UK Online Retail Sales Data Analysis

Project Background

This project analyzes sales data from a UK-based online retail store, with the aim of identifying sales trends, customer behavior patterns, and return patterns. The insights generated can help the business improve sales performance and enhance customer satisfaction.

Data Cleaning

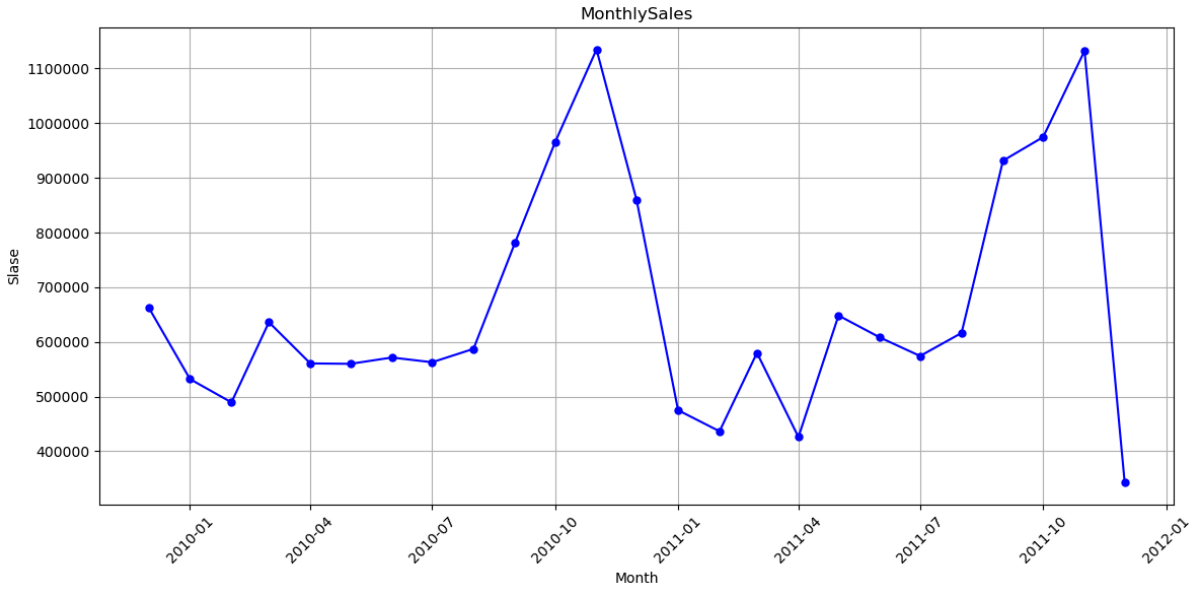
1. **The data size:** 1,067,371 rows and 8 columns.
2. **Missing values:**
    - Description: 4,380 missing values
    - Customer ID: 243,007 missing values
3. **Data transformations:**
    - Converted InvoiceDate to datetime type
    - Converted Customer ID to string format
4. **Null values handling:**
    - Remove all rows with missing Customer ID
    - Final dataset size after cleaning: 824,364 rows
5. **New feature:**
    - Calculated Sales as Quantity * UnitPrice

6. **Return order detection:**
    - Identified return orders based on invoice number starting with ‘C’
7. **Date range:**
    - From December 1,2009 to December 1,2011.

