldos.

## Discusión

El trabajo de campo realizado permitió responder los objetivos específicos planteados, evaluar la cantidad de segmentos de clientes persona natural obtenidos al usar información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana, conocer las características de los cinco segmentos de clientes obtenidos y determinar las decisiones que las organizaciones del sector bancario colombiano pueden tomar al segmentar a sus clientes persona natural usando información conductual y por beneficios buscados.

Estos resultados permitieron identificar qué tipos de segmentos de clientes persona natural se obtienen al usar información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana, encontrando que al aplicar el procedimiento estadístico K-medias, se encontraron cinco segmentos con características diferenciadas según la información analizada, permitiendo establecer una estrategia de oferta de productos y servicios, enfocada en los canales que usan los clientes de cada uno de los segmentos y teniendo en cuenta la percepción que tienen con respecto a los costos, a los servicios y su recomendación del banco. El haber utilizado distintos tipos de información permitió caracterizar los segmentos de manera más completa que al haber utilizado un solo tipo de información, logrando tener una visión más completa de los clientes persona natural y posibilitando tomar decisiones más precisas. Esto repercutirá en la fidelización, en la retención y en la satisfacción de los clientes persona natural.

Como se comentó en el marco teórico, las iniciativas de segmentación de clientes en otros países e instituciones del sector bancario se han centrado en la recomendación del uso de herramientas como la minería de datos y el Big Data para la segmentación de clientes, en la segmentación de clientes según una variable de interés para la organización, según el uso de sus canales, según la tenencia y el uso de ciertos productos y servicios y según los beneficios buscados por los clientes en la relación con su banco. Los resultados obtenidos nos permiten concluir que el aplicar procedimientos de minería de datos usando distintas fuentes de información disponibles sobre los clientes, lo cual se puede lograr usando herramientas de Big Data, es una necesidad para las organizaciones del sector bancario colombiano si no quieren quedar relegadas tecnológicamente y perder competitividad y rentabilidad.

En cuanto a las limitaciones de los resultados, se debe tener en cuenta que estos son específicos de la información utilizada, para aplicar este procedimiento la organización debe contar con la información de sus clientes, personas capacitadas en aplicar procedimientos estadísticos y de minería de datos, contar con las herramientas tecnológicas para realizar esta labor y apertura por parte de la alta gerencia para llevar a la práctica los resultados encontrados.

Por último, también se debe tener en cuenta que para obtener los resultados se uso una de las técnicas más importantes y más usadas dentro del análisis de conglomerados, la técnica estadística K-medias, pero se podrían usar otros procedimientos y comparar los resultados obtenidos.

## **Conclusiones y Trabajo Futuro**

A continuación se presentan las conclusiones y el trabajo futuro que se podría realizar a partir de los resultados obtenidos.

### **Conclusiones**

Con respecto al objetivo específico de evaluar la cantidad de segmentos de clientes persona natural obtenidos al usar información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana, se concluye que en el caso de la información analizada se obtuvieron cinco segmentos, el segmento más grande es el 1 con el 33% de los clientes, seguido por los segmentos 2 y 5 con el 21% de los clientes, el segmento 4 con 14% de los clientes y el segmento 3 con el 11% de los clientes.

En cuanto a conocer las características de los segmentos de clientes persona natural de la banca colombiana usando información conductual y por beneficios buscados, los clientes de cada segmento tienen características diferenciales, encontrando que el segmento 1 se puede denominar el segmento de clientes oportunidad para crédito y CDT, todos los clientes del segmento 2 tienen CDT y su saldo promedio en este producto es el más alto de todos los segmentos, el segmento 3 se puede denominar el segmento de clientes móviles dado que ninguno de los clientes de este segmento utilizan la página Web del banco y su proporción de uso de las oficinas de servicio del banco es el más bajo de todos los segmentos, ninguno de los clientes del segmento 4 tiene tarjeta de crédito pero todos tienen crédito y el segmento 5 se puede denominar el segmento de clientes del activo dado que todos los clientes de este segmento tienen créditos y tarjetas de crédito y sus saldos promedio en estos productos son los más altos de todos los segmentos.

