
Optimized Exploratory Data Analysis Report

1. Introduction

This report analyzes online retail transaction data to uncover insights into sales trends, product performance, and customer behavior. Key columns include:

- **InvoiceNo:** Transaction identifier.
- **InvoiceDate:** Transaction date and time.
- **StockCode:** Item identifier.
- **Description:** Product description.
- **Quantity:** Units sold per transaction.
- **UnitPrice:** Price per unit.
- **Country:** Customer location.

Objectives

- Analyze sales trends.
- Identify top products and regions.
- Provide insights for inventory and customer segmentation.

2. Data Overview

Dataset Summary

- **Shape:** (541,909 rows, 8 columns).
- **Key Columns:** All relevant data types are as expected.
- **Preview:** | InvoiceNo | InvoiceDate | StockCode | Description | Quantity | UnitPrice | Country |
-----|-----|-----|-----|-----|-----|-----
-----| 536365 | 2010-12-01 08:26 | 85123A | WHITE HANGING... | 6 | 2.55 |
United Kingdom|

Issues Identified

- Missing values:
 - **Description:** 1,454 missing entries (filled as `Missing`).
 - **CustomerID:** 135,080 missing entries (replaced with `Unknown`).
 - Outliers in **Quantity** and **UnitPrice**.
-

3. Data Cleaning

Steps Taken

1. **Handling Missing Values:**
 - Description: Replaced with "Missing".
 - CustomerID: Replaced with "Unknown".
 2. **Derived Column:**
 - Added `TotalRevenue = Quantity × UnitPrice`.
 3. **Datetime Conversion:**
 - Converted `InvoiceDate` to datetime format.
 4. **Removed Outliers:**
 - Filtered rows with negative or zero `Quantity` and `UnitPrice`.
-

4. Exploratory Analysis

4.1 Sales Trends

- Peak sales occurred in **November**, showing seasonality.
- Monthly revenue trends indicate strong end-of-year performance.

