Diwali Sales Data Analysis

Introduction

This project aims to analyze Diwali sales data to gain insights into customer behavior and preferences. The dataset used for analysis is provided in a CSV file called "Diwali Sales Data.csv". The analysis focuses on various aspects such as gender, age group, state, marital status, occupation, product category, and product ID. The project utilizes Python libraries such as NumPy, pandas, matplotlib, and seaborn for data manipulation, visualization, and analysis.

Data Import and Preprocessing

The first step involves importing the necessary Python libraries and reading the CSV file using the pandas library's `read_csv()` function. The encoding parameter is set to "unicode_escape" to avoid any potential encoding errors. The data frame's shape is examined using the `shape` attribute, and the first few rows of the data frame are displayed using the `head()` function. Similarly, the last few rows can be viewed using the `tail()` function. The `info()` function provides information about the data types and missing values in the data frame.

```
# Checking the data to analyze the need for cleaning areas
df.info()
```

```
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 15 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   User_ID           11251 non-null  int64  
 1   Cust_name         11251 non-null  object 
 2   Product_ID        11251 non-null  object 
 3   Gender            11251 non-null  object 
 4   Age Group         11251 non-null  object 
 5   Age               11251 non-null  int64  
 6   Marital_Status    11251 non-null  int64  
 7   State             11251 non-null  object 
 8   Zone              11251 non-null  object 
 9   Occupation        11251 non-null  object 
 10  Product_Category  11251 non-null  object 
 11  Orders            11251 non-null  int64  
 12  Amount            11239 non-null  float64
 13  Status            0 non-null      float64
 14  unnamed1          0 non-null      float64
dtypes: float64(3), int64(4), object(8)
```

Unrelated or blank columns are dropped using the `drop()` function, specifying the column names and axis. Null values are checked using the `isnull()` function, and rows with missing values are dropped using the `dropna()` function. The data type of the "Amount" column is converted to integer using the `astype()` function.

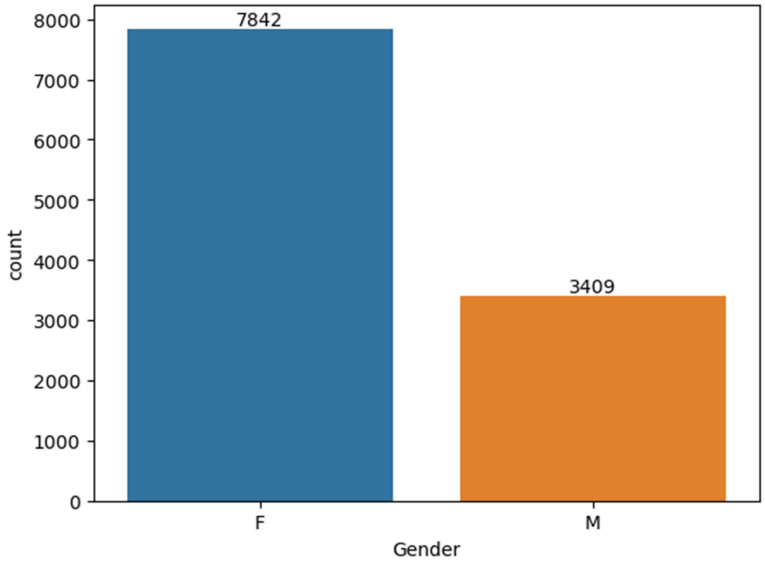
Exploratory Data Analysis

The exploratory data analysis phase involves visualizing and analyzing different aspects of the data to uncover patterns, trends, and insights.