Las características de los clientes de cada grupo permitieron determinar decisiones que podrían tomar las organizaciones del sector bancario colombiano al segmentar a sus clientes persona natural usando información conductual y por beneficios buscados, respondiendo así el tercer objetivo específico. Estas decisiones se relacionan con la oferta de productos, el canal por el que se podría hacer esta oferta e incluso el posible costo de cada producto y servicio dependiendo de la percepción de los clientes de cada segmento.

Es así como los clientes del segmento 1 pueden ser profundizados ofreciéndoles crédito y CDT con un costo adecuado para no afectar la percepción positiva que tienen los clientes de este segmento con respecto al costo de los servicios ofrecidos por el banco y la oferta se podría hacer por el canal Web. Los clientes del segmento 2 podrían ser asesorados personalmente en productos de inversión al ser los clientes con mayor saldo promedio en sus cuentas de ahorro y al tener todos un CDT. Los clientes del segmento 3 deben tener una oferta de productos móviles amplia dado que ninguno de los clientes de este segmento utilizan la página Web del banco. Los clientes del segmento 4 pueden ser profundizados ofreciéndoles tarjeta de crédito y pueden ser objeto de campañas de comunicación por medio de la aplicación móvil del banco para mejorar su percepción de los costos, de los servicios y su recomendación del banco. Por último, los clientes del segmento 5 pueden ser objeto de una campaña de educación financiera dado que todos los clientes de este segmento tienen créditos y tarjetas de crédito, sus saldos promedio en estos productos son los más altos de todos los segmentos y tienen el menor saldo promedio en sus cuentas de ahorro.

Los anteriores resultados permitieron identificar qué tipos de segmentos de clientes persona natural se obtienen usando información conductual y por beneficios buscados de los clientes persona natural de la banca colombiana. El haber utilizado distintos tipos de información, a diferencia de como se ha realizado en otros estudios y países, permitió caracterizar los segmentos de manera más completa que al haber utilizado un solo tipo de información, logrando tener una visión más completa de los clientes persona natural y posibilitando tomar decisiones más precisas. Esto repercutirá en la fidelización, en la retención y en la satisfacción de los clientes persona natural.

### **Trabajo futuro**

En cuanto al trabajo futuro, lo primero que se debe resaltar es que el procedimiento estadístico utilizado, K-medias, puede ser replicado a la información de distintas organizaciones del sector bancario colombiano y latinoamericano. Los resultados obtenidos pueden ser contrastados con los obtenidos en esta investigación.

Por otra parte, dependiendo de la disponibilidad de información sobre los clientes que tengan las organizaciones del sector bancario colombiano, se podría incluir otras

fuentes de información y aplicar el mismo procedimiento. La disponibilidad de más información sobre los clientes va a ser cada vez más común en las organizaciones a medida que implementen herramientas de Big Data.

Por último, también sería interesante que se usen otras técnicas de análisis de conglomerados, en esta investigación se utilizó una de las técnicas más importantes y más usadas dentro del análisis de conglomerados, la técnica estadística K-medias, pero sería interesante comparar los resultados usando las mismas fuentes de información pero usando otras técnicas de análisis.

