Exploratory Data Analysis(EDA)

Subtitle

Project Overview

- We have diwali sales data of a company.
- We have done Exploratory data Analysis using python
- We have used pandas, numpy, matplotlib and seaborn libraries.

Imported libraries and read csv file

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
Matplotlib is building the font cache; this may take a moment.
```

In [2]:
```
df=pd.read_csv(r'C:\Users\gaura\Desktop\Untitled Folder\Diwali Sales Data.csv', encoding='unicode_escap
```

Deriving info of csv data file

```
In [6]: df.info()
```

```
<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

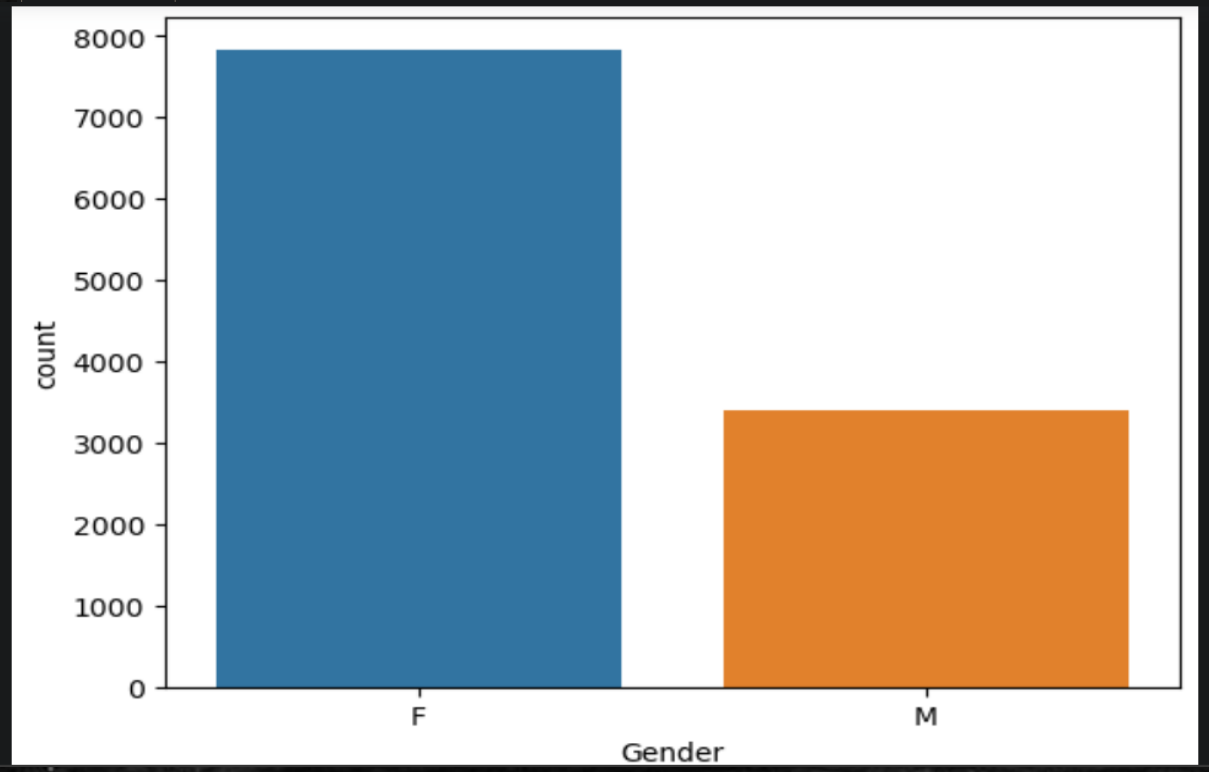
Dropping non useful columns and null values

In [7]:
```
#drop unrelated/blank columns
df.drop(['Status','unnamed1'],axis=1,inplace=True)
```

