

Retail Saviors

# SEGMENTACIÓN CLIENTES

# DESCRIPCIÓN DEL PROYECTO

Segmentación efectiva de clientes :

- \*Personalizar estrategias de marketing
- \*Mejorar la retención de clientes
- \*Maximizar el valor del cliente a lo largo del tiempo.

# OBJETIVO

Desarrollar una aplicación web interactiva que permita a los científicos de datos realizar análisis de segmentación de clientes utilizando el modelo RFM (Recencia, Frecuencia, Valor Monetario).

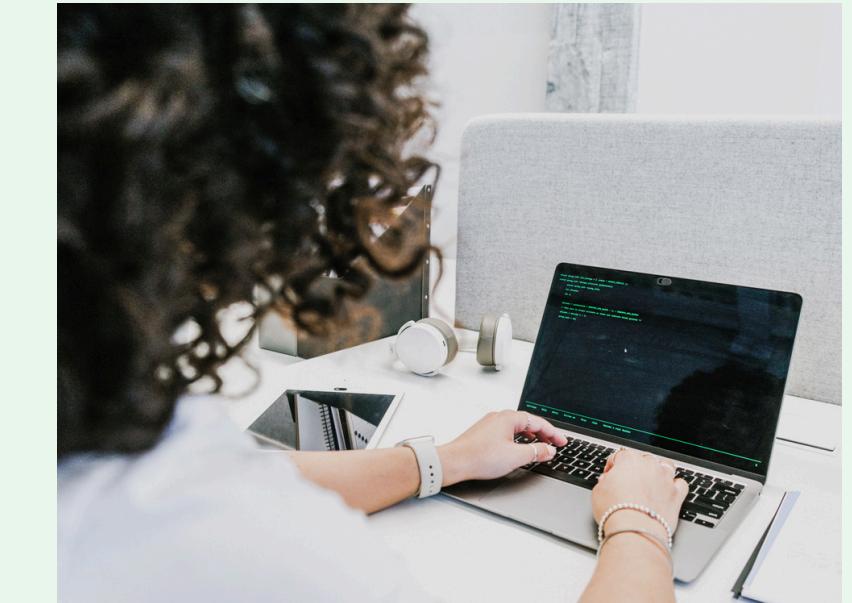
La aplicación proporcionará herramientas para cargar datos, realizar análisis exploratorio y visualizar resultados de segmentación mediante gráficos interactivos



# DATASET

- INVOICE\_NO: Número de orden
- STOCK\_CODE: Código del producto
- DESCRIPTION: Descripción del producto
- QUANTITY: Cantidad de producto seleccionado
- INVOICE\_DATE: Fecha de compra
- UNIT\_PRICE: Precio unitario
- CUSTOMER\_ID: Número de cliente
- REGION: País de compra

# HALLAZGOS



- \*5268 valores duplicados ->eliminados dado que tenemos un ID de identificación.
- \*Dos columnas con valores nulos.
- \*Observamos la cantidad de valores únicos de cada columna.
- \*Se aprecian valores negativos en las columnas QUANTITY con dos valores y UNIT\_PRICE estos valores negativos pueden representar devoluciones o descuentos por lo que se dejaran tal cual están.

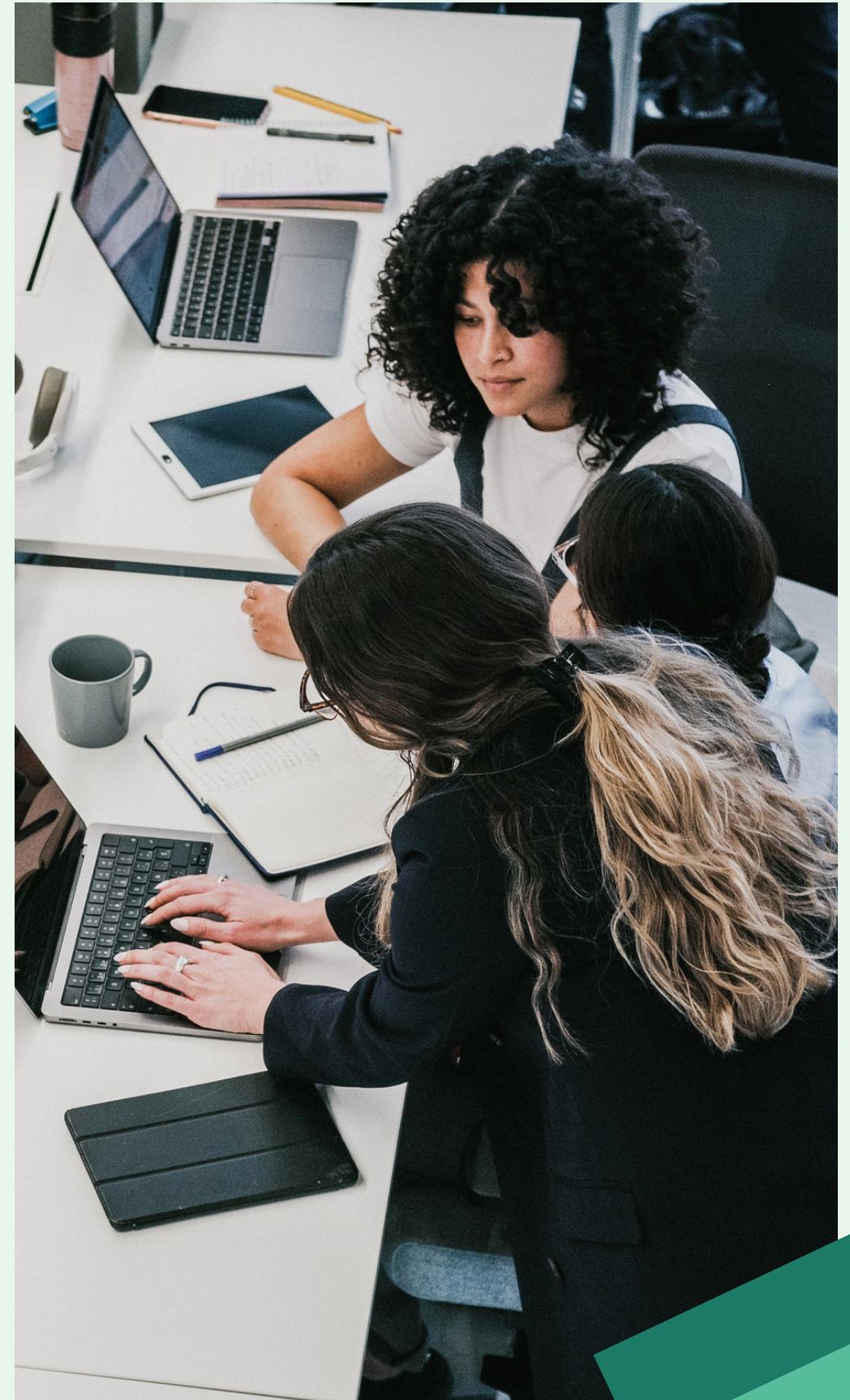
Las siguientes acciones a tomar son:

- Cambio por minusculas los nombres de las columnas.
- Eliminación de valores duplicados.
- Tratamiento de valores nulos.
- Creación de columna total y corrección de tipo de datos.

Despues del preprocesamiento de datos culminamos con un total de 401604 -> una reducción de poco más de 26%

\*Se eliminaron los datos que contenían valores nulos en la columna de customer\_id.

