

# **MODULO 3**

## **DIPLOMADO EN CIENCIA Y ANALISIS DE**



**Jean David Garcia Jaime**  
**09/09/24**

Para el proyecto del Módulo 3 de Análisis de Datos con Python de BEDU, se realizó un análisis exhaustivo de un dataset obtenido en Kaggle sobre el comportamiento de los clientes de una tienda. Este CSV ya había sido analizado previamente en el módulo anterior, pero con las nuevas herramientas adquiridas en el módulo, se buscó obtener nuevas conclusiones y profundizar en los patrones encontrados mediante un enfoque más sofisticado. A lo largo del análisis, se emplearon técnicas avanzadas de visualización, segmentación y algoritmos de predicción para proporcionar mejores resultados y estrategias.

### Proceso y Métodos Utilizados

El proceso seguido en el análisis se desglosa en las siguientes etapas:

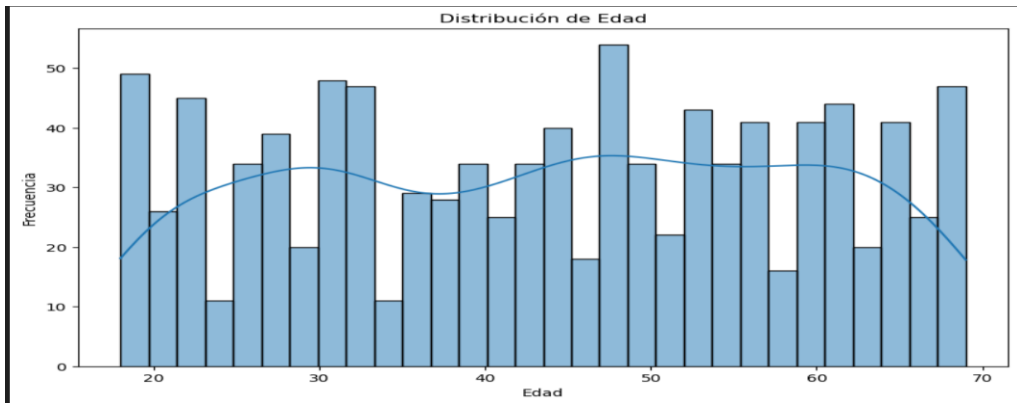
#### 1. Lectura de Datos

El análisis comenzó con la carga y revisión del dataset, asegurando que los datos estuvieran bien estructurados para su posterior análisis. Esto incluyó la modificación de variables categóricas, como género y categoría preferida, que se convirtieron en valores numéricos para facilitar su manipulación en el análisis posterior.

	id	age	gender	income	spending_score	membership_years	purchase_frequency	preferred_category	last_purchase_amount
0	1	38	Female	99342	90	3	24	Groceries	113.53
1	2	21	Female	78852	60	2	42	Sports	41.93
2	3	60	Female	126573	30	2	28	Clothing	424.36
3	4	40	Other	47099	74	9	5	Home & Garden	991.93
4	5	65	Female	140621	21	3	25	Electronics	347.08
5	6	31	Other	57305	24	3	30	Home & Garden	86.85
6	7	19	Other	54319	68	5	43	Clothing	191.72
7	8	43	Male	108115	94	9	27	Groceries	734.56
8	9	53	Male	34424	29	6	7	Sports	951.71
9	10	55	Female	45839	55	7	2	Electronics	821.18

#### 2. Primeros Histogramas para Conocer el Comportamiento

Inicialmente, se generaron histogramas para comprender mejor la distribución de las variables y el comportamiento general de los clientes. Estas visualizaciones permitieron identificar patrones iniciales como la variación en las categorías preferidas, la distribución de las edades y la frecuencia de compra.





### 3. Estrategias en Base a los Primeros Resultados

Con base en los primeros histogramas, se plantearon algunas estrategias de ventas preliminares. Por ejemplo, se observó que los clientes de mayor edad tienden a gastar más por transacción, mientras que los clientes jóvenes realizan compras con mayor frecuencia. Esto permitió sugerir campañas dirigidas a cada grupo, incentivando la compra frecuente en clientes jóvenes y ofreciendo productos de mayor valor a clientes mayores.