In [11]:
```
#drop null values
df.dropna(inplace=True)
```

Data analysis: count customers a/c to gender

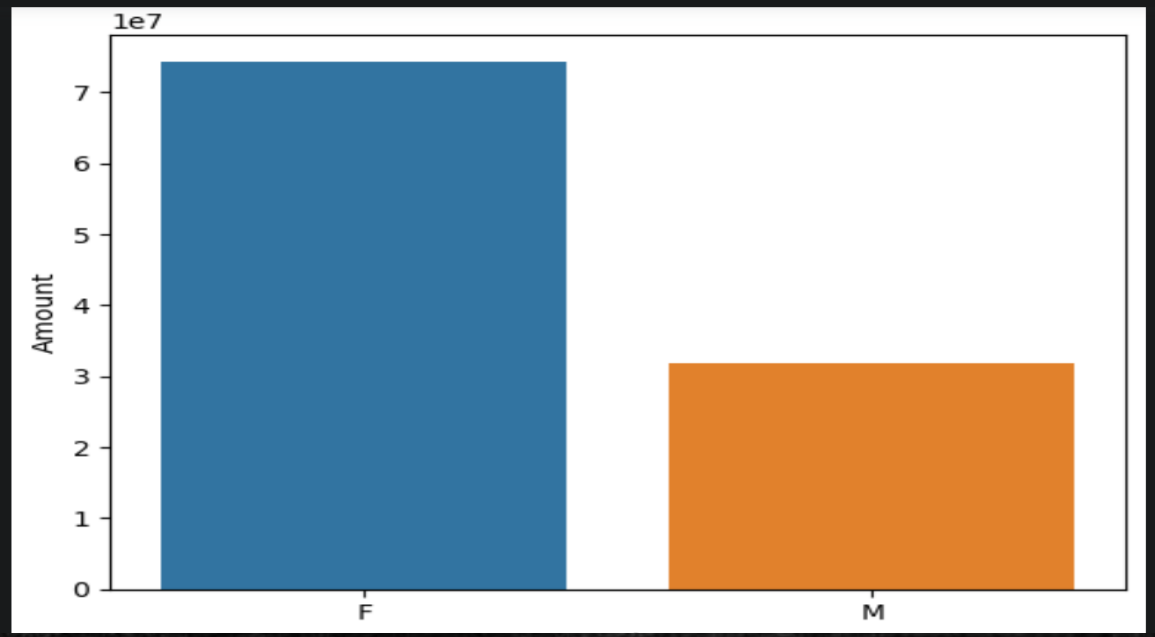
```
In [15]: ax=sns.countplot(x='Gender',data=df)

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



Female customers are much higher than male

Total Amount Spent by each gender

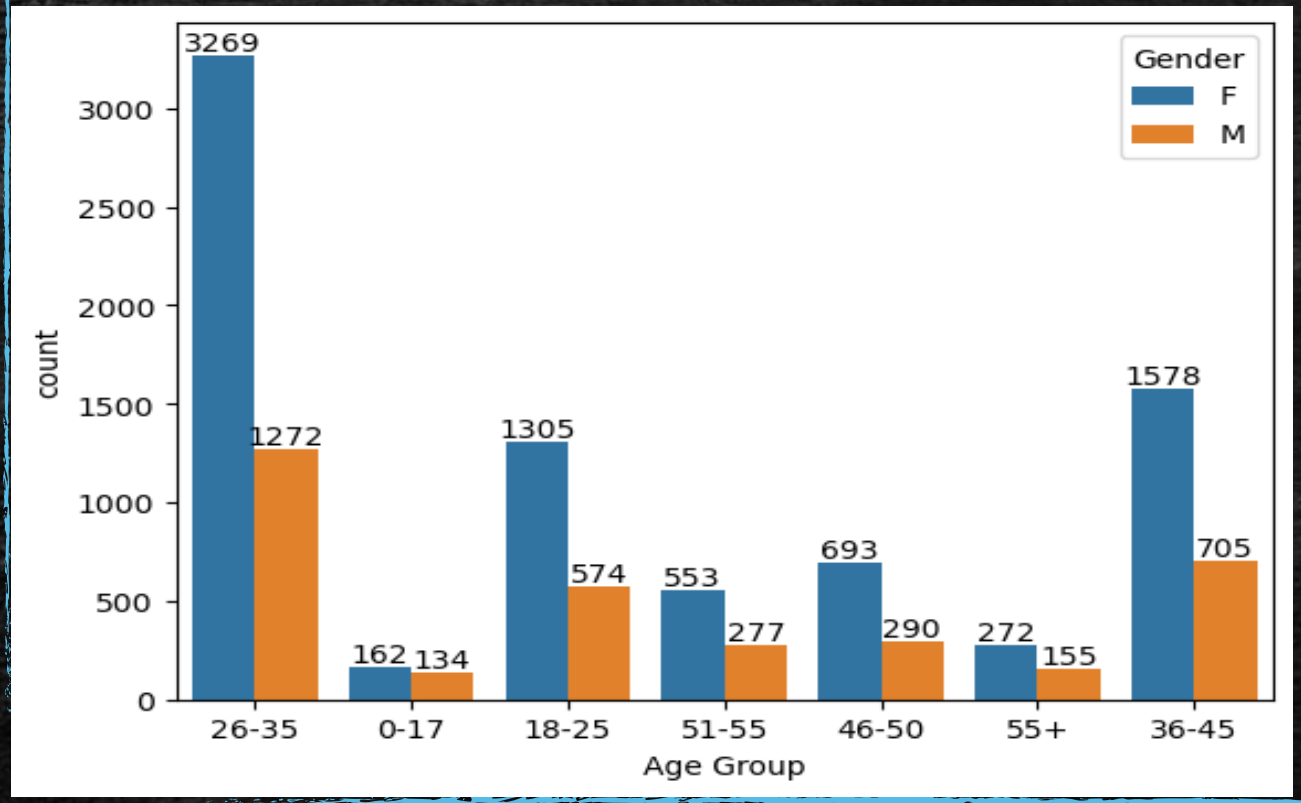
```
In [16]: sales_gen=df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=Fals

