Diwali sales analysis.

Presentation by Vishal Raj

Objectives:

The goal of this project is to analyze Diwali sales data to uncover key insights related to customer behaviour, sales trends, and business performance. The findings can help businesses optimize their sales strategies, improve customer targeting and increasing revenue.

Importing Libraries and Modules.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2] ✓ 1.2s Python

```
df=pd.read_csv(r"C:\Users\gupta\Downloads\Diwali Sales Data.csv",encoding='unicode_escape')
```

[3] ✓ 0.0s Python

```
df.head()
```

[4] ✓ 0.0s Python

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	ur
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.0	NaN	
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0	NaN	
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0	NaN	
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0	NaN	
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0	NaN	

Checking the info

Provide a quick overview of the DataFrame , including number of rows, data types and memory usage.

```
df.info()
```

[6] ✓ 0.0s Python

```
<class 'pandas.core.frame.DataFrame'>
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)
memory usage: 1.3+ MB
```

Data Cleaning

To drop unwanted columns in Pandas and checking for NULL values.

```
[7]  df.drop(['Status','unnamed1'],axis=1,inplace=True)  #drop unrelated column
     ✓ 0.0s                                                                Python

[8]  df.isnull().sum() #checking null values
     ✓ 0.0s                                                                Python
```

```
User_ID             0
Cust_name           0
Product_ID          0
Gender              0
Age Group           0
Age                 0
Marital_Status      0
State               0
Zone                0
Occupation          0
Product_Category    0
Orders              0
Amount             12
dtype: int64
```

Drop NULL values and change the data types.

```
df.dropna(inplace=True) #drop null values
```

[28] ✓ 0.0s Python

+ Code + Markdown

Change data type

```
df['Amount']=df['Amount'].astype('int')
df['Amount'].dtypes
```

[11] ✓ 0.0s Python

```
dtype('int64')
```

Identifying duplicates entries.

```
df=df.drop_duplicates()
```

[91] ✓ 0.0s Python

```
print("Number of duplicates entries", df.duplicated().sum())
```

[73] ✓ 0.0s Python

```
Number of duplicates entries 0
```

Exploratory Data Analysis

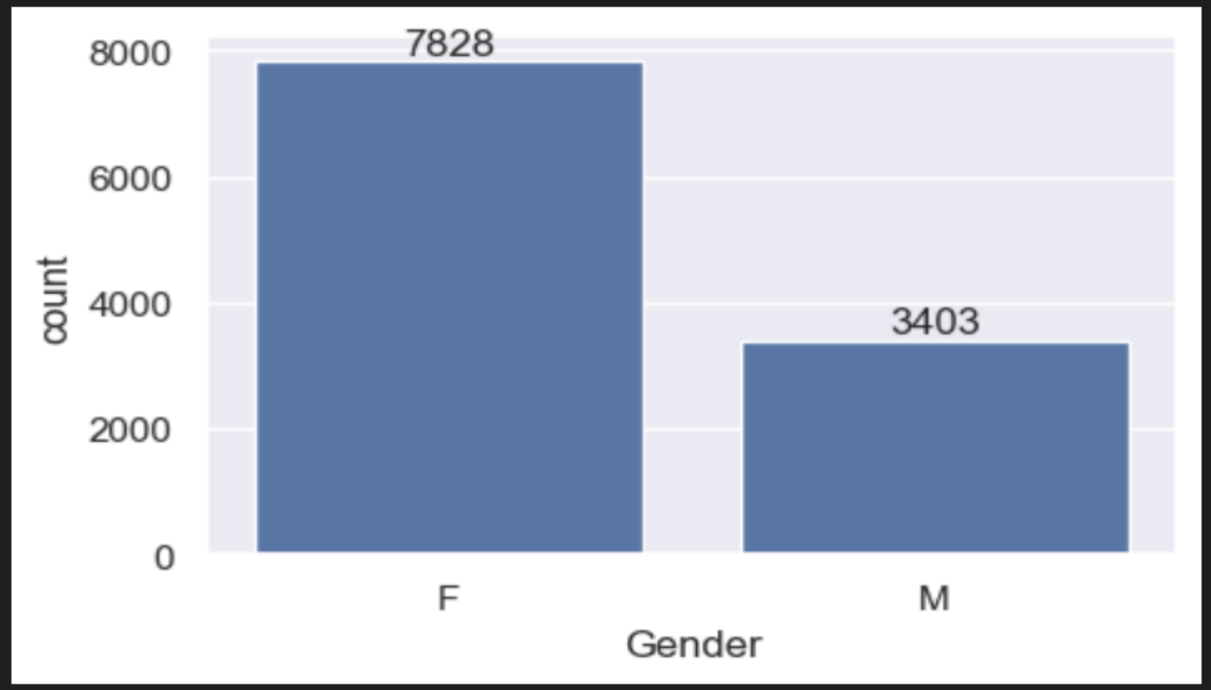
Exploratory data analysis (EDA) help to understand data characteristics. It identifies patterns, trends, and correlations. It ensures informed decision-making.