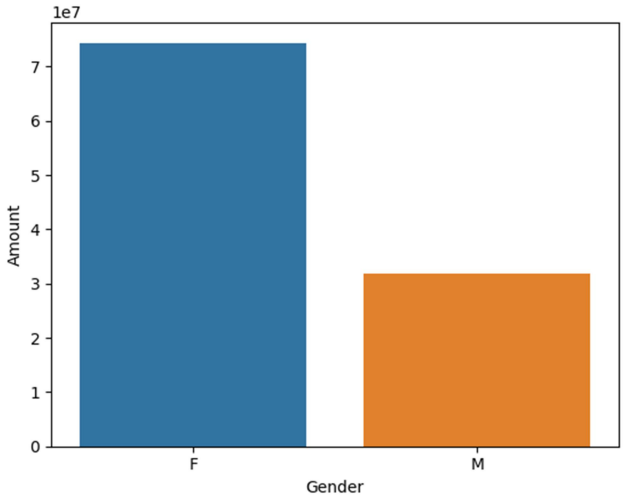
Gender Analysis

The analysis begins with examining the gender distribution of the buyers. A bar chart is plotted using the `countplot()` function from the seaborn library to visualize the count of buyers for each gender. Additionally, a bar chart is created to compare the total amount spent by each gender.

```
# plotting a bar chart for Gender and it's count
```




```
# plotting a bar chart for gender vs total amount
```

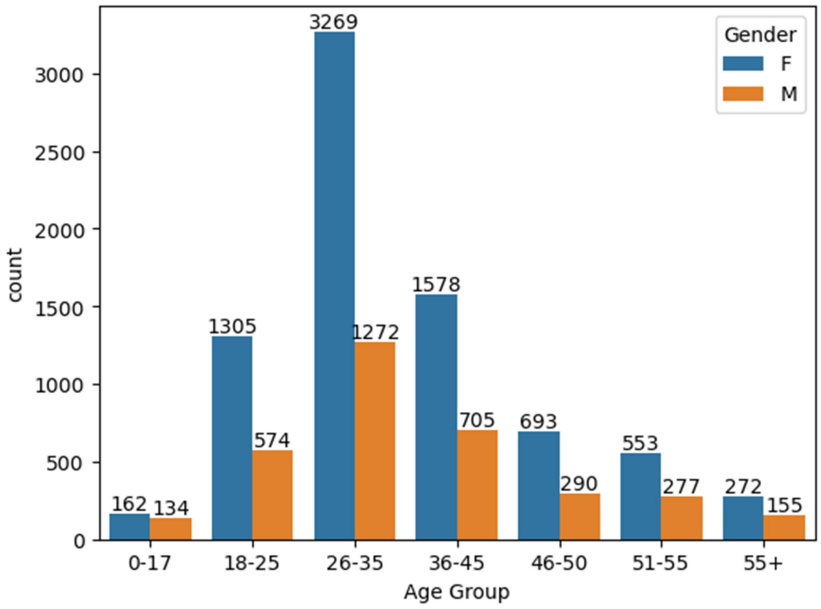


From above graphs we can see that most of the buyers are females and even the purchasing power of females are greater than men.

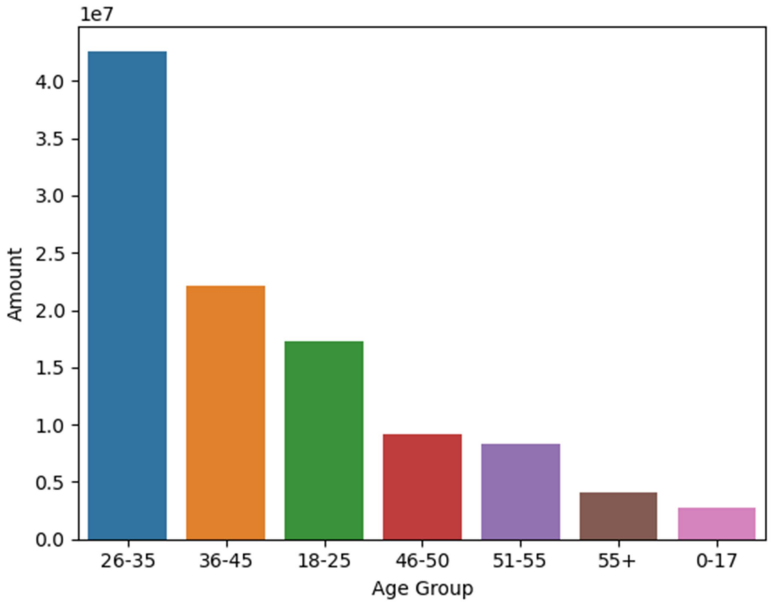
Age Group Analysis

The next analysis focuses on the age groups of the buyers. A bar chart is created to visualize the count of buyers for each age group, categorized by gender. Another bar chart is generated to compare the total sales amount for each age group.

```
#which gender in age group is more active in shopping
```




```
#Total Amount vs Age Group
```

