Summary Report

EDA on Online Retail Dataset

👨‍💼 **Name: Nikhil Sandip Rokade**

🏢 **Organization: Heal Bharat**

🗓️ **Date: 16-05-2025**

📝 **Task: Day 1 – Exploratory Data Analysis (EDA)**

🧩 1. Introduction

As part of my internship at Heal Bharat, I worked on an exploratory data analysis (EDA) task using the *Online Retail* dataset. This dataset contains one year of transaction records from a UK-based online store and includes details like product descriptions, invoice numbers, customer IDs, and purchase dates.

The aim of this task was to explore the dataset thoroughly to understand how the business performs — identify sales patterns, customer behaviors, and key revenue-driving products and countries. This kind of analysis helps businesses make more informed decisions in areas like marketing, inventory management, and customer retention.

⚙️ 2. Methodology

To begin, I loaded and explored the dataset using Python, mainly working with pandas, matplotlib, and seaborn in a Jupyter Notebook.

Here's a quick overview of the steps I followed:

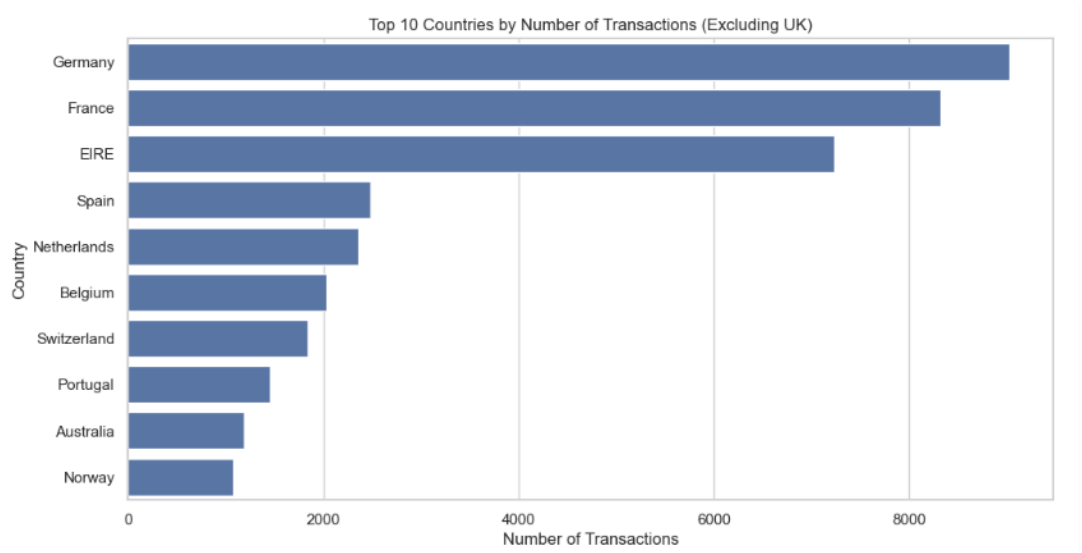
1. **Loading the Data**
   I loaded the Excel file (Online Retail.xlsx) and checked its structure — looking at the number of rows/columns, data types, and a few sample entries.
2. **Cleaning the Data**
   - I removed rows with missing customer IDs, since these couldn't be linked to any buyer.
   - Duplicates were dropped to avoid skewed results.
   - I filtered out rows with negative quantities, which usually represent returned or canceled items.
   - A new column called TotalPrice was created by multiplying Quantity with UnitPrice to represent transaction value.
3. **Feature Engineering and Time Analysis**
   - I converted invoice dates to proper datetime format.
   - Created a new column for monthly sales analysis.

4. **Visual Exploration**
   I used visualizations to highlight patterns and trends — such as top countries, best customers, popular products, and monthly sales performance.