Q1. Which gender has a higher shopping frequency based on the distribution?

```
plt.figure(figsize=(5,3))
ax=sns.countplot(x='Gender',data=df)

for bars in ax.containers:
    ax.bar_label(bars)
```

[95] ✓ 0.2s

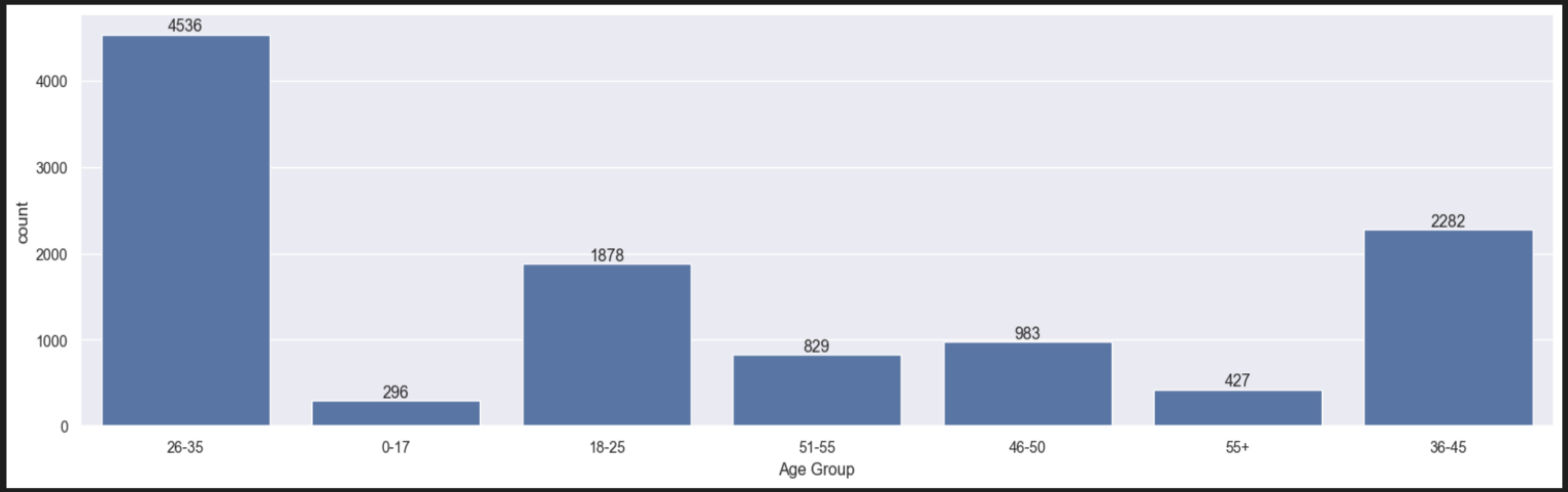


Q2. Which age group has a higher shopping frequency based on the distribution?

```
ax=sns.countplot(x='Age Group',data=df)

for bars in ax.containers:
    ax.bar_label(bars)
```

