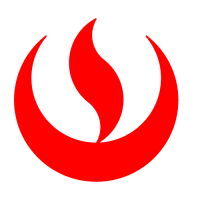
**UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS**

**FACULTAD DE INGENIERÍA**

**CARRERA PROFESIONAL**

**DE CIENCIAS DE LA COMPUTACIÓN**



**Asignatura:**

**CC57- APLICACIONES DE DATA SCIENCE**

**Sección: CC73**

**Trabajo Parcial**

**Sistema de recomendaciones para un E- commerce**

**Autores**

Franck Goñas Lopez – u201421134

**Profesor**

Carlos Fernando Montoya Cubas

Lima, abril de 2024

**CONTENIDO**

[1. Descripción del caso de uso 3](#_30j0zll)

[2. Descripción del conjunto de datos (dataset) 4](#_1fob9te)

[3. Análisis exploratorio de los datos (EDA) 4](#_3znysh7)

[4. Propuesta de Modelización 10](#_2et92p0)

[5. REFERENCIAS BIBLIOGRÁFICAS 11](#_1t3h5sf)

[6. ANEXOS 12](#_l19bdzymncpe)

# Descripción del caso de uso

En el contexto del comercio electrónico y las tiendas minoristas, uno de los retos más significativos es mantener la relevancia y personalización en la interacción con cada cliente. Con una vasta cantidad de productos disponibles y una diversidad de preferencias de los consumidores, es esencial para las empresas entender y anticipar las necesidades de sus clientes para mejorar la experiencia de compra y optimizar las ventas. El problema fundamental es cómo maximizar la eficacia del proceso de venta cruzada y la venta sugerida de productos sin abrumar al cliente con opciones irrelevantes. Un sistema efectivo debe ser capaz de predecir las preferencias futuras y presentar recomendaciones de producto.

Este escenario presenta una oportunidad significativa para desarrollar un sistema de recomendación que no solo aumente la satisfacción y fidelidad del cliente, sino que también impulse las ventas incrementales a través de recomendaciones personalizadas y contextualmente apropiadas.

# Descripción del conjunto de datos (dataset)

El dataset usado en el presente trabajo se extrajo de la plataforma web Kaggle, siendo más específico de <https://www.kaggle.com/datasets/carrie1/ecommerce-data/data>