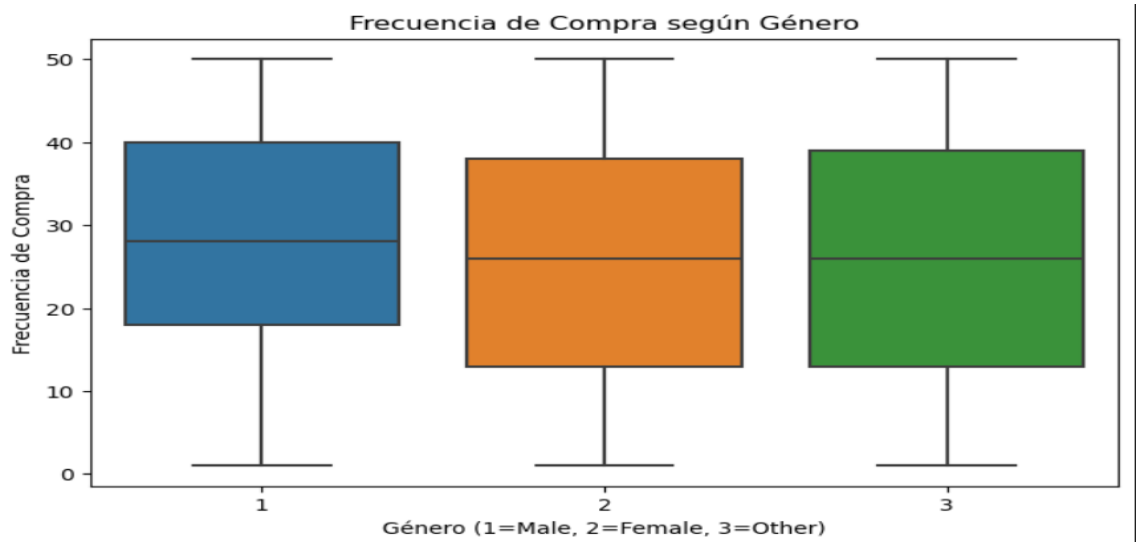
### 4. Segmentación con K-Means

Para profundizar en el comportamiento de los clientes, se implementó el algoritmo de segmentación K-Means, dividiendo a los clientes en varios grupos con características similares. Este método permitió identificar segmentos de clientes con preferencias de compra similares, lo que facilitó la creación de estrategias de marketing dirigidas. Por ejemplo, ciertos clusters mostraban preferencias por categorías específicas de productos o patrones de compra que no eran evidentes en los análisis iniciales.

	age	income	spending_score	membership_years	Frecuency
G1	56.384892	89828.870504	26.503597	5.359712	24.352518
G2	27.493506	92973.961039	28.839827	5.086580	28.160173
G3	44.469880	81269.489960	72.433735	8.232932	32.899598
G4	44.148760	90145.809917	76.938017	3.115702	21.194215

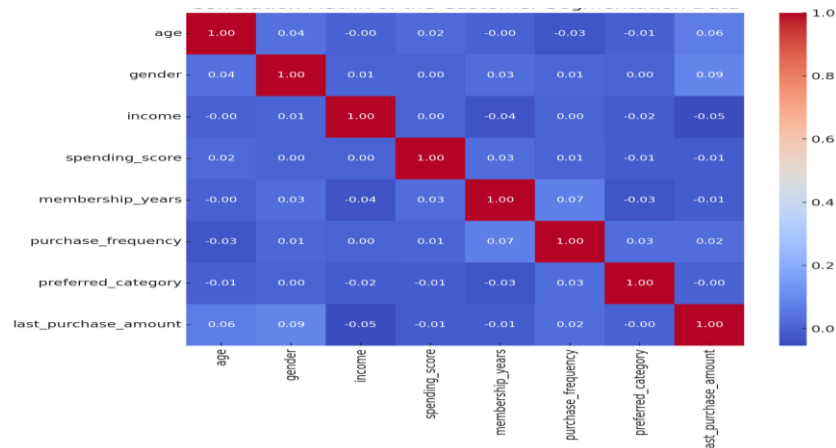
### 5. Segundo Análisis: Visualización de Datos en Búsqueda de Patrones

Posteriormente, se realizaron visualizaciones más detalladas para buscar patrones más complejos en los datos. Se usaron gráficos de dispersión, diagramas de caja y otras visualizaciones avanzadas para observar la relación entre diversas variables como el gasto, la edad, las categorías preferidas, y el género. Estas visualizaciones confirmaron que ciertos segmentos de clientes, como los interesados en las categorías de Home & Garden y Electronics, tienden a realizar compras de mayor valor.



