Customer Behavior & Revenue Outlier Report

Title: Outlier Analysis Project

Author: Mahmoud Mohamed Mohamed Abdullh

Date: 2025/11/16

Table of Contents

1.Project Objective

The goal of this project is to identify high-value customers (VIPs) and detect outliers in sales transactions of an e-commerce retail store. By analyzing transaction data, returns, and unusual purchases, we aim to generate actionable insights for business decisions and targeted marketing.

2.Data Description

The dataset consists of e-commerce transactions including sales, returns, and high-value outliers. Key columns include:

- **InvoiceNo** → Invoice number
- **StockCode** → Product code
- **Description** → Product name
- **Quantity** → Sold or returned quantity
- **InvoiceDate** → Transaction date and time
- **UnitPrice** → Price per unit
- **CustomerID** → Customer identifier
- **Country** → Customer country
- **Revenue** → Calculated as Quantity × UnitPrice
- **Year / Month / Day / Time** → Derived columns for analysis

After cleaning, the data was divided into three main tables:

1. **Clean_Data** – contains all valid transactions.
2. **Returns** – contains all negative transactions (returns).
3. **Outliers** – contains transactions with unusually high revenue.

3. Data Cleaning Steps

The following steps were implemented:

1. Removed rows with missing `CustomerID`.
2. Deleted irrelevant descriptions such as `"POSTAGE"` and `"Discount"`.
3. Converted `InvoiceDate` to datetime format.
4. Created derived columns: Time, Year, Month, Day, Month Abbreviation.
5. Calculated `Revenue = Quantity × UnitPrice`.
6. Separated returns (`Quantity < 0`) into a dedicated table.
7. Used Numpy to detect outliers in `Revenue`.
8. Ensured no duplicates exist.
9. Formatted columns for Power BI dashboard integration.

4.Outlier Detection Methodology

To identify VIP customers:

Metric:

- `TotalPrice = Quantity × UnitPrice`

Methods:

1. **Z-Score:** Customers with Z-Score > 3 considered outliers.
2. **IQR (Interquartile Range):**
    a. Outlier if `Revenue > Q3 + 1.5 × IQR`
3. **Optional RFM Score:** Recency, Frequency, Monetary analysis to rank VIPs.

5. Analytical Findings

Key insights from the dashboard and analysis:

- **VIP Customers:** Top 10 customers contribute the majority of high-value transactions.
- **Revenue Distribution:** Most transactions are within normal range; few high-value outliers significantly affect total revenue.
- **Returns Analysis:** Returns are concentrated among a small number of products.
- **Customer Trends:** VIP customers tend to purchase frequently and in larger quantities.
- **Country Analysis:** Certain countries show a higher concentration of VIP customers.

Visualizations in the Power BI dashboard include:

1. **Outlier Detection Summary** – cards, boxplots, top 10 tables, VIP revenue KPI.
2. **Customer Profile** – individual customer trends, time series of purchases.
3. **Country Outliers** – maps and comparisons by country.

6. Recommendations

1. **Targeted Marketing:** Focus campaigns on VIP customers to increase retention and upsell.
2. **Inventory Management:** Monitor high-value product stock closely to avoid stockouts.
3. **Returns Management:** Analyze return patterns to minimize losses.
4. **Expansion Opportunities:** Focus on countries with high VIP density for promotions and localized strategies.
5. **Ongoing Monitoring:** Update the outlier analysis regularly to detect changes in customer behavior.

7. Appendix / Dashboard Screenshots

Sales Performance Dashboard

Total Revenue	Total Quantity	Number of Invoices	Number of Customers	Avg Revenue per Invoice
4.37M	3M	361.50K	4196	12.10