Data Analysis & Visualization

1. Monthly Sales Trends

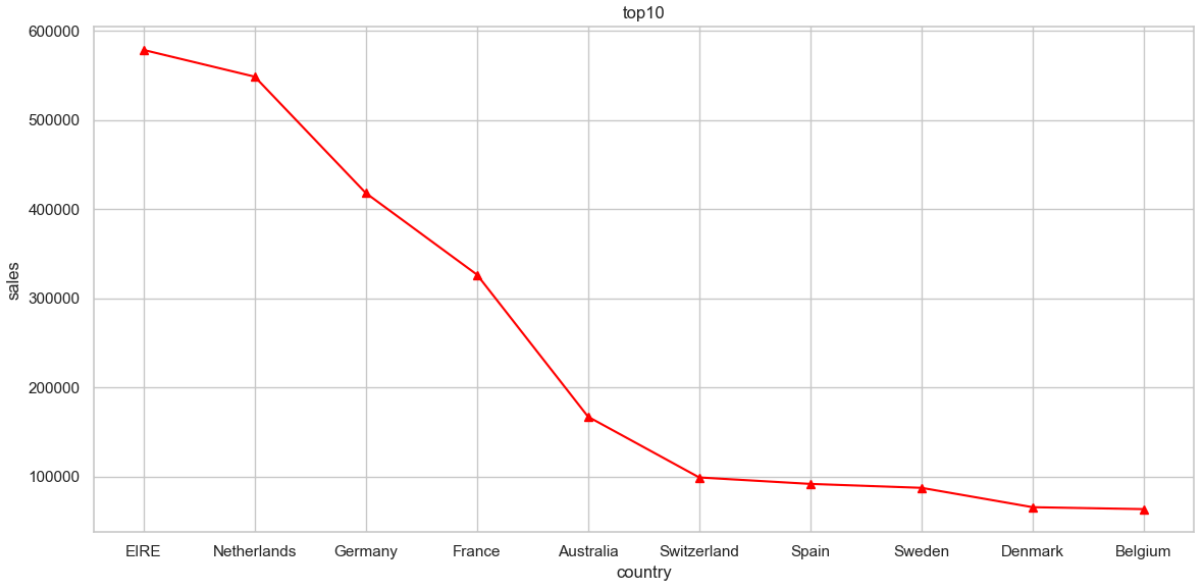
📈 Plot: Monthly sales over time



Finding:
Augster, September, October,Novenber consistently show higher sales.
Recomendation:
Increase the production and stuff during peak months; scale down in low seasons to reduce costs.

2. Sales by Country

🌍 Plot: Top 10 countries by total sales



Finding:

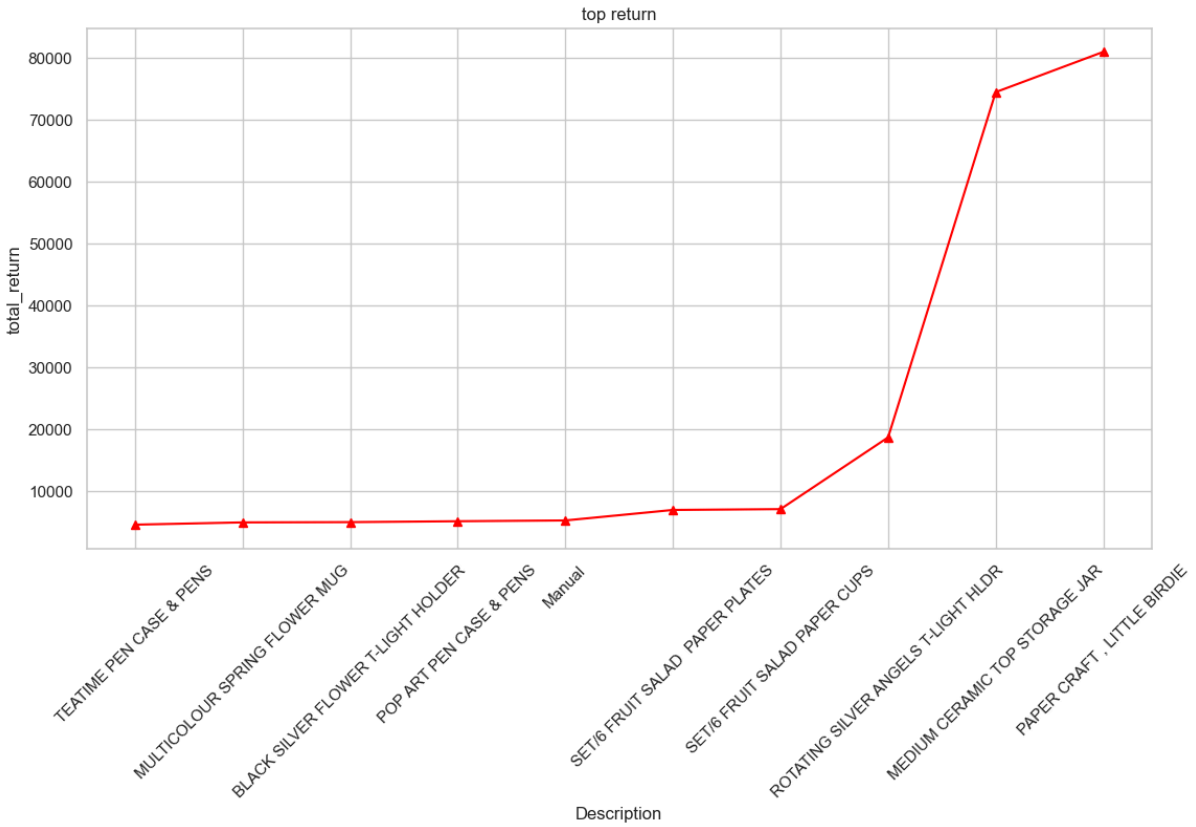
- Top 3 countries: Ireland(EIRE), Nethrland, Germany -each with sales exceeding £400,000
- France and Australia also show strong potential.