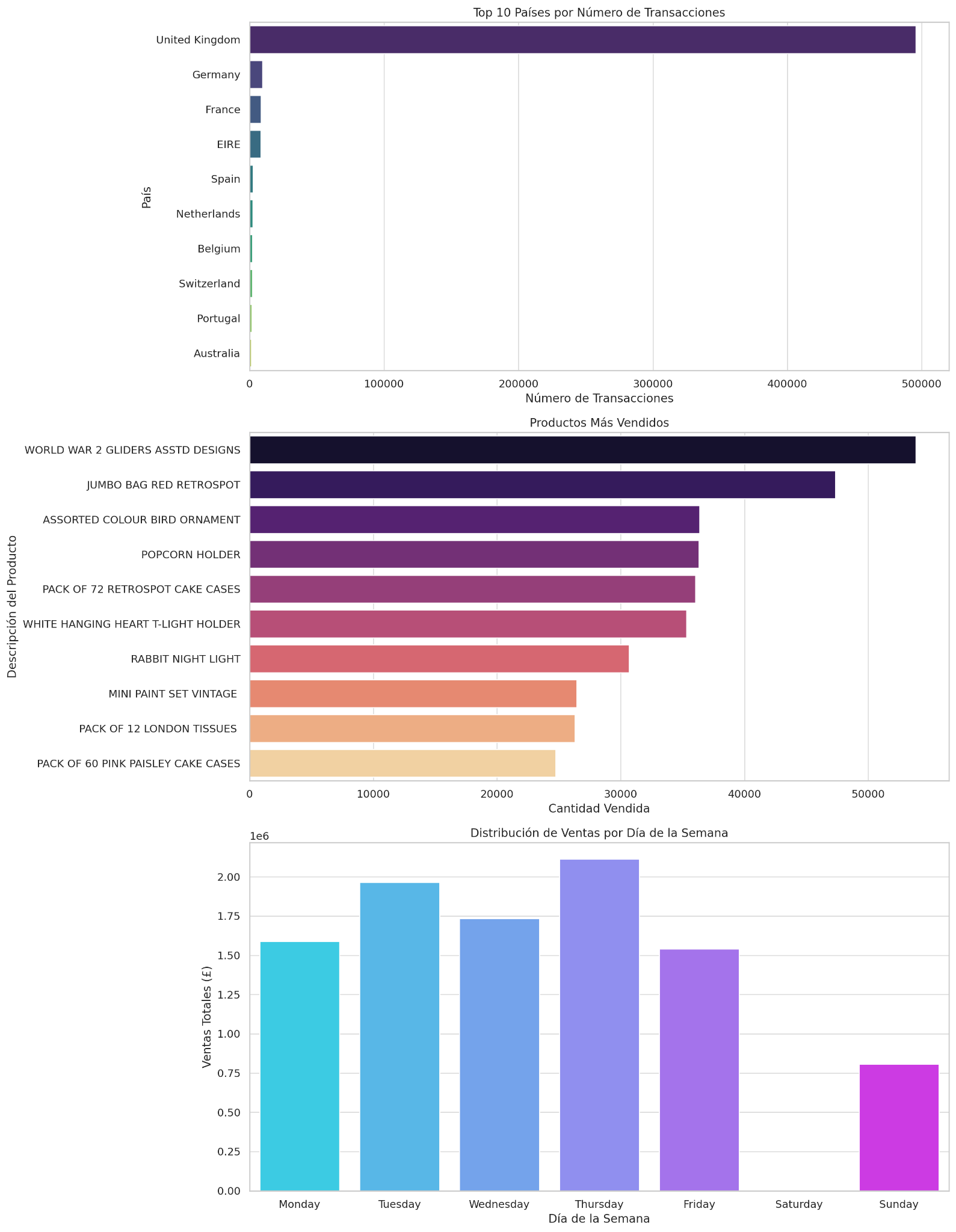
El conjunto de datos contiene 541,909 registros distribuidos en 8 columnas.

| Variable | Tipo de Dato | Descripción |
| --- | --- | --- |
| InvoiceNo | int | Número de factura o transacción. Identificador único para cada transacción. |
| StockCode | int | Código del producto. Código único asignado a cada producto diferente en el inventario. |
| Description | text | Descripción textual del producto. Puede contener detalles como color, material, etc. |
| Quantity | int | Cantidad del producto en la transacción. Los números negativos podrían indicar devoluciones. |
| InvoiceDate | date | Fecha y hora de la factura. Formato de fecha y hora cuando se realizó la transacción. |
| UnitPrice | float | Precio por unidad del producto en GBP. |
| CustomerID | int | Identificador único del cliente. |
| Country | text | País del cliente. Indica la ubicación geográfica desde donde se realizó la transacción. |

