

**Aplicación de ciencia de datos y analítica para conocer los comportamientos financieros de los clientes y desarrollar ofertas personalizadas en una compañía de administración de activos**

John Alexander Vallejo Quintero

Asesor

Eduardo Sanchez Sandoval

Universidad Nacional Abierta y a Distancia UNAD  
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI  
Especialización en Ciencia de Datos y Analítica  
2025

### **Dedicatoria**

A mis padres, por ser mi mayor fuente de inspiración, por su apoyo incondicional en cada etapa de mi vida, y por enseñarme con su ejemplo el valor del esfuerzo, la disciplina y la integridad.

Este logro también es suyo.

## **Agradecimientos**

Quiero expresar mi profundo agradecimiento a todas las personas que, de una u otra forma, hicieron posible la realización de este proyecto de grado.

En primer lugar, agradezco a mis padres, por su apoyo constante, su amor incondicional y por ser mi motor en cada uno de los desafíos que he enfrentado. Sin su respaldo, este camino no habría sido posible.

Agradezco también a los docentes y tutores de la Universidad Nacional Abierta y a Distancia – UNAD, por su guía académica, sus valiosos aportes y su compromiso con la formación integral.

Su acompañamiento fue fundamental para el desarrollo riguroso de este trabajo.

Extiendo mi gratitud a mis colegas y compañeros de estudio, por los espacios de intercambio, colaboración y motivación mutua durante el proceso. Asimismo, agradezco a la organización donde actualmente laboro, por inspirarme a abordar una problemática relevante del sector, y por brindarme las herramientas y el entorno necesario para aplicar lo aprendido de manera práctica.

Finalmente, agradezco a la vida por permitirme crecer profesional y personalmente a través de este proyecto, que representa no solo un logro académico, sino también un paso firme hacia mi desarrollo como científico de datos y analista estratégico.

## Resumen

En un entorno financiero cada vez más competitivo, las compañías del sector de administración de activos enfrentan el desafío de comprender con mayor profundidad los comportamientos financieros de sus clientes para ofrecer soluciones personalizadas y efectivas. Este trabajo de grado implementó un enfoque basado en ciencia de datos y analítica avanzada para segmentar clientes de pensiones voluntarias de tres países latinoamericanos: Chile, Colombia y México. A partir de una base de datos representativa, se aplicaron técnicas como análisis exploratorio de variables demográficas y financieras, cruces para identificar patrones de comportamiento y modelos de clusterización (K-means) para generar segmentos accionables.

Los principales hallazgos revelaron que factores como edad, nivel educativo, tipo de inversión y país de residencia inciden significativamente en la frecuencia de inversión y en el valor promedio invertido. Se identificaron seis perfiles distintos de clientes, incluyendo desde jóvenes de bajo ahorro hasta adultos mayores con alto patrimonio y estrategias de inversión agresivas.

Con base en esta segmentación, se formularon recomendaciones específicas orientadas a mejorar la experiencia del cliente y optimizar las estrategias de inversión según cada perfil. La metodología empleada permitió validar cómo la aplicación de herramientas de machine learning y análisis multivariado puede aportar valor estratégico a las compañías del sector financiero, facilitando la toma de decisiones informadas y centradas en el cliente.

Este proyecto demuestra que la integración de enfoques analíticos robustos no solo contribuye a la eficiencia operativa, sino que también fortalece la personalización de la oferta y la fidelización en un mercado cada vez más exigente. En conclusión, el uso estratégico de los datos representa una oportunidad clave para transformar la relación entre las administradoras de activos y sus clientes, apoyando una visión más sostenible y adaptativa del negocio en el largo

plazo.

***Palabras claves:*** Analítica predictiva, ciencia de datos, comportamientos financieros, segmentación de clientes, satisfacción del cliente

## Abstract

In an increasingly competitive financial environment, asset management companies face the challenge of gaining a deeper understanding of their clients' financial behaviors in order to offer more personalized and effective solutions. This thesis implemented a data science and advanced analytics approach to segment clients of voluntary pension plans from three Latin American countries: Chile, Colombia, and Mexico. Using a representative dataset, techniques such as exploratory analysis of demographic and financial variables, cross-variable analysis to identify behavioral patterns, and clustering models (K-means) were applied to generate actionable customer segments.

The main findings revealed that factors such as age, educational level, investment type, and country of residence significantly influence investment frequency and average invested amount. Six distinct client profiles were identified, ranging from low-saving young individuals to older adults with high wealth and aggressive investment strategies.

Based on this segmentation, specific recommendations were formulated to improve customer experience and optimize investment strategies for each profile. The methodology validated how the application of machine learning tools and multivariate analysis can generate strategic value for companies in the financial sector by facilitating informed, customer-centered decision-making.

This project demonstrates that the strategic use of robust analytical approaches contributes not only to operational efficiency but also to a more personalized and loyalty-driven service model in an increasingly demanding market. In conclusion, the strategic use of data represents a key opportunity to transform the relationship between asset managers and their clients, supporting a more sustainable and adaptive long-term business vision.

**Keywords:** Client satisfaction, customer segmentation, data science, financial behaviors, predictive analytics.

## Tabla de Contenido

Introducción .....	13
Planteamiento del Problema .....	15
Justificación .....	17
Objetivos .....	19
Objetivo General .....	19
Objetivos Específicos .....	19
Marco de Referencia .....	20
Marco Teórico .....	20
Ciencia de Datos .....	20
Analítica Predictiva .....	20
Segmentación de Clientes .....	21
Big Data .....	22
Modelos de Aprendizaje Automático .....	23
Clusterización en Machine Learning .....	23
Metodología .....	25
Tipo de Estudio .....	25
Método .....	25
Carga de los Datos .....	26
Limpieza y Filtrado .....	26
Análisis Exploratorio y Transformación .....	27
Modelado y Segmentación .....	27
Resultados .....	28

Resultados Objetivo 1 .....	28
Análisis Exploratorio Inicial .....	28
Resultados Objetivo 2 .....	34
Justificación Metodológica de los Cruces .....	34
Resultados Objetivo 3 .....	38
Preprocesamiento para Clustering .....	38
Método del Codo para Determinar el Número Óptimo de Clústers .....	38
Aplicación del Algoritmo KMeans .....	39
Análisis de los Clústers .....	39
Gráficas de los Clústers .....	41
Resultados Objetivo 4 .....	43
Perfilamiento de los Clúster y Recomendaciones .....	43
Clúster 0 Prejubilados con Alto Patrimonio y Perfil Conservador .....	44
Clúster 1 Jóvenes de Bajos Ingresos y Ahorro Incipiente .....	44
Clúster 2 Adultos Mayores con Bajo Valor Acumulado .....	45
Clúster 3 Adultos con Alta Capacidad de Ingreso y Apetito de Riesgo .....	45
Clúster 4 Jóvenes Estructurados con Ingresos Medios .....	46
Clúster 5 Seniors con Muy Alto Patrimonio y Perfil Agresivo .....	46
Conclusiones .....	48
Recomendaciones .....	50
Referencias Bibliográficas .....	52
Apéndices .....	54

**Lista de Tablas**

<b>Tabla 1</b> <i>Descripción de Variables</i> .....	29
<b>Tabla 2</b> <i>Edad Promedio de la Cartera</i> .....	31
<b>Tabla 3</b> <i>Patrones de Inversión y Factores Clave por Variable</i> .....	36
<b>Tabla 4</b> <i>Clusters</i> .....	40

## Lista de Figuras

<b>Figura 1</b> <i>Distribución de Clientes (por Edad y País)</i> .....	30
<b>Figura 2</b> <i>Distribución de Ingresos (por Clientes y País)</i> .....	32
<b>Figura 3</b> <i>Distribución de Valor Invertido (por Clientes y País)</i> .....	33
<b>Figura 4</b> <i>Método del Codo</i> .....	39
<b>Figura 5</b> <i>Edad vs Valor Invertido (por Cluster)</i> .....	41
<b>Figura 6</b> <i>Comparación de Clústers (Radar Chart)</i> .....	42

### **Lista de Apéndices**

**Apéndice A** *Enlace Código Python Proyecto de Grado* ..... 54

**Apéndice B** *Enlace Video Sustentación Proyecto de Grado* ..... 54

## Introducción

En un entorno económico y social cada vez más dinámico, las compañías del sector de administración de activos enfrentan el desafío de comprender profundamente los comportamientos financieros de sus clientes para ofrecer soluciones más personalizadas y eficaces. Esta necesidad se vuelve aún más crítica en productos como las pensiones voluntarias, donde las decisiones de ahorro e inversión dependen fuertemente del perfil, las expectativas y los hábitos individuales de cada afiliado. En este contexto, la ciencia de datos y la analítica emergen como herramientas clave para transformar grandes volúmenes de información en conocimiento útil y accionable.

