

Diwali Sales Data - Exploratory Data Analysis Report

1. Project Overview

This report presents the Exploratory Data Analysis (EDA) conducted on the Diwali Sales dataset using Python in Jupyter Notebook. The objective of this analysis is to understand customer purchasing behavior, identify key revenue drivers, and uncover patterns across demographics and product categories. The analysis helps in identifying high-performing customer segments and products to support better business decision-making.

2. Dataset Summary

Total Rows: 11251

Total Columns: 15

Key Columns:

- User_ID
- Cust_name
- Product_ID
- Gender
- Age Group
- Age
- Marital_Status
- State
- Zone
- Occupation
- Product_Category
- Orders
- Amount

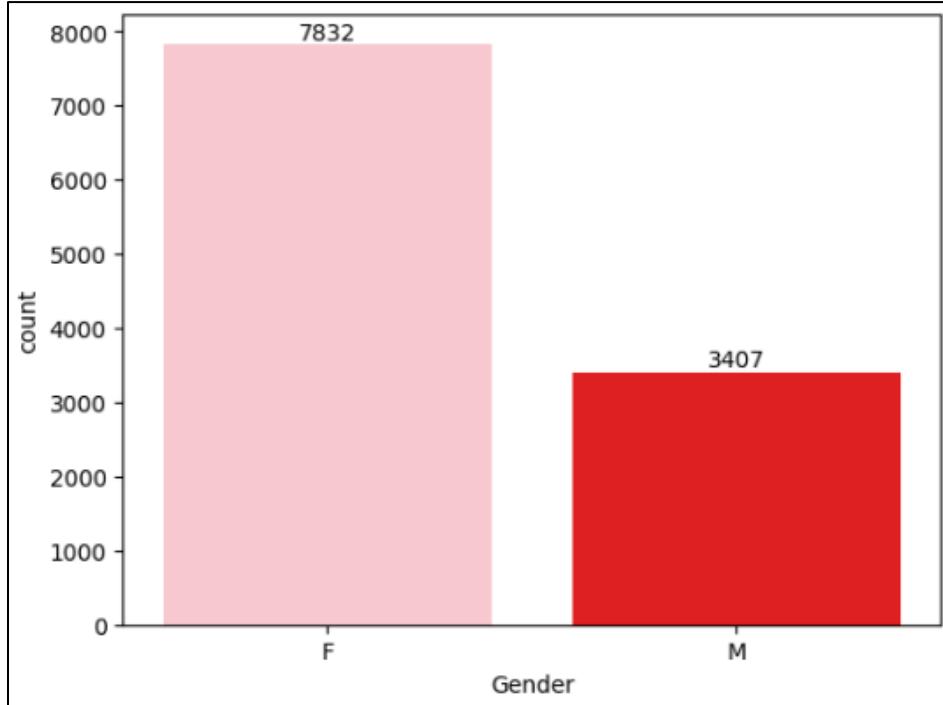
3. Data Cleaning and Preparation

- **Data Loading:** Imported the Diwali Sales dataset using pandas in Jupyter Notebook.
- **Initial Exploration:** Used df.columns to review column names and df.head() to preview the dataset structure.
- **Dropping Unnecessary Columns:** Removed unrelated columns such as 'Status' and 'unnamed1' using df.drop() to clean the dataset.
- **Handling Missing Values:** Checked for null values using pd.isnull(df).sum(). Identified missing values in the Amount column.
- **Data Type Conversion:** Converted the Amount column to integer type using .astype('int') to ensure accurate numerical analysis.
- **Value Replacement:** Replaced numeric values in the Marital_Status column (0 and 1) with meaningful labels ('Married' and 'Single') for better readability.