         sns.barplot(x='Gender',y='Amount',data=sales_gen)
```



Amount Spent is much higher by female than male.

Number of orders a/c To age group

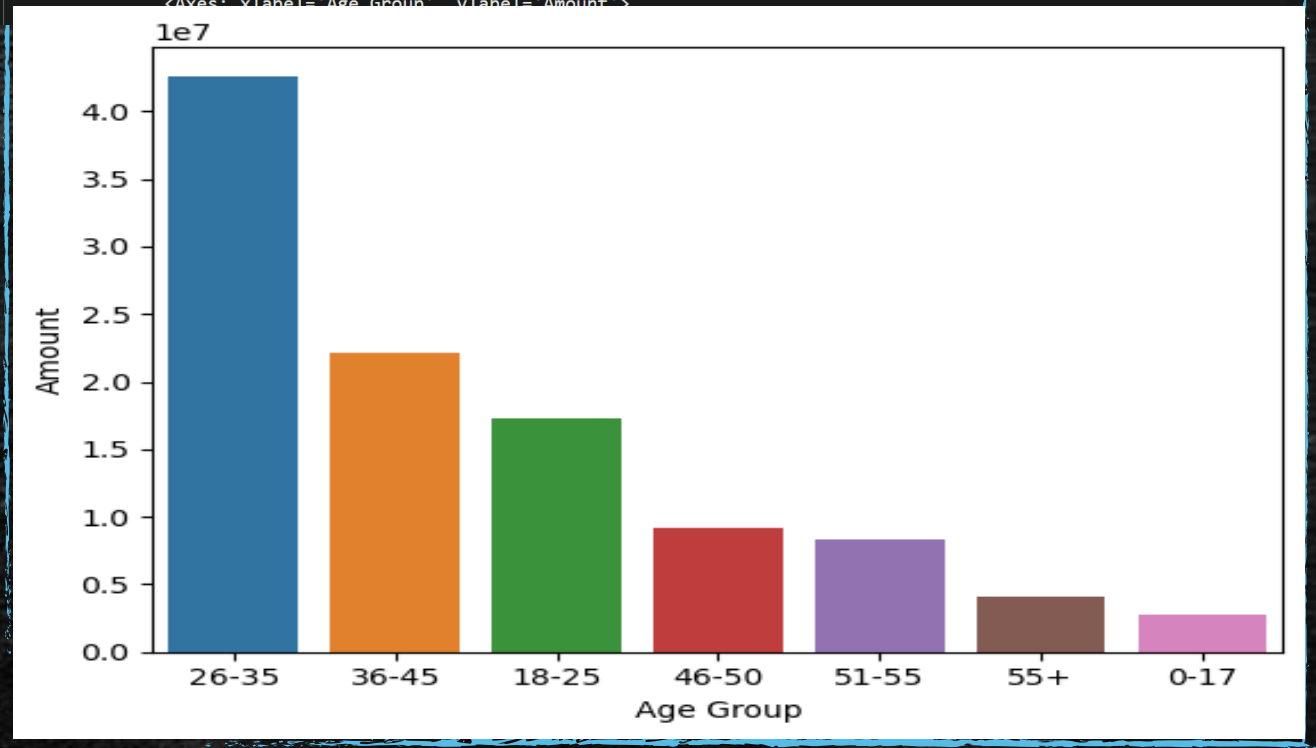
```
In [18]: ax=sns.countplot(data=df,x='Age Group',hue='Gender')