El presente proyecto de grado tiene como propósito demostrar cómo la aplicación de técnicas analíticas puede mejorar la comprensión del comportamiento financiero de los clientes en las compañías del sector de administración de activos, con el fin de diseñar estrategias personalizadas que optimicen sus carteras de inversión y fortalezcan su experiencia. La pregunta que guía este trabajo es: ¿cómo puede la ciencia de datos y la analítica proporcionar a una compañía del sector de administración de activos las herramientas necesarias para analizar y comprender profundamente los comportamientos financieros de sus clientes, permitiendo así desarrollar ofertas personalizadas que optimicen tanto la experiencia del cliente como la estrategia de inversión?

Para responder a esta pregunta, se plantearon como objetivos específicos analizar los hábitos actuales de los clientes utilizando variables demográficas y transaccionales, identificar patrones y factores clave que influyen en la toma de decisiones financieras, segmentar a los clientes con base en sus características comunes, y proponer recomendaciones estratégicas orientadas a mejorar la gestión comercial y de inversión de la compañía.

Metodológicamente, se desarrolló un enfoque cuantitativo y exploratorio, que inició con la limpieza y consolidación de bases de datos simuladas representativas de cinco países latinoamericanos. Se realizó un análisis exploratorio de variables como edad, ingresos, valor invertido, nivel educativo, género y tipo de inversión. Posteriormente, se aplicó una técnica de clustering no supervisado (K-means) para construir un modelo de segmentación, identificando seis grupos de clientes con perfiles financieros claramente diferenciados.

Los resultados obtenidos no solo permitieron caracterizar a la base de clientes de forma más precisa, sino también generar recomendaciones concretas para mejorar la personalización de productos, la estrategia de relacionamiento y la propuesta de valor. Además, se demostró el valor que tiene la analítica para tomar decisiones estratégicas fundamentadas en datos, habilitando a las organizaciones del sector financiero para evolucionar hacia modelos de gestión centrados en el cliente.

Este proyecto busca aportar tanto al ámbito académico como al contexto empresarial, evidenciando cómo la ciencia de datos puede ser aplicada de forma práctica para resolver problemáticas reales y generar ventajas competitivas sostenibles.

## Planteamiento del Problema

La globalización y el avance de la tecnología han incrementado exponencialmente el volumen de datos generados a nivel mundial, transformando la manera en que las empresas operan y toman decisiones (Davenport & Harris, 2007). En el sector financiero, las organizaciones han comenzado a utilizar herramientas analíticas avanzadas para comprender y anticipar las necesidades de sus clientes. Según un informe de McKinsey (2023), más del 70 % de las empresas líderes en servicios financieros a nivel global han implementado estrategias basadas en ciencia de datos para personalizar sus ofertas y optimizar sus operaciones.

En el contexto colombiano, se ha evidenciado un crecimiento importante en la adopción de tecnologías de datos en el sector financiero. Estudios del Banco de la República (2023) señalan que el sistema financiero nacional presenta niveles de liquidez y solvencia superiores a los estándares internacionales, lo que refleja una gestión eficiente y una implementación progresiva de tecnologías para la administración de riesgos y operaciones. Asimismo, la Superintendencia Financiera de Colombia, a través de su iniciativa "Estadísticas dinámicas" lanzada en julio de 2024, ha impulsado el uso de herramientas interactivas para el análisis y la toma de decisiones basadas en datos (Superintendencia Financiera de Colombia, 2024).

Sin embargo, a nivel organizacional, las compañías del sector de administración de activos enfrentan el reto de mantenerse competitivas en un entorno altamente dinámico y centrado en el cliente. El problema central radica en la dificultad para ofrecer una propuesta de valor verdaderamente personalizada, debido a la falta de conocimiento profundo sobre los comportamientos y hábitos financieros de sus clientes. Esta limitación impide identificar patrones clave en las decisiones de inversión y segmentar de forma precisa a los clientes según sus necesidades específicas. Como consecuencia, se generan estrategias de inversión menos

eficaces, disminuye la satisfacción del cliente y se incrementa el riesgo de pérdida de competitividad y fidelización.

El impacto de esta problemática no solo se refleja en los resultados financieros, sino también en la percepción que los clientes tienen de la organización. Una oferta poco alineada con sus expectativas puede generar desconfianza y migración hacia competidores más innovadores. Además, la falta de segmentación adecuada limita el diseño de estrategias adaptadas a los distintos perfiles de clientes, restringiendo el potencial de crecimiento.

Frente a este panorama, resulta crucial abordar esta problemática para garantizar la sostenibilidad y competitividad en el mercado financiero. En este contexto, el presente proyecto busca responder a la siguiente pregunta de investigación:

¿Cómo puede la ciencia de datos y la analítica proporcionar a las compañías del sector de administración de activos las herramientas necesarias para analizar y comprender profundamente los comportamientos financieros de sus clientes, permitiendo así desarrollar ofertas personalizadas que optimicen tanto la experiencia del cliente como la estrategia de inversión?

## Justificación

La implementación de este proyecto contribuye a fortalecer las capacidades analíticas de las compañías del sector de administración de activos, permitiéndoles avanzar en el uso estratégico de la información para comprender mejor a sus clientes y tomar decisiones más informadas. Si bien no constituye por sí sola una ventaja competitiva definitiva, sí representa un paso importante hacia la incorporación de prácticas basadas en datos, cada vez más necesarias en un entorno financiero dinámico y centrado en el cliente. En un entorno altamente competitivo, donde los clientes exigen soluciones cada vez más personalizadas y ajustadas a sus necesidades, la capacidad de comprender a profundidad los comportamientos y hábitos financieros se convierte en un factor clave para diseñar estrategias de inversión más efectivas y alineadas con las expectativas del cliente.

Ejemplos concretos del sector muestran cómo la personalización basada en datos puede traducirse en beneficios operativos y estratégicos. Por ejemplo, la AFP Habitat en Chile ha implementado modelos de segmentación avanzados que le permiten enviar comunicaciones diferenciadas y recomendaciones de inversión adaptadas al perfil de riesgo y ciclo de vida de sus afiliados. Esto ha contribuido a mejorar los niveles de *engagement*, aumentar el uso de canales digitales y optimizar la gestión comercial. Casos como este evidencian que, si bien el uso de ciencia de datos no garantiza por sí solo una ventaja competitiva, sí constituye una capacidad clave para adaptarse a un entorno cada vez más centrado en el cliente y orientado por la información (AFP Habitat, 2022).

La implementación de modelos predictivos y técnicas avanzadas de segmentación también habilita la anticipación de tendencias y comportamientos futuros, optimizando no solo las estrategias de inversión, sino también la experiencia general del cliente. Esto repercute

positivamente tanto en los resultados financieros de la organización como en su posicionamiento como un actor innovador, confiable y centrado en el cliente dentro del sector financiero. Abordar este desafío resulta fundamental para garantizar la sostenibilidad operativa y fortalecer la relación con los clientes, consolidando así una posición de liderazgo en el mercado.

Además, este proyecto contribuye al desarrollo del conocimiento en el campo de la ciencia de datos aplicada a las finanzas, promoviendo el uso de modelos predictivos, segmentación avanzada y análisis de grandes volúmenes de datos para la toma de decisiones estratégicas. La aplicación de estos enfoques permite explorar nuevas metodologías en la personalización de servicios financieros, una línea aún en evolución dentro de la industria. Los hallazgos y resultados derivados de este estudio podrían servir como base para investigaciones futuras, ampliando las posibilidades de uso de la ciencia de datos para optimizar la gestión de activos y mejorar la experiencia del cliente en otras organizaciones del sector.

## **Objetivos**

### **Objetivo General**

Implementar técnicas de ciencia de datos y analítica para analizar los comportamientos y hábitos financieros de los clientes de una compañía del sector de administración de activos, con el fin de diseñar estrategias personalizadas que optimicen sus carteras de inversión.

### **Objetivos Específicos**

Analizar los comportamientos financieros actuales de los clientes, utilizando datos históricos demográficos y transaccionales disponibles en una compañía del sector de administración de activos.

Identificar patrones de inversión y factores clave que influyen en las decisiones financieras de los clientes.

Desarrollar un modelo de segmentación que clasifique a los clientes según sus características y necesidades financieras.

Proponer recomendaciones personalizadas, basadas en los análisis realizados, que permitan optimizar las estrategias de inversión y mejorar la experiencia del cliente.

## **Marco de Referencia**

### **Marco Teórico**

#### ***Ciencia de Datos***

La ciencia de datos es un campo interdisciplinario que combina estadística, informática, matemáticas y conocimiento del dominio para analizar e interpretar grandes volúmenes de datos, con el objetivo de extraer conocimiento útil y apoyar la toma de decisiones informadas (Provost & Fawcett, 2013). En los últimos años, esta disciplina se ha consolidado como una herramienta esencial en sectores como el financiero, donde la complejidad de los datos y la necesidad de respuestas ágiles exigen capacidades analíticas avanzadas.