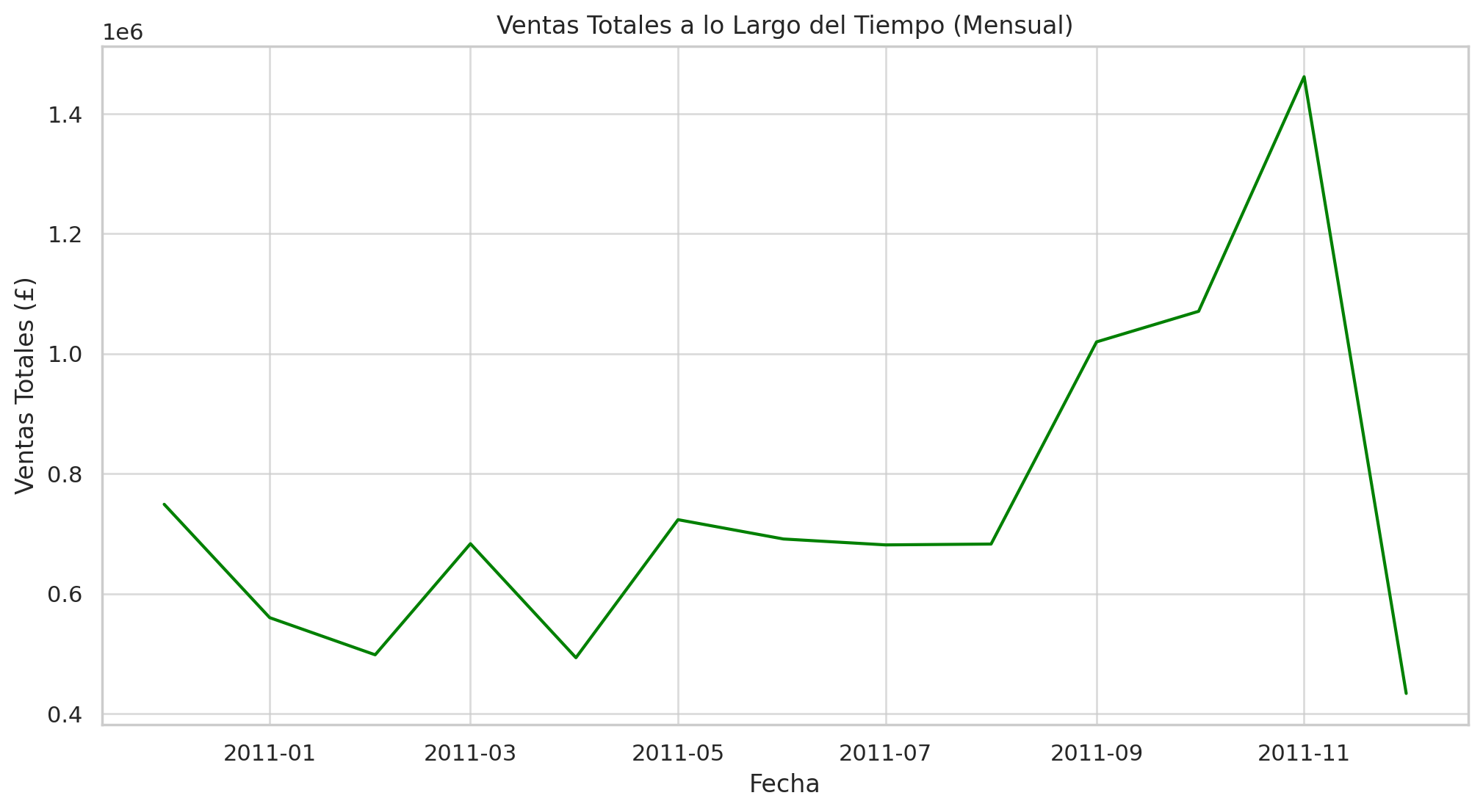
# Análisis exploratorio de los datos (EDA)

Se puede acceder al código en el Anexo 1.

