



**Universidad Internacional San Isidro Labrador, Costa Rica**

Especialización de Data Science.

**PROYECTO No. 1: Data Mining o Minería de Datos**

**Tema :**

Modelo de segmentación de clientes en el mercado (dataset: Mall Customer Segmentation Data).

**Profesor: Dr. Samuel Saldaña Valenzuela.**

**Estudiante Jhonnatan Arroyo Viquez .**

## Tabla de contenido

<b>Introducción .....</b>	<b>3</b>
Objetivo general y específicos .....	7
Justificación .....	8
<b>Marco Metodológico .....</b>	<b>9</b>
Marco Teórico .....	10
Implementación Practica .....	12
Conclusiones .....	13
Bibliografía .....	35

## **Introducción**

**Con este trabajo se pretende dar muestras del conocimiento adquirido en el curso de data science , si bien es cierto no desde una perspectiva de tecnicismo o grado académico de Universidad, el curso libre me permite enriquecer el vocabulario y familiarizar con el tema tan complejo, para desde otros cursos o técnicos no empezar de cero y se pueda dar un mejor desenvolvimiento en el tema.**

**La intención es utilizar este proyecto como una oportunidad para aplicar los conocimientos aprendidos de manera práctica, permitiendo una mayor profundización en el tema y evitando partir desde cero en futuros cursos o proyectos técnicos. Al abordar un problema real o simulado con las habilidades adquiridas, se**

**pretende consolidar la comprensión de los conceptos y técnicas clave del campo de la ciencia de datos, lo que, a su vez, contribuirá a un mejor desenvolvimiento en el tema en contextos profesionales o académicos futuros.**

## **Pregunta de Investigación**

**¿Cómo varían los patrones de compra y las preferencias entre diferentes grupos de edad de los clientes del centro comercial, y cómo pueden estas diferencias informar estrategias de marketing dirigidas a cada segmento de edad?**

**Van Variando de acuerdo a un segmento generacional que reúne grupos con intereses a fines y a la vez se diferencian de otros**

**grupos , debido a tiempos y tendencias diferentes de actualizaciones según sus años .**

**Baby Boomers (nacidos entre 1946 y 1964):**

**Prefieren la experiencia de compra en la tienda física.**

**Valoran la calidad, la seguridad y la atención al cliente.**

**Son menos influenciables por las tendencias y más leales a las marcas establecidas.**

**Generación X (nacidos entre 1965 y 1980):**

**Valoran la conveniencia y la calidad.**

**Son propensos a la lealtad a la marca, pero también buscan ofertas.**

**Prefieren una experiencia de compra sin complicaciones.**

**Millennials (nacidos entre 1981 y 1996):**

**Buscan experiencias únicas y personalizadas.**

**Prefieren marcas con valores y propósitos claros.**

**Están dispuestos a gastar más en productos de calidad.**

**Generación Z (nacidos después de 1997):**

**Son digitales nativas y prefieren la conveniencia de comprar en línea.**

**Valoran la autenticidad y la sostenibilidad en las marcas.**

**De esta manera se ve una afinidad de gustos de las distintas generaciones por segmentos y se da una oferta de servicios más personalizada, dando un objetivo específico a sus necesidades por categoría de población generacional.**

**Permitiendo a su vez tener mapeado como el gusto va cambiando a medida pasan los años .**

### **Objetivo General:**

**Aplicar conocimientos curso Data Science.**

#### **Objetivos específicos**

**Realizar un análisis exploratorio de los datos para comprender la distribución y características de los clientes en el conjunto de datos.**

**Aplicar técnicas de segmentación de clientes para dividir la base de clientes en grupos homogéneos según características , comportamentales y de preferencias.**

## **Justificación**

**Este trabajo investigativo busca sentar las bases conceptuales, necesarias para guiar el análisis de datos subsiguiente en las fases posteriores del modelo CRISPDM. Al comprender completamente los objetivos y las necesidades del negocio en esta fase inicial, se garantizará que el análisis de datos posterior esté alineado con los objetivos estratégicos de la empresa y pueda generar insights significativos y aplicables para mejorar la segmentación de clientes y las estrategias de marketing en el centro comercial.**

**En la fase de comprensión del negocio, se analiza la situación actual del centro comercial, sus objetivos estratégicos y las necesidades de sus clientes. Esto implica comprender las tendencias de compra, los hábitos de consumo y las preferencias de los clientes.**

**Al comprender mejor el negocio, se pueden identificar áreas de oportunidad para mejorar la experiencia del cliente y optimizar las operaciones del centro comercial.**

**La idea es segmentar los grupos y encontrar características iguales o desiguales que permitan, etiquetar y asociar ciertos patrones**



para las estrategias o estabilidad del negocio, gracias a la información interpretada y descubierta.

## **Marco Metodológico**

**Se utilizarán para su debida aplicación**

**Selección de Datos:** Selecciona los datos necesarios para el análisis, como registros de transacciones y perfiles de clientes, asegurando que sean representativos y completos.

- **Preparación de Datos:** Limpia y prepara los datos eliminando valores faltantes o inconsistentes, y asegurándote de que estén en un formato adecuado para el análisis.
- **Análisis Exploratorio de Datos:** Explora los datos utilizando gráficos y estadísticas simples para entender su distribución y patrones básicos.
- **Segmentación de Clientes:** Utiliza una herramienta de segmentación simple, como K-Means, para dividir a los clientes en grupos basados en características demográficas y de comportamiento.
- **Análisis de Patrones de Compra y Preferencias de Productos:** Examina los patrones de compra y preferencias de productos dentro de cada segmento utilizando tablas de frecuencia y gráficos de barras.
- **Desarrollo de Estrategias de Marketing Personalizadas:** Utiliza los insights obtenidos para diseñar estrategias de marketing específicas para cada grupo de clientes, como ofertas personalizadas o campañas dirigidas.

## **Marco Teórico**

**El Customer Service un término en inglés se refiere a los servicios prestados al cliente en todas las etapas del proceso de compra. Es un área cada vez más amplia, directamente relacionada con la calidad de la experiencia.**

**Algunos factores que influyen:**

**Es la facilidad de acceso al centro comercial, así como la disponibilidad de estacionamiento y servicios de transporte, pueden afectar la conveniencia y la satisfacción del cliente.**

**Los clientes esperan sentirse seguros y protegidos mientras visitan el centro comercial. La presencia de personal de seguridad, la iluminación adecuada, la vigilancia por cámaras y otras medidas de seguridad contribuyen a crear un entorno seguro para los clientes.**

**El ambiente físico del centro comercial, incluyendo la limpieza, el mantenimiento y la comodidad de las instalaciones, puede influir en la experiencia del cliente y en su percepción sobre el nivel de servicio ofrecido.**

**La calidad y la actitud del personal de atención al cliente tienen un impacto significativo en la experiencia del cliente. La cortesía, la amabilidad, la paciencia y la capacidad para resolver problemas son aspectos clave que influyen en la percepción del cliente sobre el servicio recibido.**

## Conceptos Teóricos Fundamentales

**Segmentación de Mercado:** Este concepto se refiere a la división de un mercado en grupos más pequeños y homogéneos de clientes que comparten características similares. La segmentación de mercado es fundamental para comprender las necesidades y preferencias de los diferentes grupos de clientes y para adaptar las estrategias de marketing de manera efectiva.

