



Informe de segmentación de clientes para RetailMax (K-Means)

Fecha: 8 de octubre de 2025

Autores: Jesús M. Vásquez L. – jesus2407019@hybridge.education

Clase: Inteligencia Artificial - Aprendizaje No Supervisado - Reporte de Actividad

Introducción

RetailMax contrató un consultor de ciencia de datos para ayudarle a entender mejor su base de clientes. El objetivo era segmentar a los clientes en grupos con características de compra y demográficas similares de modo que se pudieran diseñar campañas de marketing más efectivas. Para ello se utilizó un enfoque de aprendizaje no supervisado (K-Means), ya que no se contaba con etiquetas previamente definidas y el reto consistía en descubrir patrones ocultos en los datos. El conjunto de datos contiene 200 clientes e incluye su identificador, sexo, edad, ingreso anual (en miles de dólares) y una puntuación de gasto de 1 a 100 que resume su nivel de compras en la tienda.

Metodología

El análisis se llevó a cabo en dos etapas: (1) exploración de datos y (2) construcción de clusters mediante K-Means. Siguiendo las recomendaciones del notebook original, se eligieron como variables de entrada la edad, el ingreso anual y la puntuación de gasto porque juntas capturan el poder adquisitivo, las preferencias de consumo y el estilo de vida de cada cliente. Se descartó el identificador por no contener información útil y el sexo se dejó fuera en una primera etapa para trabajar únicamente con variables numéricas. Los datos se estandarizaron con `StandardScaler` y se entrenó un modelo K-Means con cinco clusters (`n_clusters=5`) y la inicialización `k-means++` para mejorar la estabilidad de los centroides.

Resultados

Después de entrenar el modelo, cada cliente fue asignado a uno de cinco grupos. La siguiente tabla resume el tamaño y las características de cada cluster:

Cluster	Clientes	Edad prom.	Ingreso prom. (k\$)	Puntuación de gasto prom.	% Hombres	% Mujeres	Descripción
0	47	55.6	54.4	48.9	42.6 %	57.4 %	Clientes mayores , con ingresos moderados y gasto medio. Probablemente son personas jubiladas o de mediana edad que compran de forma estable y valoran la calidad y el servicio.
1	40	32.9	86.1	81.5	45.0 %	55.0 %	Adultos jóvenes con alto ingreso y alto gasto . Constituyen el segmento premium/VIP con potencial para productos exclusivos y programas de lealtad.
2	54	25.2	41.1	62.2	40.7 %	59.3 %	Segmento joven con ingresos medios-bajos pero niveles de gasto relativamente altos. Muestran afinidad por promociones y experiencias entretenidas.

Cluster	Clientes	Edad prom.	Ingreso prom. (k\$)	Puntuación de gasto prom.	% Hombres	% Mujeres	Descripción
3	20	46.3	26.8	18.4	40.0 %	60.0 %	Cientes de ingreso bajo y bajo gasto . Visitan la tienda ocasionalmente; son sensibles al precio y responden a descuentos.
4	39	39.9	86.1	19.4	51.3 %	48.7 %	Individuos de ingreso muy alto pero poca disposición a gastar . Representan una oportunidad de conversión a través de promociones, paquetes y cross-selling.

La distribución muestra que los grupos 0 y 2 concentran la mayor parte de los clientes (47 y 54 respectivamente), mientras que el cluster 3 es el más pequeño. La edad media varía desde 25 años (cluster 2) hasta 55 años (cluster 0), lo que sugiere diferencias en etapas de vida y motivaciones de compra. La proporción entre hombres y mujeres es relativamente balanceada en todos los grupos, aunque las mujeres predominan ligeramente en los clusters de gasto medio y alto.

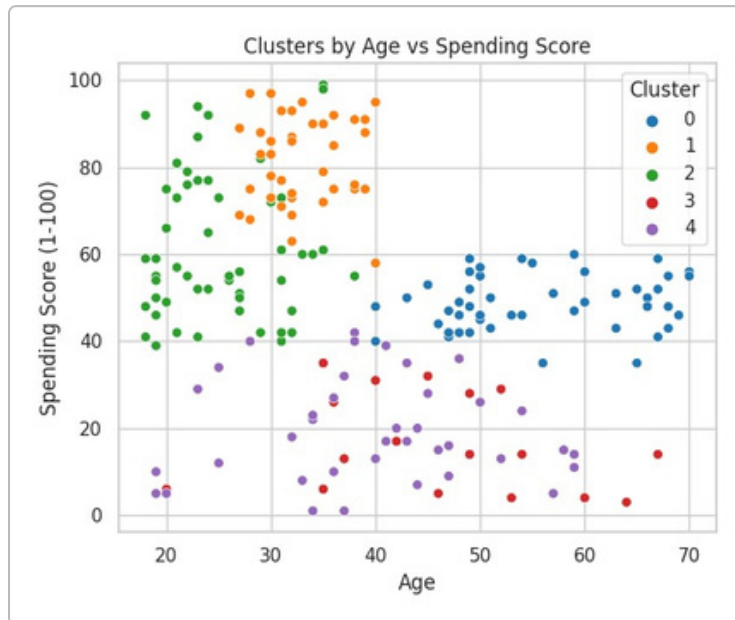
Visualizaciones

En las siguientes gráficas se observan las relaciones entre las variables y la estructura de los clusters:

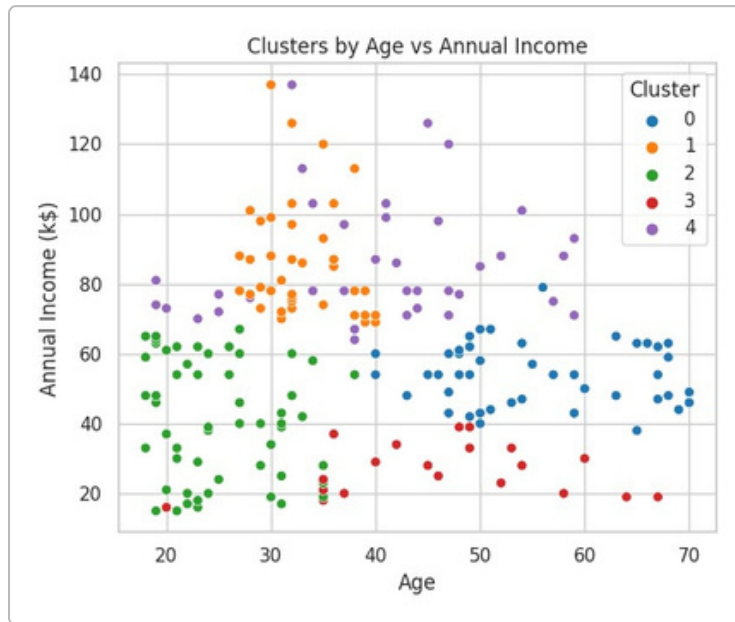
1. **Ingreso anual vs. puntuación de gasto:** permite identificar claramente tres regiones: clientes con alto ingreso y alto gasto (cluster 1), alto ingreso y bajo gasto (cluster 4) y los demás grupos en posiciones intermedias. La dispersión sugiere que el poder adquisitivo no garantiza necesariamente un mayor gasto.



1. **Edad vs. puntuación de gasto:** muestra que los clientes jóvenes (clusters 2 y 1) tienden a tener puntuaciones de gasto más altas. Los clientes mayores (cluster 0) presentan un rango medio de gasto y los clientes de ingreso alto pero poco gasto (cluster 4) abarcan una edad intermedia.



1. **Edad vs. ingreso anual:** revela que los ingresos más altos se concentran entre 30 y 50 años (clusters 1 y 4), mientras que los clientes jóvenes tienen ingresos más variados. El cluster 3 tiene ingresos notablemente bajos independientemente de la edad.



Implicaciones de marketing

El análisis sugiere estrategias diferenciadas para cada segmento:

- **Cluster 1 – Clientes VIP:** Con altos ingresos y gasto elevado, estos clientes valoran el servicio personalizado, el acceso prioritario y experiencias exclusivas. Se recomienda implementar un programa de membresía premium con atención personal, preventas y recompensas exclusivas. Además, utilizar mensajes de lujo y calidad en la comunicación para mantener su lealtad.
- **Cluster 2 – Jóvenes entusiastas:** Aunque sus ingresos son moderados, gastan significativamente. Son sensibles a tendencias y promociones creativas. Estrategias efectivas incluyen campañas en redes sociales, marketing de influencers y programas de referidos. Ofrecer productos de moda y experiencias de compra gamificadas puede incrementar su fidelidad.
- **Cluster 0 – Adultos de ingreso medio:** Con edades mayores y gasto moderado, este grupo aprecia la fiabilidad y el valor. Promociones de fidelidad, descuentos por volumen y comunicación centrada en la calidad del producto pueden aumentar el ticket promedio. El canal ideal son newsletters y programas de recompensas basados en la frecuencia de compra.
- **Cluster 4 – Alto ingreso, bajo gasto:** Aunque cuentan con recursos, gastan poco. Podrían estar explorando la tienda o compran sólo productos básicos. Se sugiere campañas de cross-selling, paquetes premium y demostraciones para mostrar el valor de productos de mayor margen. Una estrategia de retargeting digital enfocada en destacar beneficios y estatus puede incentivar su gasto.
- **Cluster 3 – Sensibles al precio:** Estos clientes tienen ingresos bajos y puntaje de gasto reducido. Representan un segmento menos rentable pero pueden aumentar sus compras con promociones

agresivas, programas de cupones y ofertas económicas. La comunicación debe centrarse en el ahorro y el valor.

Conclusiones y siguientes pasos

El proyecto demuestra que el **aprendizaje no supervisado** puede revelar patrones útiles de comportamiento y apoyar la toma de decisiones de marketing. La segmentación obtenida ofrece una base objetiva para diseñar campañas dirigidas, optimizar la asignación de recursos publicitarios y mejorar la experiencia del cliente. Para profundizar en la personalización, se recomienda incorporar variables adicionales (frecuencia de visita, categorías de productos adquiridos, canales de compra) y actualizar periódicamente los modelos para captar cambios en las preferencias. También sería conveniente utilizar técnicas de validación como el índice de silueta o el método del codo para ajustar el número óptimo de clusters y evaluar la estabilidad de los segmentos.