Python Diwali Sales Analysis

1. Introduction

Diwali is one of the biggest festive seasons in India, marked by high consumer spending across multiple product categories.

This project analyzes Diwali sales data using Python to understand **customer demographics, purchasing behavior, top-selling products, and regional performance** through data cleaning and exploratory data analysis (EDA).

2. Objective

The objectives of this project are:

- Load and inspect the Diwali sales dataset
- Clean and preprocess raw data
- Perform exploratory data analysis (EDA)
- Identify key customer segments and sales trends
- Visualize insights using Python libraries

3. Dataset Description

- **File Type:** CSV
- **Total Records (after cleaning):** 11,239
- **Total Columns:** 13

Key Attributes:

- User_ID, Customer Name
- Gender, Age, Age Group, Marital Status

- State, Zone, Occupation
- Product Category, Product ID
- Orders, Amount

4. Tools & Technologies

- **Python**
- **Pandas & NumPy** – data cleaning and transformation
- **Matplotlib & Seaborn** – data visualization
- **Jupyter Notebook** – analysis environment

5. Data Cleaning Summary

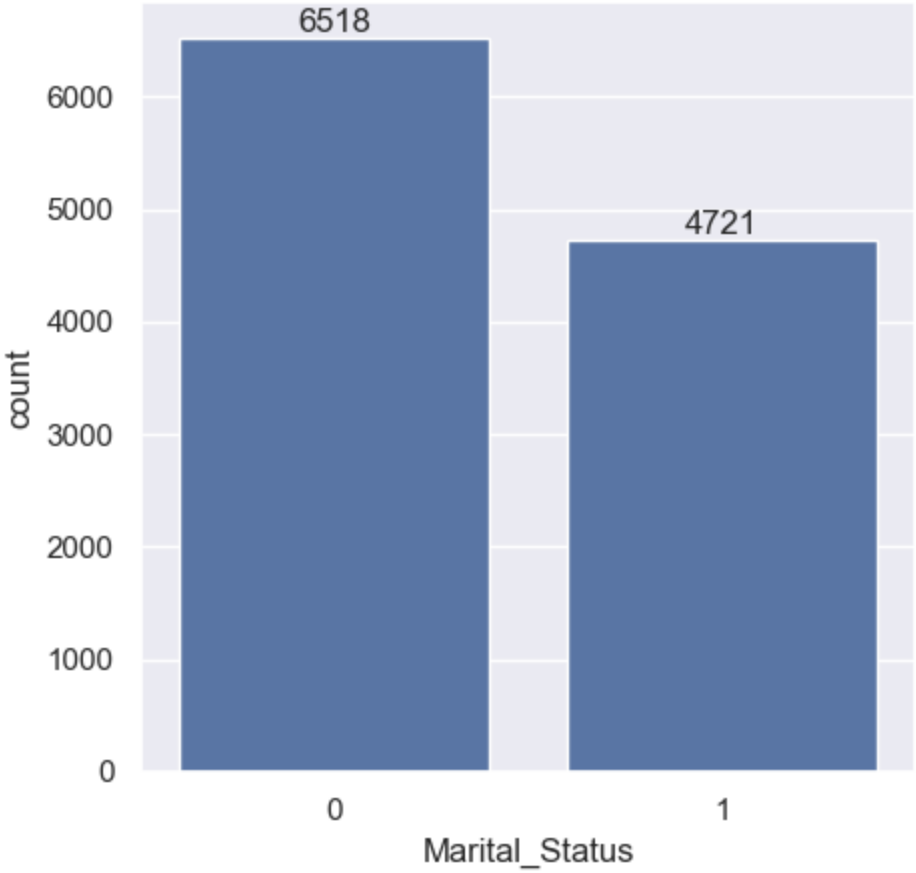
- Removed empty columns (`Status`, `unnamed1`)
- Handled missing values in the `Amount` column
- Converted `Amount` data type from float to integer
- Ensured no null values remained after cleaning

✔ Final dataset was clean, consistent, and ready for analysis.

6. Exploratory Data Analysis (EDA)

6.1 Marital Status Distribution

Marital status was analyzed to understand purchasing behavior.