Recommendation:

Focus marketing and logistics on top-performing regions while exploring growth in emerging markets.

3. The Return Order Analysis

↩️ Plot: Top 10 most returned products



Finding:

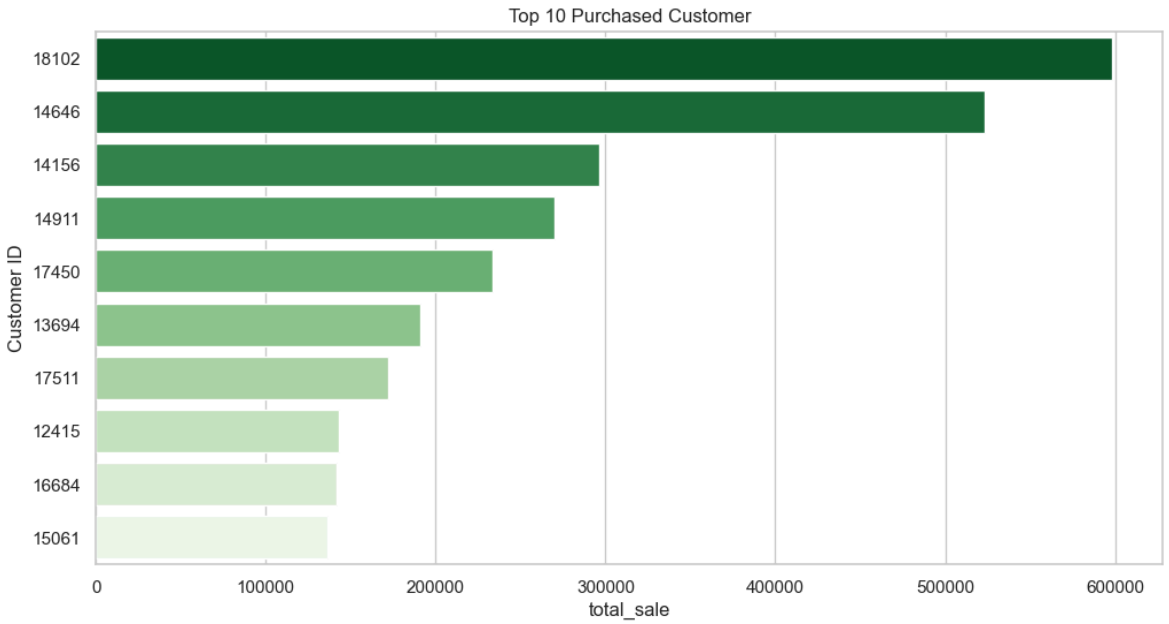
- Return Rate: **17.61%**
- Most returned items:
  - ❖ Medium Ceramic Top Storage Jar
  - ❖ Paper Craft,Left Birdie