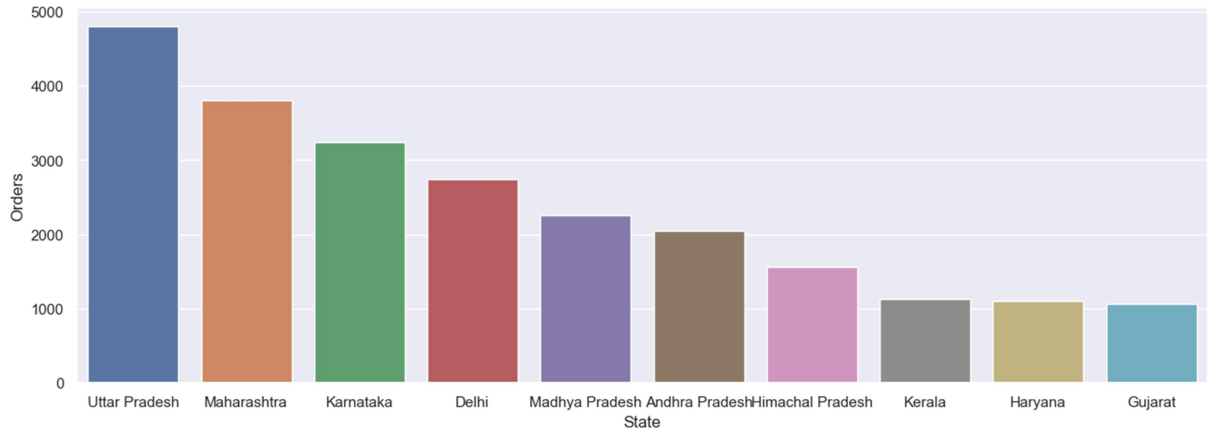


From above graphs we can see that most of the buyers are of age group between 26-35 yrs female

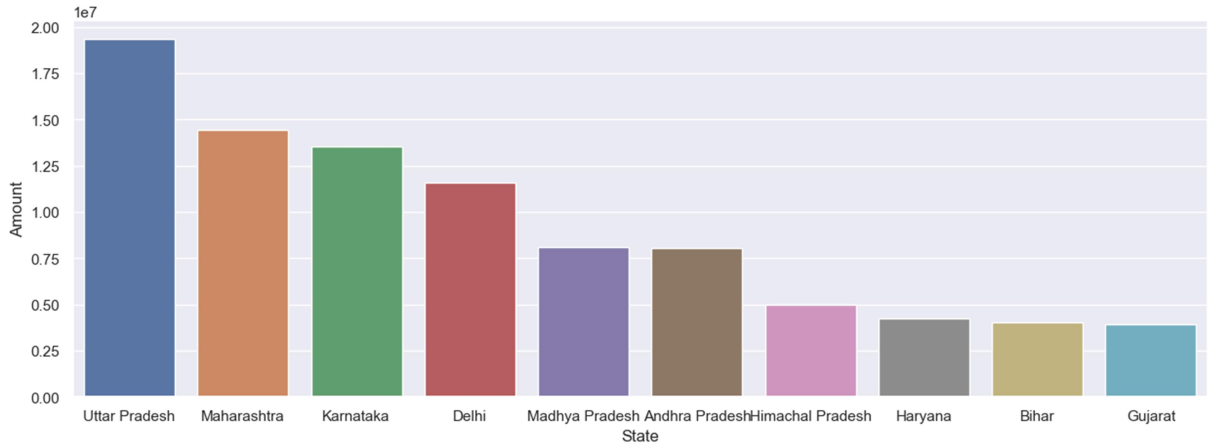
State Analysis

The analysis shifts towards the states from which the orders originate. Two bar charts are created to show the total number of orders and the total sales amount from the top 10 states.

```
#Total number of orders from top 10 states
```




```
#Total amount/sales from top 10 states
```

