E-Commerce Data Analysis Report

1. Dataset Overview

- Dataset downloaded from Kaggle (Ecommerce Data).
- Contains ~536,641 transaction records.
- Includes columns: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country.
- Multiple countries included in the dataset.
- Revenue column created as: Revenue = Quantity × UnitPrice.

2. Data Cleaning & Preprocessing Steps

- Checked dataset shape and data types using df.shape and df.info().
- Identified missing values using df.isnull().sum().
- Removed duplicate records using df.drop_duplicates().
- Converted InvoiceDate to datetime format.
- Handled negative Quantity values (returns identified).
- Created new column 'Revenue' for business analysis.

3. Feature Engineering

- Calculated Revenue per transaction.
- Computed customer purchase frequency (unique invoices per CustomerID).
- Calculated total customer spending (sum of Revenue per CustomerID).
- Analyzed country-wise distribution using value_counts().

4. Key Business Insights

- High transaction volume dataset suitable for sales trend analysis.
- Customer segmentation possible based on frequency and spending behavior.
- Some transactions contain negative quantities indicating product returns.
- Certain countries contribute significantly higher sales volume.

- Revenue column enables profitability and customer lifetime value analysis.

5. Business Applications

- Customer segmentation and loyalty targeting.
- Sales forecasting and demand prediction.
- Identifying high-value customers.
- Market expansion analysis by country.
- Return rate and product performance analysis.