# ONLINE RETAIL EXPLORATORY DATA ANALYSIS (EDA)

Rhea George

# PROJECT SUMMARY

This project explores over 500,000 online transactions made by customers of a UK-based retailer between 2010–2011. It analyzes customer behavior, product performance, and revenue trends to uncover actionable business insights.

The project applies core data analysis techniques to simulate how analysts extract value from real-world e-commerce datasets.

# PROJECT GOALS

1. 
   - **Preprocessing and cleaning transactional data**
   - **Analyzing KPIs like revenue, returns, and customer spend**

2. 
   - **Visual storytelling with clear business metrics**
   - **Insights ready for dashboards and business review**

3. 
   - **Supports strategy through RFM segmentation and country-wise analysis**
   - **Helps stakeholders identify top customers and product demand trends**

# DATASET OVERVIEW

- **Source: UCI Machine Learning Repository**

- **Size: ~540,000 transactions**

# ANALYSIS PIPELINE

**Data Cleaning**

- Removed nulls in CustomerID
- Filtered out returns (InvoiceNo starting with 'C')
- Created TotalPrice = Quantity × UnitPrice

**Time-Series Analysis**

- Monthly revenue trends
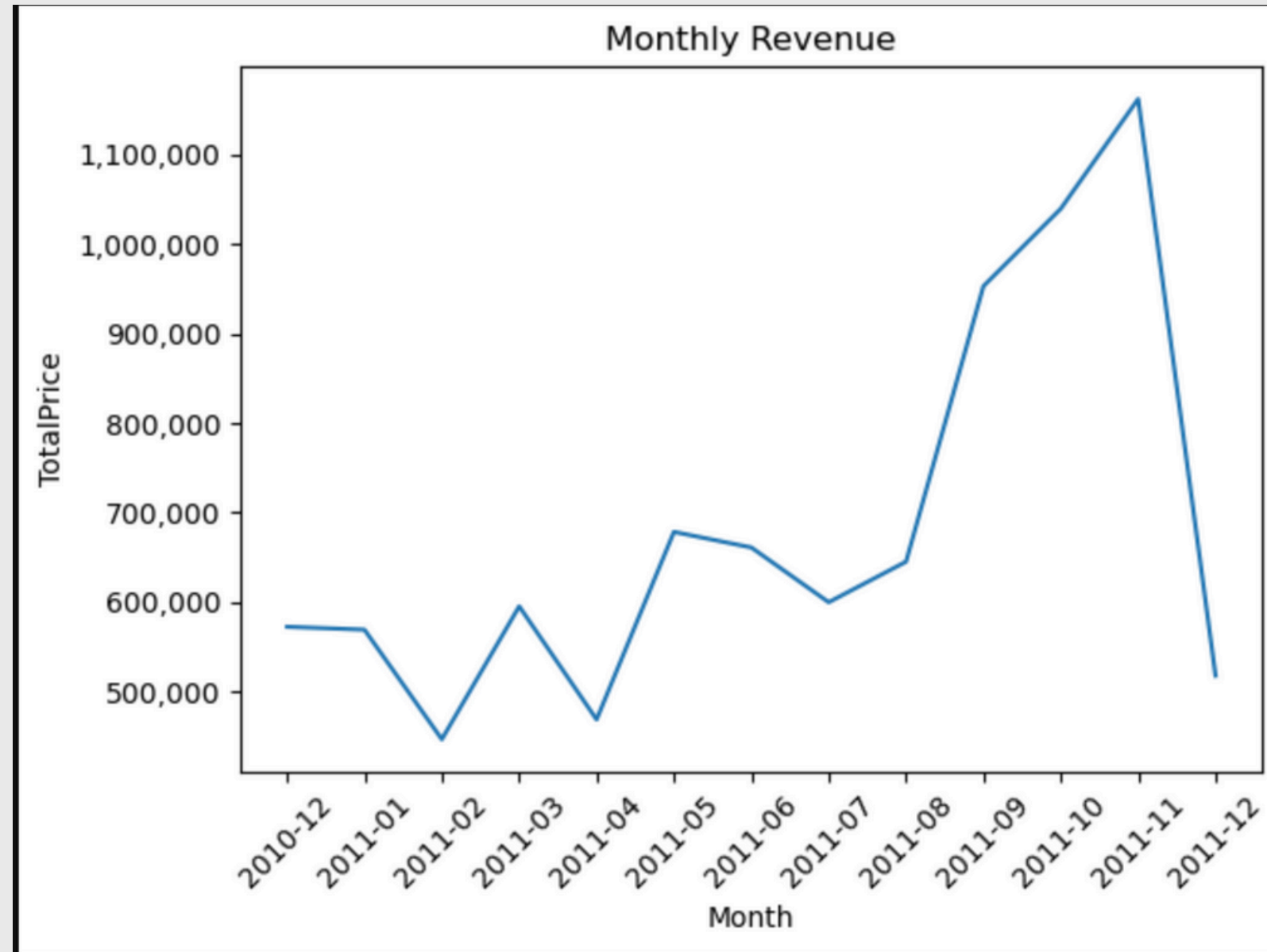- Hourly & daily order patterns