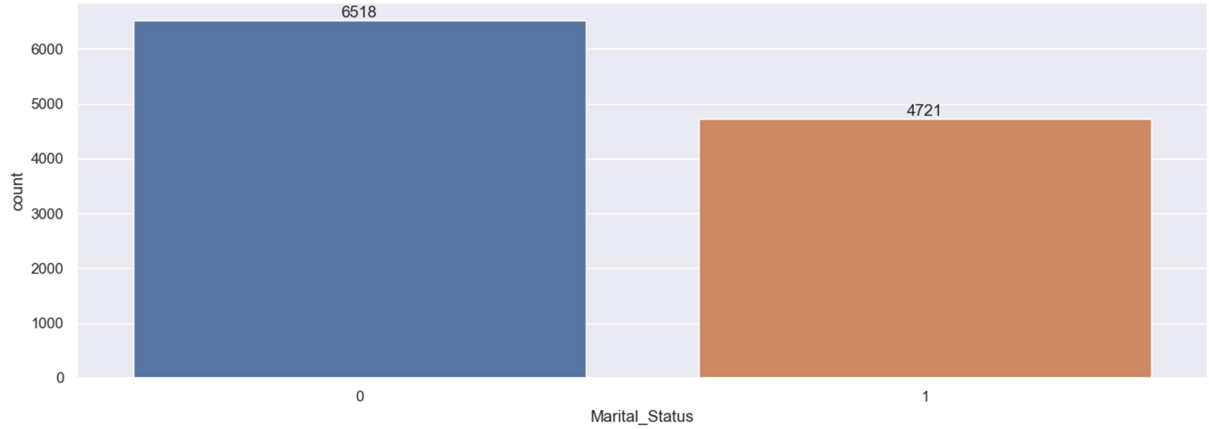


From above graphs we can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively

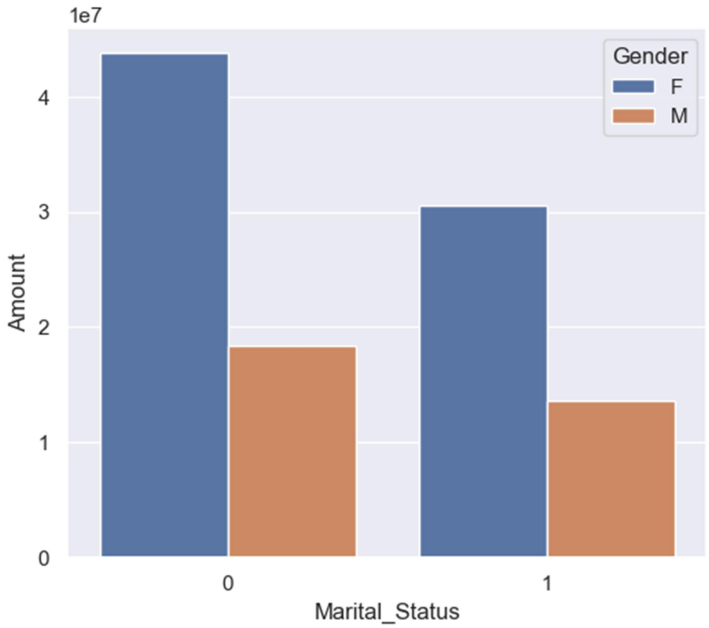
Marital Status Analysis

The marital status of the buyers is analyzed next. A bar chart is plotted to display the count of buyers for each marital status. Another bar chart is generated to compare the total sales amount for each marital status, categorized by gender.

```
#count of marital status (zero - Married, one - Unmarried)
```




```
#count of marital status amoung gender
```

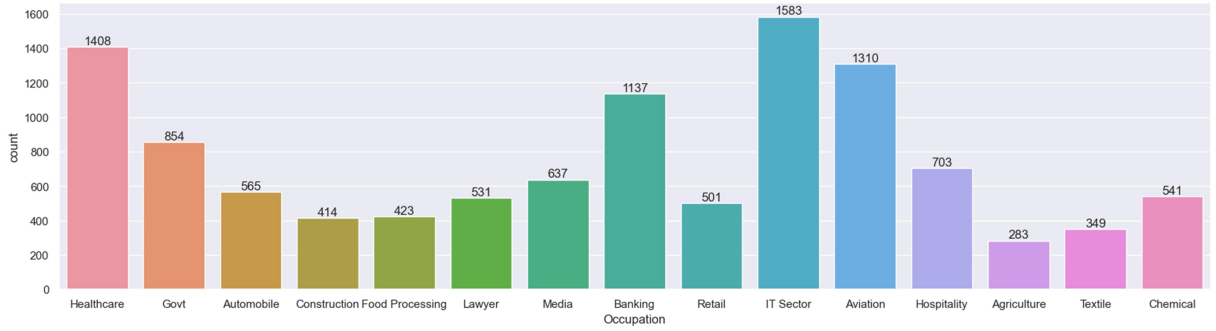


From above graphs we can see that most of the buyers are married (women) and they have high purchasing power