4. Exploratory Data Analysis

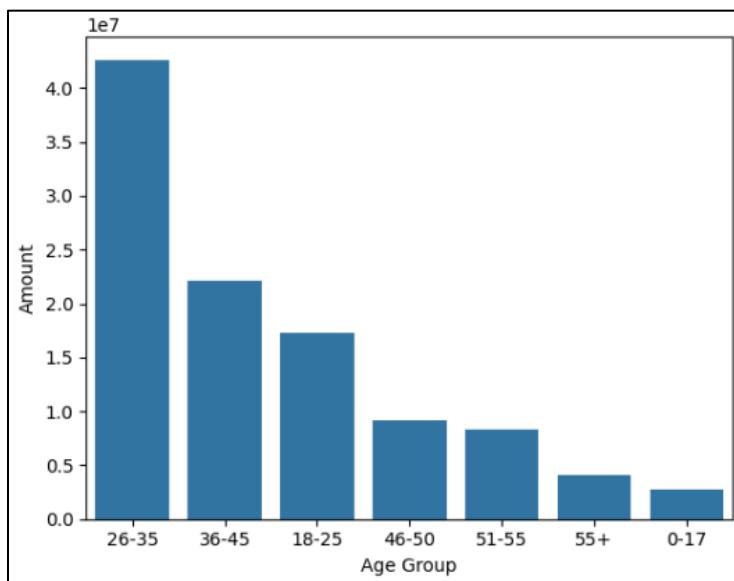
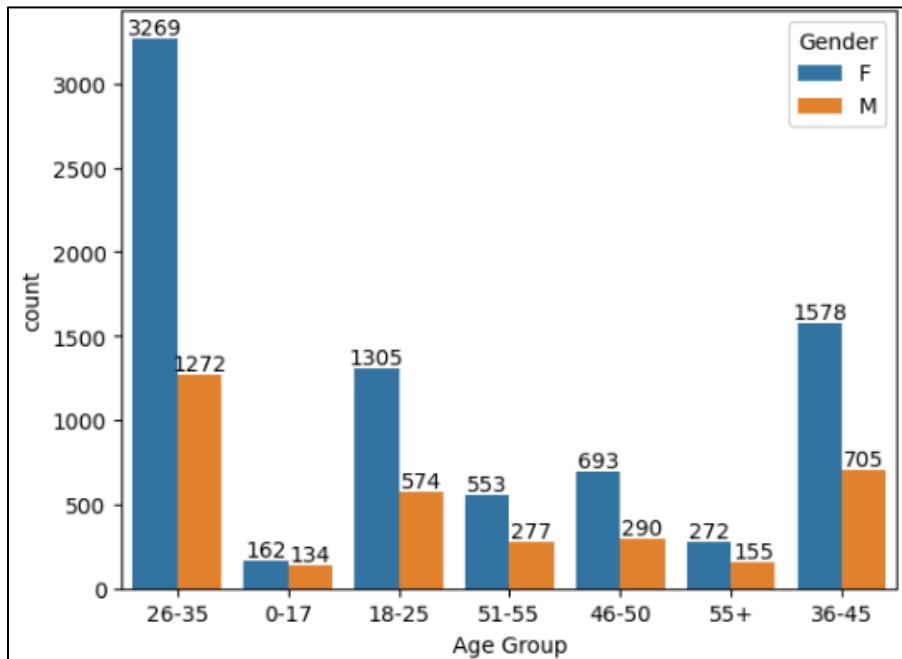
The following analysis was performed using pandas, numpy, matplotlib, and seaborn:

- Gender-wise purchasing behavior analysis.



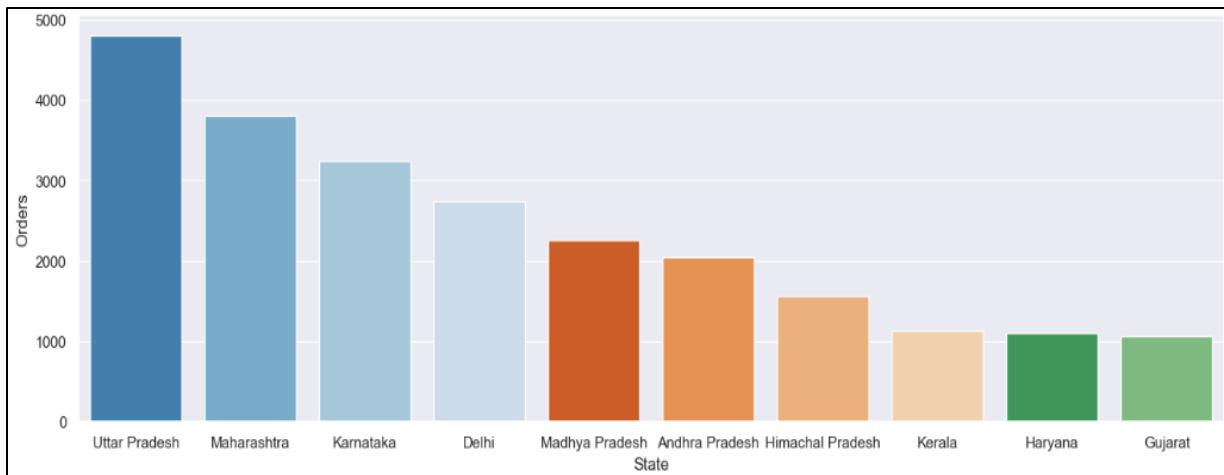
- From above graphs we can see that most of the buyers are females and even the purchasing power of females are greater than men

- Age group distribution of customers.