En particular, las compañías del sector de administración de activos encuentran en la ciencia de datos una oportunidad para comprender mejor a sus clientes, sus hábitos financieros y patrones de inversión, lo que facilita la generación de propuestas de valor más relevantes y personalizadas. Según Marr (2018), el uso inteligente de los datos puede mejorar la eficiencia operativa, reducir riesgos y aumentar la satisfacción del cliente.

Asimismo, el enfoque iterativo y basado en evidencia de la ciencia de datos impulsa una cultura de mejora continua dentro de las organizaciones. En este contexto, la capacidad de recolectar, transformar, analizar y comunicar hallazgos derivados de los datos se convierte en un activo estratégico clave para responder a los desafíos del mercado actual y adaptarse a las preferencias cambiantes de los usuarios.

#### ***Analítica Predictiva***

La analítica predictiva es una rama avanzada de la ciencia de datos que utiliza modelos estadísticos, minería de datos y técnicas de aprendizaje automático para predecir resultados futuros basados en datos históricos (Siegel, 2016). Esta técnica ha demostrado ser altamente

efectiva en múltiples industrias, particularmente en el sector financiero, donde permite anticipar eventos críticos como el abandono de clientes, fluctuaciones del mercado o comportamientos de inversión. Según Davenport y Harris (2007), la analítica predictiva no solo permite tomar decisiones informadas, sino también adelantarse a los cambios del entorno y actuar de manera proactiva.

En el contexto de las compañías de administración de activos, la analítica predictiva puede ser utilizada para identificar clientes con mayor propensión a incrementar sus aportes, anticipar necesidades de productos financieros, o incluso personalizar recomendaciones basadas en patrones históricos de comportamiento. Esta capacidad de anticipación puede traducirse en una mejora significativa en la experiencia del cliente, así como en una asignación más eficiente de los recursos comerciales y de asesoría.

Además, la analítica predictiva permite simular escenarios futuros y evaluar la efectividad de distintas estrategias antes de ser implementadas, lo cual constituye una ventaja crucial en un sector caracterizado por la volatilidad y la necesidad de decisiones basadas en evidencia (Mayer-Schönberger & Cukier, 2014).

### ***Segmentación de Clientes***

La segmentación de clientes es una estrategia de marketing y análisis que permite dividir a los usuarios en grupos homogéneos basados en características comunes como edad, ingresos, comportamiento financiero o perfil de riesgo. Según Kotler y Keller (2012), una segmentación efectiva posibilita el diseño de estrategias personalizadas que incrementan la satisfacción y fidelización del cliente. En el sector financiero, esta práctica resulta esencial para identificar las distintas necesidades de inversión y diseñar soluciones adaptadas a cada perfil.

La segmentación también mejora la eficiencia de los recursos empresariales, ya que

permite dirigir campañas específicas a segmentos con mayor probabilidad de conversión, minimizar el riesgo de abandono y aumentar la rentabilidad de los clientes existentes. De acuerdo con Wedel y Kamakura (2000), una segmentación bien estructurada constituye la base para una propuesta de valor centrada en el cliente y sustentada en datos objetivos.

En este proyecto, la segmentación se implementa a través de técnicas de clusterización no supervisada, que permiten descubrir agrupaciones naturales dentro de la población sin necesidad de etiquetas predefinidas. Este enfoque facilita el análisis exploratorio, la identificación de patrones ocultos y la generación de insights accionables para la toma de decisiones estratégicas.

### ***Big Data***

El concepto de Big Data hace referencia al manejo y análisis de grandes volúmenes de datos que superan la capacidad de los sistemas tradicionales de procesamiento. Esta característica se define comúnmente por las 5V: volumen, velocidad, variedad, veracidad y valor (Gandomi & Haider, 2015). En el sector financiero, el Big Data permite identificar tendencias, prever comportamientos, detectar fraudes y optimizar procesos internos mediante el análisis continuo de flujos masivos de información.

Según Marr (2018), el uso de Big Data en entornos empresariales ha transformado la forma en que las organizaciones toman decisiones, pasando de modelos reactivos a modelos predictivos y prescriptivos. Las compañías que logran integrar eficazmente estas tecnologías pueden ofrecer servicios personalizados, mejorar la experiencia del cliente y ganar eficiencia operativa.

En el caso de las compañías que administran activos, el Big Data es clave para capturar información de millones de interacciones de clientes, analizar historiales de inversión y construir perfiles de comportamiento detallados. Esta capacidad de análisis permite adaptar la oferta de

productos a las necesidades reales del cliente, anticipar cambios de comportamiento y fortalecer la relación con el usuario mediante una experiencia más relevante y personalizada (Fosso Wamba et al., 2015).

### ***Modelos de Aprendizaje Automático***

El aprendizaje automático (machine learning) es una rama de la inteligencia artificial que permite a los sistemas aprender automáticamente a partir de los datos, sin necesidad de ser programados explícitamente. Estos modelos son capaces de identificar patrones, realizar predicciones y tomar decisiones basadas en información histórica, y se han vuelto esenciales en sectores que manejan grandes volúmenes de datos, como el financiero (Goodfellow, Bengio & Courville, 2016).

En el contexto del análisis de clientes, los modelos de machine learning permiten detectar comportamientos atípicos, prever el abandono de clientes, recomendar productos personalizados y optimizar carteras de inversión. Jordan y Mitchell (2015) destacan que una de las principales ventajas del aprendizaje automático es su capacidad para adaptarse a contextos dinámicos, ajustando sus predicciones a medida que se actualizan los datos.

Dentro de este proyecto, el uso de modelos no supervisados, como la clusterización, ha sido fundamental para descubrir estructuras ocultas en la base de datos. Además, los resultados obtenidos pueden complementarse con modelos supervisados para desarrollar soluciones analíticas completas que incluyan tanto clasificación como predicción.

### ***Clusterización en Machine Learning***

La clusterización es una técnica de aprendizaje automático no supervisado que permite agrupar observaciones en subconjuntos (clúster) basados en su similitud, sin que exista una variable objetivo predeterminada. Esta técnica es particularmente útil cuando se busca explorar y

segmentar datos complejos, como los comportamientos financieros de clientes, permitiendo encontrar patrones que de otra forma pasarían desapercibidos (Han, Kamber & Pei, 2011).

Entre los algoritmos de clusterización más conocidos se encuentran el K-means, el análisis jerárquico y DBSCAN. El algoritmo K-means, por ejemplo, busca particionar el conjunto de datos en K grupos tal que cada observación pertenezca al grupo más cercano a su centroide, optimizando así la homogeneidad interna (Géron, 2019). Este enfoque es ideal cuando se trabaja con variables cuantitativas y grandes volúmenes de datos, como los utilizados en este proyecto.

En el ámbito de las compañías que administran activos, la clusterización permite descubrir segmentos de clientes con características y necesidades similares, lo que resulta crucial para diseñar estrategias de inversión diferenciadas, mejorar la asignación de recursos comerciales y elevar la satisfacción del cliente. Según Xu y Tian (2015), este tipo de segmentación basada en datos empíricos fortalece la toma de decisiones estratégicas al proporcionar una comprensión más profunda de la base de clientes.

## **Metodología**

### **Tipo de Estudio**

El presente estudio es de tipo cuantitativo, no experimental y exploratorio-descriptivo. Se fundamenta en el análisis de bases de datos que reflejan las características demográficas y financieras de clientes en el segmento de pensiones voluntarias de varios países latinoamericanos. El enfoque cuantitativo permite identificar patrones y relaciones estadísticas entre variables clave, como edad, ingresos, valor invertido y tipo de inversión, sin manipular ninguna de ellas.

A su vez, la naturaleza exploratoria del estudio radica en el objetivo de comprender los comportamientos y hábitos financieros de los clientes, sin partir de hipótesis rígidas, sino más bien permitiendo que los datos revelen agrupaciones o tendencias relevantes. El carácter descriptivo complementa este enfoque, ya que busca caracterizar perfiles de clientes y aportar insumos que sirvan para la toma de decisiones estratégicas dentro de una compañía de administración de activos.

### **Método**

La metodología aplicada combina técnicas de análisis estadístico exploratorio y modelamiento no supervisado. En primer lugar, se realizó una limpieza y unificación de cinco bases de datos representativas de los países analizados. Posteriormente, se filtró la muestra para incluir únicamente a los clientes con inversiones en productos voluntarios.

El análisis comenzó con una exploración de variables clave mediante medidas de tendencia central, dispersión y visualizaciones (histogramas, diagramas de dispersión y cruces bivariados), con el fin de identificar patrones preliminares de comportamiento.