**Comportamiento del Consumidor:** Se refiere al estudio de cómo los individuos toman decisiones de compra y cómo influyen diversos factores en estas decisiones, como las características demográficas, psicográficas y situacionales. Comprender el comportamiento del consumidor es crucial para desarrollar estrategias de marketing que satisfagan las necesidades y deseos de los clientes.

**El Data Mining, o minería de datos,** desempeña un papel fundamental en el Modelo de Segmentación de Clientes en el mercado, especialmente cuando se utiliza un conjunto de datos como el "Mall Customer Segmentation Data".

**Análisis de Datos:** Este concepto se refiere al proceso de examinar, limpiar, transformar y modelar datos con el objetivo de descubrir información útil, llegar a conclusiones y apoyar la toma de decisiones.

**El análisis de datos implica usar técnicas estadísticas y de minería de datos para identificar patrones y relaciones en los datos.**

- **Personalización en Marketing:** Se refiere a la adaptación de las estrategias de marketing a las características y preferencias individuales de los clientes. La personalización en marketing implica el uso de datos sobre el comportamiento y las preferencias de los clientes para ofrecer mensajes, productos o servicios personalizados que sean relevantes y atractivos para cada cliente.

## **Implementación Práctica**

**La implementación práctica de este trabajo de investigación implicaría una serie de pasos, desde la recolección y análisis de datos hasta el desarrollo e implementación de estrategias de marketing personalizadas basadas en los insights obtenidos del análisis de datos.**

**El objetivo final sería mejorar la comprensión de los clientes y aumentar la efectividad de las estrategias de marketing en el mercado minorista.**

**Utilizando los datos disponibles para generar información que permita conocer el negocio , generar opiniones y tomar decisiones enfocadas en el mejoramiento del negocio.**

## **Conclusiones:**

### **Primera Parte**

**En Primera instancia se identifica el tipo de negocio los servicios que ofrece.**

**Se obtiene en automático que las características de estos clientes son diferentes considerando que al tener varios segmentos edades , generacionales los gustos y modas de estos clientes son diversas.**

**Por lo que la oferta y demanda de este negocio se ve beneficiada de diferentes puntos de vista en gustos generacionales.**

**El negocio es un mall me da entender que tiene una ubicación geográfica muy transitada ,independientemente de la provincia o estado así no tenga el dato exacto , ya que este tipo de negocio se subdivide en oferta de servicios variados como áreas de comidas, parqueos, servicios financieros, entretenimiento, supermercados,etc, lo que alimenta la idea de que sus comerciantes no se arriesgarían a un lugar poco transitado.**

**Basado en la diversidad de negocios que contiene permite la apertura de visitantes de todas las etapas de la vida podemos sugerir desde el niño hasta el adulto mayor.**

**De esta manera podemos apuntar a enfoques de comportamientos distinguidos de acuerdo a la clasificación de generación.**