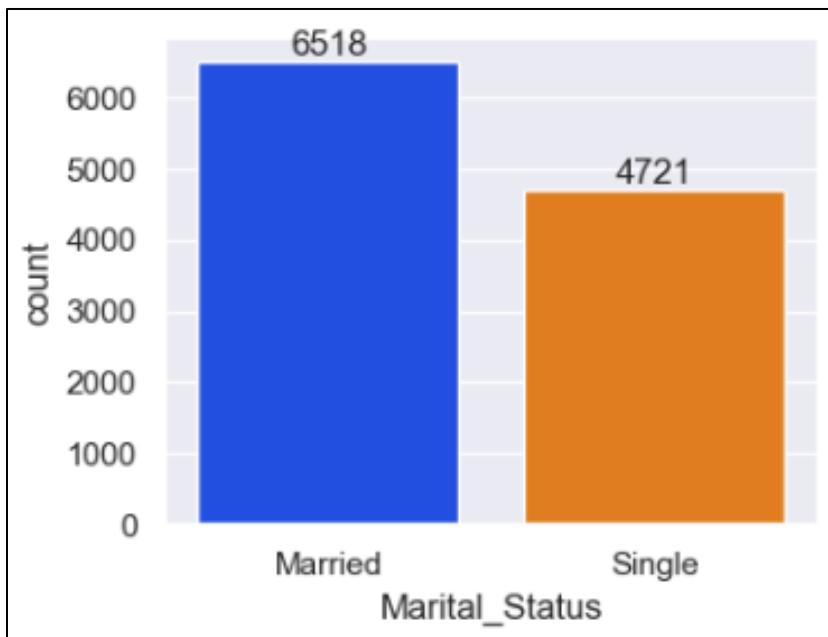
- From above graphs we can see that most of the buyers are of age group between 26-35 yrs and they are mostly female

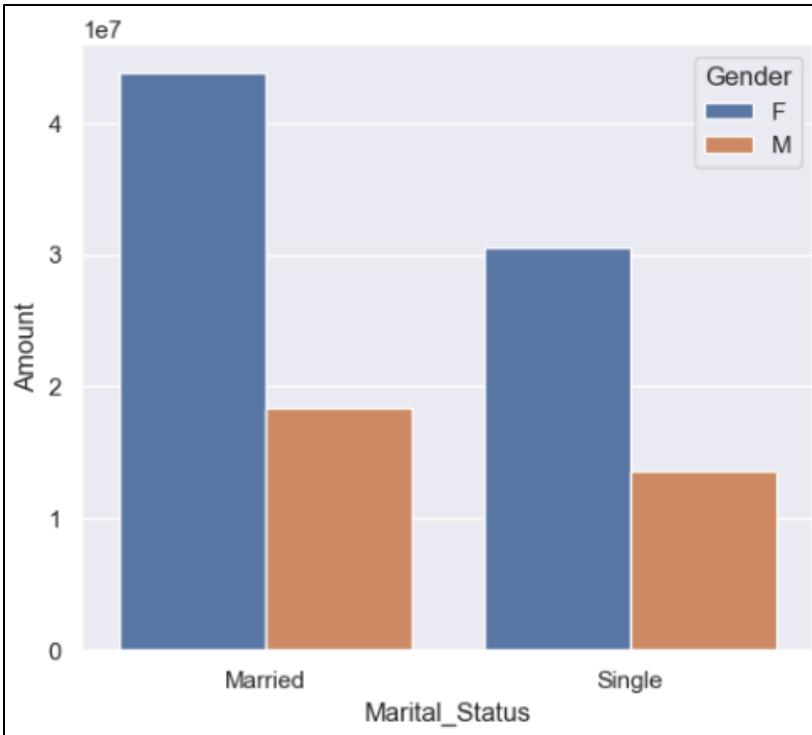
- Sales distribution across different states.