4.2 Product Performance

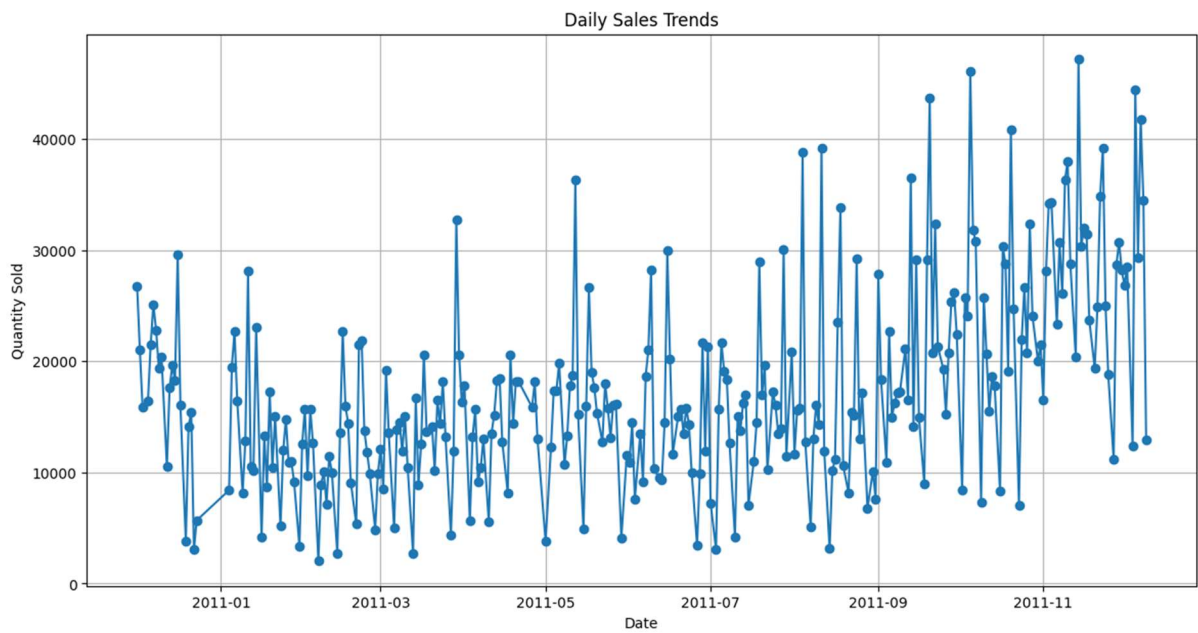
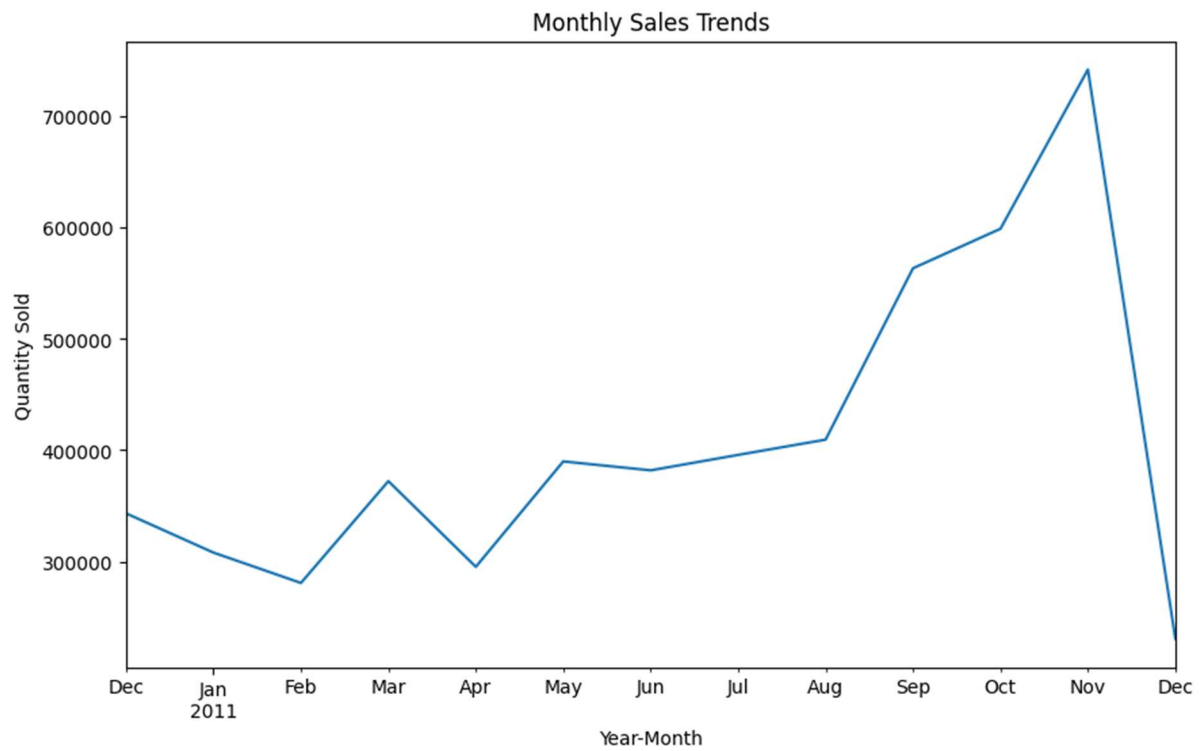
- **Top Products by Quantity Sold:**
 - WORLD WAR 2 GLIDERS ASSTD DESIGNS.
- **Top Products by Revenue:**
 - DOTCOM POSTAGE: £206,245.48.
 - REGENCY CAKESTAND 3 TIER: £164,762.19.