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



Maximum customers are female of 26-35 age group

Amount a/c to age groups

```
In [20]: #Total amount vs Age group
         sales_age=df.groupby(['Age Group'], as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=F
         sns.barplot(x='Age Group',y='Amount',data=sales_age)
```

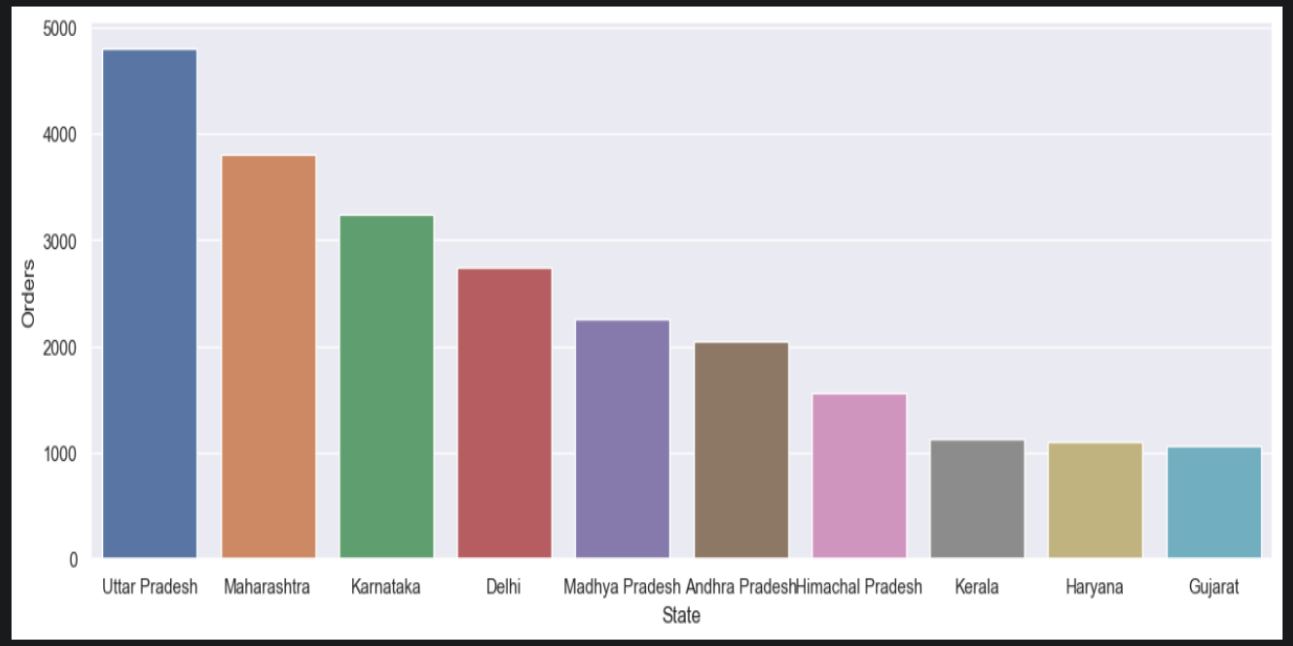


Total amount spent by age group 26-35 is highest

Total orders from top 10 states

```
sales_state=df.groupby(['State'], as_index=False)['Orders'].sum().sort_values(by='Orders',ascending=Fal

#sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data=sales_state,x='State',y='Orders')
```