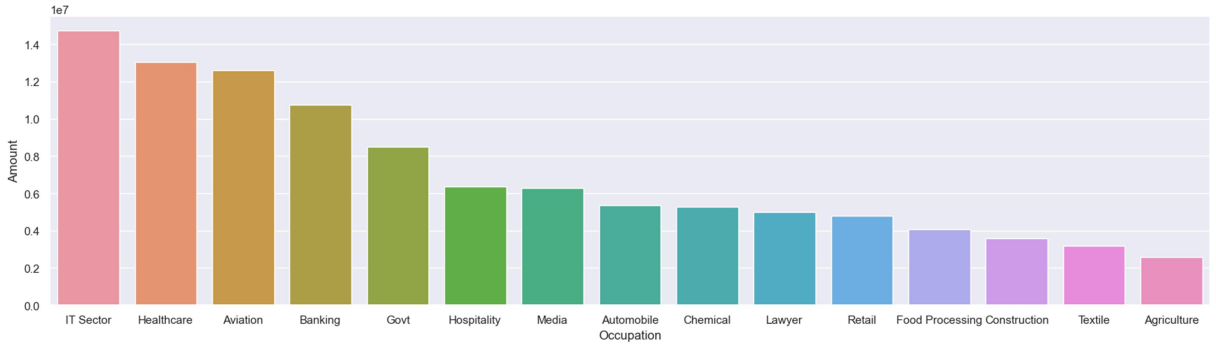
Occupation Analysis

The analysis then focuses on the occupation of the buyers. A bar chart is created to visualize the count of buyers for each occupation. Additionally, a bar chart is generated to compare the total sales amount for each occupation.

```
# count of Occupation
```




```
#Total amount/sales from Occupation
```

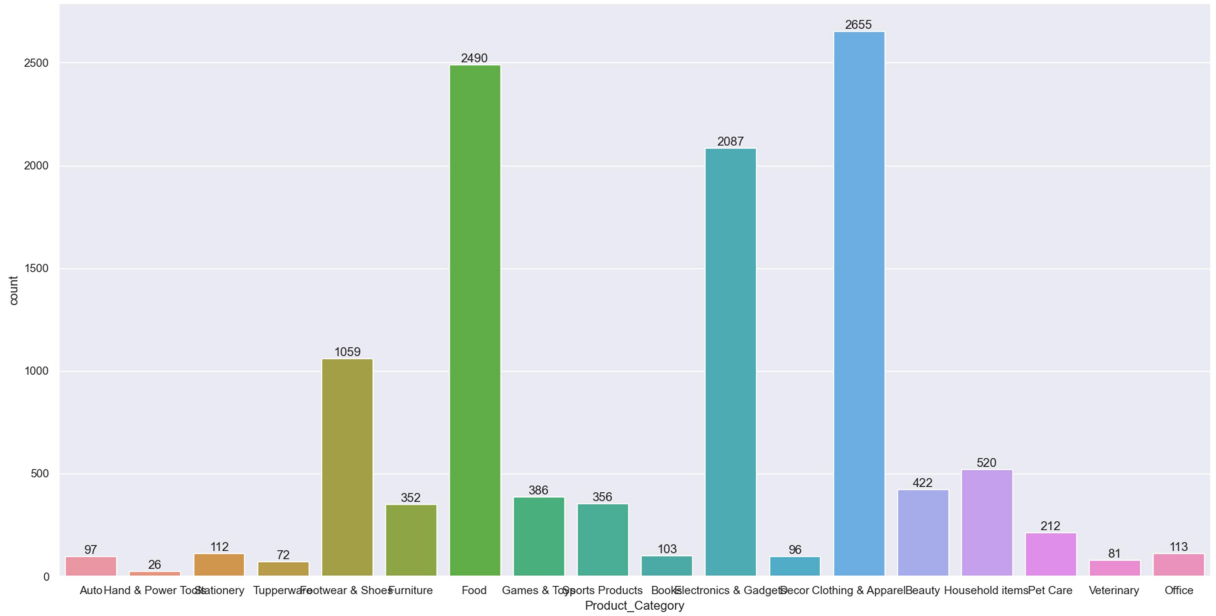


From above graphs we can see that most of the buyers are working in IT, Healthcare and Aviation sector