4.3 Regional Insights

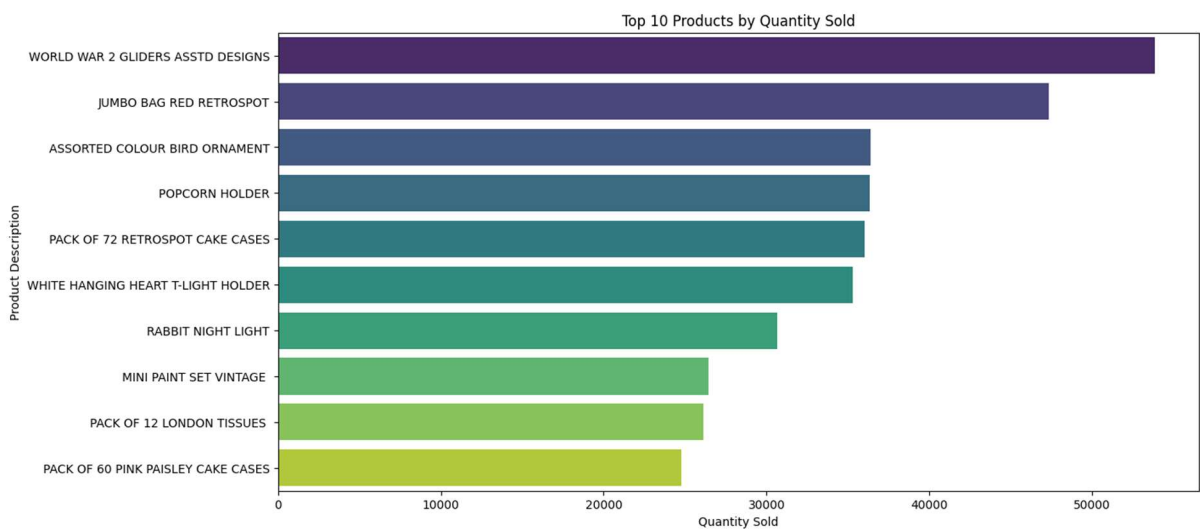
- **Top Countries by Revenue:**
 - United Kingdom: £8,209,930.48 (80%+ of total revenue).
 - Netherlands: £284,661.54.
 - EIRE: £263,276.82.
-