Posteriormente, se aplicó una técnica de clustering utilizando el algoritmo K-means,

precedida por la codificación y estandarización de variables relevantes. Para determinar el número óptimo de clústeres, se empleó el método del codo, evaluando el punto de inflexión en la inercia total. Una vez agrupados los clientes, se caracterizó cada clúster con base en su promedio de edad, ingresos, valor invertido y otras variables categóricas como género y tipo de inversión.

Este enfoque permitió identificar segmentos de clientes con comportamientos financieros similares, lo cual sirvió como base para formular recomendaciones estratégicas personalizadas orientadas a mejorar la gestión y el diseño de productos de inversión.

A continuación, se presenta una descripción general del proceso metodológico llevado a cabo en el desarrollo del presente proyecto aplicado.

### **Carga de los Datos**

El proyecto inició con la recopilación de bases de datos en formato .csv correspondientes a clientes de pensiones voluntarias en cinco países: Uruguay, Chile, Colombia, México y Perú. Para su procesamiento, se utilizó el lenguaje Python y la librería pandas, asegurando una lectura adecuada mediante el manejo de codificaciones y delimitadores. A cada archivo se le añadió una columna adicional llamada "Pais", facilitando así los análisis comparativos entre geografías. Todos los archivos fueron integrados en un único Dataframe denominado df\_all.

### **Limpieza y Filtrado**

Posteriormente, se ejecutó una limpieza de los datos, que incluyó la eliminación de valores nulos y registros inconsistentes. Luego, se filtró el conjunto de datos para conservar exclusivamente los registros correspondientes al negocio de Pensiones Voluntarias, foco principal del análisis, lo que redujo el universo a tres países: Chile, México y Colombia. Adicionalmente, se excluyó la variable Rentabilidad, por depender de factores exógenos no atribuibles directamente al comportamiento del cliente.

## **Análisis Exploratorio y Transformación**

Con la base consolidada (`df_filtrado`), se realizó un análisis exploratorio de datos (EDA), examinando estadísticas descriptivas, distribuciones y relaciones entre variables clave.

Asimismo, se aplicaron técnicas de transformación como escalado y codificación de variables categóricas para preparar los datos para el modelado.

## **Modelado y Segmentación**

Finalmente, se implementó un modelo de clustering utilizando el algoritmo K-means, con el fin de segmentar a los clientes en grupos homogéneos según características financieras y demográficas. Este proceso incluyó la selección del número óptimo de clústers mediante el método del codo y la posterior interpretación de los perfiles resultantes.

Este flujo metodológico permitió abordar de manera estructurada los objetivos planteados, garantizando coherencia entre las fases técnicas y el propósito analítico del estudio.

## Resultados

### Resultados Objetivo 1

Analizar los comportamientos financieros actuales de los clientes, utilizando datos históricos demográficos y transaccionales disponibles en una compañía del sector de administración de activos.

#### *Análisis Exploratorio Inicial*

Con el propósito de conocer mejor la composición y características generales de la base de datos, se llevó a cabo un análisis exploratorio de datos (EDA, por sus siglas en inglés), que constituye una etapa fundamental en todo proyecto de ciencia de datos. Esta fase tuvo como objetivo principal identificar patrones generales, valores atípicos, relaciones preliminares entre variables y posibles sesgos en la información, elementos esenciales para la posterior definición de estrategias analíticas y modelos predictivos.

Inicialmente, se consolidó una base de datos integrada (df\_filtrado) que incluye exclusivamente registros del negocio de Pensiones Voluntarias, filtrando previamente por país (Chile, México y Colombia) y eliminando valores nulos o inconsistentes. A partir de este dataset limpio, se seleccionaron las variables más representativas para el análisis, tanto numéricas como categóricas.

Las variables numéricas (Edad, Ingresos en USD, Valor Invertido y NPS) fueron evaluadas mediante medidas de tendencia central (media, mediana), dispersión (desviación estándar, rango intercuartílico), y distribución (histogramas y gráficos de densidad). Este proceso permitió detectar comportamientos como asimetrías, curtosis y outliers, fundamentales para ajustar correctamente los modelos posteriores. Por ejemplo, el Valor Invertido mostró una distribución fuertemente sesgada hacia la derecha, lo cual fue confirmado gráficamente mediante

histogramas con colas largas.

En paralelo, se exploraron variables categóricas (como Nivel Educativo, Tipo de Producto, Frecuencia de Inversión y Tipo de Inversión) utilizando gráficos de barras, tablas de frecuencias y cruces con otras variables relevantes, lo que permitió comprender la prevalencia de cada categoría, así como identificar diferencias según el país o grupo etario.

A continuación, se detallan las variables definitivas con las cuales se procede a hacer el análisis exploratorio inicial:

**Tabla 1**

*Descripción de Variables*

Nombre Variable	Descripción	Tipo de dato
Pais	País de residencia del cliente	Categórica
Negocio	Tipo de negocio o segmento (Obligatorio, Voluntario, etc.)	Categórica
Rango_Edad	Grupo etario al que pertenece el cliente (Ej. 18-25, 26-35, etc.)	Categórica
Edad	Edad exacta del cliente	Numérica
Genero	Género del cliente (Masculino, Femenino, Otro)	Categórica
Nivel_Educativo	Nivel educativo alcanzado (Básica, Secundaria, Universitaria, Posgrado, etc.)	Categórica
Ingresos_USD	Ingresos mensuales del cliente en dólares estadounidenses	Numérica
Tipo_Inversion	Tipo de inversión preferida (Ej. conservadora, moderada, agresiva)	Categórica
Valor_Invertido	Monto total invertido por el cliente en USD	Numérica
Frecuencia_Inversion	Frecuencia con la que el cliente realiza inversiones (Mensual, Trimestral, etc.)	Categórica
Tipo_Producto	Tipo de producto financiero adquirido, en este caso está	Categórica

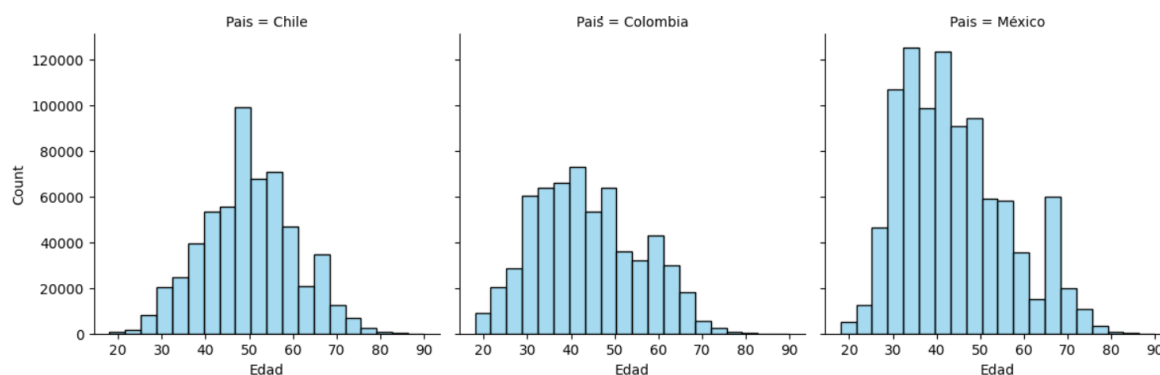
Nombre Variable	Descripción	Tipo de dato
NPS	variable nos indica si el tipo de producto es con complemento pensión o si es flexible, Net Promoter Score (medida de lealtad del cliente, escala -100 a 100)	Numérica

En la Tabla 1 se presenta la descripción de las variables contenidas en el dataset utilizado para el análisis. Esta tabla incluye tanto variables demográficas (como Rango\_Edad, Género, País, Nivel Educativo), como variables de comportamiento financiero (Ingresos\_USD, Tipo de Inversión, Valor Invertido, Frecuencia de Inversión, entre otras). Su identificación clara y su clasificación por tipo de dato (categórica o numérica) resultan fundamentales para seleccionar las técnicas de análisis adecuadas.

Luego se procede a analizar comportamientos y distribuciones de variables relevantes para el análisis como lo son la edad del cliente y la cantidad:

### Figura 1

#### *Distribución de Clientes (por Edad y País)*



En la Figura 1, se presenta la distribución de clientes según la edad y el país, lo que

permite visualizar la estructura demográfica de la cartera analizada. Se identifica una distribución tipo campana (normal), lo cual sugiere una concentración significativa de clientes en rangos de edad media, con menores proporciones en los extremos jóvenes y mayores.