Year

2010

2011

Month

Apr Jun

Aug Mar

Dec May

Feb Nov

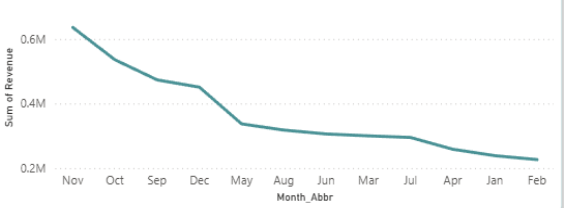
Jan Oct

Jul Sep

Country

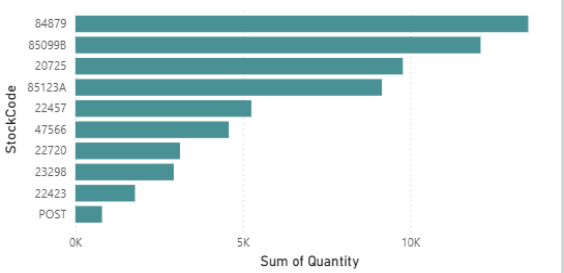
All

Revenue by Month

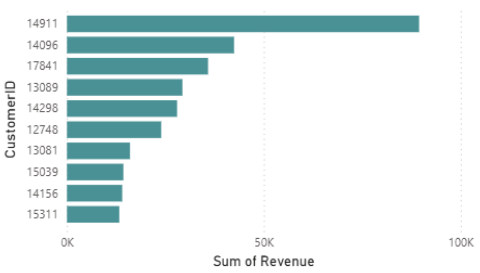


Revenue by Country

Top 10 Products



Top 10 Customers



Product Returns Analysis Dashboard

Total Returns	Avg Return Value	Return Rate	Number of Return Invoices
608.69K	68.61	13.92%	8872

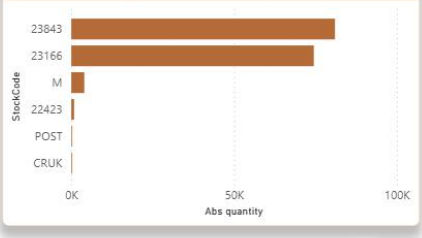
Year: All

Month: All

Country: All

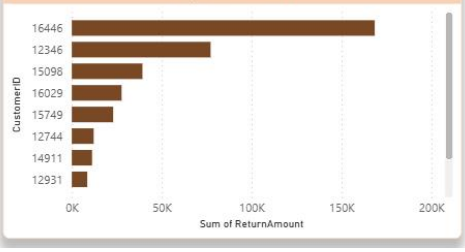
Returns by Month

Top Returned Products

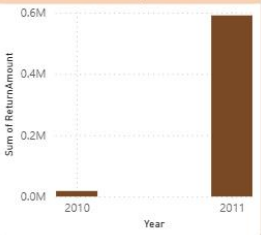


Return by Country

Top Customers



Return by Year



High Value Customers & Outlier Transactions

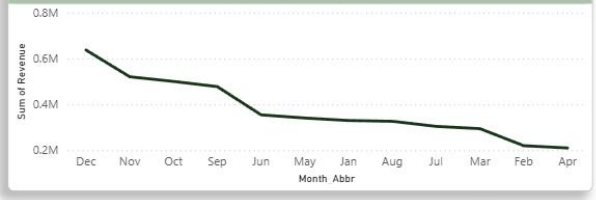
Total Revenue	Number of Transactions	Avg Revenue	Total Quantity
4.51M	31.23K	144.52	2M

Year: All

Month: All

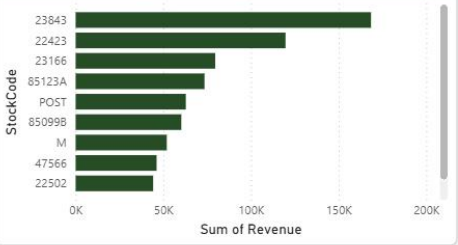
Country: All

Revenue by Month

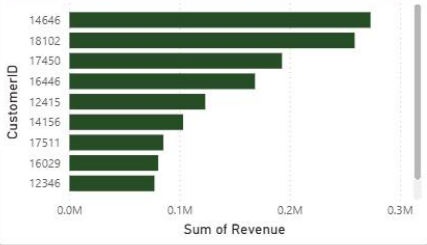


Revenue by Country

Top Prodect by Revenue



Top Customers by Revenue



Normal Rev Vs Outliers Rev