[123] ✓ 0.2s Python



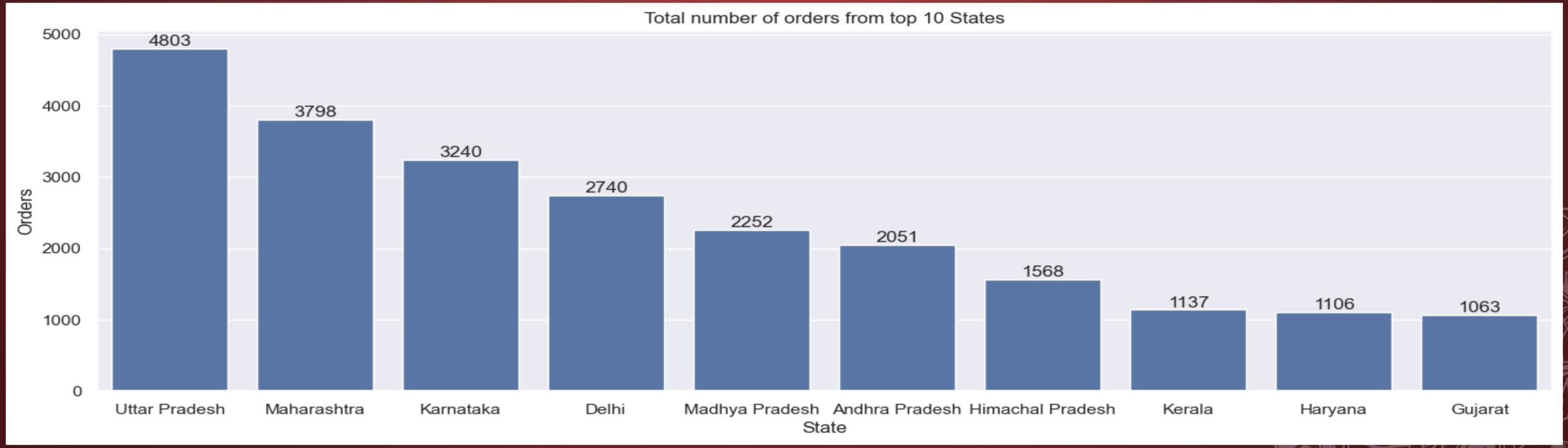
Q3. Which state has the highest number of orders?

```python
# total number of orders from top 10 states.
sales_state=df.groupby(['State'],as_index=False)['Orders'].sum().sort_values(by='Orders',ascending=False).head(10)
sns.set(rc={'figure.figsize':(16,5)})
ax=sns.barplot(x='State',y='Orders',data=sales_state)

plt.title("Total number of orders from top 10 States")

for bars in ax.containers:
    ax.bar_label(bars)
```