## Referencias

- Alfansi, L., & Sargeant, A. (2000). *Market segmentation in the Indonesian banking sector: the relationship between demographics and desired customer benefits*. The International Journal of Bank Marketing, 18(2), 64-74.
- Aryuni, M., Didik Madyatmadja, E. & Miranda, E. (2018). *Customer Segmentation in XYZ Bank Using K-Means and K-medoids Clustering*. Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018 8528086, 412-416.
- Berger, R. (2017). *Industry 4.0, New Industrial Revolution. How Europe will succeed*. Recuperado de [http://www.iberglobal.com/files/Roland\\_Berger\\_Industry.pdf](http://www.iberglobal.com/files/Roland_Berger_Industry.pdf)
- Bhuvana, M., Thirumagal, P.G. & Vasantha, S. (2016). *Big data analytics - a leveraging technology for Indian commercial banks*. Indian Journal of Science and Technology, 9(32), 98643. doi: 10.17485/ijst/2016/v9i32/98643
- Blanco, L. (2004). *Probabilidad. Colombia: Universidad Nacional de Colombia*.
- Chawla, D. & Joshi, H. (2017). *Consumer perspectives about mobile banking adoption in India - a cluster analysis*. International Journal of Bank Marketing, 35(4), 616-636.
- Díaz, L. G. (2002). *Estadística Multivariada: inferencia y métodos. Colombia: Universidad Nacional de Colombia*.
- Dibb, S. (1998). *Market segmentation: strategies for success*. Marketing Intelligence & Planning, 16(7), 394-406.
- Dimitriadis, S.(2011). *Customers' relationship expectations and costs as segmentation variables: Preliminary evidence from banking*. Journal of Services Marketing, 25(4), 294-309.
- Ferrel, O. C. y Hartline, M. D. (2018). *Estrategia de marketing*. México: Cengage.

Gil, J.M., Aguilera, A., Suárez, L. (2020). *A Bibliometric Analysis of China-Latin America Economic and Political Relations*. *Latin American Policy*, 0(0), 1-23.

Goyat, S. (2011). *The basis of market segmentation: a critical review of literature*. *European Journal of Business and Management*, 3(9), 45-54.

Hernández-Sampieri, R. y Mendoza, C. R. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. Ciudad de México: McGraw-Hill Interamericana.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical learning*. New York: Springer.

Jayasree, J. & Balan, J. (2013). *A Review on Data Mining in Banking Sector*. *American Journal of Applied Sciences*, 10(10), 1160-1165. doi: 10.3844/ajassp.2013.1160.1165

Kaur, G., Sharma, R.D. & Mahajan, N. (2014). *Segmentation of bank customers by loyalty and switching intentions*. *Vikalpa*, 39(4), 75-89.

Kotler, P., & Keller, K. (2015). *Marketing Management*. Harlow: Pearson.

Lambin, J., Gallucci, C. y Sicurello, C. (2009). *Dirección de marketing: gestión estratégica y operativa del mercado*. México: McGraw-Hill.

Machauer, A. & Morgner, S. (2001). *Segmentation of bank customers by expected benefits and attitudes*. *The International Journal of Bank Marketing*, 19(1), 6-18.

Mehdiabadi, A., Tabatabeinasab, M., Spulbar, C., Karbasii, A. & Birau, R. (2020). *Are We Ready for the Challenge of Banks 4.0? Designing a Roadmap for Banking Systems in Industry 4.0*. *International Journal of Financial Studies*, 8(32), 1-28. doi: 10.3390/ijfs8020032

Mihova, V. & Pavlov, V. (2018). *A customer segmentation approach in commercial banks*. *AIP Conference Proceedings*, 2025, 030003-1-030003-8. doi: 10.1063/1.5064881

Ogwueleka, F.N., Misra, S., Colomo-Palacios, R. & Fernandez, L. (2015). *Neural network and classification approach in identifying customer behavior in the banking sector: A case study of an international bank*. Human Factors and Ergonomics In Manufacturing, 25(1), 28-42. doi: 10.1002/hfm.20398

Olcese, A., Rodriguez, M.A. y Alfaro, J. (2008). *Manual de la empresa responsable y sostenible: conceptos, ejemplos y herramientas de la responsabilidad social corporativa o de la empresa*. España: McGraw-Hill.

Rojas, G.D., Ramírez, C.H. y Vélez, J.D. (2013). *Gerencia de la responsabilidad social en las organizaciones de hoy*. Colombia: Universidad de Ibagué.