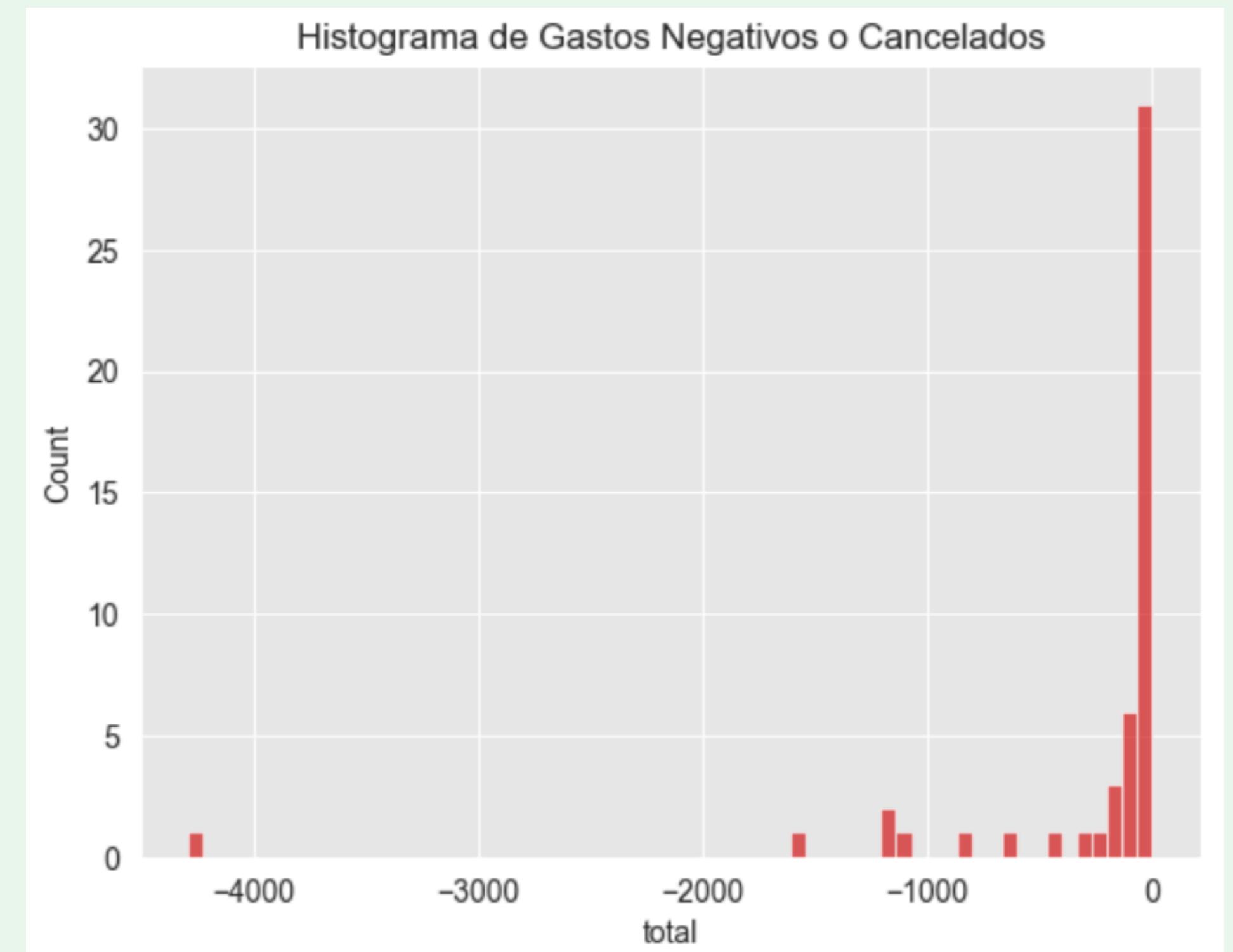
\*Se creo una nueva columna llamada total para tener la cantidad comprada por producto, esta columna nos servirá para la creación de nuestra tabla RFM -> 9 columnas



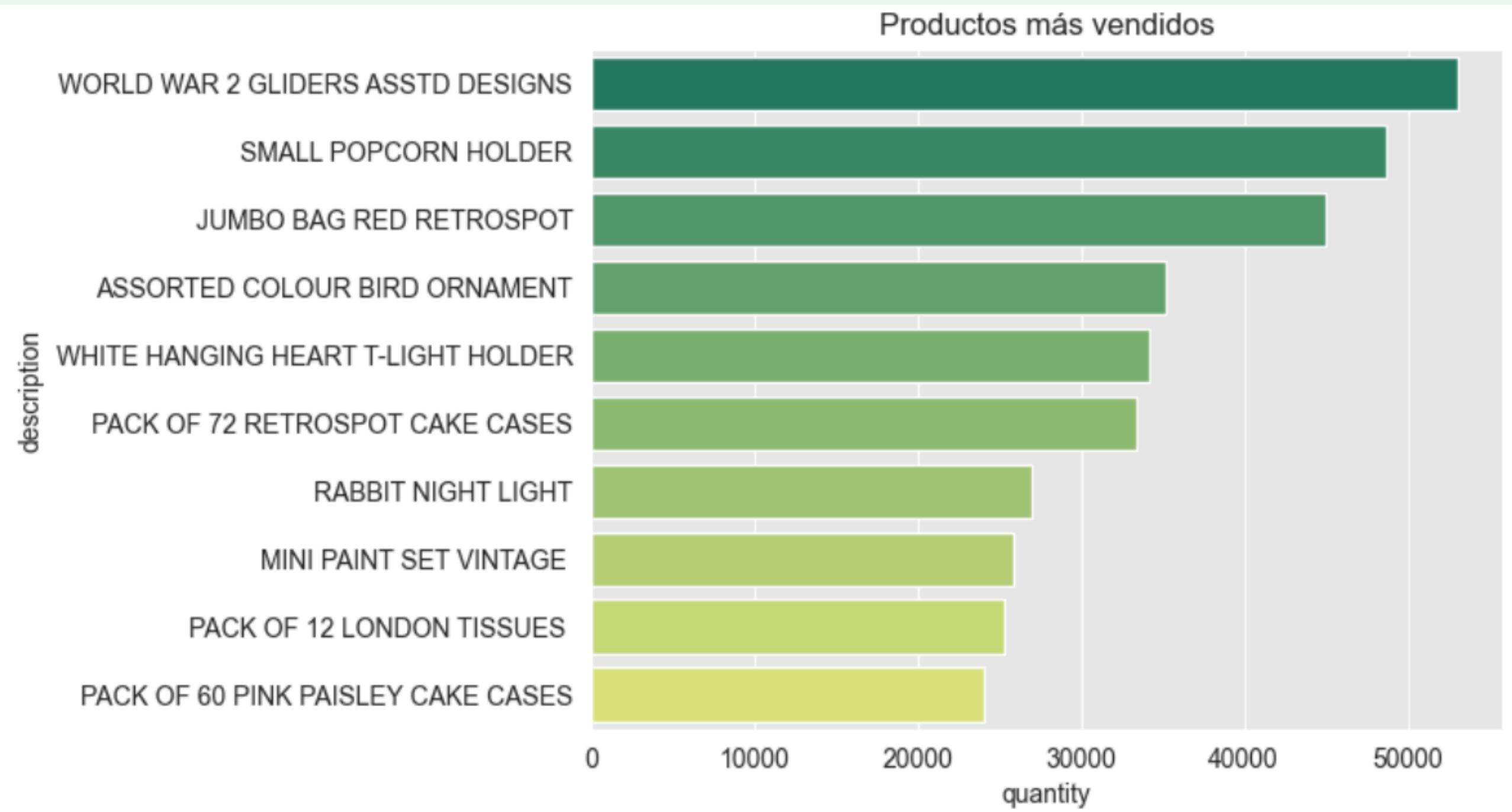
# EDA (ANÁLISIS ESTADÍSTICO DE DATOS).

El gasto negativo se aplica a los clientes que han realizado cancelaciones de sus pedidos,

Los clientes que se muestran en su mayoría, pertenecen a gastos en 0, significando que 31 clientes cancelaron todas sus facturas.