## 6. Mapas de Correlación en Búsqueda de Relaciones

Se generaron mapas de correlación para identificar relaciones entre las variables. Estos mapas permitieron confirmar que no había una correlación fuerte entre el ingreso y la frecuencia de compra, sugiriendo que factores como el gasto promedio y las categorías preferidas podrían tener más influencia sobre el comportamiento del cliente. Sin embargo, se encontraron correlaciones débiles, como la relación entre los años de membresía y la frecuencia de compra, lo que llevó a sugerir programas de lealtad para clientes antiguos.



## 7. Algoritmos de Predicción: Bosques Aleatorios y Support Vector Machines

Para predecir el comportamiento de compra de los clientes, se aplicaron dos algoritmos de machine learning: Bosques Aleatorios y Support Vector Machines (SVM). Estos métodos se usaron para intentar predecir la frecuencia de compra de los clientes con base en variables como la edad, el ingreso, el género, el gasto, y la categoría preferida.

Bosques Aleatorios: Este modelo presentó un error cuadrático medio (MSE) de 220.7, lo que indicaba que el modelo no era completamente preciso en la predicción de la frecuencia de compra. Se observaron ciertos patrones, pero aún había margen de mejora.

Support Vector Machines (SVM): Al aplicar SVM, el modelo mejoró ligeramente, con un MSE de 209.29. Aunque las predicciones fueron más precisas que las de Bosques Aleatorios, el error seguía siendo significativo, lo que indicaba que la frecuencia de compra puede depender de otros factores no incluidos en el análisis o que las relaciones entre variables son complejas.

Mean Squared Error del modelo: 218.1325515		
	Real	Predicción
521	13	21.38
737	42	28.63
740	36	27.52
660	27	25.24
411	14	29.00

- Resultados y Conclusiones

El análisis reveló varios patrones clave sobre el comportamiento de compra de los clientes:

**Frecuencia de Compra:** Se encontró una ligera correlación entre la frecuencia de compra y factores como la categoría preferida y los años de membresía, pero en general, los ingresos no parecieron influir significativamente en la frecuencia con la que los clientes compran.

**Gasto por Categoría:** Los clientes que prefieren las categorías de Electronics y Home & Garden tienden a gastar más en sus compras, lo que sugiere que las campañas de marketing dirigidas a estos segmentos deben enfocarse en ofrecer productos de mayor valor.

**Diferencias por Edad y Género:** Los clientes jóvenes tienden a realizar compras más frecuentemente, mientras que los clientes mayores gastan más en cada transacción. Los hombres realizan compras más frecuentes que las mujeres, pero el gasto por compra es menor.

Estrategias de Ventas en Base a los Resultados

**Segmentación por Edad y Género:**

Implementar campañas para incentivar la compra recurrente en clientes jóvenes con descuentos por compras frecuentes.

Ofrecer productos premium y experiencias exclusivas para clientes mayores que tienden a gastar más por transacción.

**Promociones basadas en Categorías:**

Para categorías como Groceries y Sports, ofrecer descuentos por suscripción o recompensas para incentivar compras más frecuentes.

Enfocar las campañas de marketing de Home & Garden y Electronics en destacar productos de mayor valor, con promociones que fomenten la compra de artículos costosos.

**Programas de Lealtad:**

Desarrollar programas de fidelización para clientes con mayor antigüedad en la tienda, recompensando su lealtad con ofertas exclusivas, lo que puede aumentar su frecuencia de compra a lo largo del tiempo.

En resumen, el análisis avanzado del dataset proporcionó valiosos insights sobre el comportamiento de los clientes y permitió diseñar estrategias de marketing específicas que podrían mejorar tanto la frecuencia de compra como el valor de cada transacción.