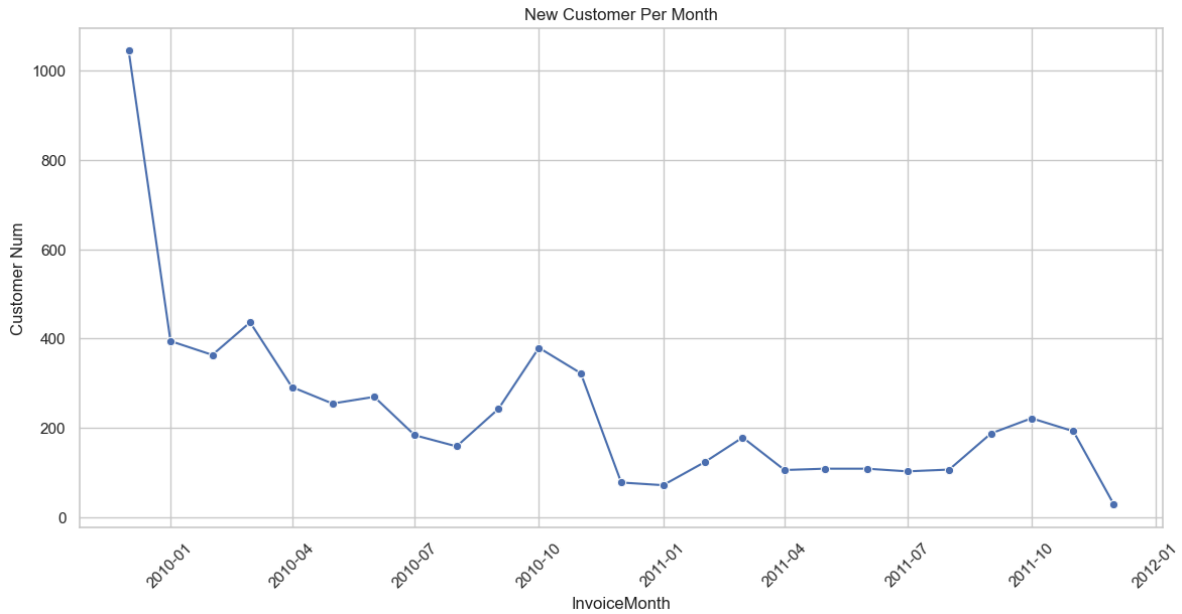
Recommendation:

Investigate quality issuers or customer dissatisfaction with these products and improve accordingly.

4. Customer Behaviour Analysis

🛒 Plots: Top purchasing customers and new customers per month





Finding:

- Top customers:
  Customer 18102 and Customer 14646 contribute significantly more than others.
- New customer count drops sharply after 2010

Recommendation:

- Implement retention strategies for top customers.
- Analyze causes of customer decline and improve acquisition strategy

Conclusion & Recommendations

1. Address high return rates for key products by improving quality or setting clearer product expectations.
2. Focus on top-performing countries(EIRE, Netherlands, Germany) for sales and marketing expansion.
3. Build loyalty programs or personalized offers for high-value customers(18102, 14646).
4. Plan resource allocation to align with seasonal peaks in August to November.

Tool & Skill

1. **Panda** for data cleaning, transformation, and aggregation
2. **Matplotlib/seaborn** for data visualization
3. **Jupyter Notebook** for documentation and interactive analysis