Project: Online Retail Sales EDA

Problem Statement

In this project, you are tasked with performing Exploratory Data Analysis (EDA) on a real-world E-commerce retail dataset. The goal is to analyze customer purchasing behavior, identify the top-selling products, uncover seasonal trends, and segment customers based on their buying patterns. You will also handle common data preprocessing tasks such as missing values, duplicate entries, and outliers.

This project is designed for beginners to build a strong foundation in data analysis using Pandas, Matplotlib, and Seaborn, and to learn how to convert insights into visual dashboards using Power BI.

Dataset

We will use the Online Retail Dataset available on Kaggle, which contains real-world e-commerce transactions for a UK-based online retail store.

Download from Kaggle: https://www.kaggle.com/datasets/lakshmi25npathi/online-retail-dataset

Dataset Details:

- Covers transactions from December 2010 to December 2011
- Includes over 500,000 rows
- Each row represents a single product purchase in an invoice

Columns Included:

- InvoiceNo: Invoice number. Unique identifier for each transaction
- StockCode: Product/item code
- Description: Product name
- Quantity: Number of products purchased
- InvoiceDate: Date and time of purchase
- UnitPrice: Price per product

- CustomerID: Unique ID assigned to each customer

Project: Online Retail Sales EDA

- Country: Country name where the customer resides

These columns provide rich information for analyzing sales, product performance, customer behavior, and geographic trends.

Project Goals

- Clean and preprocess the data
- Perform univariate, bivariate, and multivariate analysis
- Identify trends and customer segments
- Visualize insights using charts and graphs
- Create a Power BI dashboard to present your findings

Tasks Breakdown

1. Data Cleaning:
- Handle missing values (especially CustomerID and Description)
- Remove duplicate rows
- Handle canceled orders (where InvoiceNo starts with 'C')
- Remove or cap outliers in Quantity and UnitPrice
- Convert InvoiceDate to proper datetime format

2. Exploratory Data Analysis (EDA):

General Overview:
- How many unique products are there?
- What is the total number of transactions?
- How many unique customers are there?

- What countries do the customers belong to?

Project: Online Retail Sales EDA

Product Analysis:

- What are the top 10 selling products by quantity?
- What are the top 10 revenue-generating products?
- Which products have negative or zero unit prices?

Customer Analysis:

- Which customers bought the most products?
- What is the distribution of purchases per customer?
- Are there any loyal customers?

Time Series & Seasonal Trends:

- What are the monthly sales trends?
- What are the daily or weekly patterns of transactions?
- Identify peak sales months and slow seasons.

Country Analysis:

- Which countries have the highest number of orders?
- Revenue comparison across countries

RFM (Recency, Frequency, Monetary) Segmentation:

- Segment customers based on their:
  - Recency: How recently a customer made a purchase
  - Frequency: How often they purchased
  - Monetary: How much they spent in total

Power BI Dashboard

After completing the analysis, convert your findings into an interactive Power BI dashboard.

Suggested Dashboard Pages:

Project: Online Retail Sales EDA

1. Sales Overview

 - KPIs: Total Revenue, Number of Invoices, Total Customers

- Monthly Sales Trend (Line Chart)

2. Top Products

 - Top 10 Products by Quantity and Revenue (Bar Chart)

- Product Category Breakdown (Pie Chart or Treemap)

3. Customer Segments

 - RFM Segmentation (Table + Scatter Plot)

 - Top 10 Loyal Customers

4. Geographical Insights

 - Sales by Country (Map or Bar Chart)

Deliverables

- A Jupyter Notebook with:

 - Cleaned data

 - Well-commented code

 - Visualizations and insights

- A Power BI Dashboard (.pbix file)