E-commerce Sales Data Analysis

1. Introduction

The objective of this project is to perform a complete data analysis workflow on an E-commerce sales dataset. The analysis includes data loading, cleaning, exploration, visualization, and extracting meaningful business insights using Python libraries such as Pandas and Matplotlib.

This project demonstrates practical data analytics skills including handling real-world data, creating professional charts, and writing analytical insights.

2. Dataset Description

The dataset contains sales transaction records of an e-commerce platform. The main columns used in this analysis are:

- Customer_ID: Unique order number
- Product: Category of the product
- Quantity: Number of items sold
- Price: Price per unit
- Date: Date of purchase

The dataset is stored in CSV format and loaded using the Pandas library.

3. Tools & Technologies Used

- Python
- Pandas (Data manipulation)
- Matplotlib (Data visualization)
- VS Code

4. Data Loading and Cleaning

Steps Performed:

- Loaded the CSV file using Pandas
- Checked for missing values
- Filled missing numerical values using mean

- Removed duplicate records
- Converted date columns into proper datetime format

Data Validation:

- Ensured correct data types
- Verified no null values remained after cleaning

---

5. Exploratory Data Analysis (EDA)

During EDA, the following analysis was performed:

- Total sales calculation
- Sales by product category
- Monthly sales trends
- Identification of best-selling category

A new column **Total_Sales** was created using: Total_Sales = Quantity × Price

---

6. Data Visualization

At least two different chart types were created as required.

6.1 Bar Chart – Sales by Product Category

A bar chart was used to compare total sales across different product categories.

Purpose:

- Identify top-performing product categories

Observation:

- Electronics and Fashion categories generated the highest revenue.

Screen Shot of Bar Chart Code using Pandas/Matplotlib and Result Image1

```
C: > Users > shubh > Desktop > Arena projects file document > 4th week project > code >  bar.py > ...
  1   import pandas as pd
  2   import matplotlib.pyplot as plt
  3
  4   # 1.Load data
  5   df = pd.read_csv(r"C:\Users\shubh\Desktop\Arena projects file document\4th week project\code\Sales_Data.csv")
  6
  7
  8   # 2. Data Cleaning
  9   df.drop_duplicates(inplace=True)
 10   df['Quantity'].fillna(df['Quantity'].mean(), inplace=True)
 11   df['Price'].fillna(df['Price'].mean(), inplace=True)
 12
 13   # 3. Feature Engineering
 14   df['Total_Sales'] = df['Quantity'] * df['Price']
 15
 16   # 4. Bar Chart: Total Sales by Product Category
 17   category_sales = df.groupby('Product')['Total_Sales'].sum()
 18
 19   plt.figure()
 20   category_sales.plot(kind='bar')
 21   plt.title("Total Sales by Product Category")
 22   plt.xlabel("Product Category")
 23   plt.ylabel("Total Sales")
 24   plt.xticks(rotation=45)
 25   plt.tight_layout()
 26   plt.show()
```

```
PROBLEMS    OUTPUT    DEBUG CONSOLE    TERMINAL    PORTS

        ^^^^^^^^^^^^^^^^^
    )
    ^
  File "C:\Users\shubh\AppData\Roaming\Python\Python314\site-packages\pandas\core\groupby\grouper.p
    raise KeyError(gpr)
KeyError: 'Product_Category'
PS C:\Users\shubh> & "C:/Program Files/Python314/python.exe" "c:/Users/shubh/Desktop/Arena projects
PS C:\Users\shubh> & "C:/Program Files/Python314/python.exe" "c:/Users/shubh/Desktop/Arena projects
```

6.2 Line Chart – Monthly Sales Trend

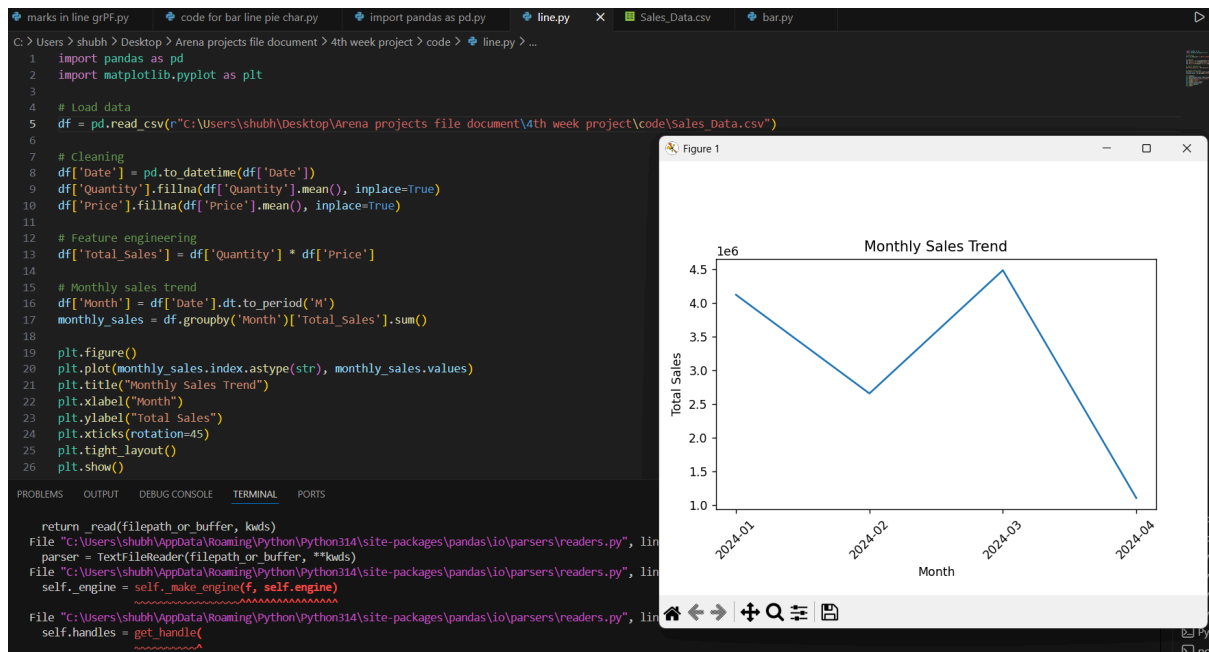A line chart was created to visualize sales trends over time.

Purpose:

- Analyze seasonality and monthly performance

Observation:

- Sales increased significantly during festive months.

Screen Shot of Line Chart Code using Pandas/Matplotlib and Result Image2

6.3 Pie Chart – Sales Distribution

A pie chart shows the percentage contribution of each category to total sales.

Screen Shot of Pie Chart Code using Pandas/Matplotlib and Result Image3