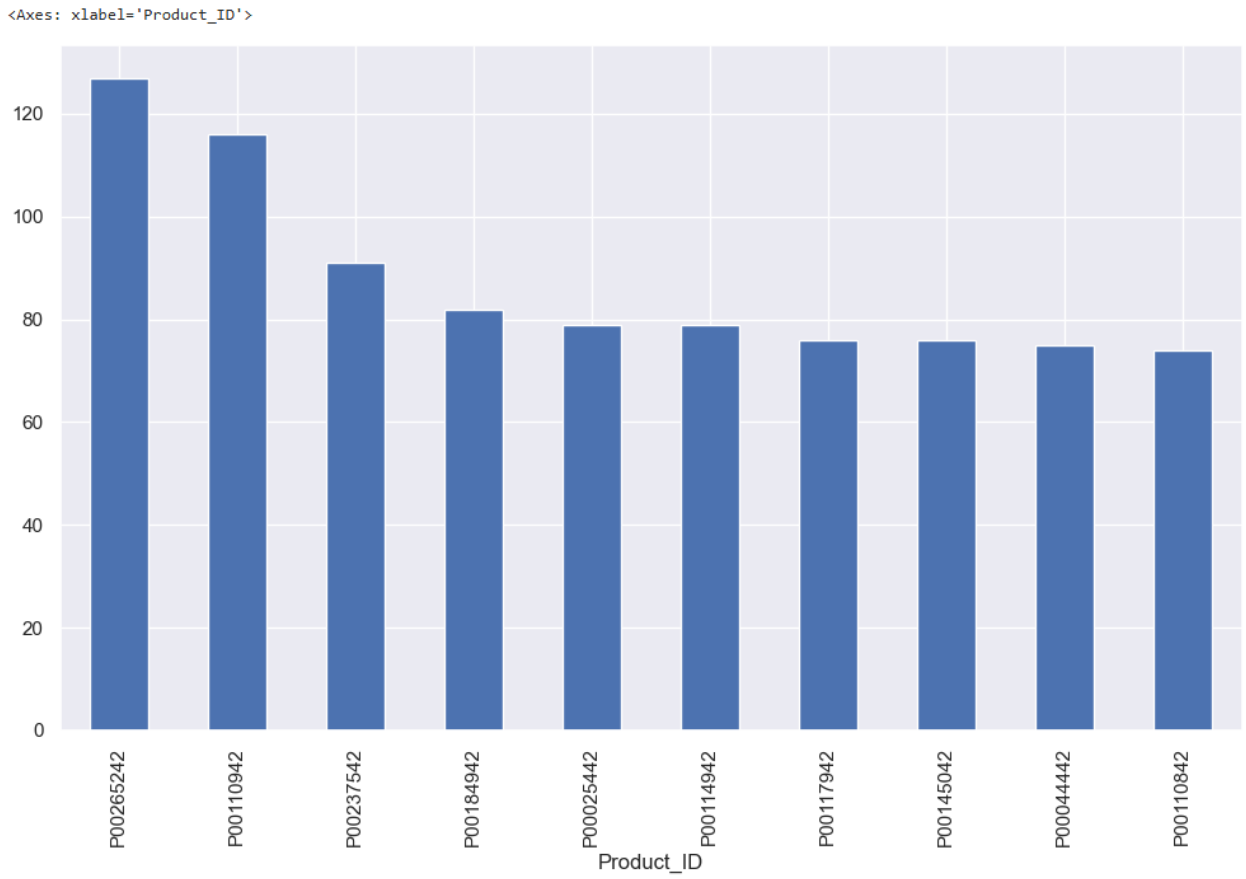


Insight:

- Married customers made more purchases compared to unmarried customers.

6.2 Top Products by Orders

The most frequently ordered products were identified.