# PRODUCTOS MAS COMPRADOS

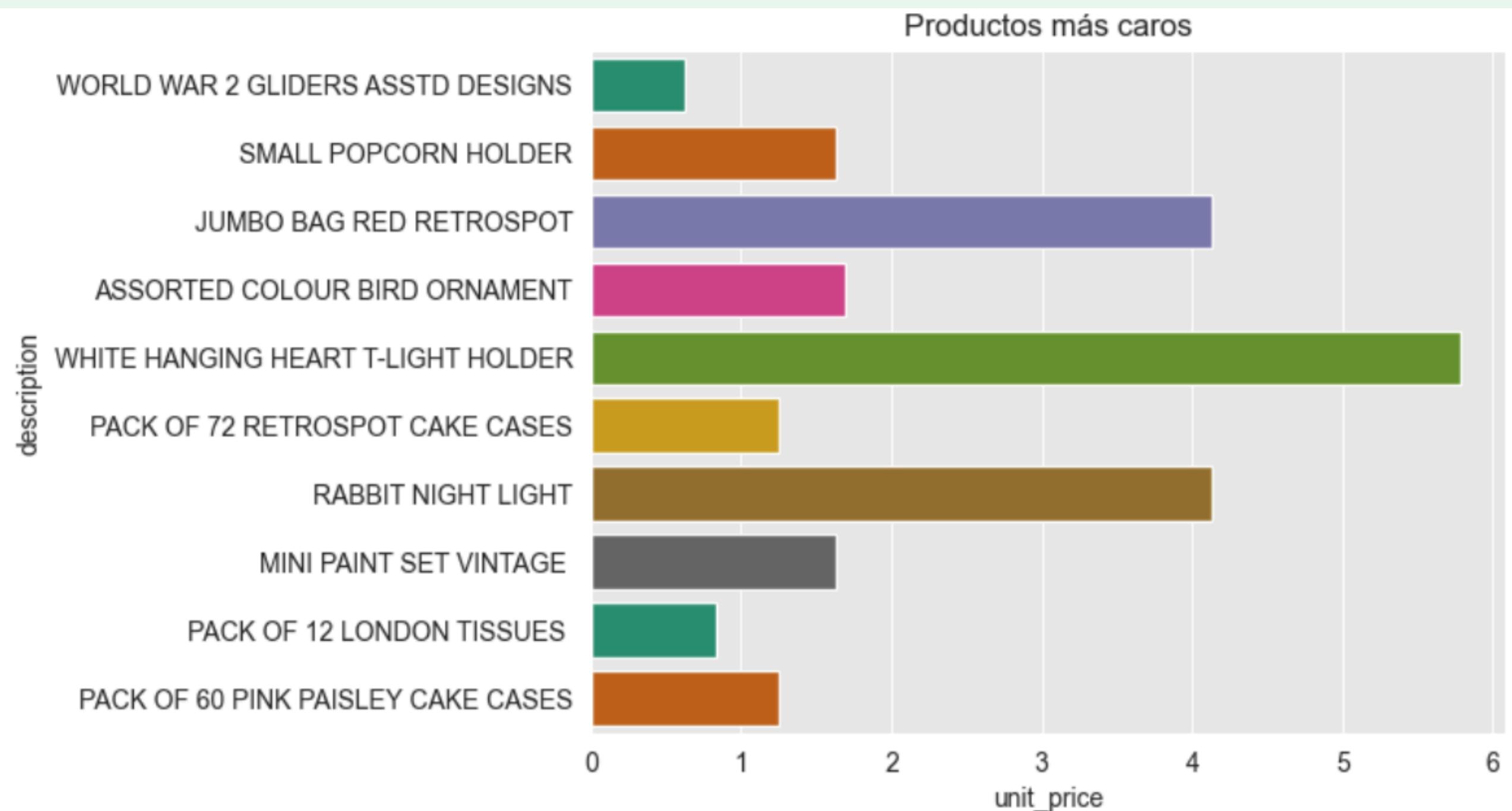


\*Los 10 productos que han sido adquiridos con mayor frecuencia.

\*El producto más solicitado es el producto en la parte superior: 'World War 2 Gliders Asstd Designs' con más de 50,000 ventas totales.

# PRECIOS DE LOS PRODUCTOS MÁS COMPRADOS

El producto mas vendido tiene el menor precio



# PRODUCTOS MAS COSTOSOS

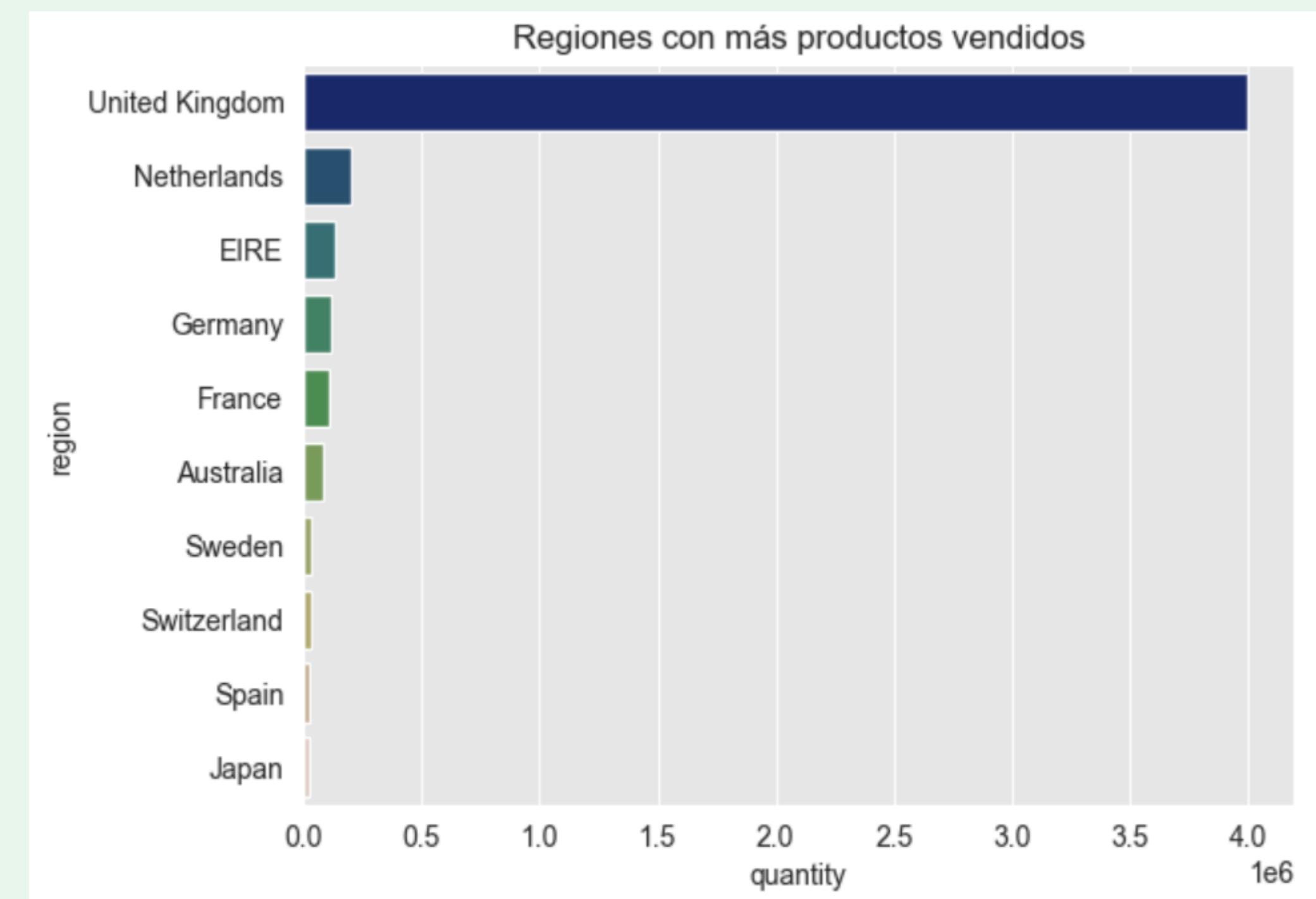
stock_code	description	unit_price
M	Manual	38970.00
POST	POSTAGE	8142.75
D	Discount	1867.86
DOT	DOTCOM POSTAGE	1599.26
CRUK	CRUK Commission	1100.44
22502	PICNIC BASKET WICKER SMALL	649.50
22655	VINTAGE RED KITCHEN CABINET	295.00
22656	VINTAGE BLUE KITCHEN CABINET	295.00
22826	LOVE SEAT ANTIQUE WHITE METAL	195.00
22828	REGENCY MIRROR WITH SHUTTERS	165.00

\*El producto 'Manual' contiene un precio demasiado elevado.