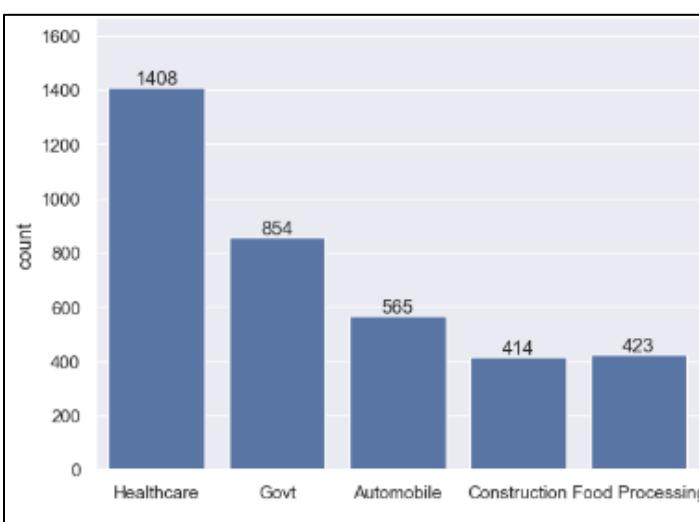
- From above graphs we can see that most of the orders are from Uttar Pradesh, Maharashtra and Karnataka respectively

- Marital status analysis and its impact on purchasing.



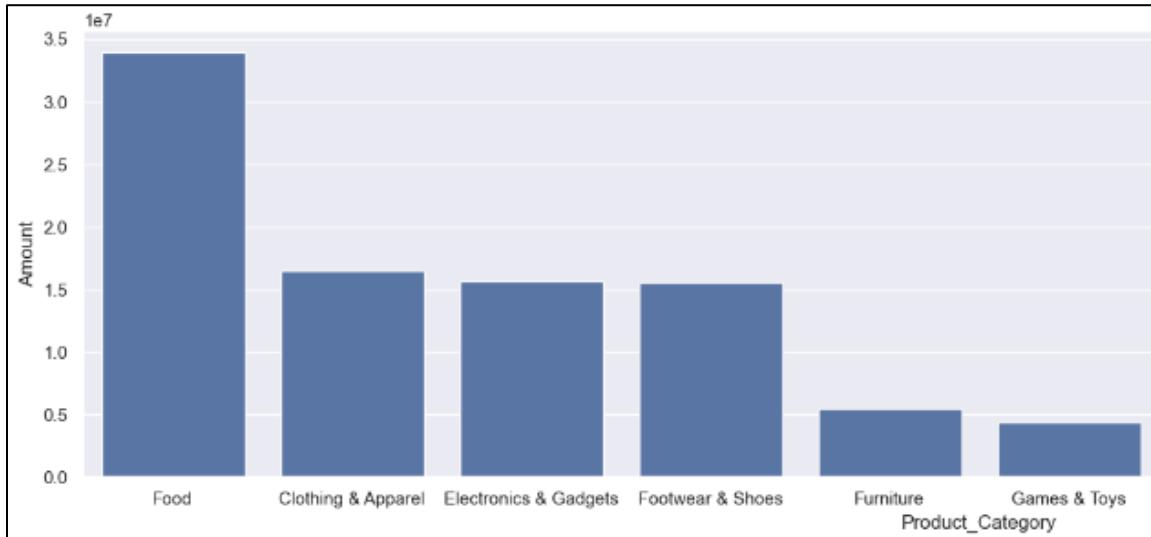


- From above graphs we can see that most of the buyers are married (women) and they have high purchasing power
- Occupation-based sales analysis.



- From above graphs we can see that most of the buyers are working in IT, Healthcare and Aviation sector

- Product category performance analysis.



- **From above graphs we can see that most of the sold products are from Food, Clothing and Electronics category**

5. Key Insights

1. Female customers contributed more towards overall sales compared to male customers.
2. The 26–35 age group showed the highest purchasing activity.
3. Certain states contributed significantly higher sales compared to others.
4. Married customers showed higher purchasing power in comparison to unmarried customers.
5. Specific occupations such as IT, Healthcare, and Aviation showed strong purchasing patterns.
6. Product categories like Food, Clothing, and Electronics performed strongly in terms of revenue.

6. Conclusion

The exploratory analysis provided a clear understanding of customer demographics and purchasing trends during the Diwali season. The insights suggest that marketing efforts should focus on high-performing age groups, occupations, and states to maximize revenue. Product strategies can be optimized based on top-performing categories to improve overall business performance.