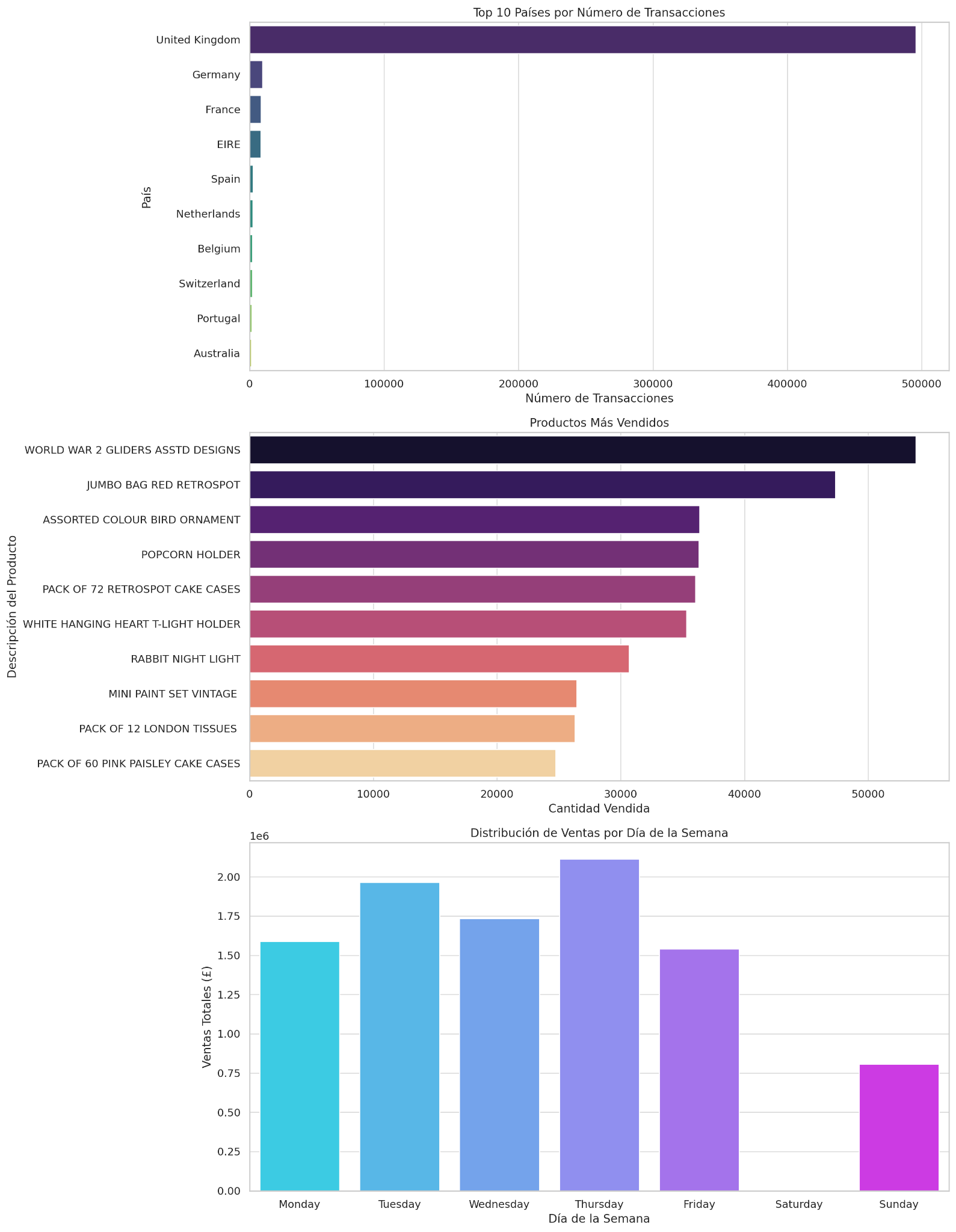
Para el análisis de datos se comenzó explorando la distribución de la las ventas en la variable país.