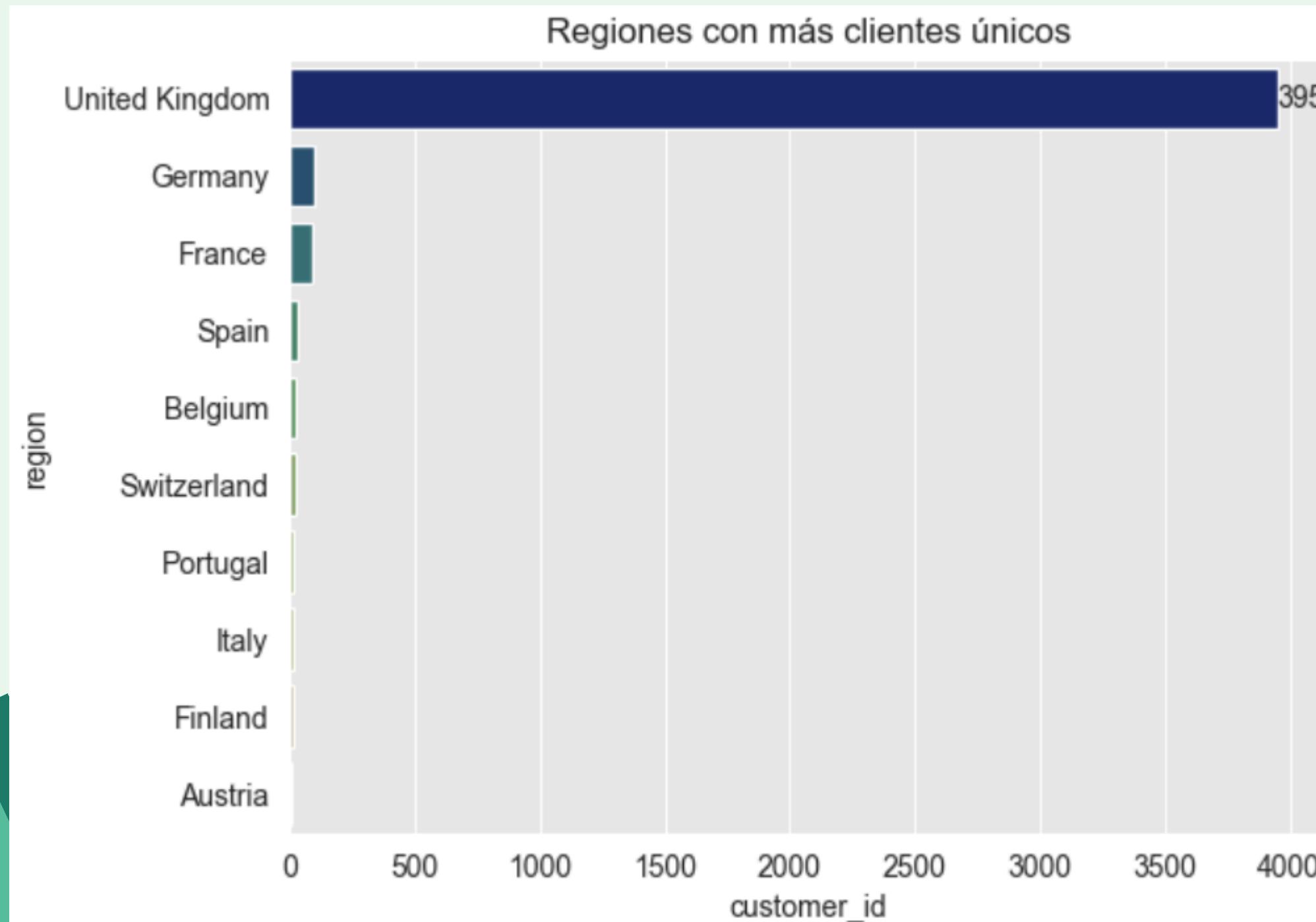
\*El producto: 'PICNIC BASKET WICKER SMALL' tiene el mayor precio unitario, con un valor de \$649.50.

# REGIÓN CON MÁS PRODUCTOS VENDIDOS

La región del Reino Unido  
sobrepasa a todo el resto de  
regiones en ventas.



# REGIONES CON CLIENTES ÚNICOS

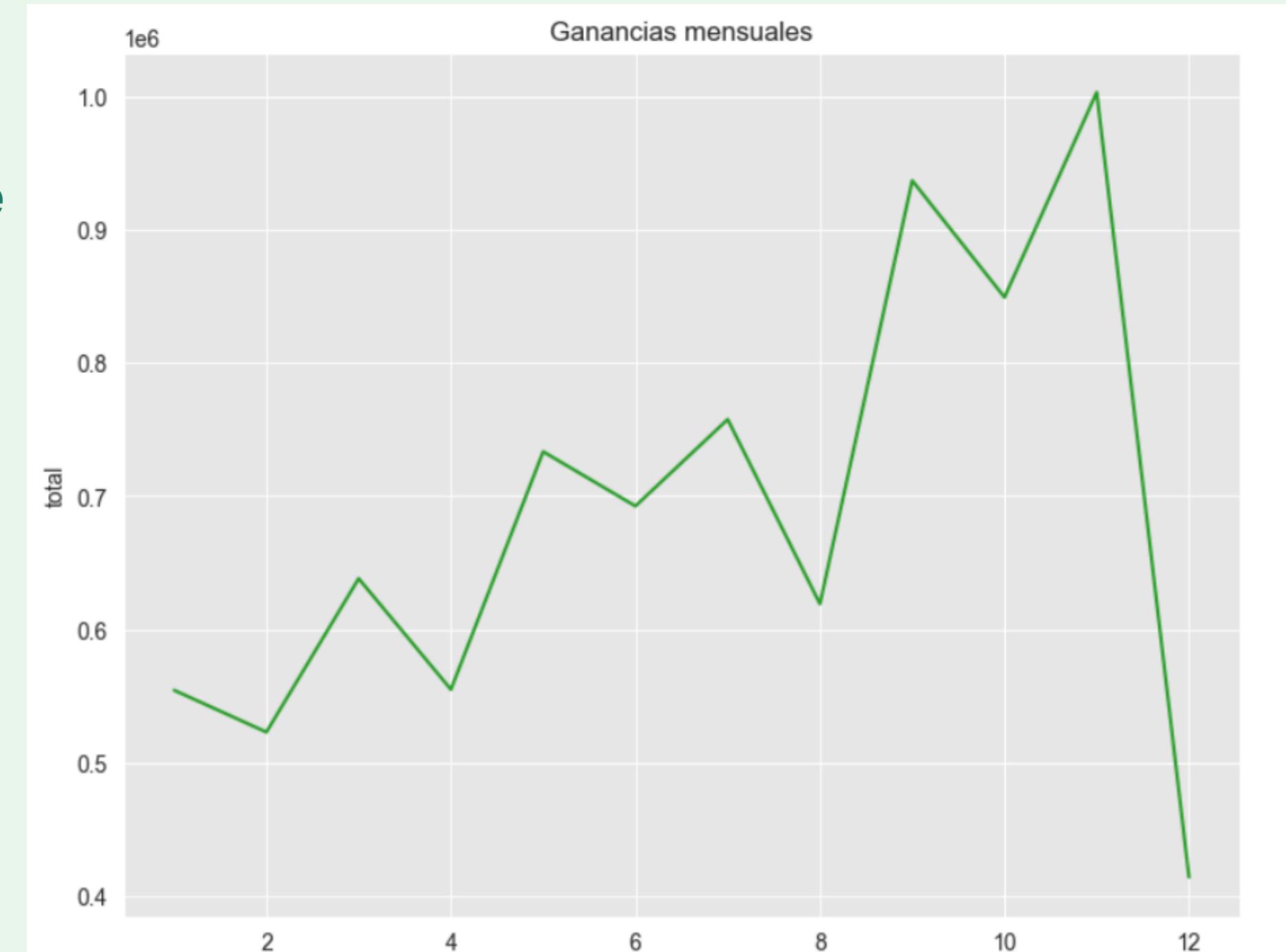


La región del Reino Unido tiene la mayor cantidad de usuarios únicos y se observa la gran diferencia que existe en relación al resto de regiones. publicitado las ventas.