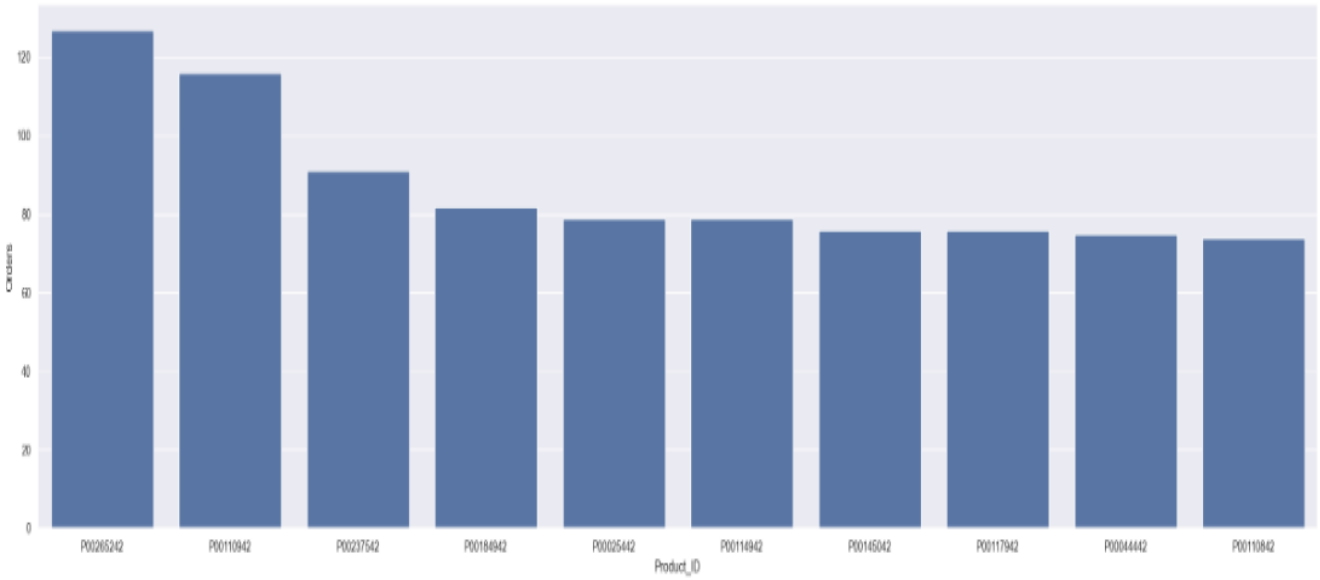
Insight:

- A small group of products generated the highest number of orders, indicating strong product demand.

6.3 Product Orders – Bar Chart

Another visualization highlighting product-wise order distribution.

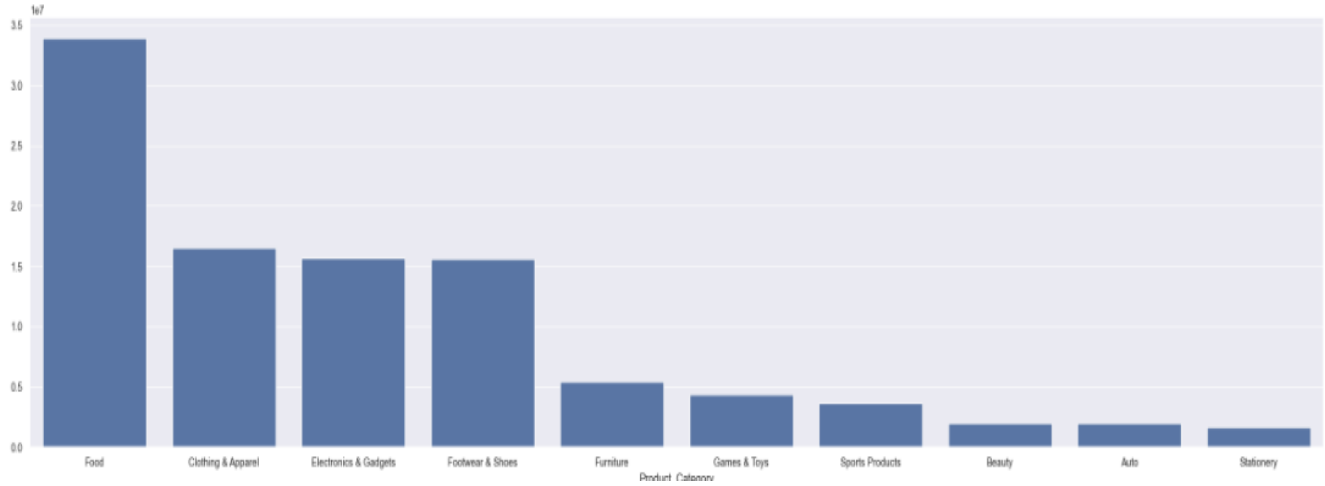
<Axes: xlabel='Product_ID', ylabel='Orders'>



6.4 Sales by Product Category

Total sales amount was analyzed across product categories.