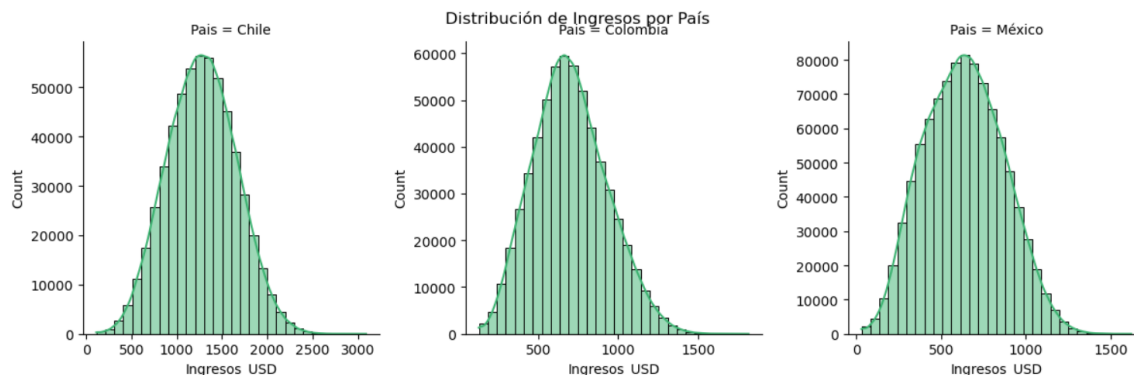
Esta forma de distribución es útil para entender en qué etapas del ciclo de vida financiera se encuentran los clientes, aspecto clave para diseñar estrategias diferenciadas. Si bien la tendencia general es consistente en los tres países incluidos (Chile, Colombia y México), se observan ligeras variaciones en la media y la dispersión, reflejo de las particularidades sociodemográficas de cada contexto.

Las edades promedio por país se resumen en la Tabla 2, permitiendo una comparación más precisa entre mercados. Estos resultados sirven como base para interpretar cómo varía el comportamiento financiero en función de la edad y cómo puede ajustarse la oferta según la etapa de vida predominante en cada país.

**Tabla 2**

*Edad Promedio de la Cartera*

País	Edad Promedio de la Cartera
Chile	49.81
Colombia	43.56
México	44.02

**Figura 2***Distribución de Ingresos (por Clientes y País)*

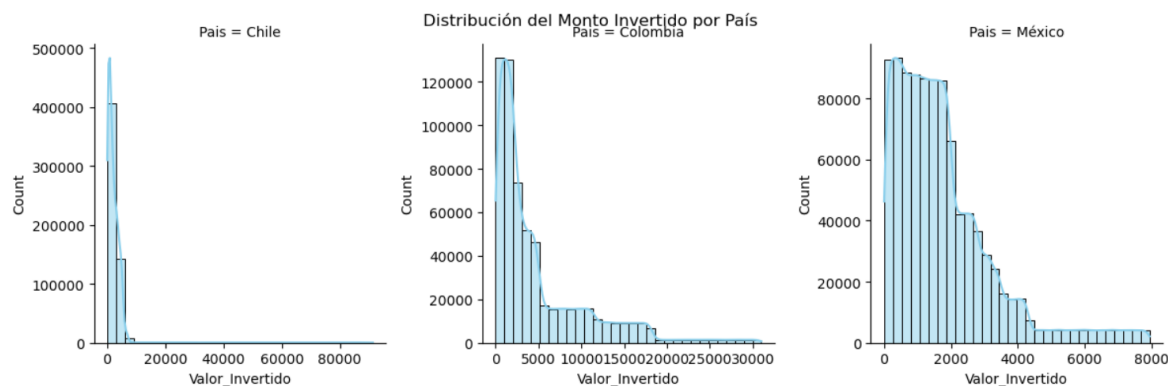
En la Figura 2, se representa la distribución de los ingresos mensuales (en USD) de los clientes por país, lo que permite identificar patrones importantes asociados a la capacidad económica de los distintos segmentos.

En general, se observa que la mayoría de los clientes se concentran en niveles de ingresos medios, sin embargo, la forma de la distribución varía significativamente entre países:

- Chile presenta una distribución más simétrica, con una mayor dispersión en niveles altos de ingreso.
- Colombia muestra una concentración más marcada en ingresos medios-bajos, con una caída más rápida hacia los niveles altos.
- México exhibe una distribución levemente sesgada hacia la izquierda, reflejando una mayor proporción de clientes con ingresos relativamente bajos.

### Figura 3

#### *Distribución de Valor Invertido (por Clientes y País)*



En la Figura 3 se representa la distribución del Valor Invertido por cliente, expresado en dólares estadounidenses (USD), para los tres países considerados en el análisis: Chile, Colombia y México. Esta figura es clave para comprender las diferencias en comportamiento financiero respecto al ahorro individual.

Se observa que las tres distribuciones son notoriamente asimétricas, con una alta concentración de clientes en los tramos de inversión más bajos. Este patrón es particularmente evidente en México y Colombia, donde la mayoría de los clientes mantiene niveles bajos de inversión, lo cual puede estar asociado a factores estructurales como la menor educación financiera, el limitado acceso a productos especializados o una menor capacidad de ahorro mensual.

Por otro lado, tanto Chile como Colombia presentan colas alargadas hacia la derecha, lo que indica la presencia de un segmento reducido de clientes que manejan montos significativamente altos invertidos. Este hallazgo sugiere la existencia de nichos de alto patrimonio dentro de estas carteras, lo cual representa una oportunidad para diseñar estrategias diferenciadas y productos financieros personalizados.

Estas observaciones, en conjunto con los resultados previos sobre ingresos (ver Figura 2), sugieren que, aunque los ingresos de los clientes tienden a concentrarse en rangos medios, las decisiones de inversión y la capacidad real de acumular capital varían de forma importante entre países. Factores como la madurez del sistema previsional, el nivel de asesoría recibida y la cultura de ahorro previsional podrían explicar parte de esta heterogeneidad.

## **Resultados Objetivo 2**

Identificar patrones de inversión y factores clave que influyen en las decisiones financieras de los clientes

Con el fin de avanzar en la comprensión de los hábitos financieros de los clientes, se llevó a cabo un análisis cruzado entre variables sociodemográficas y de comportamiento de inversión. Este ejercicio tuvo como propósito descubrir patrones diferenciales que pudieran orientar tanto el diseño de ofertas personalizadas como la construcción de una posterior segmentación con algoritmos no supervisados.

### ***Justificación Metodológica de los Cruces***

El análisis cruzado entre variables sociodemográficas (como edad, nivel educativo, género y país) y comportamentales (como frecuencia y valor promedio de inversión) se fundamenta en la necesidad de comprender en mayor profundidad los factores que influyen en los hábitos financieros de los clientes. Esta estrategia permite no solo describir la población, sino también identificar patrones de comportamiento heterogéneos que no serían visibles mediante un análisis univariado.

Desde el enfoque del marketing relacional y la economía conductual, se ha demostrado que las decisiones financieras no dependen únicamente del ingreso o la capacidad económica, sino que están fuertemente influenciadas por factores demográficos, socioculturales y

contextuales (Lusardi & Mitchell, 2014; OECD, 2020). Por ejemplo, la edad puede estar asociada con diferentes horizontes de planificación (ahorro vs. desacumulación), el nivel educativo con la comprensión de instrumentos financieros, el género con sesgos de comportamiento o brechas de acceso, y el país con factores macroeconómicos, regulatorios y de madurez del sistema previsional.

Además, en estudios de segmentación y comportamiento del consumidor financiero, el cruce entre dimensiones sociodemográficas y decisiones de inversión ha sido una técnica ampliamente utilizada para generar perfiles accionables, que luego alimentan estrategias de marketing, asesoría y desarrollo de productos (Kotler & Keller, 2012; Wedel & Kamakura, 2000).

En el caso del presente proyecto, estos cruces permiten construir una base empírica sólida que justifica el uso posterior de técnicas de clusterización. Es decir, al identificar que existen diferencias significativas entre los grupos en términos de comportamiento financiero, se evidencia que no todos los clientes responden de la misma forma ante las mismas condiciones, y por lo tanto resulta pertinente aplicar algoritmos no supervisados para descubrir agrupaciones naturales más precisas y con sentido estratégico.

A continuación, los resultados obtenidos del cruce de variables demográficas con variables de comportamientos financieros:

**Tabla 3***Patrones de Inversión y Factores Clave por Variable*

Variable	Categoría	Frecuencia más común	Valor Prom. Invertido (USD)
Rango_Edad	menor_25	Semestral	1,110.07
	26_35	Anual	1,178.67
	36_45	Anual	2,399.49
	46_jub	Semestral	3,059.67
	mayor_jub	Semestral	7,659.86
Nivel_Educativo	Básico	Anual	2,567.87
	Medio	Semestral	3,069.82
	Superior	Semestral	2,920.11
Género	Femenino	Semestral	2,686.79
	Masculino	Semestral	3,113.83
Tipo Inversión	Conservadora	Anual	3,151.61
	Moderada	Anual	2,779.92
	Agresiva	Semestral	2,912.44
País	Chile	Mensual	2,830.65
	Colombia	Semestral	4,883.61
	México	Anual	1,760.12

La Tabla 3 resume los principales patrones de comportamiento financiero de los clientes, obtenidos al cruzar variables demográficas y de perfil de inversión con dos indicadores clave: la

frecuencia más común de inversión y el valor promedio invertido (USD).