# ANALYSIS PIPELINE

**Customer Segmentation**

- RFM (Recency, Frequency, Monetary) analysis
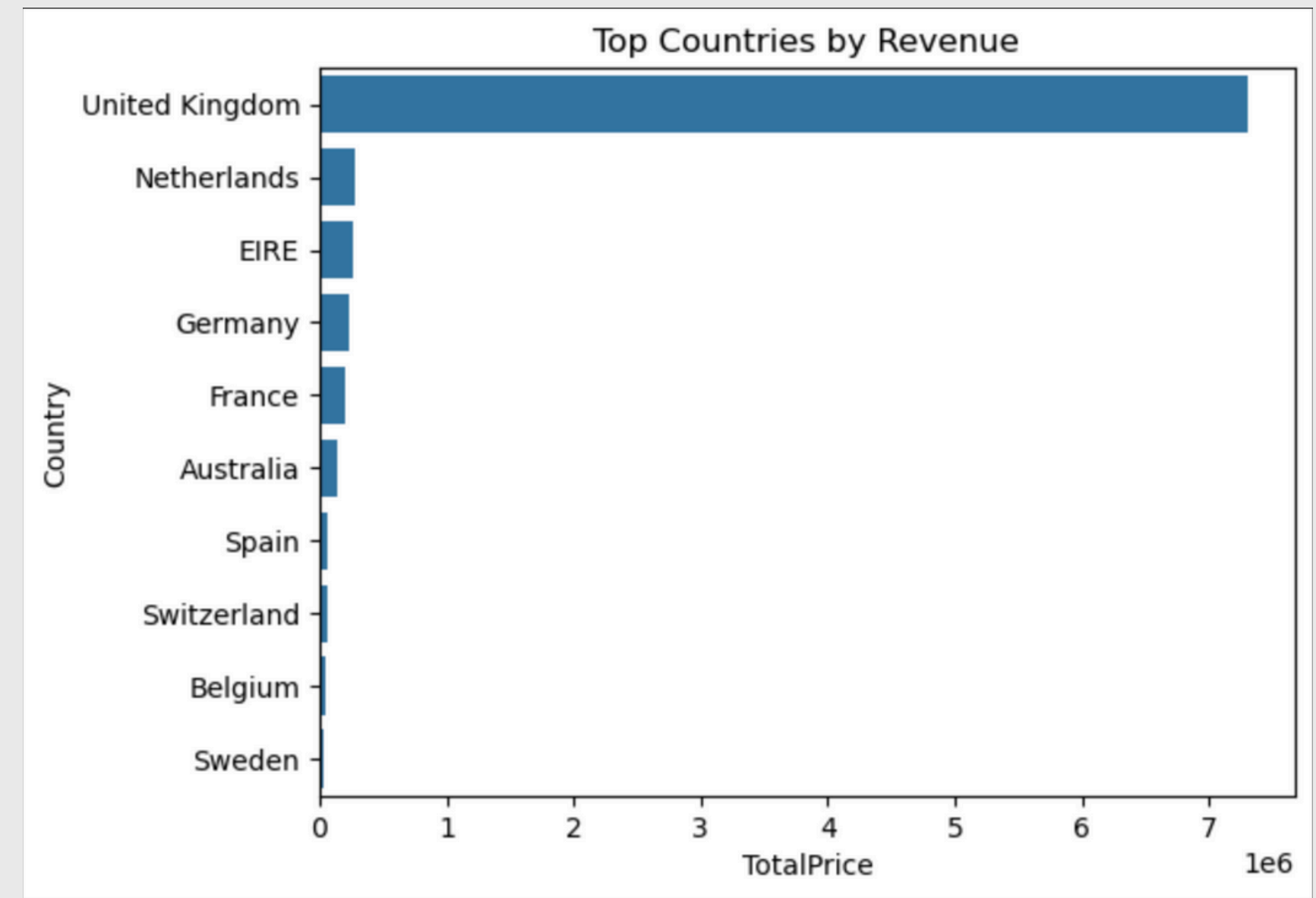- RFM heatmap & bar plots