**Se cuenta con los elementos o variables de Genero.**

**Edad.**

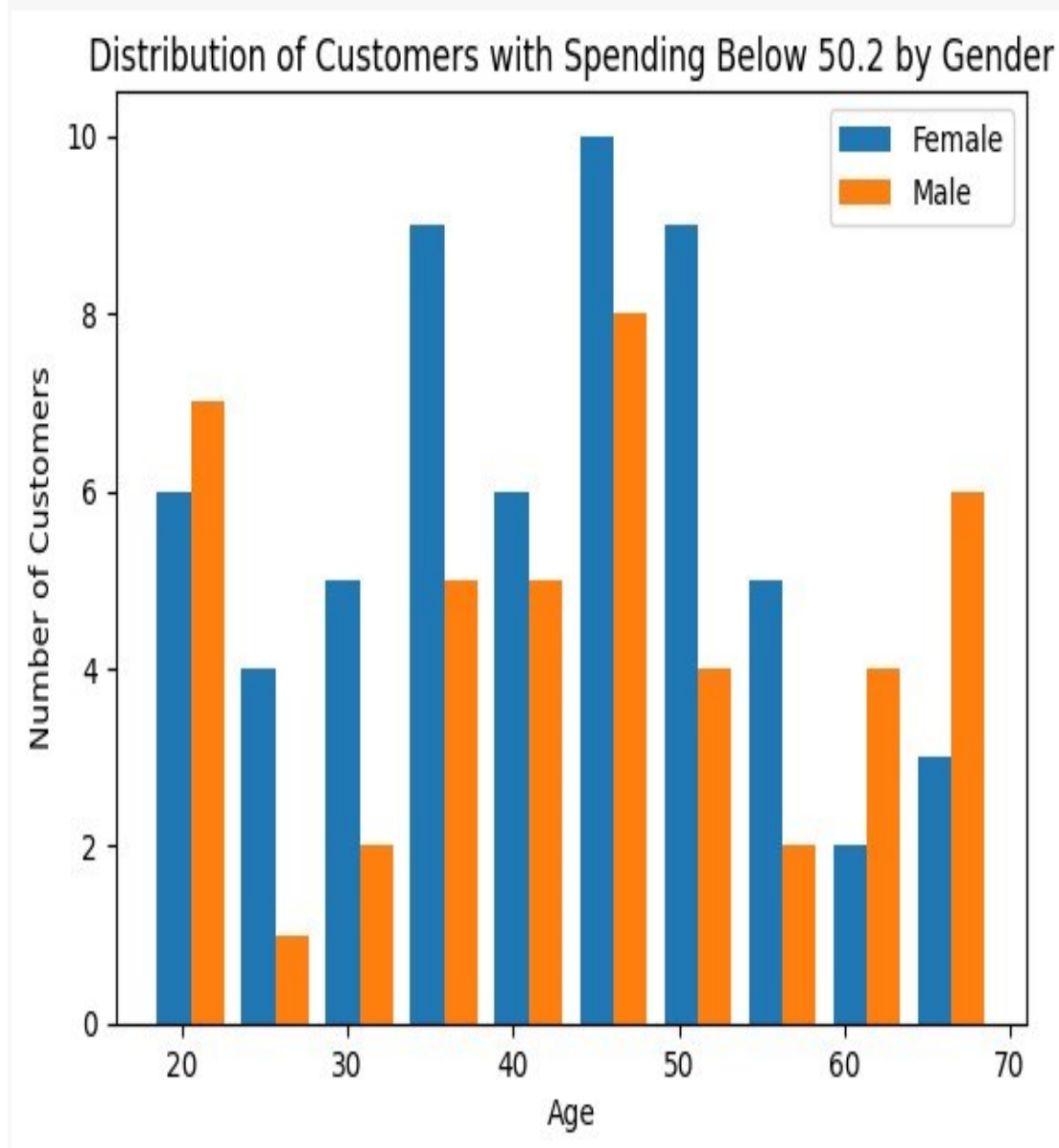
**Ingresos y gastos.**

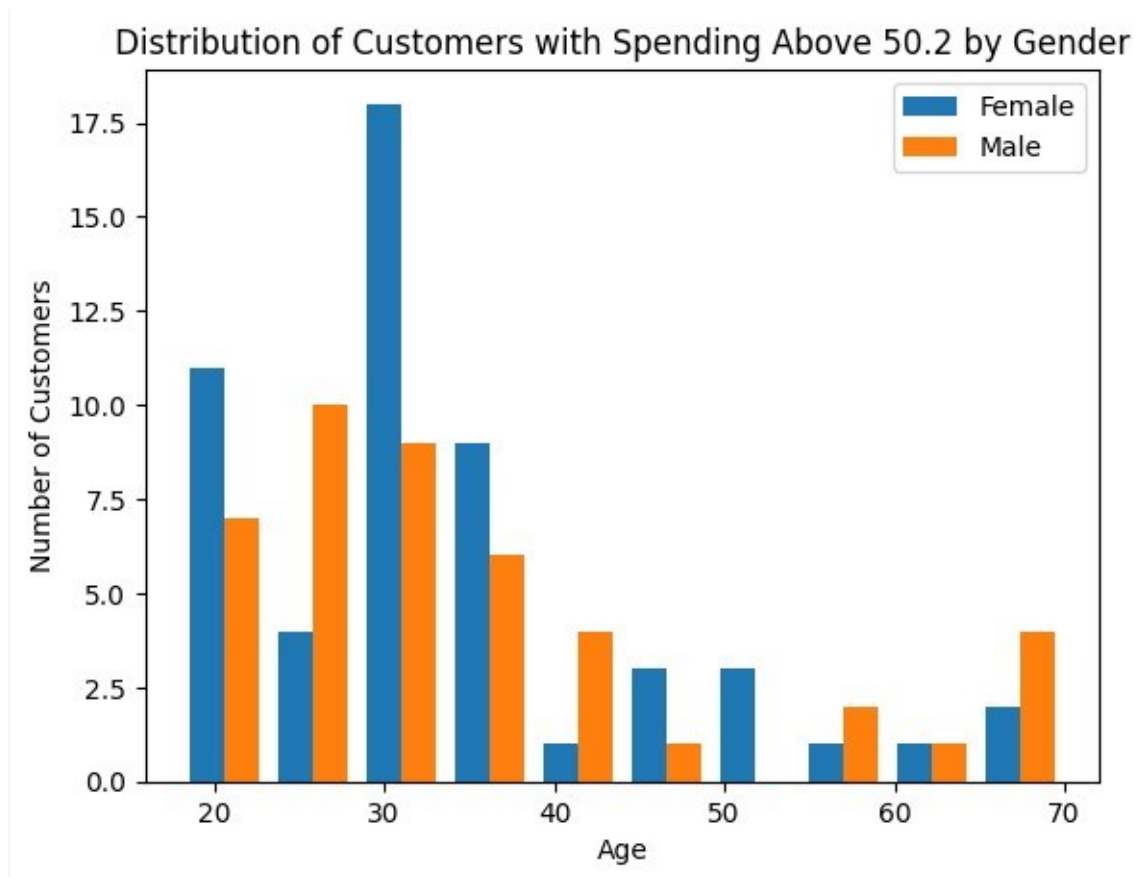
**Total, genero mujer 112. Total genero hombre 88.**