El análisis por dimensiones clave (edad, educación, género, tipo de inversión y país) evidencia que los hábitos financieros y decisiones de inversión de los clientes están fuertemente influenciados por factores demográficos y contextuales, más allá del ingreso per se.

- En términos de edad, los clientes mayores de 46 años tienden a invertir montos significativamente más altos y a hacerlo con menor frecuencia (anual o semestral), lo que puede reflejar una planificación a largo plazo o mayor estabilidad financiera.
- Por nivel educativo, quienes tienen educación media o superior invierten más y prefieren modalidades semestrales, posiblemente por mayor exposición a conceptos financieros y mejor estructuración de su ahorro.
- El género también presenta diferencias: los hombres invierten en promedio más que las mujeres, aunque ambos coinciden en preferir inversiones semestrales.
- En cuanto al tipo de fondo, los clientes conservadores tienden a realizar mayores inversiones, lo que puede interpretarse como una preferencia por seguridad con capacidad de mayor aporte.
- Finalmente, el contexto geográfico tiene un peso relevante: Colombia muestra el valor promedio más alto invertido, mientras que México presenta una cartera más conservadora en términos de valor y frecuencia, lo cual puede estar vinculado a factores económicos o regulatorios locales.

Estos hallazgos respaldan la importancia de segmentar y personalizar estrategias de inversión según las características y patrones observados, lo que permitirá a la compañía mejorar la experiencia del cliente y optimizar su oferta de valor.

### **Resultados Objetivo 3**

Desarrollar un modelo de segmentación que clasifique a los clientes según sus características y necesidades financieras.

#### ***Preprocesamiento para Clustering***

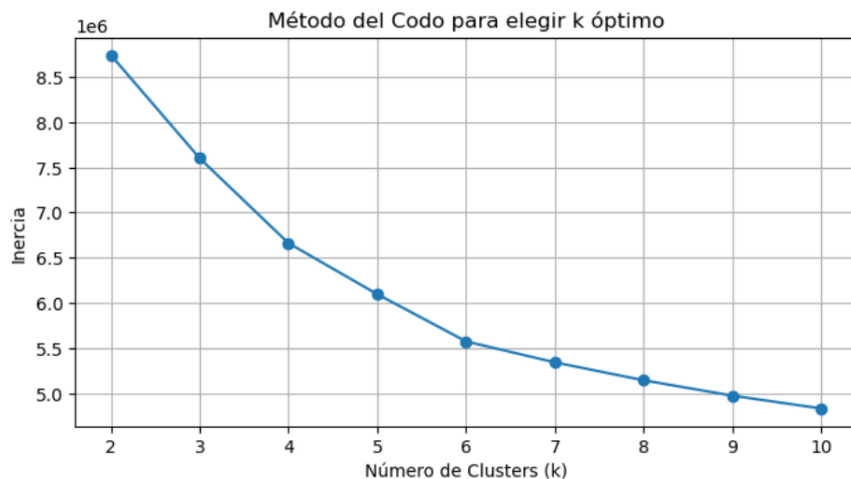
Previo a aplicar técnicas de clustering, se seleccionaron cuidadosamente las variables más relevantes desde el punto de vista analítico. Las variables categóricas fueron transformadas mediante One-Hot Encoding, convirtiéndolas en variables numéricas binarias, lo que permite que los algoritmos de clustering funcionen adecuadamente. Por su parte, las variables numéricas fueron estandarizadas con StandardScaler para eliminar el efecto de la escala y asegurar que todas las dimensiones tengan la misma importancia en la distancia euclidiana utilizada por el algoritmo KMeans.

#### ***Método del Codo para Determinar el Número Óptimo de Clústers***

Para decidir cuántos grupos (clúster) formar, se aplicó el método del codo, que consiste en graficar la suma de los errores cuadráticos dentro de cada grupo (WCSS) en función del número de clúster. Se analizó el punto de inflexión donde la mejora empieza a ser marginal, lo cual indica el valor óptimo de  $k$ . El gráfico mostró una inflexión clara en  $k = 6$ , lo que sugiere que dividir a los clientes en seis segmentos maximiza la explicación de la varianza sin sobreajustar el modelo.

## Figura 4

### *Método del Codo*



La Figura 4 muestra el resultado de aplicar el método del codo (elbow method) para determinar el número óptimo de clúster (k) en el modelo de segmentación basado en el algoritmo K-means.

### *Aplicación del Algoritmo KMeans*

Con el valor óptimo de k definido, se procedió a entrenar un modelo KMeans utilizando la matriz de variables estandarizadas. Cada cliente fue asignado a uno de los seis clúster generados, y esta información se incorporó como una nueva columna en el Dataframe. Este proceso permite agrupar clientes con características similares y proporciona una base sólida para formular estrategias personalizadas.

### *Análisis de los Clústers*

Se calculó un resumen estadístico por cada clúster, incluyendo la edad media, ingreso promedio, valor invertido, NPS medio, género predominante y tipo de inversión dominante. Este análisis cuantitativo permitió identificar diferencias marcadas entre los grupos, por ejemplo,

clientes jóvenes con ingresos bajos pero buena rentabilidad, o adultos mayores con patrimonios altos y estrategias conservadoras. Estos hallazgos son cruciales para construir perfiles diferenciados que reflejen las verdaderas necesidades y comportamientos del cliente.

**Tabla 4**

*Clústers*

Clúster	Edad Prom.	Ingresos Prom. (USD)	Valor Invertido Prom. (USD)	NPS Prom.	Tipo Inversión
0	57.0	711.1	14,688.5	40.2	Conservadora
1	37.0	552.3	1,943.2	40.1	Moderada
2	58.6	684.7	2,141.4	40.1	Moderada
3	50.7	1,497.4	2,203.1	40.2	Agresiva
4	37.5	875.5	1,878.4	40.2	Moderada
5	68.6	1,286.7	64,842.4	40.3	Agresiva

La Tabla 4 presenta un resumen consolidado de los seis clústers identificados a partir de la aplicación del algoritmo K-means, utilizando variables clave como edad, ingresos, valor invertido, tipo de inversión, frecuencia y nivel educativo. Esta tabla recoge los promedios y características predominantes de cada grupo, lo que permite una interpretación cualitativa y estratégica de los perfiles de clientes.

El modelo de segmentación permitió identificar seis perfiles financieros claramente diferenciados. Por ejemplo, el clúster 5 se destaca por incluir clientes de mayor edad con alta capacidad de inversión, perfil agresivo y el mayor valor invertido, posiblemente conformado por jubilados o altos patrimonios. Por otro lado, los clústeres 1 y 4 agrupan a personas más jóvenes,

de ingresos moderados y menor monto invertido, lo que indica un público aún en etapa de acumulación temprana.

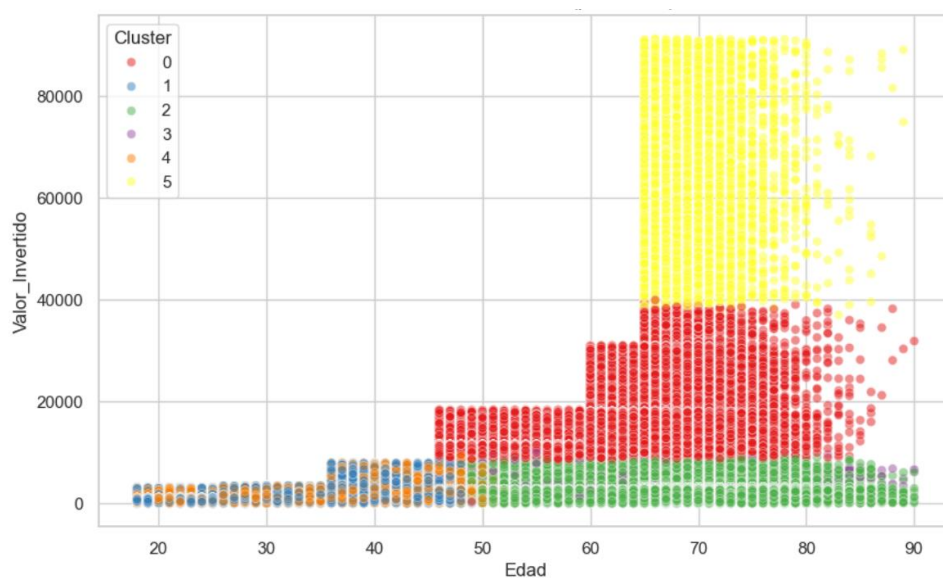
Además, se evidencia cómo el tipo de inversión preferido varía según la edad y el capital disponible, reflejando distintos niveles de aversión al riesgo y objetivos financieros. Estos hallazgos ofrecen una base sólida para diseñar estrategias diferenciadas por segmento, tanto en oferta de productos como en acompañamiento personalizado.