Smeureaunu, I., Ruxanda, G. & Badea, L.M. (2013). *Customer segmentation in private banking sector using machine learning techniques*. Journal of Business Economics and Management, 14(5), 923-939.

Smith, W.R. (1956). *Product differentiation and market segmentation as alternative marketing strategies*. Journal of Marketing, 21(1), 3-8.

Srivastava, A., Singh, S.K., Tanwar, S. & Tyagi, S. (2018). *Suitability of big data analytics in Indian banking sector to increase revenue and profitability*. Proceedings - 2017 3rd International Conference on Advances in Computing, Communication and Automation (Fall), ICACCA 2017 2018 - January, 1-4.

Srivastava, U. & Gopalkrishnan, S. (2015). *Impact of big data analytics on banking sector: Learning for Indian Banks*. Procedia Computer Science, 50(2015), 643-652. doi:10.1016/j.procs.2015.04.098

Sundjaja, A.M. (2013). *Analysis of customer segmentation in Bank XYZ using data mining technique*. Asian Journal of Information Technology, 12(1), 39-44. doi: 10.3923/ajit.2013.39.44

Tan, P. N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Estados Unidos: Pearson.

Villanueva, J. y De-Toro, J.M. (2017). *Marketing estratégico*. España: EUNSA.

Voican, O. (2020). *Using data mining methods to solve classification problems in financial-banking institutions*. *Economic Computation and Economic Cybernetics Studies and Research*, 54(1), 159-176. doi: 10.24818/18423264/54.1.20.11

Wedel, M. & Kamakura, W.A. (1998). *Market segmentation: conceptual and methodological foundations*. Boston: Kluwer Academic Publishers.

Xu, L. D., Xu, E. L., & Li, L. (2018). *Industry 4.0: state of the art and future trends*. *International Journal of Production Research*, 56(8), 2941-2962. doi: 10.1080/00207543.2018.1444806

**A. Anexo. Resultados estadísticos por segmento.**

Variable	Segmento				
	1	2	3	4	5
Clientes	166	103	54	70	107
% tenencia cuenta de ahorros	90%	90%	93%	94%	87%
Prom. saldo cuenta de ahorros	\$1.798.240	\$2.095.274	\$1.767.910	\$2.067.295	\$1.585.103
% tenencia CDT	0%	100%	24%	27%	12%
Prom. saldo CDT	\$0	\$5.309.118	\$886.512	\$956.966	\$348.167
% tenencia tarjeta de crédito	75%	82%	78%	0%	100%
Prom. saldo tarjeta de crédito	\$6.727.885	\$7.800.346	\$7.740.397	\$0	\$10.329.432
% tenencia crédito	0%	35%	59%	100%	100%
Prom. saldo crédito	\$0	\$16.344.068	\$31.669.416	\$45.793.890	\$56.417.435
% cliente app	77%	82%	78%	84%	78%
% cliente web	100%	97%	0%	100%	100%
% cliente oficina	47%	51%	43%	47%	50%
Prom. calificación costos	5,3	5,7	5,4	5,9	5,9
Prom. calificación servicios	5,6	5,3	6,1	5,1	5,9
Prom. recomendación banco	5	6,1	5,4	4,8	5,3

