Summary Report: EDA on Online Retail Dataset

---

1. Introduction

This report presents an Exploratory Data Analysis (EDA) of the Online Retail dataset sourced from the UCI Machine Learning Repository. The objective is to uncover meaningful insights, trends, and data quality issues to support business decisions.

---

2. Dataset Overview

Rows: ~540,000

Columns: 8

Key Features: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country

Time Range: December 2010 – December 2011

---

3. Tools and Libraries

Python 3.x

Pandas, NumPy

Matplotlib, Seaborn

Jupyter Notebook / Google Colab

---

4. Methodology

Loaded the dataset from Excel

Dropped rows with missing CustomerID

Removed duplicates

Created TotalAmount = Quantity * UnitPrice

Created time-based features like YearMonth

Performed summary statistics and visualizations

5. Visualizations

Top 10 Countries by Revenue – UK leads in sales

Most Sold Products – Frequently reordered items dominate

Monthly Sales Trend – Sales spike in November

Correlation Heatmap – Shows strong link between Quantity and TotalAmount

---

6. Key Insights

UK generates the highest revenue

November sees the highest sales

Several products are consistently best-sellers

Missing data was mostly in CustomerID, which was handled accordingly

---

7. Conclusion

This EDA successfully uncovered trends in the Online Retail dataset. These insights can guide marketing, inventory management, and sales forecasting. Future work may include customer segmentation or predictive modeling.