### ***Gráficas de los Clústers***

Para facilitar la interpretación visual de los resultados, se elaboró un gráfico tipo radar que compara los seis clúster en dimensiones clave normalizadas (Edad, Ingresos, Valor Invertido, NPS). Esta representación permite ver con claridad qué características destacan en cada grupo y dónde se ubican en relación con los demás. Es una herramienta útil tanto para la presentación de resultados como para la toma de decisiones basada en datos.

## **Figura 5**

### *Edad vs Valor Invertido (por Clúster)*

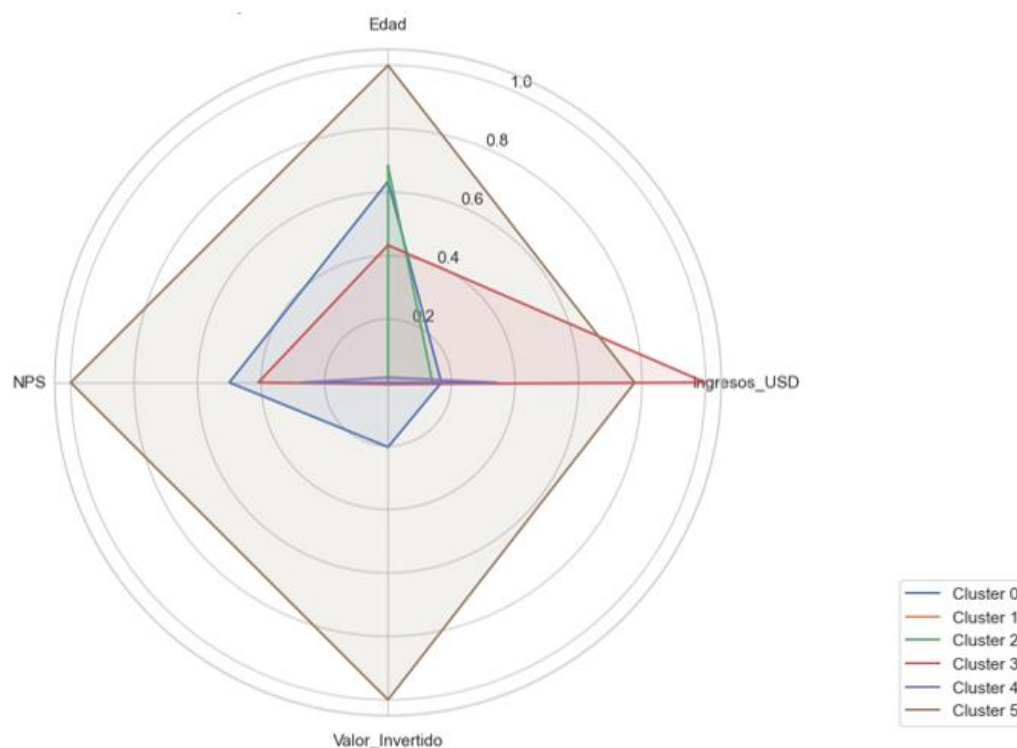


La Figura 5 ilustra la relación entre la edad de los clientes y el valor invertido promedio, desagregada por los seis clústers definidos previamente en el modelo de segmentación. Este gráfico permite visualizar claramente cómo se comporta cada grupo en función de estas dos variables clave para el negocio: la etapa del ciclo de vida financiero (edad) y la capacidad o hábito de acumulación de capital (valor invertido).

La visualización muestra una clara relación entre la edad y el valor invertido. Los clústeres con mayor edad concentran también los valores de inversión más altos, en especial el clúster 5 (amarillo), que representa clientes mayores con inversiones significativamente superiores. En contraste, los clústeres de menor edad (como el 1 y 4) muestran un comportamiento de inversión mucho más moderado, lo que evidencia una progresión natural en el ciclo de vida financiera de los clientes.

## Figura 6

### *Comparación de Clústers (Radar Chart)*



La Figura 6 presenta un Radar Chart que resume y compara las principales características de los seis clústers identificados en el proceso de segmentación, utilizando variables clave como edad, ingresos, valor invertido, NPS (Net Promoter Score) y frecuencia de inversión.

Este tipo de gráfico permite visualizar de forma intuitiva las diferencias y similitudes entre los grupos, facilitando su interpretación cualitativa y estratégica. El gráfico radar permite visualizar de forma comparativa las características promedio de cada clúster. Destaca el clúster 5 como el perfil más robusto, con los valores más altos en edad, ingresos, valor invertido y NPS, lo que sugiere un segmento consolidado y satisfecho. Por otro lado, clústeres como el 4 o 2 presentan bajos valores en la mayoría de las variables, representando probablemente perfiles más jóvenes o con menor capacidad de inversión. Esta comparación facilita la identificación de grupos prioritarios para estrategias diferenciadas.

#### **Resultados Objetivo 4**

Proponer recomendaciones personalizadas, basadas en los análisis realizados, que permitan optimizar las estrategias de inversión y mejorar la experiencia del cliente

#### ***Perfilamiento de los Clúster y Recomendaciones***

A partir de los resultados del clustering, se elaboraron descripciones cualitativas de cada grupo. Por ejemplo:

- Clúster 0: Prejubilados con alto patrimonio y perfil conservador: clientes que se aproximan a la jubilación y buscan estabilidad.
- Clúster 1: Jóvenes de bajos ingresos y ahorro incipiente: perfil de entrada que requiere educación financiera y estrategias automáticas de ahorro.
- Clúster 2: Adultos mayores con bajo valor acumulado: oportunidad para revisar estrategias pasadas y orientar hacia soluciones de protección.

- Clúster 3: Adultos con alta capacidad de ingreso y apetito de riesgo: perfil ideal para diversificación y portafolios sofisticados.
- Clúster 4: Jóvenes estructurados con ingresos medios: potencial para escalar hacia productos más avanzados.
- Clúster 5: Seniors con muy alto patrimonio y perfil agresivo: enfoque en productos de desacumulación y asesoría premium.

Para cada segmento se diseñaron recomendaciones específicas, considerando su situación financiera, etapa de vida y tolerancia al riesgo. Estas recomendaciones incluyen desde campañas de educación y automatización, hasta estrategias de diversificación o protección patrimonial. En conjunto, permiten avanzar hacia el diseño de una oferta de valor mucho más alineada con el cliente.

#### ***Clúster 0 Prejubilados con Alto Patrimonio y Perfil Conservador***

Clientes cercanos a la edad de jubilación, con buena acumulación de capital y una marcada preferencia por opciones conservadoras. A continuación, algunas recomendaciones:

- Diseñar estrategias de transición a la jubilación: Se sugiere ofrecer productos con baja volatilidad y beneficios fiscales, así como herramientas digitales de proyección de pensión y simuladores de retiro.
- Fortalecer el acompañamiento personalizado: Estos clientes valoran la estabilidad, por lo que una asesoría recurrente puede mejorar la percepción de seguridad.
- Promover la planificación intergeneracional: Dada su posición financiera, es ideal abrir la conversación sobre herencia y educación financiera para familiares.

#### ***Clúster 1 Jóvenes de Bajos Ingresos y Ahorro Incipiente***

Segmento joven con ingresos limitados, bajos montos invertidos y frecuencia de

inversión moderada a baja. A continuación, algunas recomendaciones:

- Implementar mecanismos de ahorro automatizado: Activar funcionalidades como “redondeo de compras” o “débitos programados” puede impulsar el hábito de inversión de forma casi imperceptible.
- Diseñar campañas de educación financiera gamificada: Dado su perfil digital, este grupo se beneficia de contenidos simples, visuales y accesibles desde el celular.
- Ofrecer productos flexibles con bajo umbral de entrada: Evitar barreras económicas al ingreso es clave para no excluir a estos futuros inversionistas.

### ***Clúster 2 Adultos Mayores con Bajo Valor Acumulado***

Personas de mayor edad con un nivel de acumulación bajo, probablemente con rezagos históricos en su planeación financiera. A continuación, algunas recomendaciones:

- Realizar asesorías de recuperación patrimonial: Se deben ofrecer alternativas de protección o desacumulación responsable según su horizonte financiero.
- Generar contenidos educativos sobre riesgos y cobertura: Muchos podrían no estar completamente informados sobre sus opciones ante la jubilación.
- Evaluar ayudas o incentivos solidarios si es población vulnerable: Explorar esquemas donde el Estado o la empresa puedan complementar sus aportes.

### ***Clúster 3 Adultos con Alta Capacidad de Ingreso y Apetito de Riesgo***

Adultos económicamente activos con ingresos altos y predisposición a inversiones agresivas. A continuación, algunas recomendaciones:

- Ampliar el portafolio de productos sofisticados: Ofrecerles acceso a mercados internacionales, derivados, fondos alternativos o estructurados.
- Fortalecer la asesoría en construcción de portafolios personalizados: Este grupo

suele valorar herramientas de análisis, dashboards y *reporting* de alta calidad.