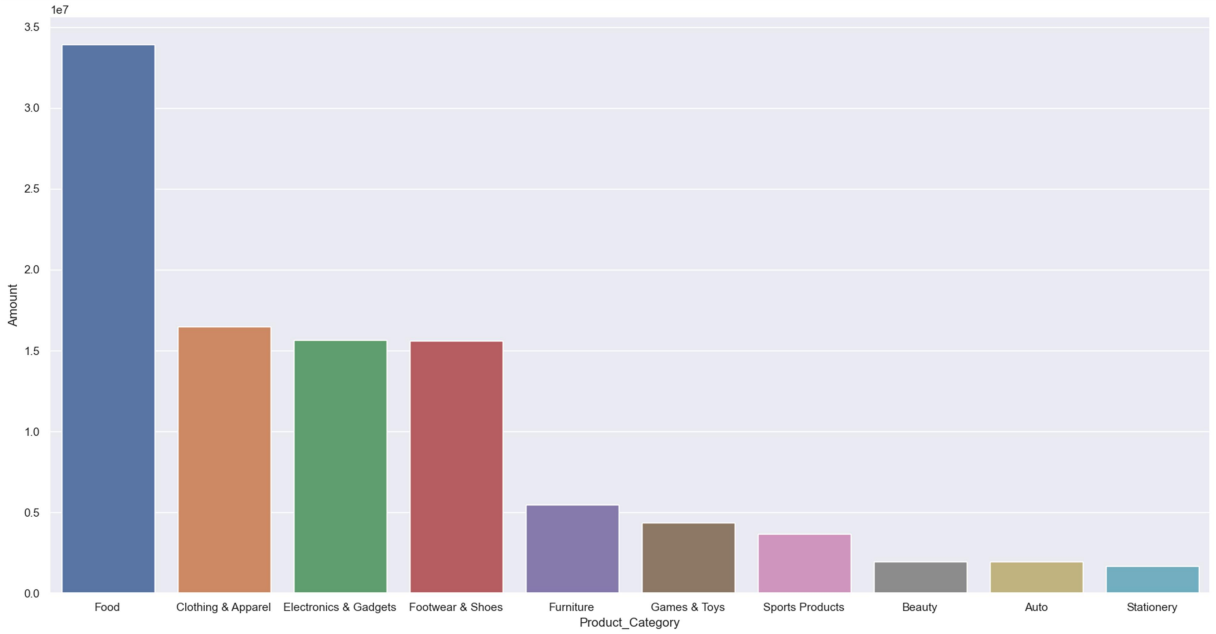
Product Category Analysis

The analysis proceeds to examine the product categories. A bar chart is plotted to display the count of products sold in each category. Another bar chart is created to compare the total sales amount for the top 10 product categories.

```
# Count of each Product Category
```




```
# Total amount/sales from Product Category
```

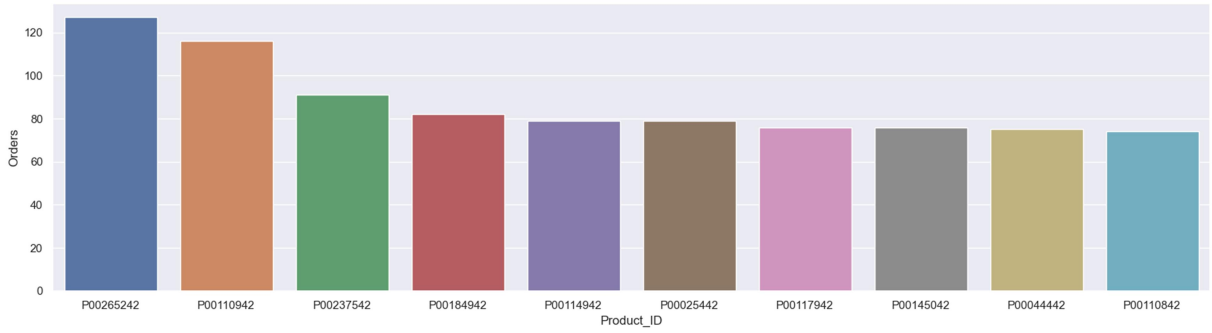


From above graphs we can see that most of the sold products are from Food, Clothing and Electronics category