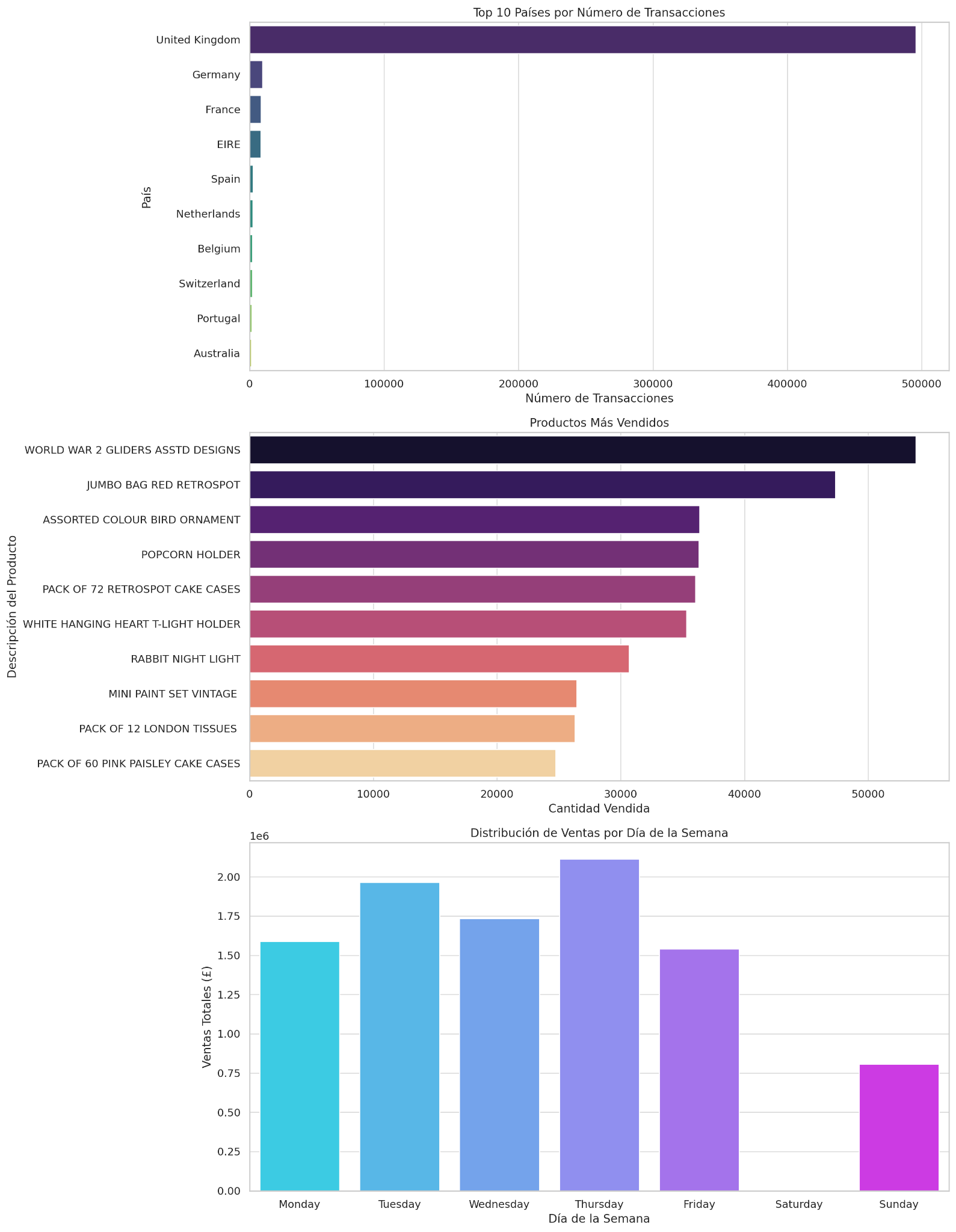
Podemos ver que la gran mayoría de compras se hicieron en United Kingdom.



Esta gráfica ayuda a observar las fluctuaciones mensuales en las ventas totales.



En esta gráfico podemos notar los 10 productos más vendidos por la empresa de e-commerce.



Esta gráfica Ilustra cuáles días de la semana generan más ventas, ayudando a planificar promociones y operaciones en función de la demanda.

# Propuesta de Modelización

Para el sistema de recomendación, luego de analizar diversos algoritmos se utilizarán Redes Neuronales Recurrentes (RNNs) debido a su eficacia en el manejo de secuencias de datos. El modelo se entrenará para reconocer patrones en las secuencias de compras de los clientes con la frecuencia de compra y las preferencias de categoría de producto.

El modelo de RNN se diseñará para procesar secuencias de compras, donde cada entrada de secuencia es un vector que representa un producto comprado, incluyendo detalles como el código del producto (codificado mediante técnicas de embedding), cantidad y precio. La salida del modelo será una lista de productos recomendados, basada en la probabilidad de que el cliente esté interesado en estos productos en su próxima compra.

# ANEXOS

Anexo1: Repositorio github

<https://github.com/franckhbz/-CC219-TP-TF-2024-1>