Retail Data Analytics Project Report

1. Project Overview

This Project was completed as part of the **TATA Data Visualisation Virtual Experience on Forage.**

This project focuses on a structured four-task data analytics assignment designed to evaluate both analytical thinking and business understanding. The assignment simulated a real-world retail analytics scenario where data-driven insights are required to support senior leadership decision-making. The objective was to clean raw transactional data, build insightful visualizations, and answer key business questions related to revenue, customers, and geographic demand using Excel and Power BI.

The tasks involved framing relevant business questions that senior stakeholders would ask a data analyst and solving multiple-choice questions based on data analytics concepts and business understanding.

2. Dataset Overview

The dataset contains retail transaction-level information including:

- Invoice Date and Number

- Customer ID

- Country

- Quantity

- Unit Price

- Description

Each row represents a single product-level transaction.

3. Tools Used

- Microsoft Excel (Data Cleaning, Conditional Logic)

- Power BI (Data Modeling, Visualization)

4. Data Cleaning and Preparation

4.1 Data Quality Issues Identified

During initial data exploration, the following issues were identified:

- Negative or zero quantity values, indicating product returns or incorrect entries

- Negative or zero unit prices, representing data entry errors

Including such records would distort revenue and demand analysis.

4.2 Cleaning Method Used (Conditional Formula Method)

To address these issues, a **conditional formula method** was applied in Excel.

A new column named Validity_Check where:

- Quantity ≥ 1

- Unit Price > 0

Helper column named **Validity_Check** was created using conditional logic to flag records as either valid or invalid. All rows marked as invalid were excluded from further analysis. This approach ensured transparency and preserved the original dataset while maintaining data accuracy.

5. Data Modeling in Power BI

After cleaning, the dataset was imported into Power BI for analysis and visualization.

A revenue measure was created using the formula:

Revenue = Quantity × Unit Price

This measure was used consistently across all visuals to ensure standardized revenue calculations.

6. Business Questions and Analysis

6.1 Question: Revenue Trend Analysis (CEO)

Objective: Analyze monthly revenue trends for the year 2011.

Visualization Used: Line Chart

Approach:

- Filtered data for the year 2011

- Aggregated revenue at a monthly level

Insight: The analysis revealed clear seasonal patterns, with higher revenue observed in the later months of the year. This insight can support forecasting, inventory planning, and strategic decision-making.

6.2 Question: Top 10 Countries by Revenue (CMO)

Objective: Identify the top 10 revenue-generating countries excluding the United Kingdom.

Visualization Used: Clustered Column Chart

Approach:

- Excluded the United Kingdom using visual-level filters

- Applied ranking to display the top 10 countries by revenue

- Displayed both revenue and quantity sold

Insight: The chart highlights high-performing international markets and shows the relationship between sales volume and revenue, helping guide regional marketing strategies.

6.3 Question: Top 10 Customers by Revenue (CMO)

Objective: Identify and rank the top 10 customers based on revenue contribution.

Visualization Used: Clustered Column Chart

Approach:

- Created a revenue-based ranking logic

- Displayed customers in descending order of revenue

Insight: A small group of customers contributes a significant share of total revenue. These customers are ideal targets for loyalty programs and personalized marketing efforts.

6.4 Question: Demand by Country (CEO)

Objective: Analyze product demand across countries excluding the United Kingdom.

Visualization Used: Treemap Chart

Approach:

- Used quantity sold as a measure of demand

- Displayed all countries in a single visual

Insight: The treemap clearly identifies high-demand regions, enabling leadership to assess potential opportunities for market expansion and investment.

7. Key Business Insights

- Data cleaning significantly improved the reliability of analysis

- Revenue shows strong seasonality

- A limited number of customers and countries drive a large portion of sales

- Geographic demand analysis highlights expansion opportunities

8. Conclusion

This project demonstrates a complete data analytics workflow, from data cleaning to business insight generation. By aligning visualizations with CEO and CMO decision-making needs, the analysis provides actionable insights that can support strategic planning, marketing optimization, and business growth.