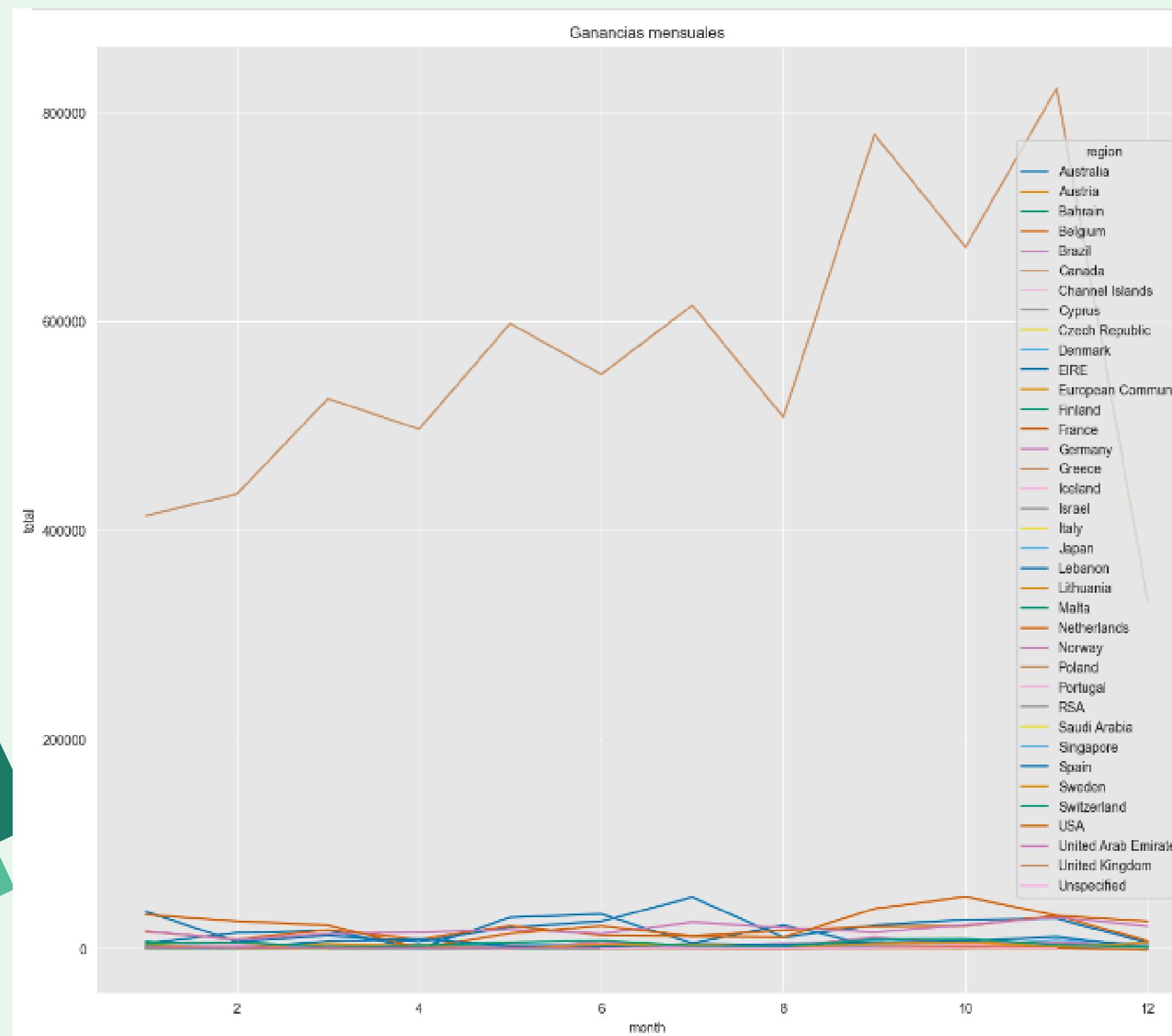
# VENTAS MENSUALES

Las ventas van en incremento, aunque con meses de grande pérdidas, aparentemente cada 3 meses.

Las ventas son más altas durante el mes de Noviembre y más bajas entre los meses de Enero y Diciembre.