## B. Anexo. Código en R para obtener la segmentación.

```
# Cargue de paquetes
```

```
library(dplyr)
```

```
library(readr)
```

```
library(ggplot2)
```

```
# Opciones de sistema
```

```
options(scipen = 999)
```

```
# Fijar semilla aleatoria
```

```
set.seed(123)
```

```
# Simulación información bancaria clientes persona natural
```

```
# Creación base de datos
```

```
bd_clientes_bancarios <- data.frame(matrix(ncol = 14 , nrow = 500))
```

```
colnames(bd_clientes_bancarios) <- c("tenencia_cuenta_ahorros",
```

```
"saldo_cuenta_ahorros", "tenencia_cdt", "saldo_cdt", "tenencia_tarjeta_credito",
```

```
"saldo_tarjeta_credito", "tenencia_credito", "saldo_credito", "cliente_app", "cliente_web",
```

```
"cliente_oficina", "calificacion_costos", "calificacion_servicios", "recomendacion_banco")
```

```
# tenencia_cuenta_ahorros
```

```
bd_clientes_bancarios$tenencia_cuenta_ahorros <- rbinom(n=500, size=1, prob=0.9)
```

```
# saldo_cuenta_ahorros
```

```
bd_clientes_bancarios$saldo_cuenta_ahorros <- round(rlnorm(500, meanlog =  
log((2000000^2)/sqrt(2000000^2+1000000^2)), sdlog =  
sqrt(log(1+(1000000^2/2000000^2))))))
```

```
bd_clientes_bancarios$saldo_cuenta_ahorros <-  
ifelse(bd_clientes_bancarios$tenencia_cuenta_ahorros == 0, 0,  
bd_clientes_bancarios$saldo_cuenta_ahorros)
```

```
# tenencia_cdt
```

```
bd_clientes_bancarios$tenencia_cdt <- rbinom(n=500, size=1, prob=0.3)
```

```
# saldo_cdt
```

```
bd_clientes_bancarios$saldo_cdt <- round(rlnorm(500, meanlog =  
log((5000000^2)/sqrt(5000000^2+3000000^2)), sdlog =  
sqrt(log(1+(3000000^2/5000000^2))))))
```

```
bd_clientes_bancarios$saldo_cdt <- ifelse(bd_clientes_bancarios$tenencia_cdt== 0, 0,  
bd_clientes_bancarios$saldo_cdt)
```

```
# tenencia_tarjeta_credito
```

```
bd_clientes_bancarios$tenencia_tarjeta_credito <- rbinom(n=500, size=1, prob=0.7)
```

```
# saldo_tarjeta_credito
```

```
bd_clientes_bancarios$saldo_tarjeta_credito <- round(rlnorm(500, meanlog =  
log((10000000^2)/sqrt(10000000^2+6000000^2)), sdlog =  
sqrt(log(1+(6000000^2/10000000^2))))))
```

```
bd_clientes_bancarios$saldo_tarjeta_credito <-  
ifelse(bd_clientes_bancarios$tenencia_tarjeta_credito == 0, 0,  
bd_clientes_bancarios$saldo_tarjeta_credito)  
  
# tenencia_credito  
  
bd_clientes_bancarios$tenencia_credito <- rbinom(n=500, size=1, prob=0.5)  
  
# saldo_credito  
  
bd_clientes_bancarios$saldo_credito <- round(rlnorm(500, meanlog =  
log((50000000^2)/sqrt(50000000^2+30000000^2)), sdlog =  
sqrt(log(1+(30000000^2/50000000^2))))))  
  
bd_clientes_bancarios$saldo_credito <- ifelse(bd_clientes_bancarios$tenencia_credito  
== 0, 0, bd_clientes_bancarios$saldo_credito)  
  
# cliente_app  
  
bd_clientes_bancarios$cliente_app <- rbinom(n=500, size=1, prob=0.8)  
  
# cliente_web  
  
bd_clientes_bancarios$cliente_web <- rbinom(n=500, size=1, prob=0.9)  
  
# cliente_oficina  
  
bd_clientes_bancarios$cliente_oficina <- rbinom(n=500, size=1, prob=0.5)  
  
# calificacion_costos  
  
bd_clientes_bancarios$calificacion_costos <- round(runif(500, min = 1, max = 10))
```

```
# calificacion_servicios
```