**Gasto total: 10040**

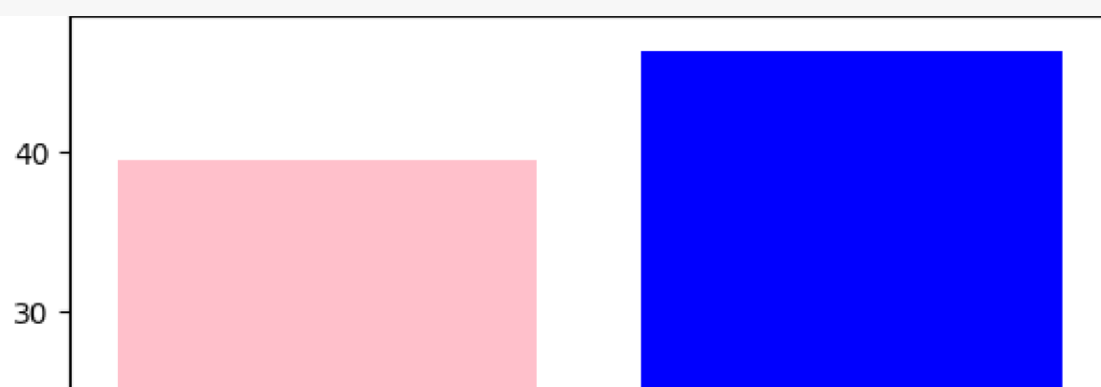
**Clientes totales: 200**

**Gasto promedio por cliente: 50.      2**

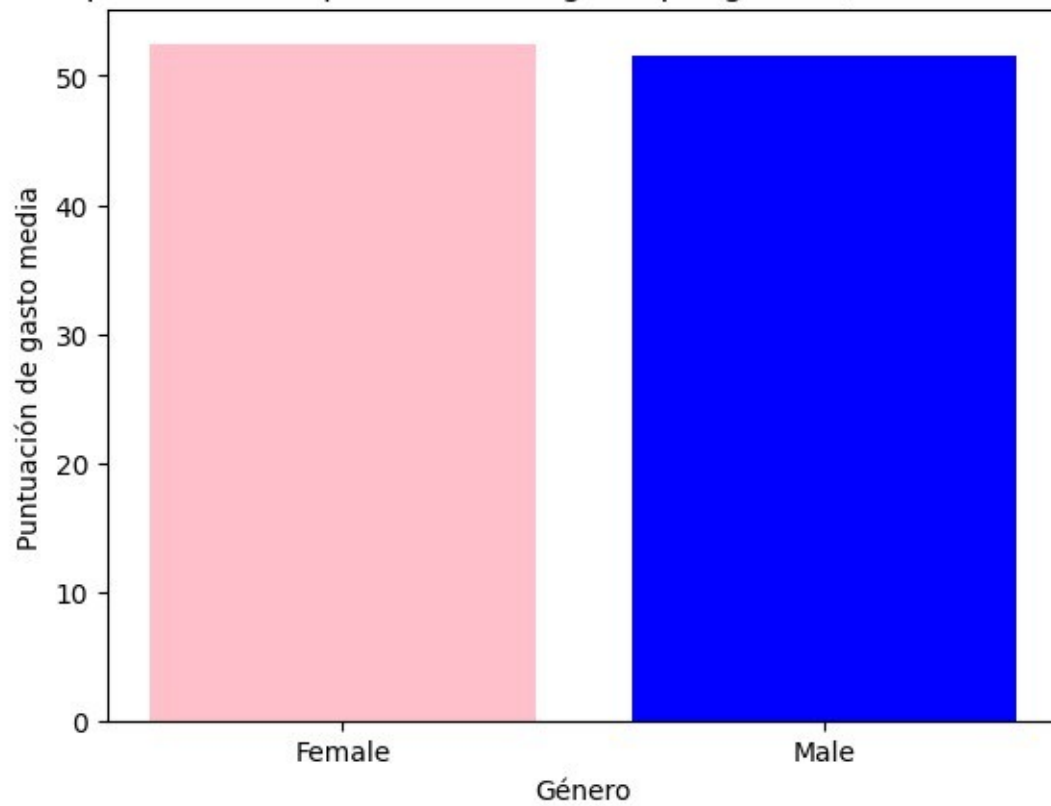




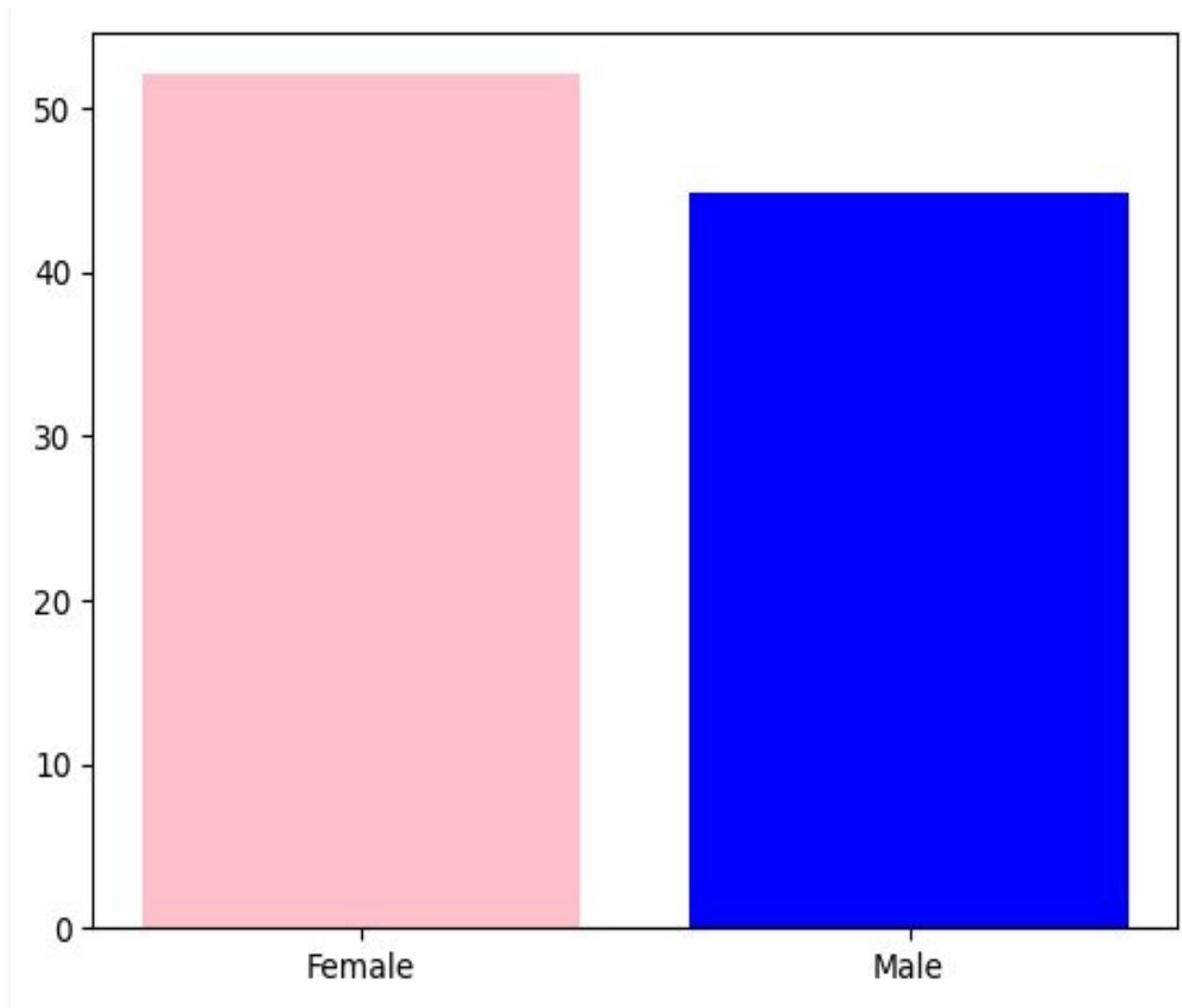
Se hacen algunas muestras aleatorias varias para determinar diferencias y persistencia en la estabilidad del resultado .



Comparación de la puntuación de gasto por género (muestra aleatoria)

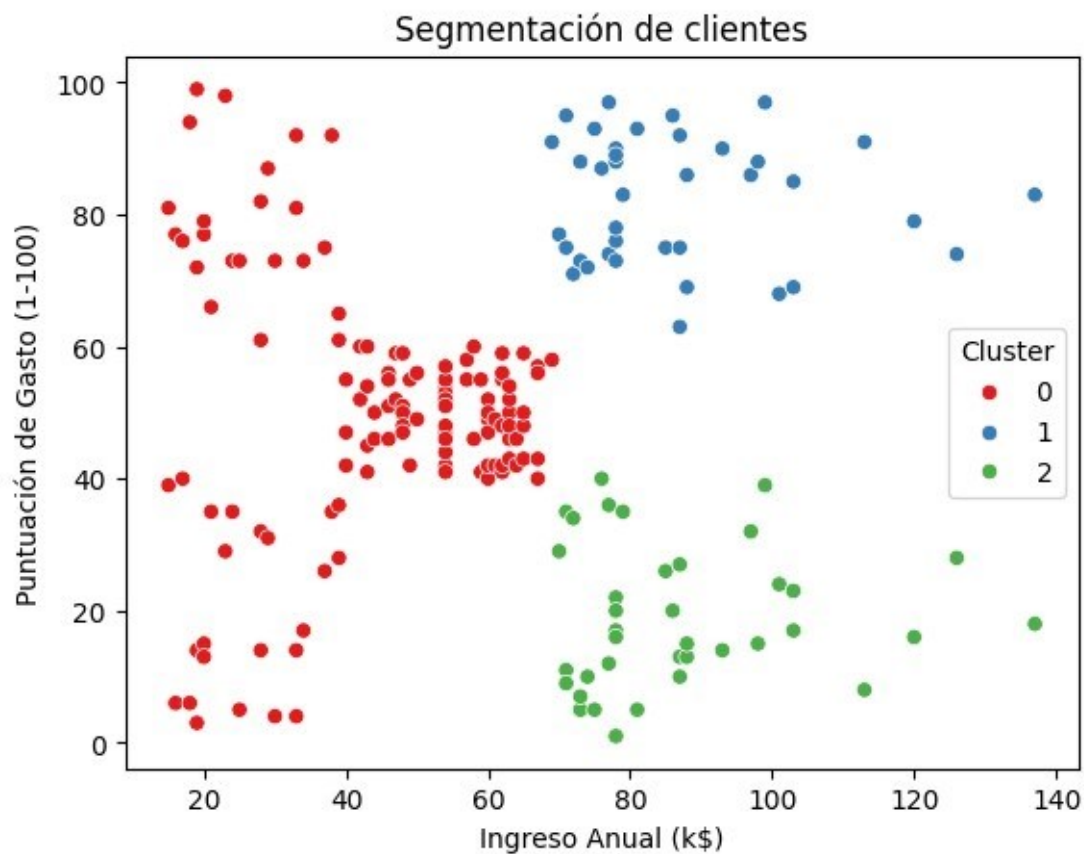






## Segunda parte

**Aplicación técnica K-meas** Se Define la cantidad de cluster bajo el método codo.



**Se definen 3 cluster.**

**Centroides de cada clúster:**

**[[44.15447154 49.82926829]**

**[86.53846154 82.12820513]**

**[87. 18.63157895]]**

**1-Primer cluster tiene un ingreso anual moderado y una puntuación de gasto media.**

**2-Segundo cluster tiene un ingreso anual alto y una puntuación de gasto alta.**

**3-Muestra una combinación peculiar de características (Ingreso Alto, Gasto Bajo).**

**Cluster 0: Este cluster tiene la mayor cantidad de clientes, con un total de 123. Esto sugiere que hay un grupo considerable de clientes que comparten características similares entre sí y diferentes de los otros clusters.**