**Product Insights**

- Most ordered vs highest revenue items
- WordCloud of product descriptions
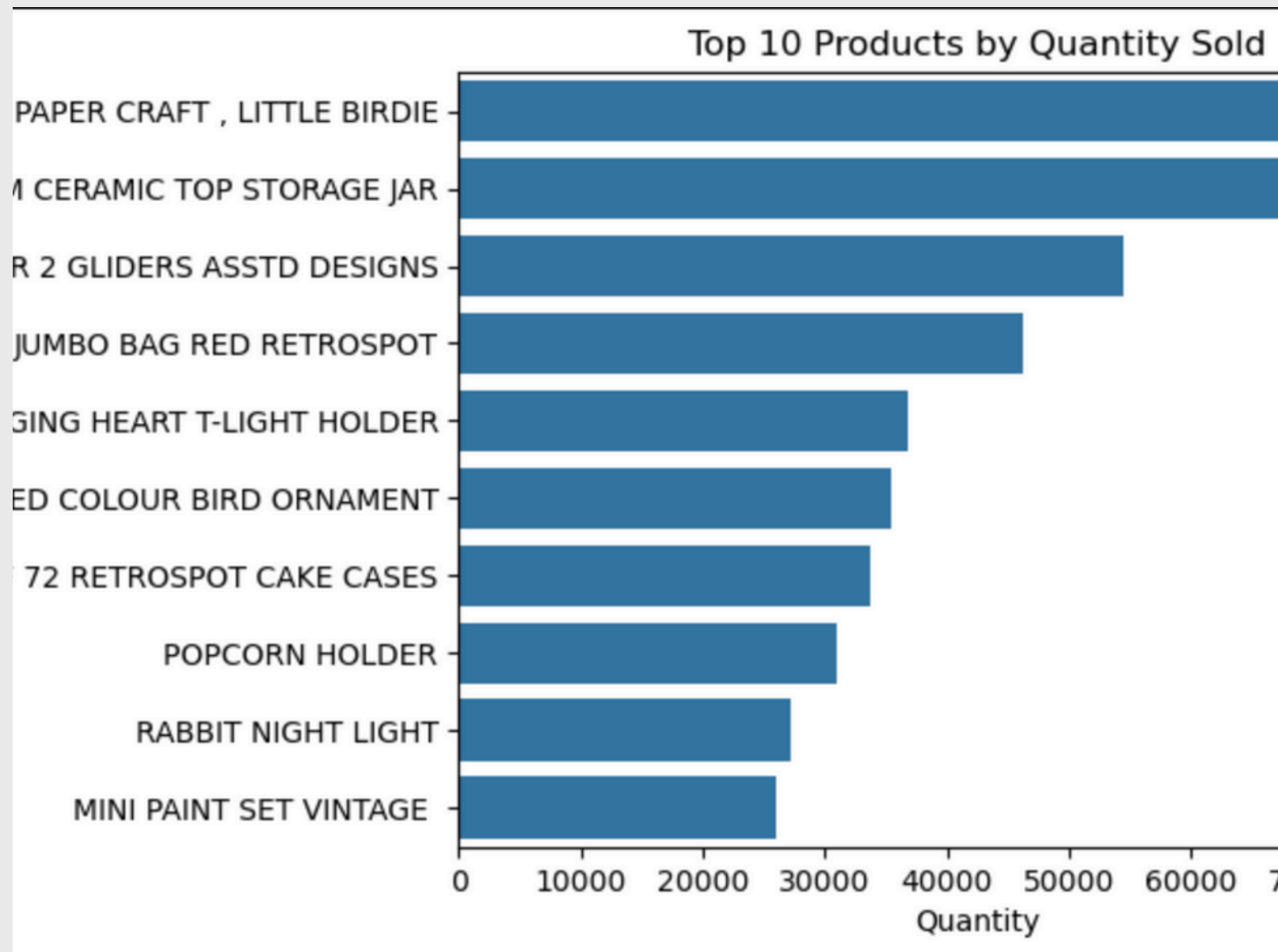
# VISUALIZATIONS