<Axes: xlabel='Product_Category', ylabel='Amount'>

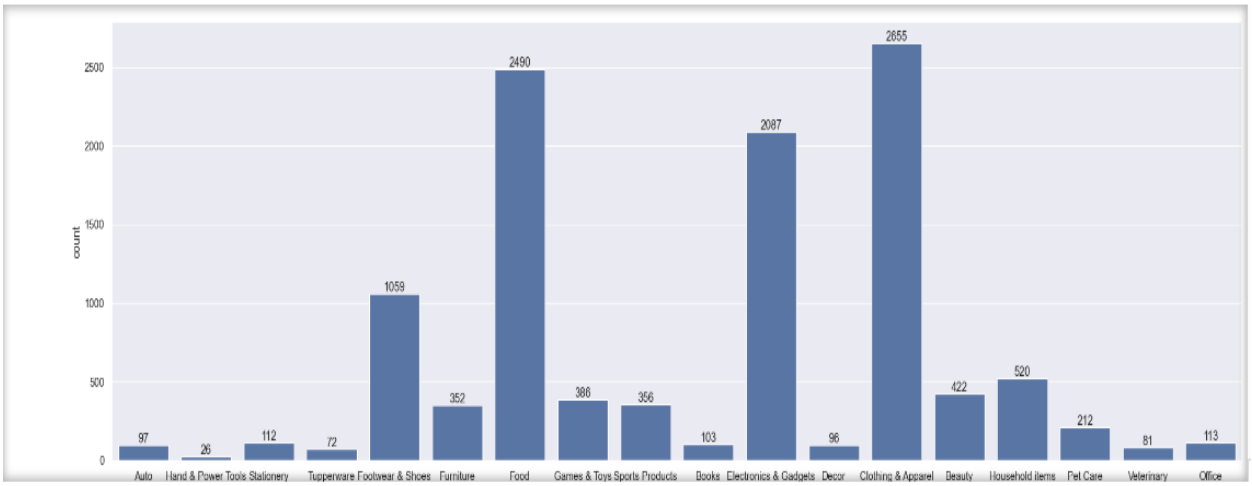


Insight:

- **Food** category generated the highest revenue
- Clothing & Apparel and Electronics also performed strongly

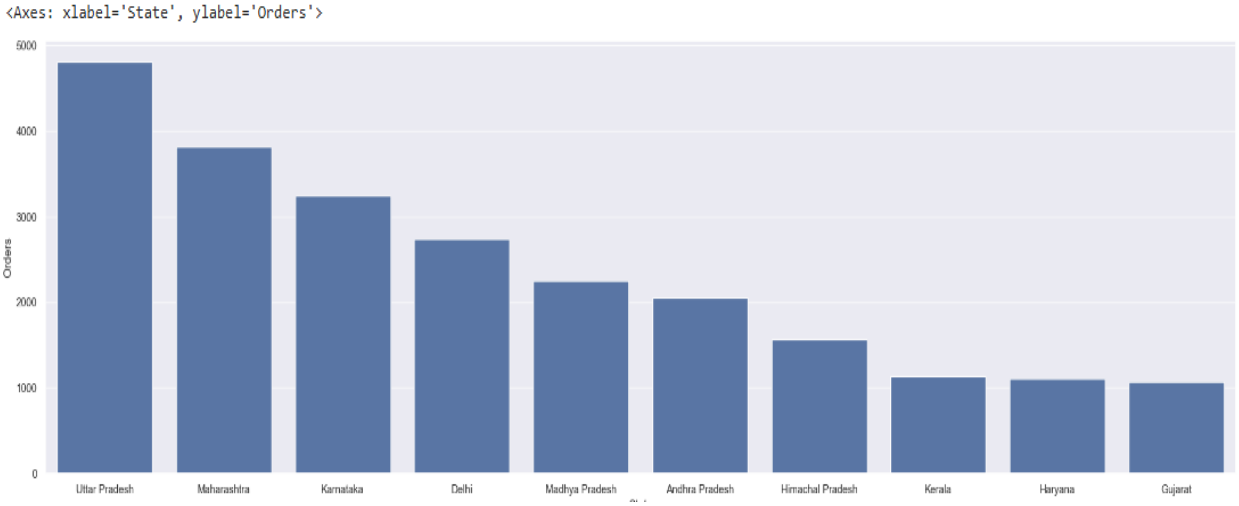
6.5 Product Category Order Count

Distribution of orders across all product categories.