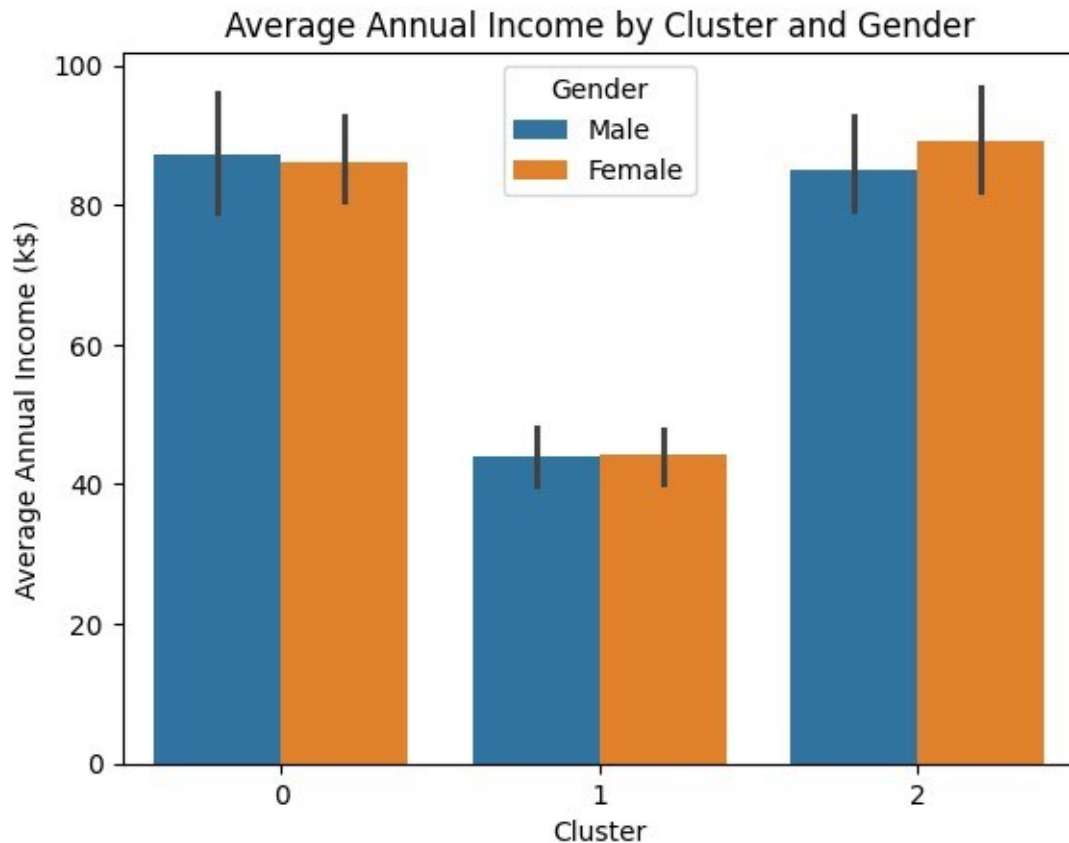
**Cluster 1: Con 38 clientes asignados, este cluster tiene una cantidad similar de clientes que el Cluster 2. Sin embargo, tiene uno menos que el Cluster 2. Esto sugiere que hay otro grupo de clientes que comparten características diferentes a los otros clusters.**

**Cluster 2: Con solo 39 clientes asignados, este cluster es el más pequeño en términos de tamaño. Esto indica que hay un grupo más pequeño de clientes que exhiben un conjunto específico de características distintivas que los separan de los demás grupos.**



**En términos de segmentación por género, aunque si hay diferencias no es muy significativo.**

**Lo que genera pesquisas homogéneas en términos de géneros y edad.**



**Se clasifica los géneros referenciados las edades con un enfoque generacional**

**Generación Z (nacidos entre 1997 y 2012).**

**Millennials o Generación Y (nacidos entre 1981 y 1996).**

**Generación X (nacidos entre 1965 y 1980).**

**Baby Boomers (nacidos entre 1946 y 1964).**

**para visualizar la cantidad de población y a su vez asociar con gustos o patrones comunes que repercuten en ingreso y gasto del mall.**

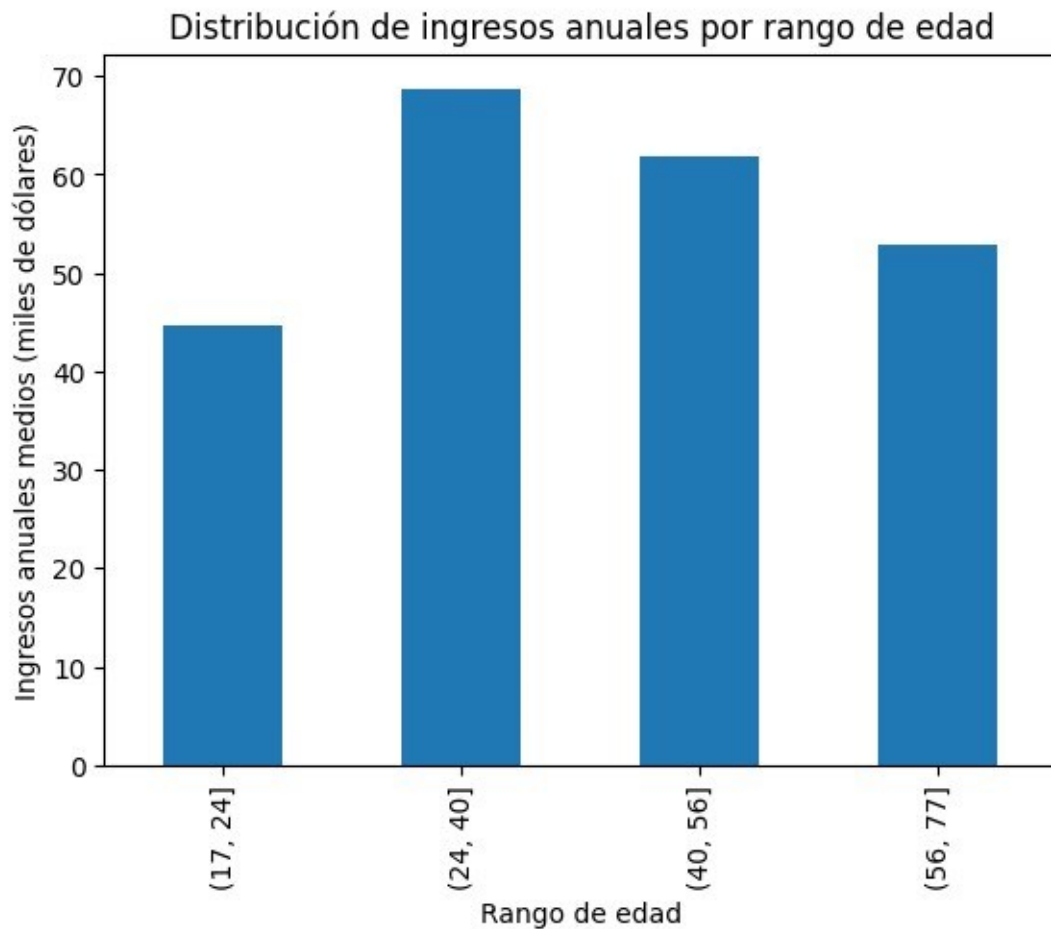
**Se identifica que la generación de 24 a 40 años es quien tiene mayores ingresos en esta etapa.**