5. Visualization

5.1 Sales Trends

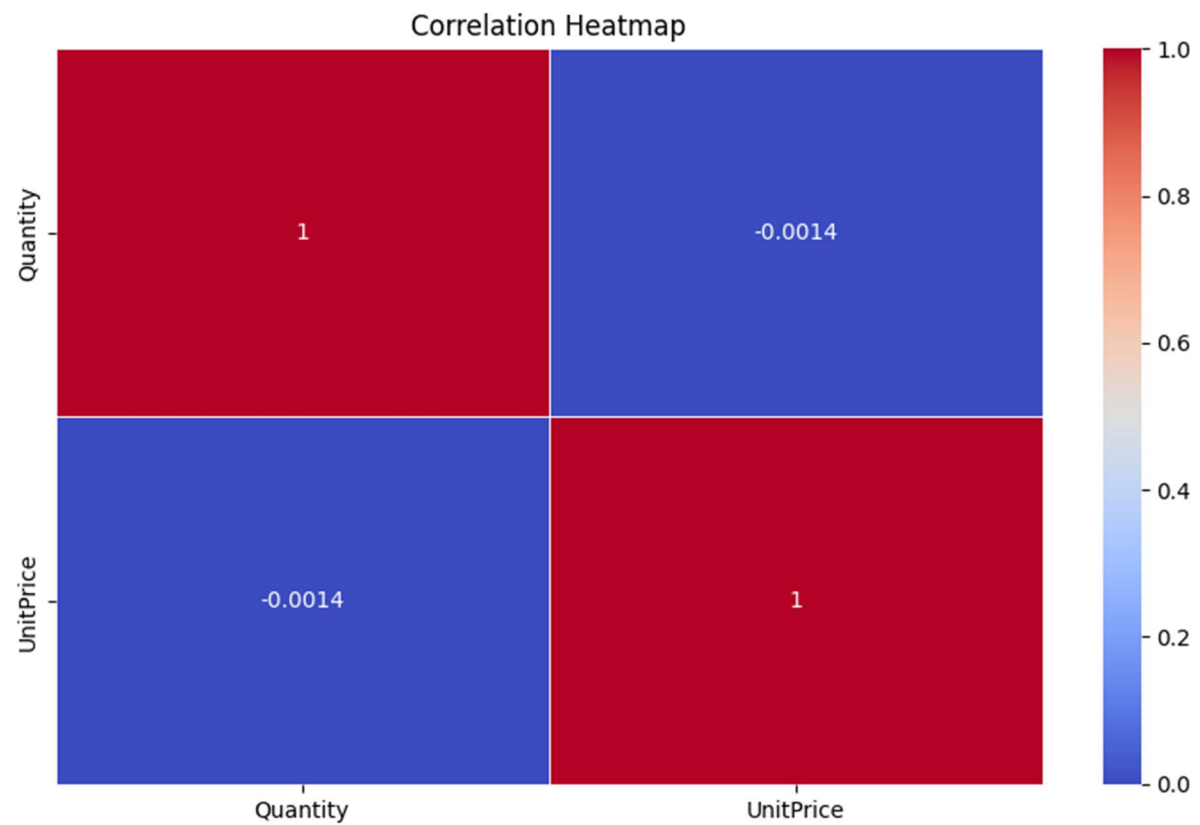


5.2 Product/Category Performance:

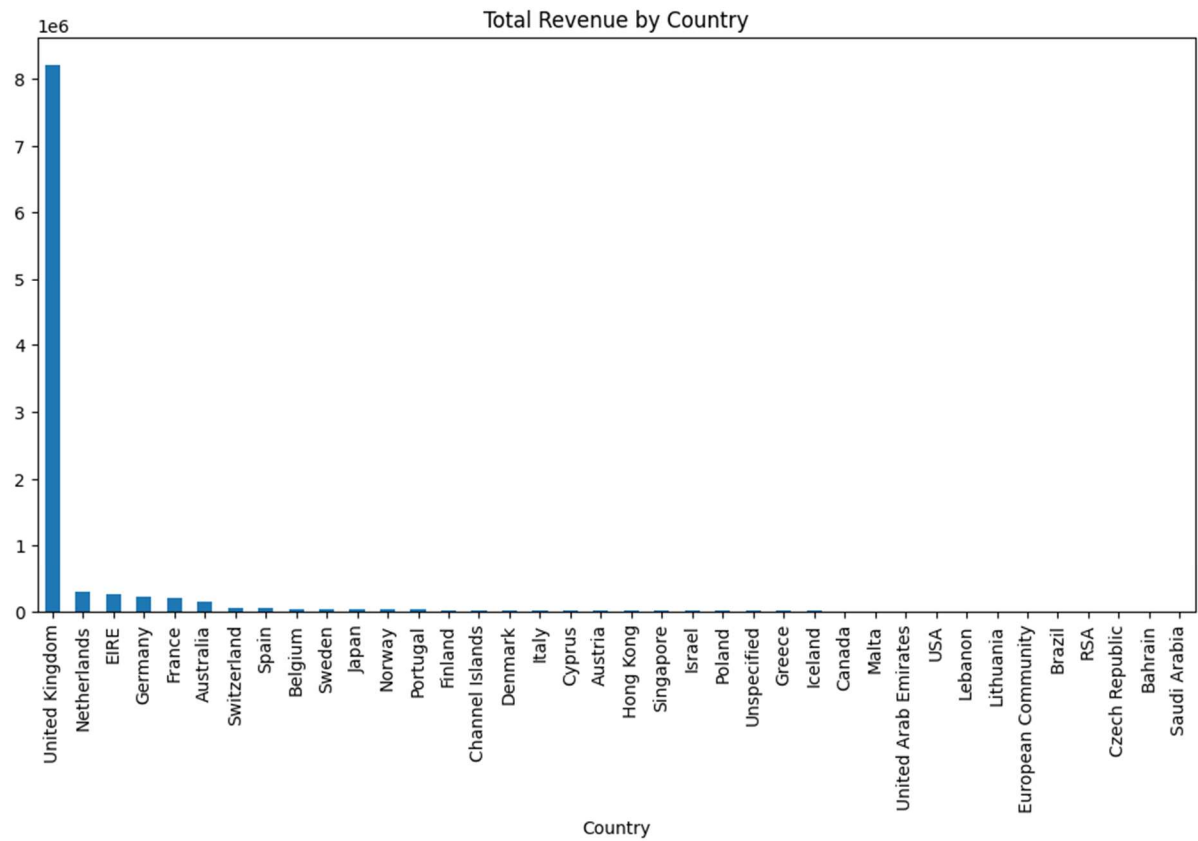


5.3 Correlation Analysis:

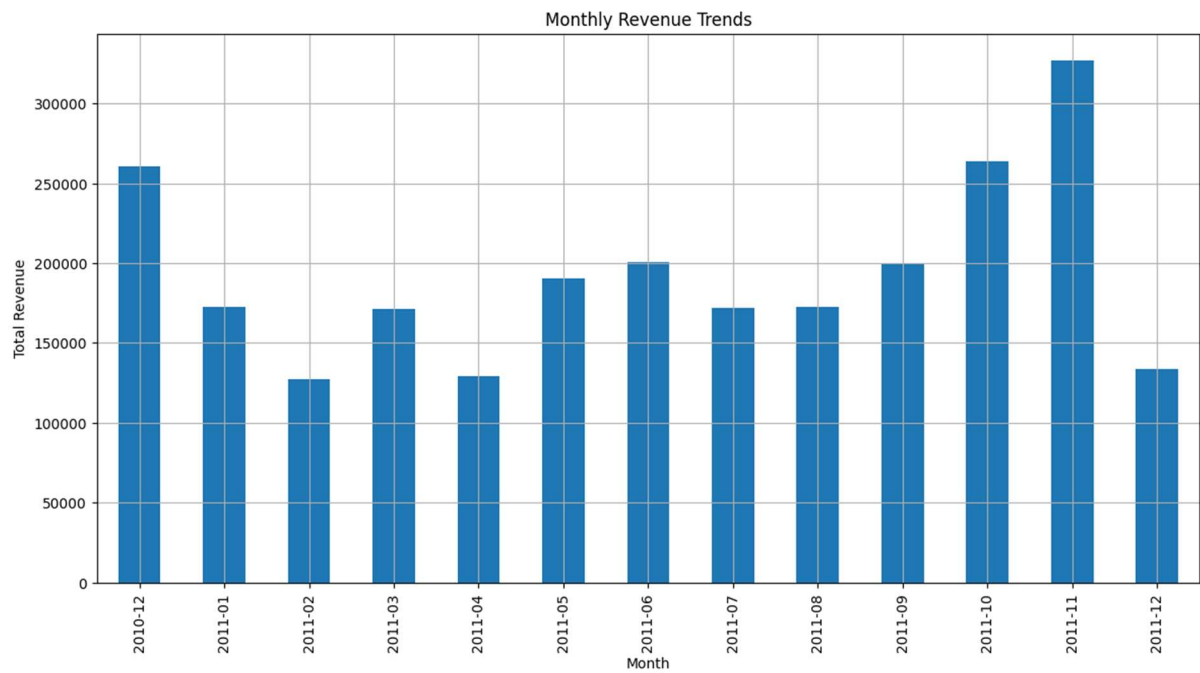
Heatmap showing a correlation between Quantity and Unit Price.



5.4 Regional Sales:



5.5 Monthly Revenue Trends:



5. Insights and Recommendations

5.1 Key Insights

1. **Top-performing product:** DOTCOM POSTAGE drives significant revenue.
2. **Regional dominance:** The United Kingdom generates the majority of revenue.
3. **Seasonal peak:** November shows the highest sales activity.

5.2 Recommendations

1. **Inventory Management:**
 - Maintain high stock for top-performing products.
2. **Regional Focus:**
 - Strengthen marketing in the UK and explore Germany for expansion.
3. **Pricing Strategies:**
 - Regularly review prices for optimization.
4. **Return Management:**
 - Investigate reasons for high returns and address them proactively.

6. Conclusion

This analysis highlights the UK as the primary market and identifies significant seasonal trends. Top-performing products and regions provide actionable insights for inventory and marketing strategies.

Next Steps:

1. Analyze customer demographics for deeper segmentation.
 2. Implement predictive modeling to forecast future trends.
-