📈 3. Key Insights

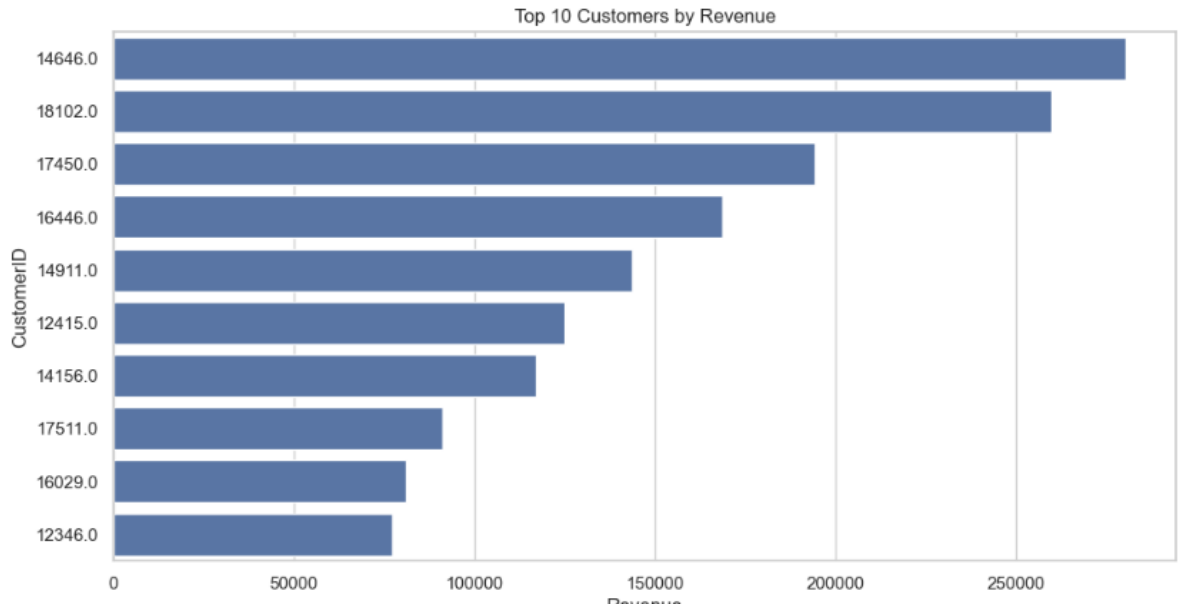
🌍 Top Countries (Excluding the UK)

Apart from the UK (which had the most transactions), countries like the Netherlands, Germany, and France stood out in terms of order volume — showing that the store had strong international sales too.



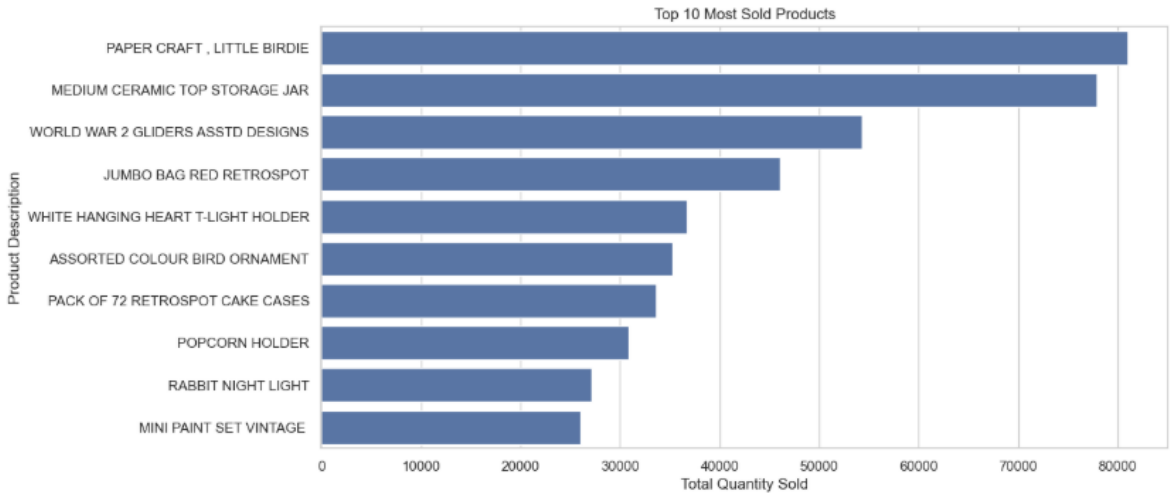
👤 Top Customers by Revenue

A small number of customers generated a large chunk of the revenue. These could be repeat buyers or bulk purchasers — suggesting the value of identifying and retaining such high-value customers.