**También esta generación es la que más gasta en el mall .**

### **Annual Income (k\$) Spending Score (1-100)**

#### **Gender Age**

<b>Female</b>	(17, 24]	42.473684
		59.947368
	(24, 40]	65.420000
		60.980000
	(40, 56]	61.909091
		35.606061
	(56, 77]	51.500000
		40.800000
<b>Male</b>	(17, 24]	47.250000
		51.875000
	(24, 40]	73.162162
		59.891892
	(40, 56]	61.764706
		32.529412
	(56, 77]	53.500000
		37.222222



**Definimos la cantidad de grupos según edad Age (17, 24] (24, 40] (40, 56] (56, 77]**

**Gender**

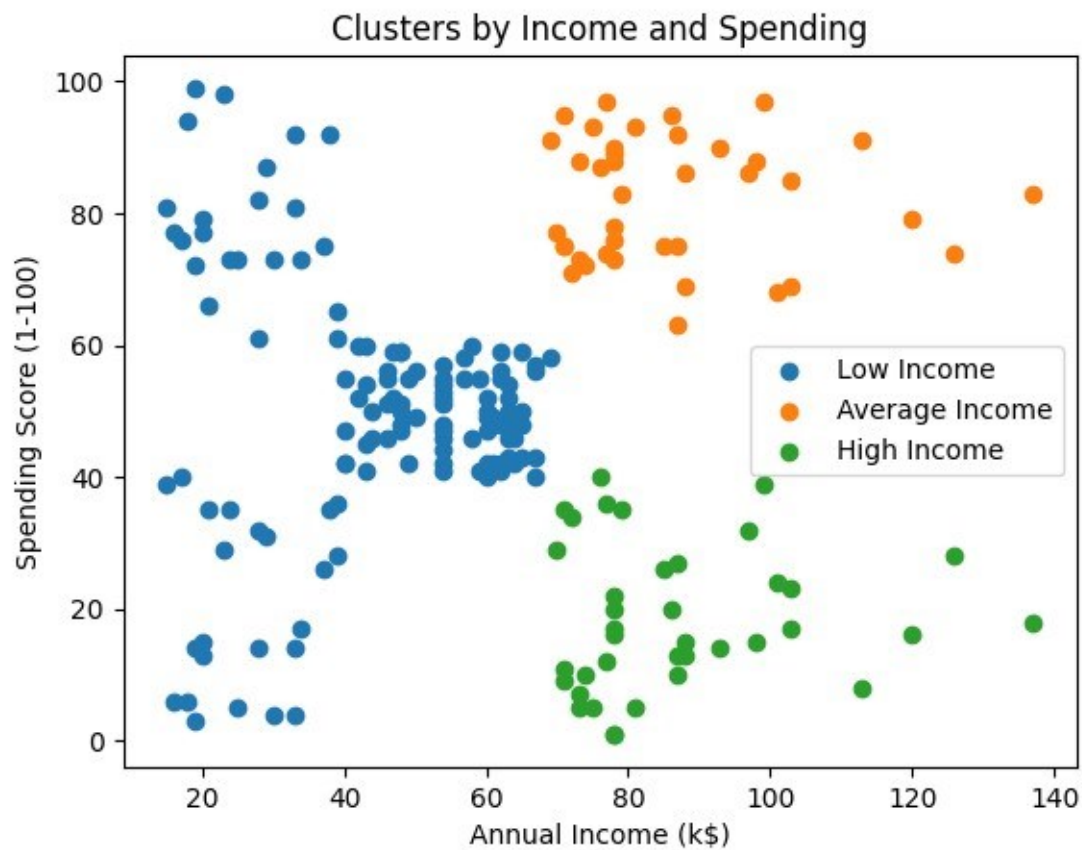
<b>Female</b>	<b>19</b>	<b>50</b>	<b>33</b>	<b>10</b>
<b>Male</b>	<b>16</b>	<b>37</b>	<b>17</b>	<b>18</b>

**Aunque no muy significativo, pero siempre marcando la diferencia mayoritaria al género femenino.**

**Porcentaje de clientes femeninos: 56.00%**

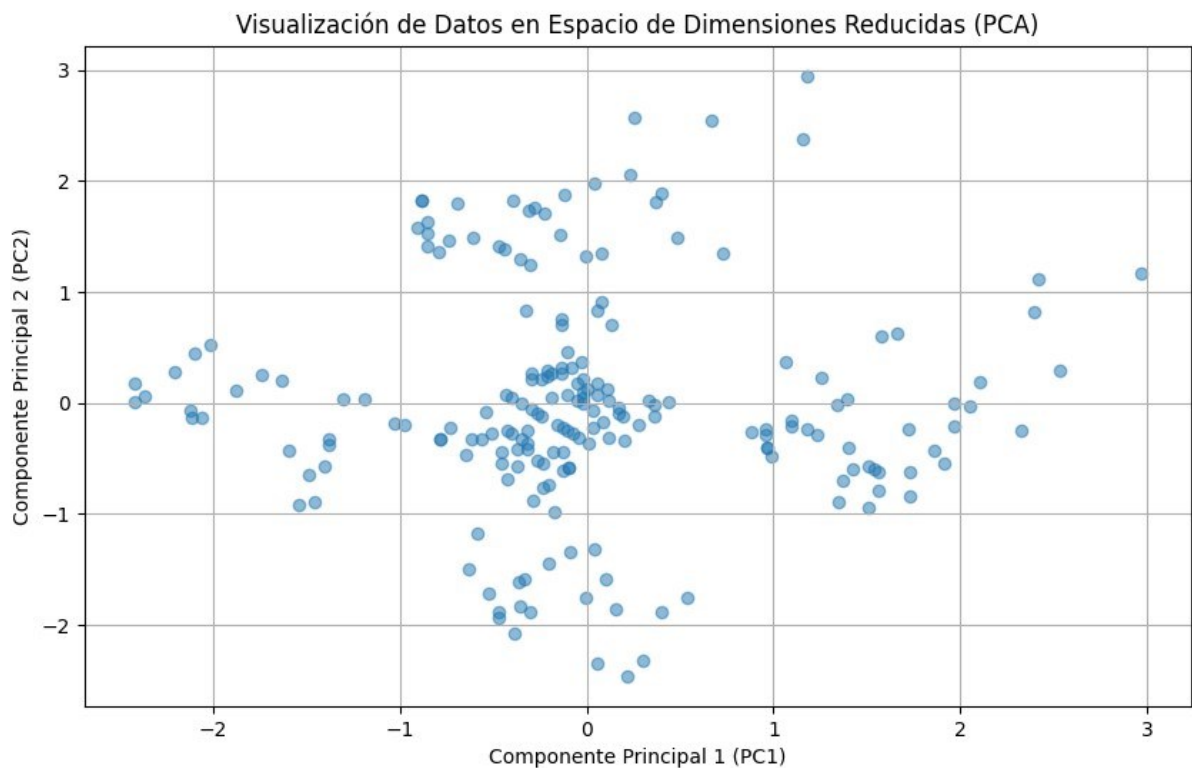
**Porcentaje de clientes masculinos: 44.00%**

