Exploratory Data Analysis on Sales Data

By Manav Sood

1. Introduction

This project performs an **Exploratory Data Analysis (EDA)** on a sales dataset to uncover trends, seasonal patterns, and key performance metrics. The goal is to identify factors affecting sales performance, including top-performing products and seasonal variations.

2. Dataset Overview

The dataset contains transactional sales data, including order details, product sales, and timestamps. Key columns include:

- **Order Number**: Unique identifier for each order
- **Quantity Ordered**: Number of units purchased
- **Price Each**: Price per unit
- **Sales**: Total revenue generated
- **Order Date**: Timestamp of the purchase
- **Quarter, Month, Year**: Derived time-based features
- **MSRP**: Manufacturer's Suggested Retail Price

3. Installation & Dependencies

To run this project, install the necessary Python libraries:

```
pip install pandas numpy seaborn matplotlib scipy
```

4. Code Explanation

The project is divided into the following key sections:

A. Data Loading & Cleaning

- The dataset is loaded using **pandas**.
- Missing values and inconsistencies are handled by dropping or imputing values where necessary.

B. Statistical Analysis

- Summary statistics (mean, median, mode, standard deviation, skewness, kurtosis) are computed for numerical columns.
- A chatbot-style interface allows the user to select which column they want statistics for.

C. Data Visualization

- **Sales Trends**: Line chart & Area chart to analyze revenue trends over time.
- **Seasonal Patterns**: Bar plots to examine how sales fluctuate across months/quarters.
- **Top Performing Products**: Pie chart & Treemap to showcase best-selling products.
- **Correlation Analysis**: Heatmap to find relationships between numerical variables.
- A chatbot interface lets users choose what type of visualization they want.

5. Usage Instructions

1. **Run the script in Jupyter Notebook.**
2. **Follow the chatbot prompts** to select an analysis type.
3. **Choose a specific category & visualization.**
4. **Interpret the insights based on the output graphs and statistics.**

6. Conclusion

This project successfully explores sales data, identifying trends, seasonal patterns, and key performance metrics. The chatbot-based interface enhances user experience, making data analysis interactive and customizable. The visualizations provide valuable insights into business performance and can aid in strategic decision-making.

This documentation serves as a guide to understanding the project workflow and how to use it effectively.