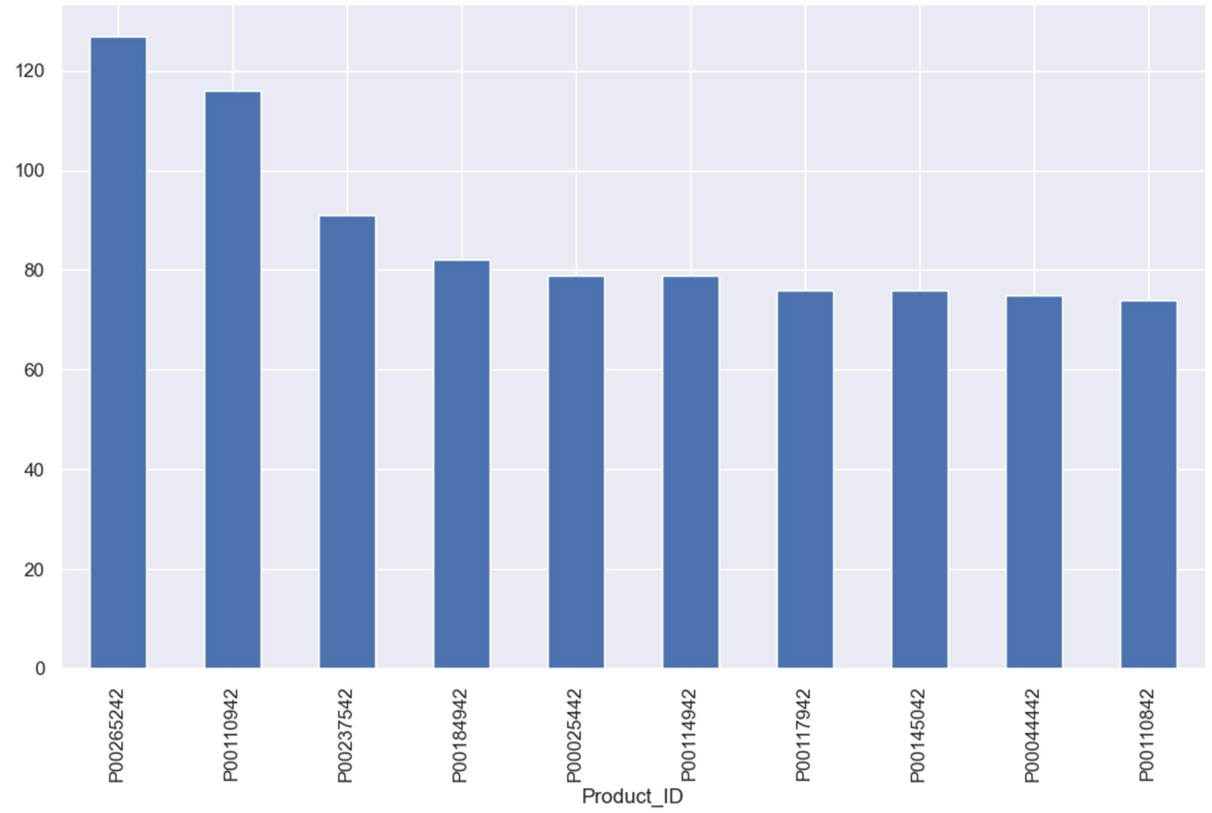
Product ID Analysis

The last analysis explores the product IDs. A bar chart is created to show the total number of orders for the top 10 product IDs. Furthermore, a bar chart is generated to visualize the top 10 most sold products.

```
#Total number of orders from top 10 product ID
```




```
#top 10 most sold products
```



Conclusion

- The majority of the buyers are females, and their purchasing power exceeds that of males.
- The most active age group in terms of shopping is females aged between 26-35 years.
- The states with the highest number of orders and total sales amount are Uttar Pradesh, Maharashtra, and Karnataka.
- Married women have a higher purchasing power compared to unmarried women.
- The most common occupations among buyers are in the IT, healthcare, and aviation sectors.
- The most popular product categories are food, clothing, and electronics.

Reference

Dataset :- https://www.kaggle.com/datasets/bishtudas/diwali-sales-dataset