**Entre los 3 grupos de clusters determinamos que el grupo con menor ingreso es quien mayor puntaje de gasto tiene podríamos pensar en gastos hormiga o por impulso de esta visualización .**





## Técnica PCA



**Se ve una clara diferenciación entre los grupos representados por los distintos colores en el gráfico, y cada grupo está concentrado en áreas específicas del gráfico, no veo que se traslapen puntos.**

**Veo una concentración muy marcada de puntos en el origen (0,0) esto puede sugerir la presencia de un grupo central o dominante en el conjunto de datos.**

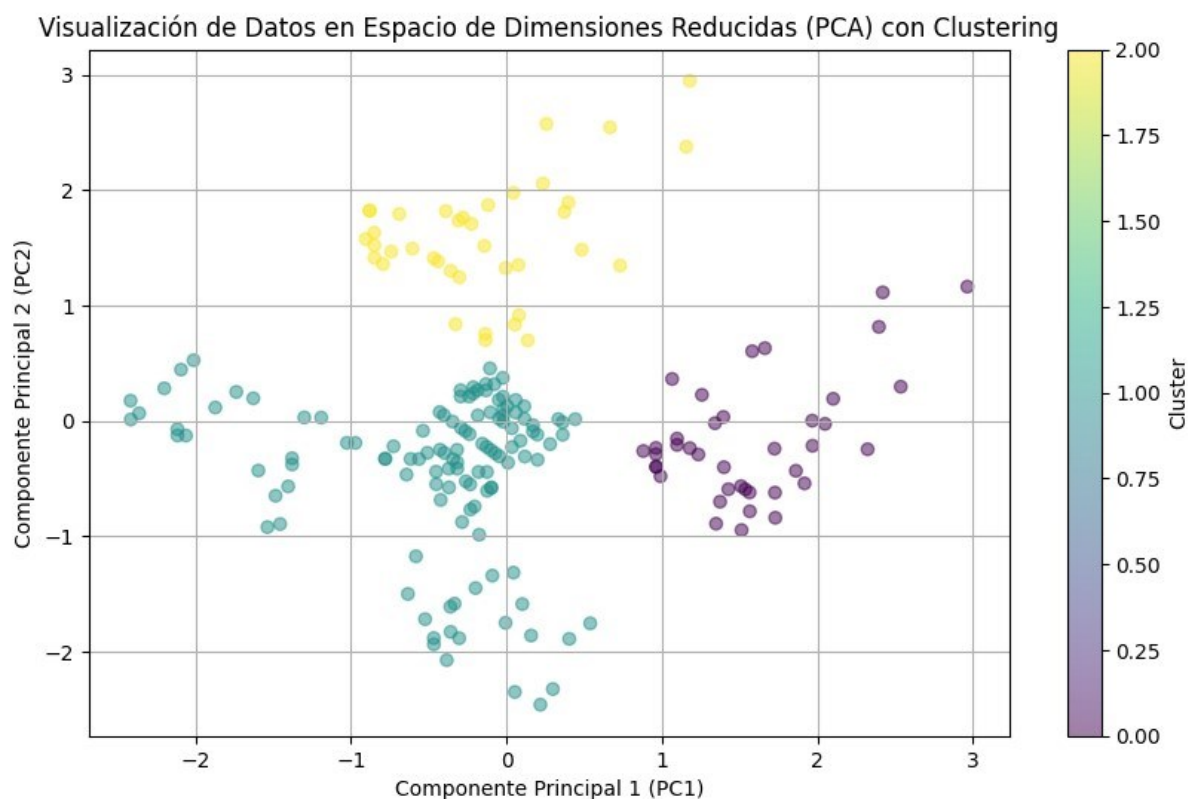
**Después de eso se observa mucha dispersión entre los puntos de diversos colores, hay una dominancia del color verde sobre el resto, y la presencia de puntos atípicos.**

**Hay 3 cluster cluster**

**0 todos sus componententes principales son del uno.**

**Cluster 1 tiene algunos componentes del uno .**

**Cluster 2 no tienen componentes del uno.**



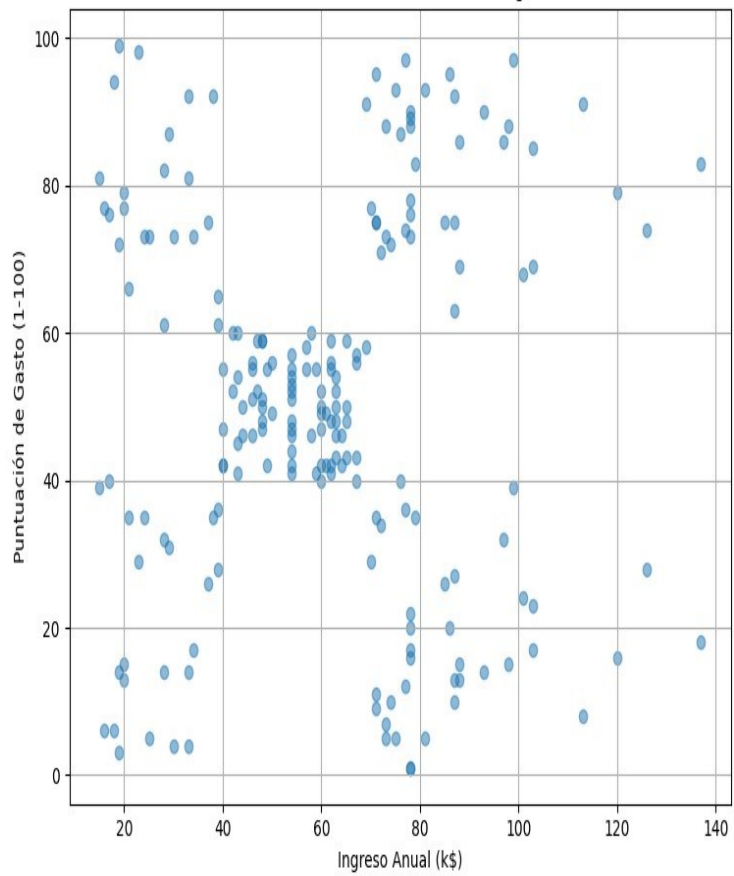
## **Tendencias**

**Cluster 1:** Aquí encontramos clientes con el ingreso anual más alto y la puntuación de gasto más alta entre todos los clusters. Este grupo, con un total de 39 clientes, está compuesto por personas con un alto poder adquisitivo y propensión a gastar.