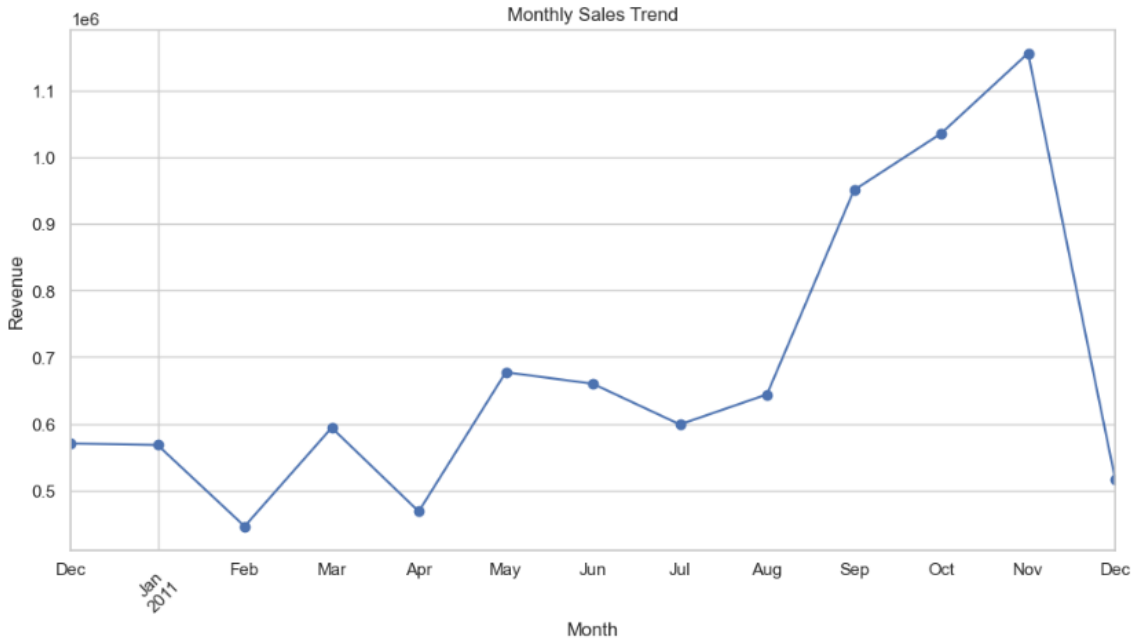
🛍️ Most Sold Products

Products like decorative items and stationery — such as heart-shaped candle holders and vintage-style accessories — were among the bestsellers in terms of quantity sold.



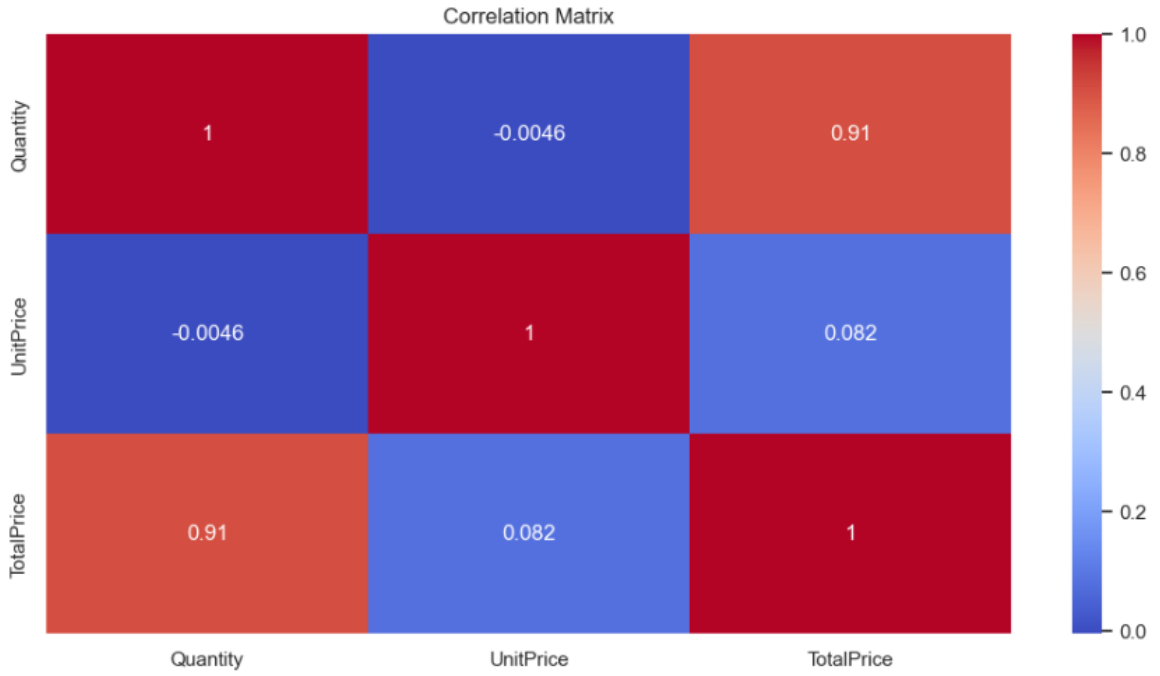
🗓️ Monthly Sales Trends

Sales showed a clear upward trend towards the end of the year, especially in November and December — likely driven by the holiday season and promotions.



📊 Correlation Insights

There was a strong correlation between Quantity and TotalPrice, which makes sense — more items sold usually mean more revenue.



🧠 4. Observations & Conclusion

- The UK led in transaction volume, but the company also had solid international demand from several European countries.
- A small group of loyal or bulk-buying customers contributed significantly to overall revenue.
- Seasonal peaks — especially during holidays — play a big role in sales, indicating a need for marketing campaigns and stock planning around those times.
- Some products are consistent top-performers, which could help guide inventory and promotion strategies.

Overall, this EDA gave me a much better understanding of how to find insights from raw data. It was a great learning experience that showed how important data-driven decisions are in real business scenarios.