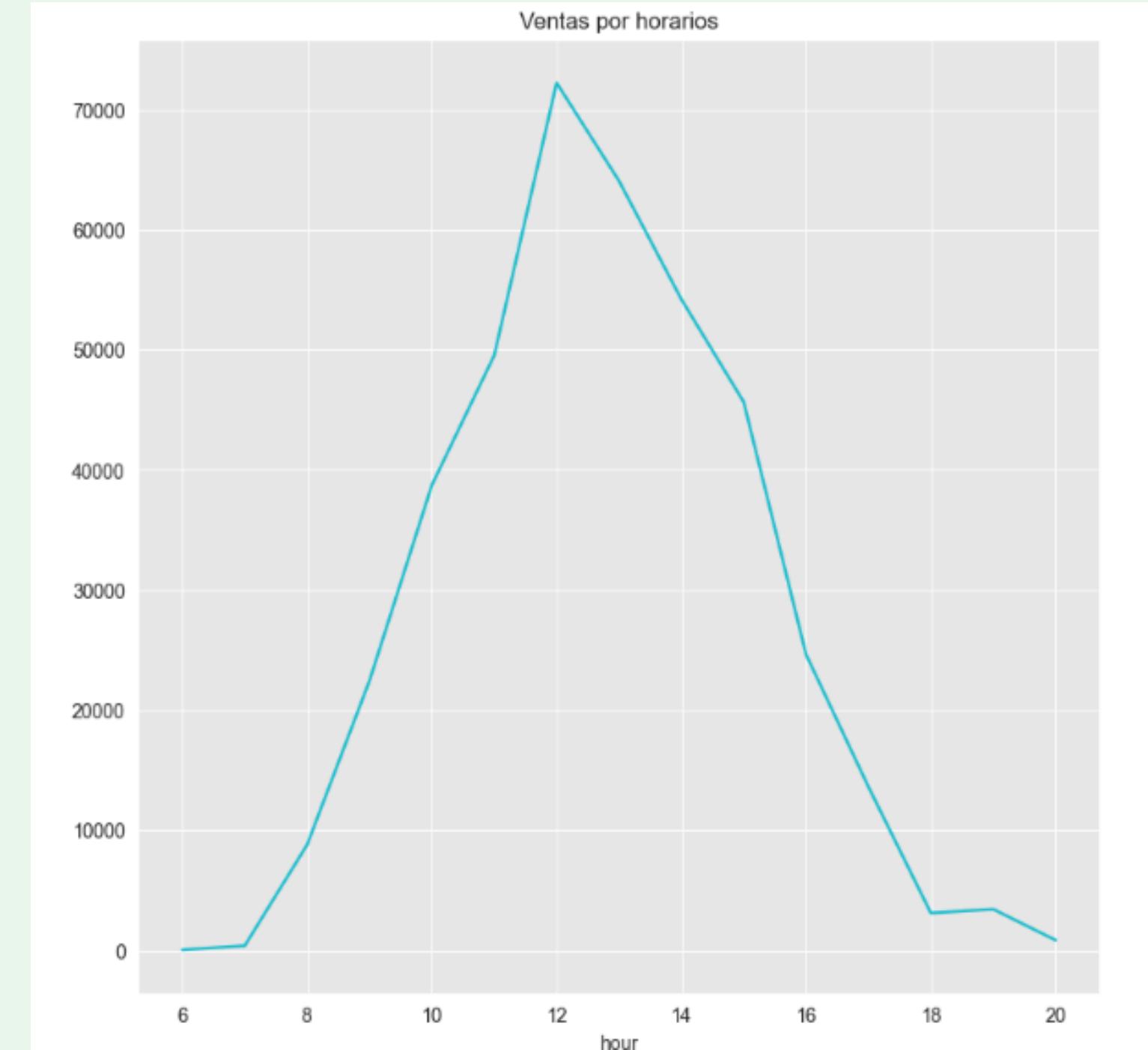
# VENTAS MENSUALES POR REGIÓN



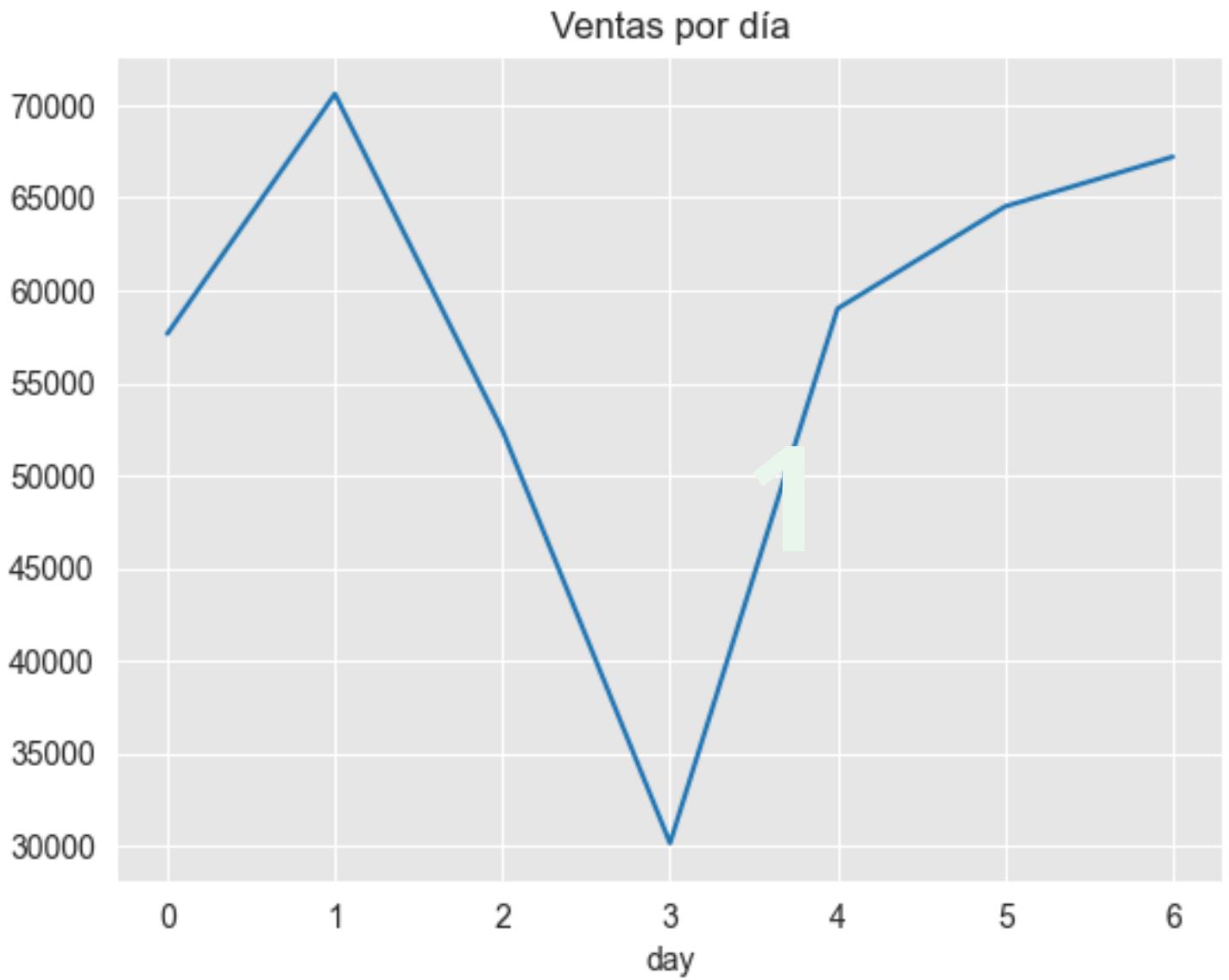
Las distintas regiones presentan picos de ventas en diferentes meses del año. Claramente la región del Reino Unido se presenta a un número de ventas mayor que el resto de las regiones,

# HORAS CON MÁS VENTAS

Como podemos observar, durante el día la actividad aumenta significativamente, es decir, entre las 7am y las 6pm aproximadamente los clientes prefieren realizar sus compras.



# Day of the week with the most sales



# Products most sold in UK



# CONSTRUCCIÓN DEL MODELO-

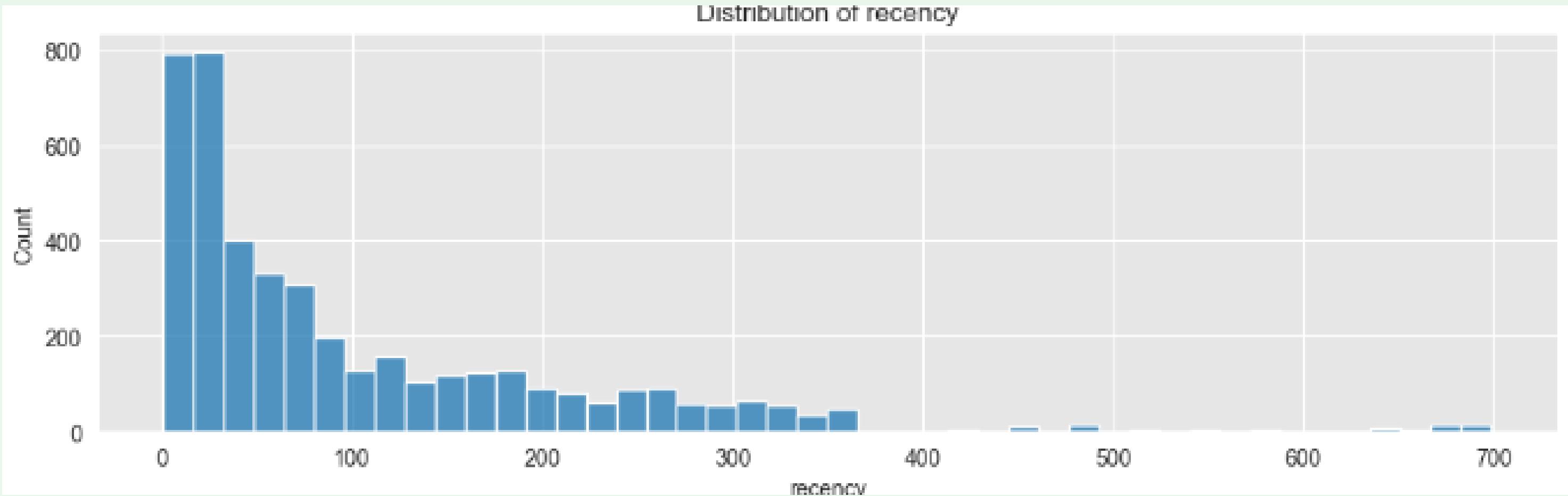
## CREACIÓN DE TABLA RFM

Primero vamos a crear una tabla con los valoresa RFM en donde cada sigla representa:

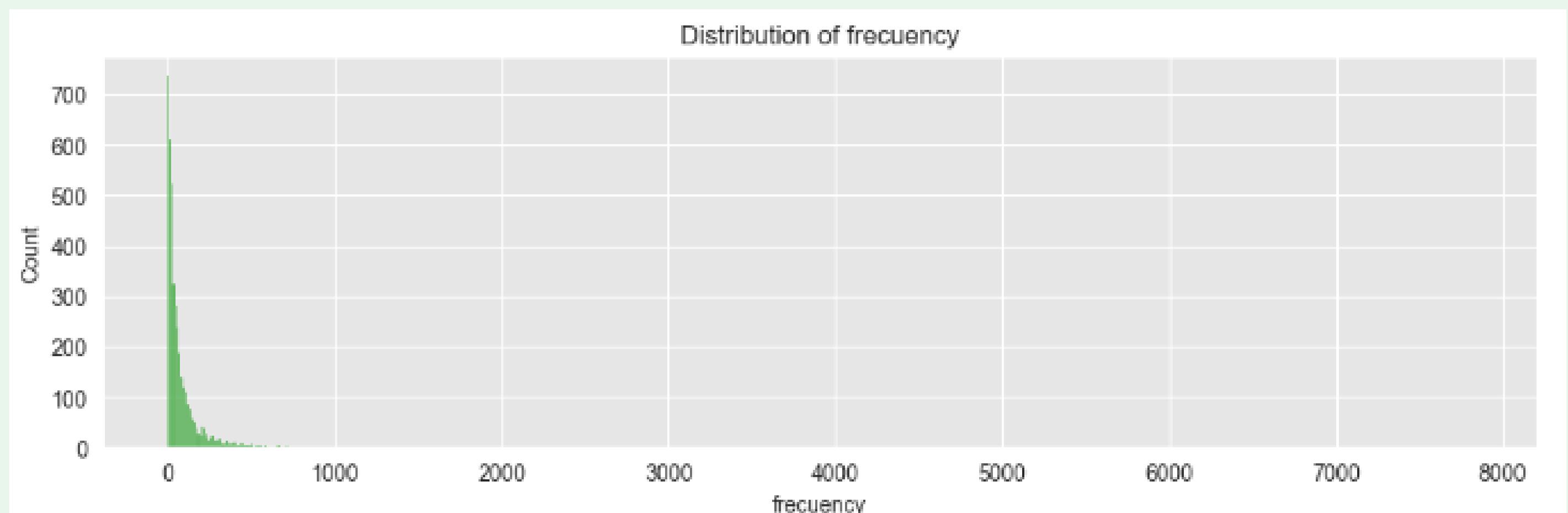
- Recency: Esta métrica nos dice la diferencia de días entre la diferencia entre la última fecha y la ultima fecha de compra\*
- Frequency: Es la cantidad de compras que ha realizado el cliente calculada haciendo la suma de todas sus compras
- Monetary: Es la cantidad de dinero gastado sumando todas sus compras totales

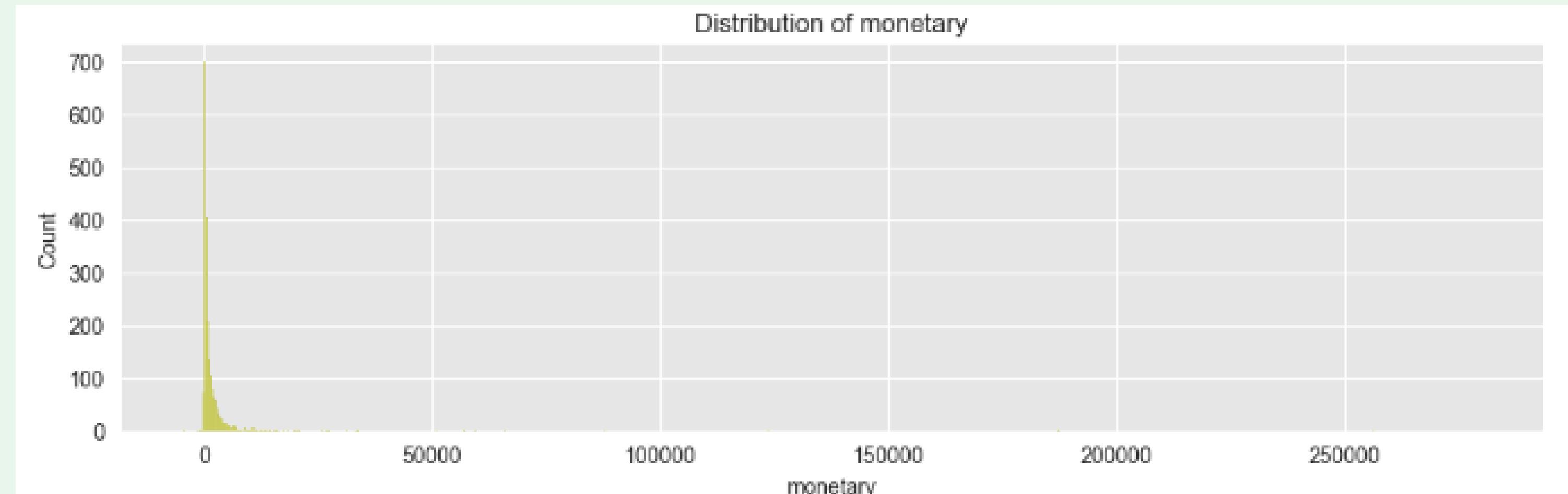
\*nota: para recency necesitamos aumentar en uno nuestra fecha máxima