```
bd_clientes_bancarios$calificacion_servicios <- round(runif(500, min = 1, max = 10))
```

```
# recomendacion_banco
```

```
bd_clientes_bancarios$recomendacion_banco <- round(runif(500, min = 1, max = 10))
```

```
# Estandarización base de datos
```

```
bd_clientes_bancarios_est <- bd_clientes_bancarios %>% scale()
```

```
dim(bd_clientes_bancarios_est)
```

```
View(bd_clientes_bancarios_est)
```

```
# Calculo suma de cuadrados total
```

```
wss <- (nrow(bd_clientes_bancarios_est)-1)*sum(apply(bd_clientes_bancarios_est,2,var))  
wss
```

```
# Escoger el número de segmentos basado en la disminución de la suma de cuadrados
```

```
for (i in 2:15) wss[i] <- sum(kmeans(bd_clientes_bancarios_est, centers=i)$withinss)  
plot(1:15, wss, type="b", xlab="Número de segmentos", ylab="Suma de cuadrados del  
error")
```

```
# La solución escogida son 5 segmentos
```

```
SOLUCION_5_SEGMENTOS <- kmeans(bd_clientes_bancarios_est, 5)
```

```
SOLUCION_5_SEGMENTOS
```

```
summary(SOLUCION_5_SEGMENTOS)
```

```
SOLUCION_5_SEGMENTOS$cluster
```

```
SOLUCION_5_SEGMENTOS$centers
```

```
SOLUCION_5_SEGMENTOS$totss
```

```
SOLUCION_5_SEGMENTOS$withinss
```

```
SOLUCION_5_SEGMENTOS$tot.withinss
```

```
SOLUCION_5_SEGMENTOS$betweenss
```

```
SOLUCION_5_SEGMENTOS$size
```

```
SOLUCION_5_SEGMENTOS$iter
```

```
SOLUCION_5_SEGMENTOS$ifault
```

```
# Una medida de bondad de ajuste
```

```
round(100*(SOLUCION_5_SEGMENTOS$betweenss/SOLUCION_5_SEGMENTOS$totss),1)
```

```
# Agregar la variable segmento a la base sin estandarizar
```

```
bd_clientes_bancarios_segmento <- data.frame(bd_clientes_bancarios,
```

```
SOLUCION_5_SEGMENTOS$cluster)
```

```
dim(bd_clientes_bancarios_segmento)
```

```
View(bd_clientes_bancarios_segmento)
```

```
# Descripción de los segmentos
```

```
des_5_kmeans <- bd_clientes_bancarios_segmento %>%
```

```
group_by(SOLUCION_5_SEGMENTOS.cluster) %>% summarise (clientes = n(),
```

```
por_tenencia_cuenta_ahorros = round(100*mean(tenencia_cuenta_ahorros),0),
```

```
pro_saldo_cuenta_ahorros = round(mean(saldo_cuenta_ahorros),0),
```

```
por_tenencia_cdt = round(100*mean(tenencia_cdt),0),
```

```
pro_saldo_cdt = round(mean(saldo_cdt),0),
```

```
por_tenencia_tarjeta_credito = round(100*mean(tenencia_tarjeta_credito),0),
```

```
pro_saldo_tarjeta_credito = round(mean(saldo_tarjeta_credito),0),
```

```
por_tenencia_credito = round(100*mean(tenencia_credito),0),
```

```
pro_saldo_credito = round(mean(saldo_credito),0),
```

```
por_cliente_app = round(100*mean(cliente_app),0),
```

```
por_cliente_web = round(100*mean(cliente_web),0),
```

```
por_cliente_oficina = round(100*mean(cliente_oficina),0),
```

```
pro_calificacion_costos = round(mean(calificacion_costos),1),
```

```
pro_calificacion_servicios = round(mean(calificacion_servicios),1),  
pro_recomendacion_banco = round(mean(recomendacion_banco),1)) %>%  
arrange(SOLUCION_5_SEGMENTOS.cluster)
```

```
View(des_5_kmeans)
```

```
write.table(des_5_kmeans, "des_5_kmeans.txt", col.names = TRUE, row.names =  
FALSE, quote = FALSE, sep = "\t", dec = ",")
```