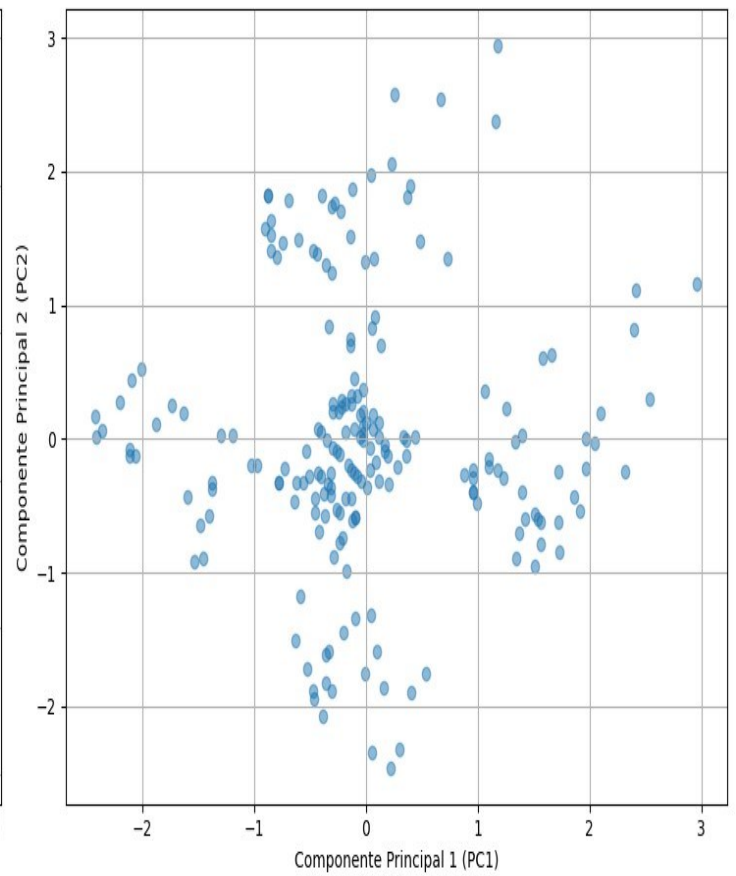
**Cluster 2:** Este cluster presenta clientes con un ingreso anual cercano al del Cluster 1, pero con una puntuación de gasto significativamente más baja. Aunque tienen un ingreso comparable al del Cluster 1, estos clientes muestran un comportamiento más conservador en términos de gasto.

**Cluster 3:** Este es el cluster más grande, con un total de 100 clientes. Presenta un ingreso anual y una puntuación de gasto moderados en comparación con los otros grupos. Estos clientes podrían representar un segmento de mercado más diverso en términos de comportamiento de compra.

Visualización de Características Originales



Visualización PCA



## **Parte Final Trabajo**

**Se trabaja el archivo disponible ( de esta manera se recolectaron datos relevantes ). Los clientes del mall se segmentan (se utilizan técnicas Kmeans y PCA), considerando rangos de edades establecidos con el fin de comprender sus gustos y preferencias según su generación a la vez estableciendo afinidades o diferencias entre grupos .**

**Se segmenta a su vez el genero femenino y masculino para establecer parámetros en términos de ingreso y puntaje de gasto en el mall.**

**Como conclusiones y recomendaciones:**

**La Personalización del Marketing enfocado en cada segmento de generación y no algo general ya que no alcanzaría para convencer a todos los clientes del Mall.**

**De esta manera se optimiza el Marketing ya que se dirige a segmentos específicos donde si surge efecto.**

**A su vez la inclusión de Productos y espacios en el mall con énfasis en cada sector generacional.**

**Los rangos de edad 17-24 son los que menos ingresos tienen sin embargo son los de mejor puntuación de gasto podemos analizar que dependen de una fuente de dinero y nos da visibilidad que si sus familiares les dan dinero para gastar en el mall es porque tienen una percepción agradable del lugar , por lo que se sugiere**

**como recomendación para el negocio no desestimar en la inversión de seguridad y la imagen de un ambiente sano y seguro para seguir generando visitas , adicionalmente se está invirtiendo para que este rango de edad una vez tengan su propia Familia , no dejen el habito de visitar el mall y lo instauren en sus hijos .**

**El grupo de 25-40 es el más grande a su vez el que mayor genera al mall por lo que se puede considerar que bajo este sector de la población del mall se ven las mayores ganancias , y de ahí en adelante se empiezan a perder ganancias , como recomendación es hacer encuestas , planes de fidelidad para comprender porque después de los 40 , disminuye esta población en el mall y poder recuperarles .**

**El modelo K-Means permitió agrupar los diferentes grupos de clientes para generar visibilidad de temas homogéneas de igual manera con la técnica PCA , de esta manera se vieron que los grupos comparten pocas igualdades siempre había un grupo grande y otros 2 menores , siempre existió diferencias marcadas , entre estas 2 grupos compartían ciertas igualdades pero mínimas , precisamente el ser mall y tener diversidad de negocios y poblaciones hace que se traduzca en esta aplicación , como recomendación de esta manera es que la naturaleza de este negocio (mall) , es ofrecer diversos servicios estar abierto a respetar y preservar servicios pasados , a su vez actualizarse en caras del presente y no dejar de tener visión a futuro para mantener su población de clientes y aumentarla .**

**De esta manera las recomendaciones le permitirán al negocio. Mejora de la Experiencia del Cliente ya que se tiene el dato y se ha**

**conceptualizado la población por géneros y rangos de edad generacionales.**

**Optimización de Operaciones Podemos ajustar los horarios según sus patrones de compra al mall y días de visita basándonos precisamente en la clasificación generacional.**

**Desarrollo de Estrategias de Marketing, dirigir campañas promocionales específicas a grupos de clientes con características similares puede aumentar la efectividad de las acciones de marketing y generar un retorno de la inversión más alto.**

**Estabilidad del Negocio, Al adaptarse continuamente a las demandas cambiantes de los clientes y mejorar la experiencia general, el centro comercial puede mantener una base de clientes leales y competitiva en el mercado.**

## Referencias Bibliográficas

**Enterarat Blog España, noviembre 2022**

**[https://www.enterat.com/actualidad/generaciones-anosedad-nombres.php#google\\_vignette](https://www.enterat.com/actualidad/generaciones-anosedad-nombres.php#google_vignette)**

**Faculta de Biblioteconomía y Documentación**

**Universitat de Barcelona Barcelona,**

**diciembre de 2006 <https://bid.ub.edu/es>**

**Federación Internacional de Asociaciones de**

**Bibliotecarios y Bibliotecas; Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2001). Directrices de la IFLA/UNESCO para el desarrollo del servicio de bibliotecas públicas. 94 p.**

**<<http://www.ifla.org/VII/s8/news/pg01-s.pdf>>. [Consulta: 404-2006].**



**¿Qué es customer service y cómo implementarlo en tu empresa? -  
Zendesk MX**

<https://www.zendesk.com.mx/blog/que-es-customer-service/service/>

**Segmentación de clientes de centros comerciales y formulación de estrategias de crecimiento**

<https://medium.com/geekculture/mall-customer-segmentation-and-forming-growth-strategies-cc4130a0f4d7>

**Transforming Malls Into Customer Hubs: 5 Strategies For Enduring Success**

<https://stefanini.com/en/insights/articles/5-ways-shopping-malls-improve-shopping-experience-for-customer>

**Centros comerciales en Miami**

[https://www.tripadvisor.es/Attractions-g34438-Activities-c26-t143-Miami\\_Florida.html](https://www.tripadvisor.es/Attractions-g34438-Activities-c26-t143-Miami_Florida.html)

**sklearn.cluster.KMedios**

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