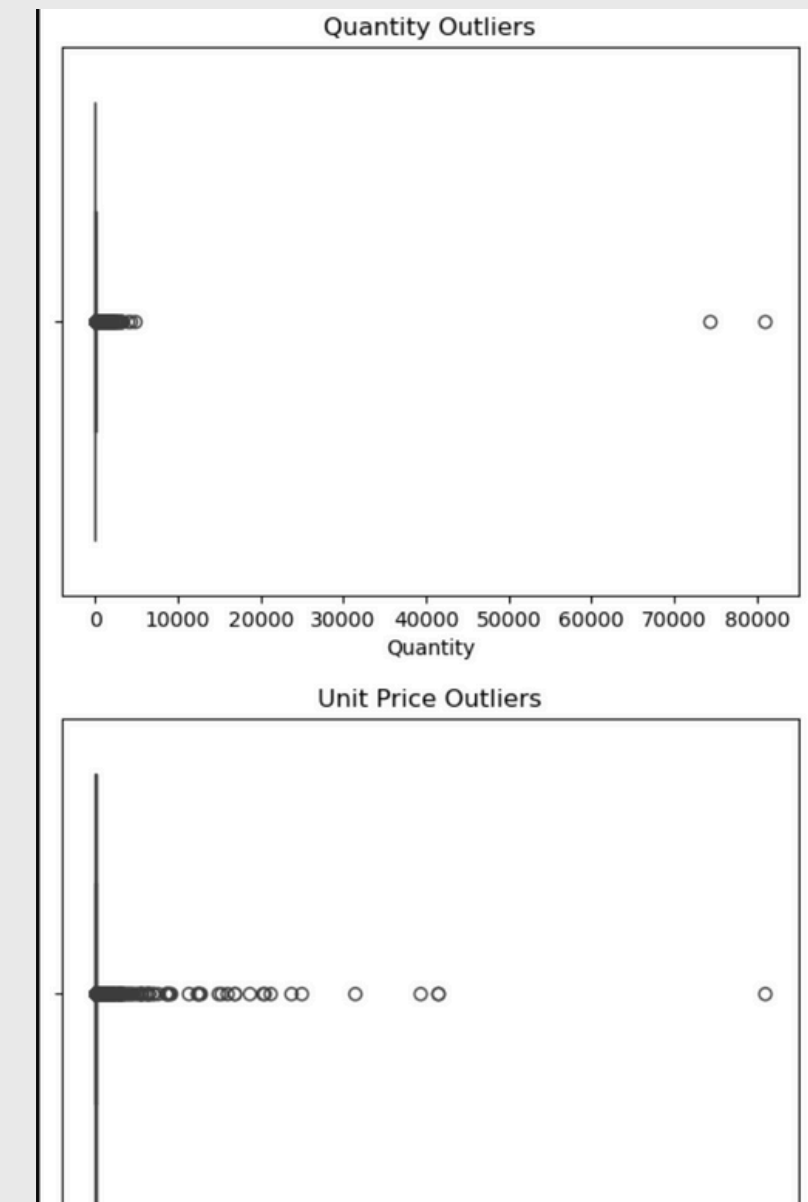Monthly Revenue Trend



Top Countries by Sale

# VISUALIZATIONS



Top 10 Products Sold



RFM Boxplots

# KEY TAKEAWAYS

- UK dominates sales but international customers drive high-value purchases
- 20% of products contribute ~80% of revenue (Pareto rule)
- High-return rate items detected via negative quantities
- RFM reveals clusters of loyal, high-value customers

# THANK YOU