[133] ✓ 0.4s Python

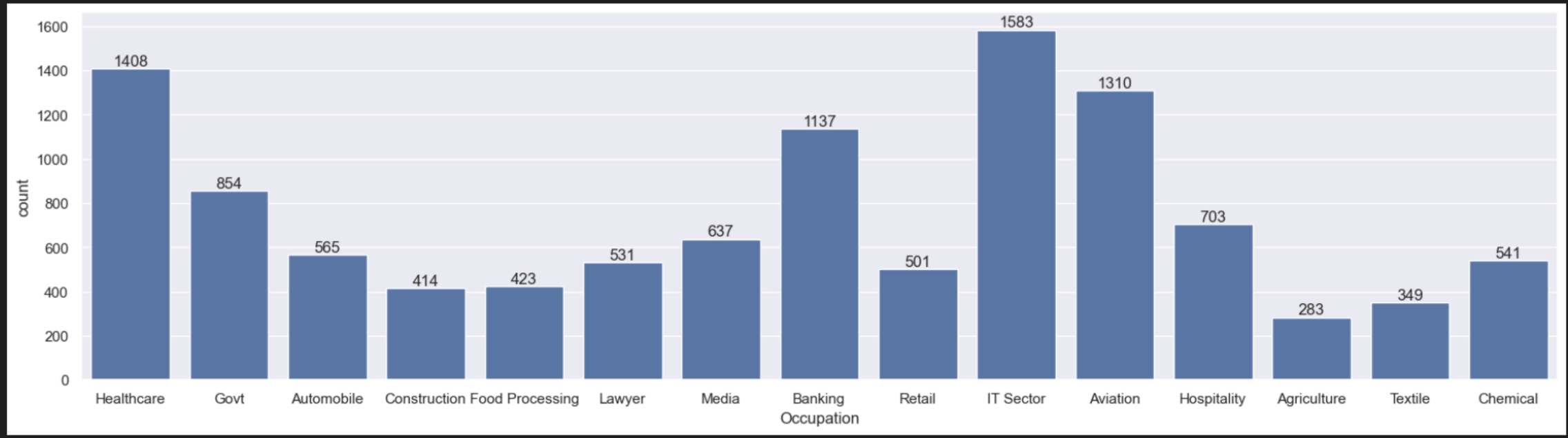


Q4. Which occupation has the highest shopping frequency?

```
ax=sns.countplot(x='Occupation',data=df)
sns.set(rc={'figure.figsize':(20,2)})

for bars in ax.containers:
    ax.bar_label(bars)
```

[86] Python



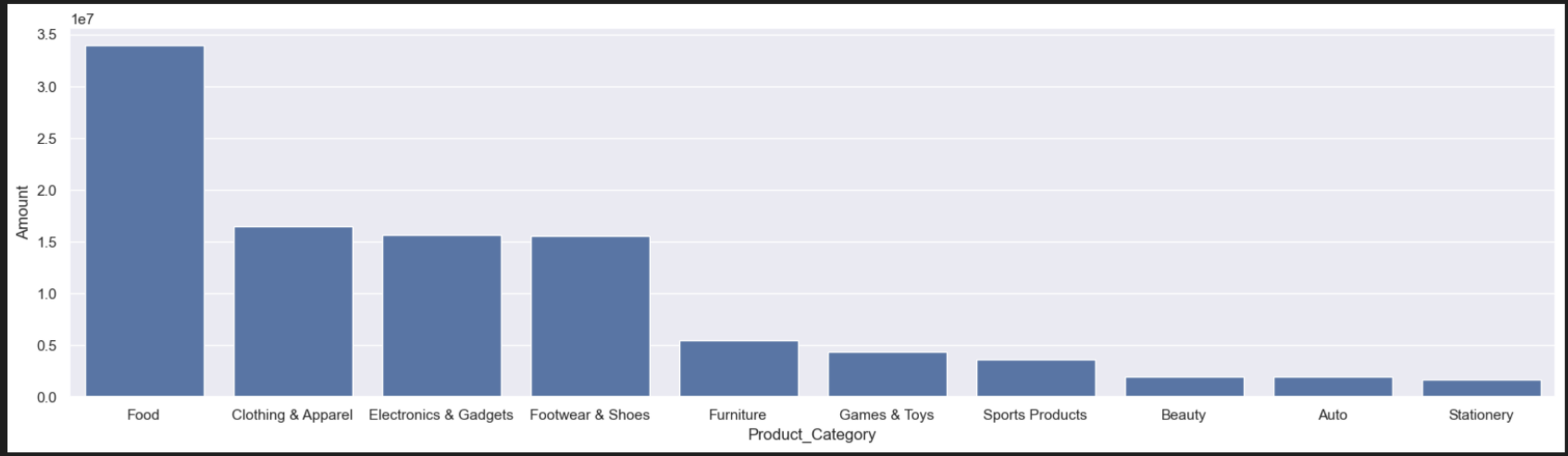
Q5. What is the top-selling product?

```
sales_state=df.groupby(['Product_Category'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False).head(10)

sns.barplot(x='Product_Category',y='Amount',data=sales_state)
```

[26] ✓ 0.1s Python

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



Q6. Which marital status group(single, married) made the most purchases during Diwali?

```python
#check the marital_status based on the amount.
sales_state=df.groupby(['Marital_Status','Gender'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False)

sns.set(rc={'figure.figsize':(5,3)})
sns.barplot(x='Marital_Status',y='Amount', hue='Gender',data=sales_state)
```

[29] ✓ 0.1s Python

Conclusion:

Married women age group 26-35 years from UP, Maharashtra and Karnataka working in IT, Aviation and Healthcare are mostly like to buy product from Food, Clothing and Electronics category.

Thank You For Your Attention