- Incentivar su fidelización con contenido exclusivo: Webinars con expertos, acceso a herramientas premium y comunidades privadas pueden reforzar el vínculo.

#### ***Clúster 4 Jóvenes Estructurados con Ingresos Medios***

Jóvenes con cierta estructura financiera, ingresos estables y nivel medio de inversión. A continuación, algunas recomendaciones:

- Fomentar la evolución hacia productos más complejos: Con acompañamiento gradual, pueden migrar a fondos balanceados, mixtos o con componentes ESG.
- Impulsar el uso de metas financieras en plataformas digitales: Esto refuerza el vínculo emocional con sus objetivos (vivienda, educación, viajes).
- Implementar incentivos por lealtad o referidos: Programas que reconozcan su constancia podrían potenciar el engagement.

#### ***Clúster 5 Seniors con Muy Alto Patrimonio y Perfil Agresivo***

Personas mayores, con gran capital acumulado y tolerancia al riesgo inusualmente alta para su edad. A continuación, algunas recomendaciones:

- Desarrollar estrategias de desacumulación planificada: Ofrecer soluciones de renta periódica, fideicomisos o gestión patrimonial multigeneracional.
- Asignar asesores patrimoniales senior con trato preferente: Este perfil espera atención exclusiva, disponibilidad y propuestas a la medida.
- Ofrecer acceso a productos alternativos y privados: Como bienes raíces, fondos de *private equity* o inversiones sostenibles de alto impacto.

La segmentación por clúster permite ir más allá de la demografía tradicional, identificando comportamientos reales y motivaciones financieras. A partir de estos perfiles, es

posible diseñar estrategias de valor personalizadas, aumentar la retención de clientes y generar mayor rentabilidad en el largo plazo, con enfoque tanto comercial como social.

## Conclusiones

Este proyecto de grado evidenció que la aplicación de técnicas de ciencia de datos y analítica avanzada permite a una compañía del sector de administración de activos obtener una comprensión profunda y estructurada de los comportamientos financieros de sus clientes. A partir de una base de datos consolidada y representativa del segmento de pensiones voluntarias en tres países latinoamericanos, fue posible identificar hallazgos relevantes que respondieron de manera clara a la pregunta de investigación planteada.

El análisis exploratorio reveló patrones de comportamiento vinculados a variables como edad, ingresos, valor invertido, frecuencia de inversión, tipo de producto y perfil de riesgo. Se observó, por ejemplo, que los clientes de mayor edad tienden a realizar inversiones más elevadas y con menor frecuencia, mientras que los perfiles más jóvenes presentan montos de inversión más bajos, hábitos menos estructurados y menor diversificación. Estas diferencias subrayan la necesidad de segmentar la base de clientes no únicamente por criterios tradicionales, sino considerando sus prácticas financieras reales.

Con base en estos patrones, se construyó un modelo de segmentación utilizando el algoritmo K-means, validado mediante el método del codo. Este modelo permitió identificar seis clústeres con características bien diferenciadas, los cuales fueron interpretados cualitativamente para orientar su aplicación estratégica. Cada grupo fue asociado a perfiles concretos, tales como prejubilados conservadores, jóvenes con ahorro incipiente o adultos con alto ingreso y apetito de riesgo, lo que facilitó la formulación de recomendaciones específicas en materia de productos, estrategias comerciales y acompañamiento financiero.

Los resultados confirman que la ciencia de datos es una herramienta clave para avanzar hacia una gestión personalizada, centrada en el cliente y basada en evidencia. Esta orientación

permite anticipar necesidades, detectar oportunidades de mejora y diseñar ofertas alineadas con el ciclo de vida financiera del cliente, promoviendo así una mayor fidelización, crecimiento patrimonial y diferenciación competitiva.

En conjunto, los hallazgos obtenidos reafirman el potencial de los enfoques analíticos como pilar fundamental para la transformación del sector financiero hacia modelos más inteligentes, humanos y sostenibles.

## Recomendaciones

A partir del desarrollo del proyecto y los resultados obtenidos, se proponen las siguientes recomendaciones orientadas tanto a la organización como a futuros ejercicios académicos similares.

En primer lugar, se recomienda que la compañía adopte de manera sistemática modelos de segmentación basados en ciencia de datos, como el clustering, para identificar perfiles de clientes con comportamientos financieros diferenciados. Esta segmentación permite diseñar estrategias comerciales, productos y servicios más alineados con las verdaderas necesidades y expectativas de cada grupo, mejorando así la experiencia del cliente y la eficiencia operativa.

En segundo lugar, se sugiere desarrollar propuestas de valor específicas para cada perfil identificado. Por ejemplo, los segmentos jóvenes con bajo nivel de ahorro requieren acompañamiento digital, automatización y educación financiera, mientras que los adultos con alta acumulación o cercanos a la jubilación pueden beneficiarse de asesoría especializada y productos de protección o desacumulación. Esta diferenciación es clave para lograr relaciones de largo plazo y mayor rentabilidad por cliente.

Como tercera recomendación, se destaca la importancia de fortalecer las capacidades internas de la compañía en el uso de herramientas analíticas y visualización de datos. Contar con equipos multidisciplinarios que comprendan tanto el negocio como las herramientas de análisis es fundamental para traducir los datos en decisiones estratégicas.

Desde la perspectiva metodológica, se recomienda asegurar una adecuada planeación del proceso de recolección y depuración de datos. En este tipo de proyectos, la calidad de la información es un factor crítico, por lo que es necesario establecer protocolos claros para la limpieza, estandarización y validación de los datos desde las primeras etapas.

También se sugiere no limitar el análisis únicamente al modelamiento, sino combinar herramientas exploratorias con algoritmos de aprendizaje automático. Esta combinación permite no solo describir los datos, sino también descubrir patrones y generar hallazgos útiles para la acción.

Por último, se recomienda documentar cuidadosamente cada etapa del proceso, desde la formulación del problema hasta la interpretación de los resultados. Esto no solo mejora la claridad del informe final, sino que permite que el análisis sea replicable y útil para otros proyectos dentro de la organización o para futuras investigaciones académicas.

### Referencias Bibliográficas

- AFP Habitat. (2022). *Memoria Anual 2022*. <https://www.afphabitat.cl/docs/default-source/memorias/memoria-integrada-2022.pdf>
- Banco de la República. (2023). *Informe de estabilidad financiera - Primer semestre de 2023*. <https://www.banrep.gov.co/es/publicaciones/informe-estabilidad-financiera-primer-semestre-2023>
- Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Press.
- Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How ‘big data’ can make big impact: Findings from a systematic review and a research agenda. *International Journal of Production Economics*, 165, 234–246. <https://doi.org/10.1016/j.ijpe.2014.12.034>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O’Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kotler, P., & Keller, K. L. (2012). *Marketing Management* (14th ed.). Pearson.

- Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5–44.  
<https://doi.org/10.1257/jel.52.1.5>
- Marr, B. (2018). *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. Wiley.
- Mayer-Schönberger, V., & Cukier, K. (2014). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.
- OECD. (2020). *OECD/INFE 2020 International Survey of Adult Financial Literacy*. Organisation for Economic Co-operation and Development.  
<https://www.oecd.org/financial/education/oecd-infe-2020-international-survey-of-adult-financial-literacy.htm>
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- Siegel, E. (2016). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* (Revised and Updated Edition). Wiley.
- Superintendencia Financiera de Colombia. (2024). *Estadísticas dinámicas del sistema financiero colombiano*. <https://www.superfinanciera.gov.co/estadisticas-dinamicas>
- Wedel, M., & Kamakura, W. A. (2000). *Market Segmentation: Conceptual and Methodological Foundations* (2nd ed.). Springer Science & Business Media.
- Xu, R., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2, 165–193. <https://doi.org/10.1007/s40745-015-0040-1>

## Apéndices

### Apéndice A

#### *Enlace Código Python Proyecto de Grado*

A continuación, se presenta el enlace en donde se encuentra ubicado el código completo para el desarrollo de este proyecto de grado aplicado:

[https://drive.google.com/file/d/1\\_919e9xbZATeyMDBOg5-CnfaVi9y3g9\\_/view?usp=sharing](https://drive.google.com/file/d/1_919e9xbZATeyMDBOg5-CnfaVi9y3g9_/view?usp=sharing)

### Apéndice B

#### *Enlace Video Sustentación Proyecto de Grado*

Además, se adjunta también en enlace en donde se encuentra alojado el video de sustentación de este proyecto de grado aplicado:

<https://drive.google.com/file/d/1QuxGBfEbgqPj7yY2bMbRlsDugC6Xba3M/view?usp=sharin>

g