6.6 State-wise Orders

Top states based on total number of orders.

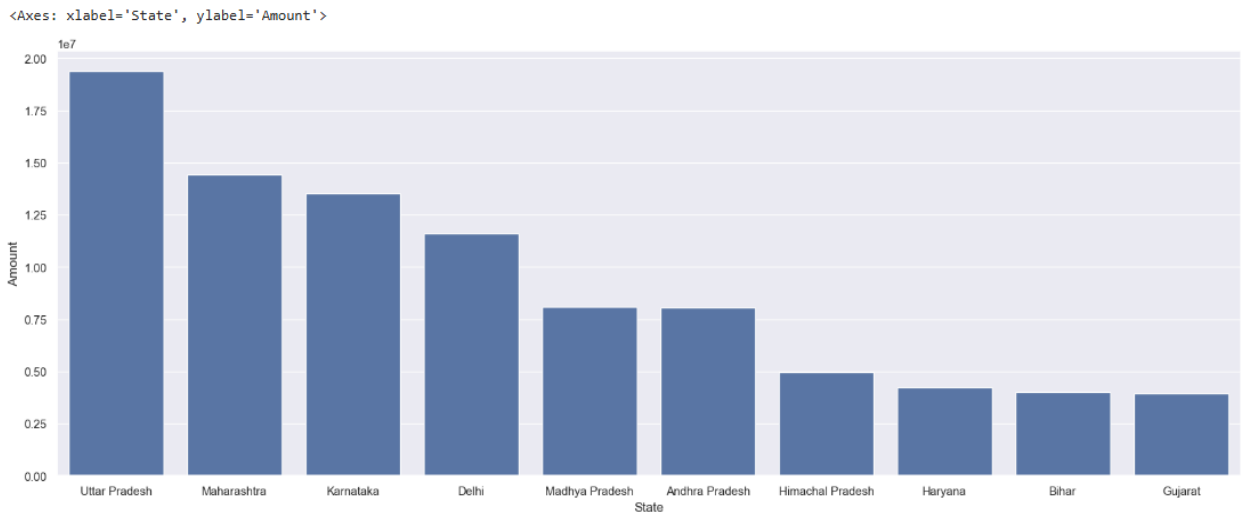


Insight:

- Uttar Pradesh, Maharashtra, and Karnataka recorded the highest order volumes

6.7 State-wise Sales Amount

Revenue contribution by state.

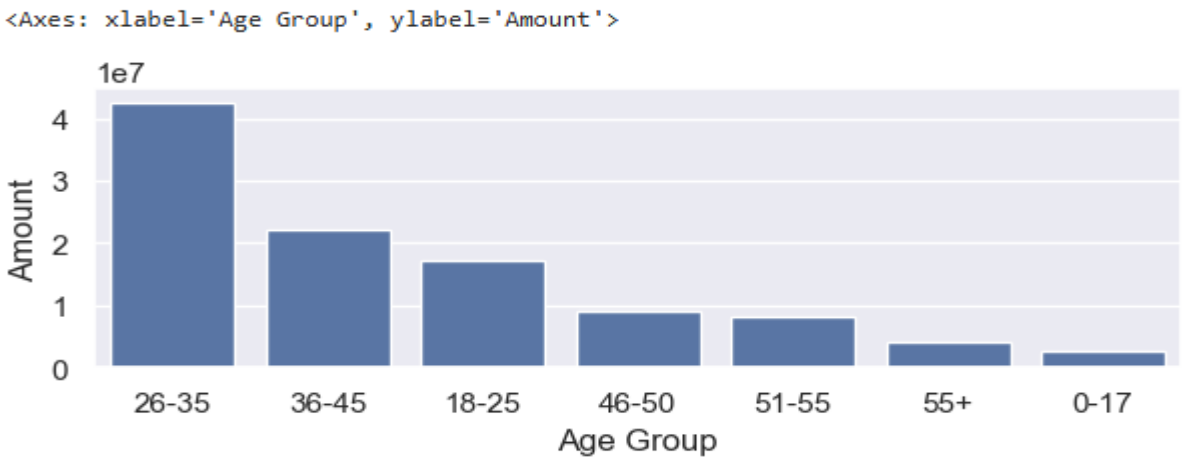


Insight:

- Uttar Pradesh leads in revenue, followed by Maharashtra and Karnataka

6.8 Sales by Age Group

Sales amount segmented by customer age group.

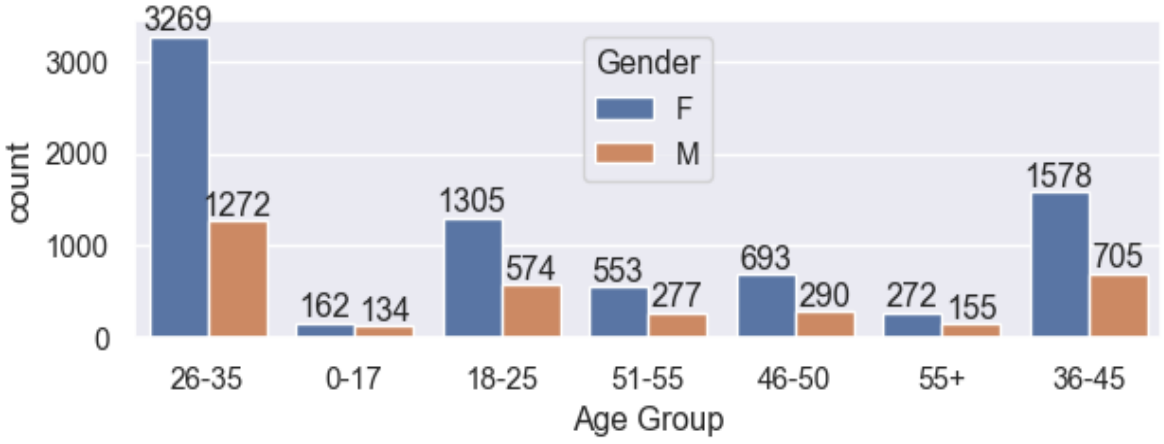


Insight:

- Customers aged **26–35** contributed the highest sales
- Followed by age groups **36–45** and **18–25**

6.9 Gender & Age Group Analysis

Comparison of male and female customers across age groups.

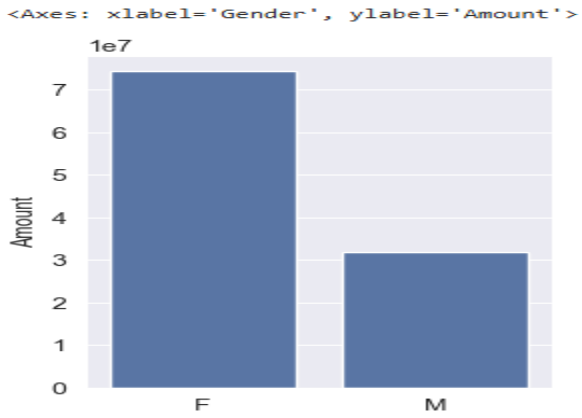


Insight:

- Female customers dominate across all age groups
- Especially strong participation from the 26–35 age group

6.10 Gender-wise Sales Contribution

Total sales amount by gender.



Insight:

- Female customers contributed significantly more to total sales than male customers

7. Key Insights & Results

- Female customers are the primary contributors to Diwali sales
- Age group **26–35 years** is the most valuable customer segment
- Food and Apparel categories generate the highest revenue
- Uttar Pradesh and Maharashtra are the top-performing states
- Married customers place more orders during the festive season

8. Conclusion

This Python-based Diwali Sales Analysis successfully transformed raw sales data into meaningful insights using data cleaning and EDA techniques.

The findings can help businesses:

- Target high-value customer segments
- Optimize product inventory
- Improve festive marketing strategies

Author

Parth Kulkarni

Aspiring Data Analyst

Python | SQL | Data Analysis | Data Visualization