Distribution of recency



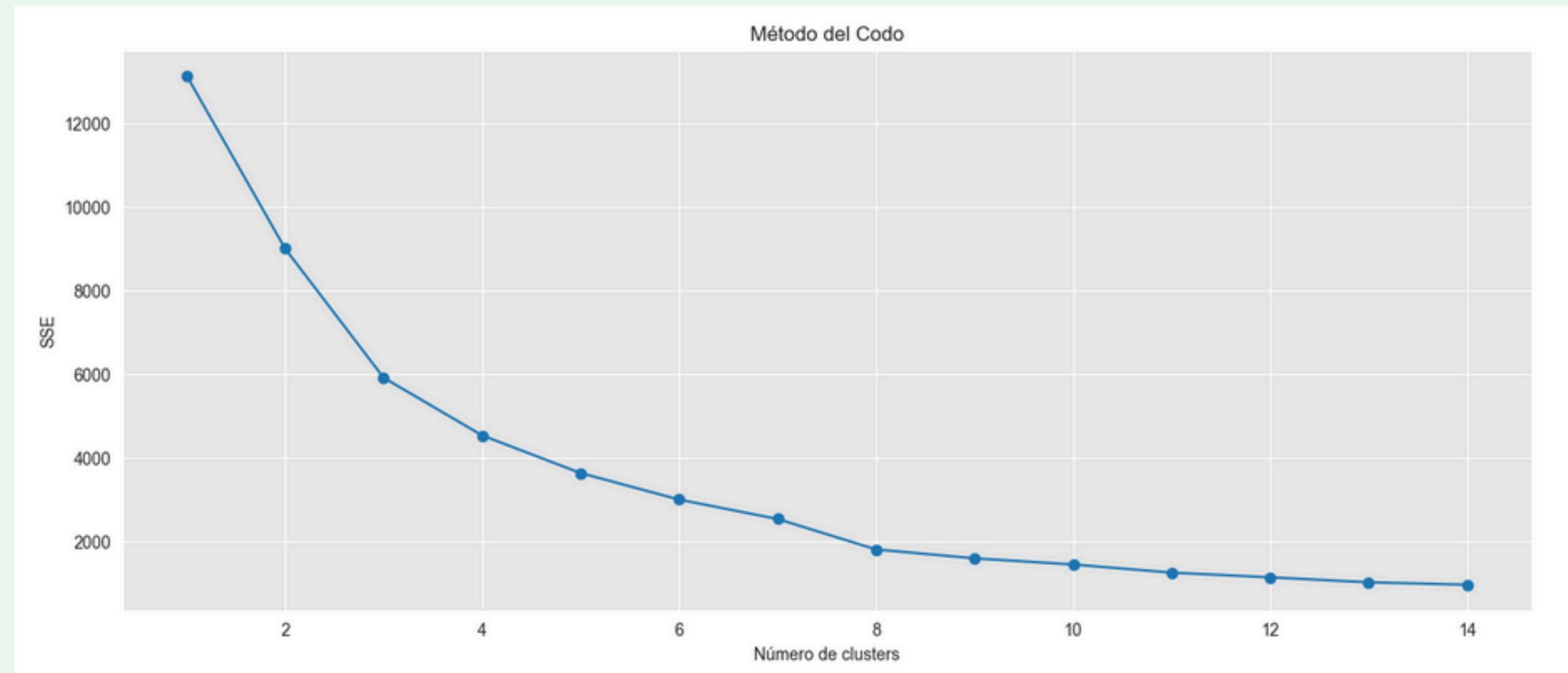
Distribution of frequency





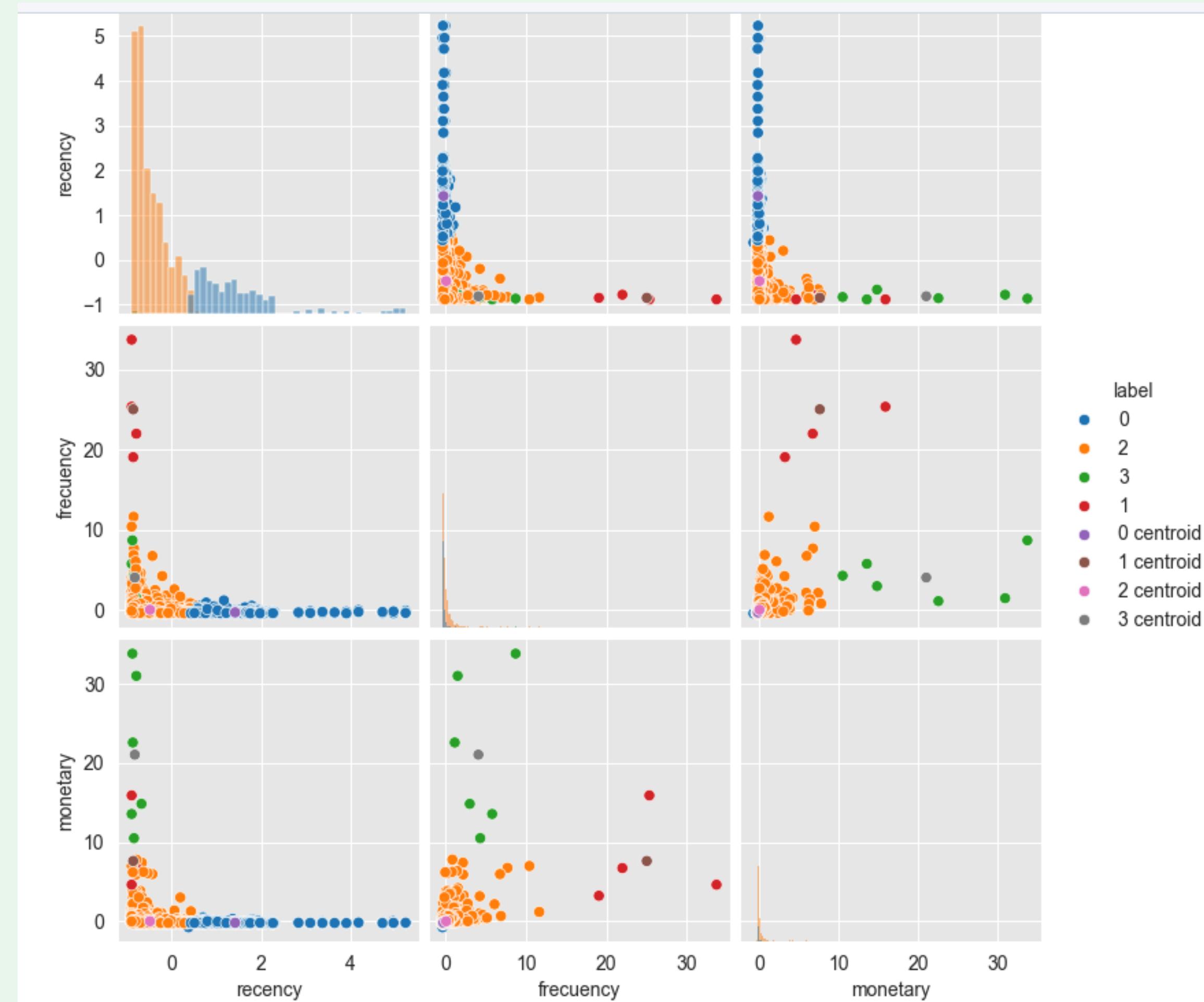
Aquí se puede notar que existen altos valores cercanos a cero en las tres columnas, esto puede indicar que hay varios clientes con compras menores a 100 días, así mismo, la cantidad de compras por cliente es menor a 250 aunque existen clientes que han realizado una gran cantidad de compras; finalmente, la mayoría de los clientes ha gastado aproximadamente menos de \$500 en sus compras.

## Método del codo para determinar el valor óptimo de clusters.



Podemos observar que el método nos dice que nuestro número óptimo está en 4 clusters

- Grupo 1 -> Clientes Platinum ya que tienen un tiempo corto de compra, con mucha compras y un buen gasto.
- Grupo 3 -> clientes Gold tardan un poquito más en comprar con alta frecuencia y tambien gastando mucho.
- Grupo 2 -> clientes Silver tardan más en comprar, compran menos y gastan mucho menos respecto a los grupos 1 y 3.
- Grupo 0 -> cliente basic son clientes que no han cuativados tantos por los productos que se ofrecen.



## Cambiaremos estos valores para una mejor comprensión de las tablas

		recency	frequency	monetary	cluster	cluster_meaning
customer_id	region					
12346	United Kingdom	328	2	0.00	0	basic
12347	Iceland	41	182	4310.00	2	silver
12348	Finland	77	31	1797.24	2	silver
12349	Italy	20	73	1757.55	2	silver
12350	Norway	313	17	334.40	0	basic

	cluster	recency	frequency	monetary	cluster_meaning
0	1	5.000000	5824.250000	64463.562500	platinum
1	3	9.000000	1022.833333	174719.183333	gold
2	2	47.792291	104.228510	1990.764054	silver
3	0	264.673933	28.562216	421.927684	basic

# CONCLUSIONES

- \*Gran cantidad de valores nulos sobre todo en la columna de `customer_id`.
- \*Una perdida de poco más de 25%.
- \*Creación de columna.
- \*Baja cantidad de usuarios con pedidos cancelados o devueltos
- \*Región del Reino Unido cuenta con una alta cantidad de clientes que representan cerca del 90% de las ventas totales.
- \*La mayoría de usuarios prefiere realizar sus compras después del medio día y en los últimos días de la semana, con un aumento en las ventas para los meses entre Septiembre y Diciembre.

# CONCLUSIONES

Para el modelo, se ha construido una tabla RFM con valores de Recency, Frequency y Monetary, se le aplicó estandarización a los datos y se utilizó el modelo de K-Means para formar 4 grupos, en donde se definieron como: "Platinum", "Gold", "Silver" y "Basic", ordenados del que proporciona mayores ganancias y con más frecuencia al peor, con menos ganancias y menor frecuencia.

# RECOMENDACIONES

- Creación de sistemas de puntos para alcanzar distintos niveles en donde vayas desbloqueando descuentos exclusivos, envíos gratis etc.
- Creación de campañas para mantener el interés de los niveles platinum y gold.
- Creación de campañas para incentivar más los clientes de nivel silver, como recordatorios de productos vistos con la aplicación de algún descuento.
- Creación de campañas para cautivar e incentivar más el nivel basic como descuentos muy grandes en la primera compra, una prueba de alguna membresía entre otra

# VISUALIZACIÓN DE DATOS

- Segmentación clientes Retail

LINK:

<https://maelmoon39.github.io/RetailSaviors/>