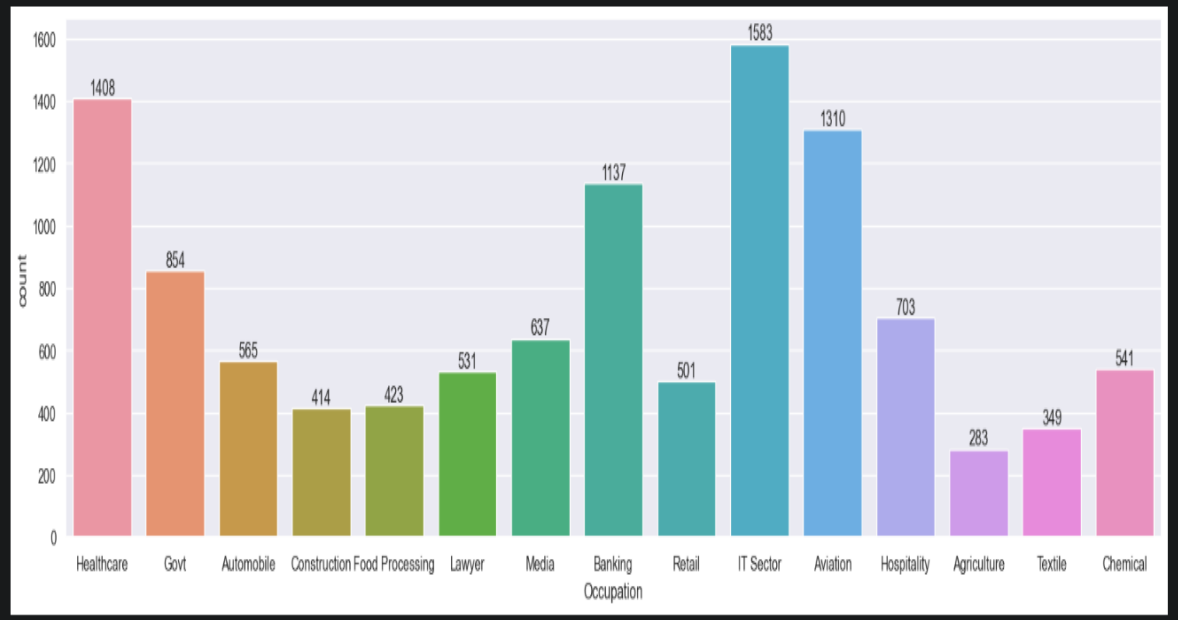
```
<Axes: xlabel='State', ylabel='Orders'>
```



Maximum orders are from Uttar Pradesh

Orders ranked a/c to Occupation

```
In [27]: sns.set(rc={'figure.figsize':(20,5)})
         ax=sns.countplot(data=df,x='Occupation')

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

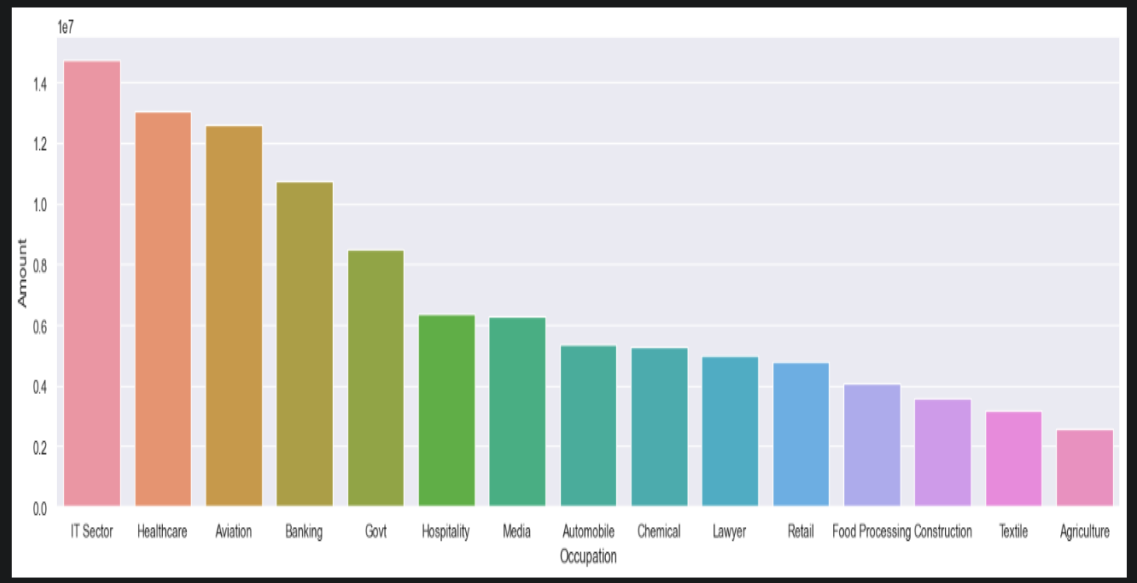


Maximum no of customers are from IT Sector.

Amount Spent a/c to Occupation

```
In [28]: sales_state=df.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascendi

         sns.set(rc={'figure.figsize':(20,5)})
         sns.barplot(data = sales_state, x='Occupation', y='Amount')
```

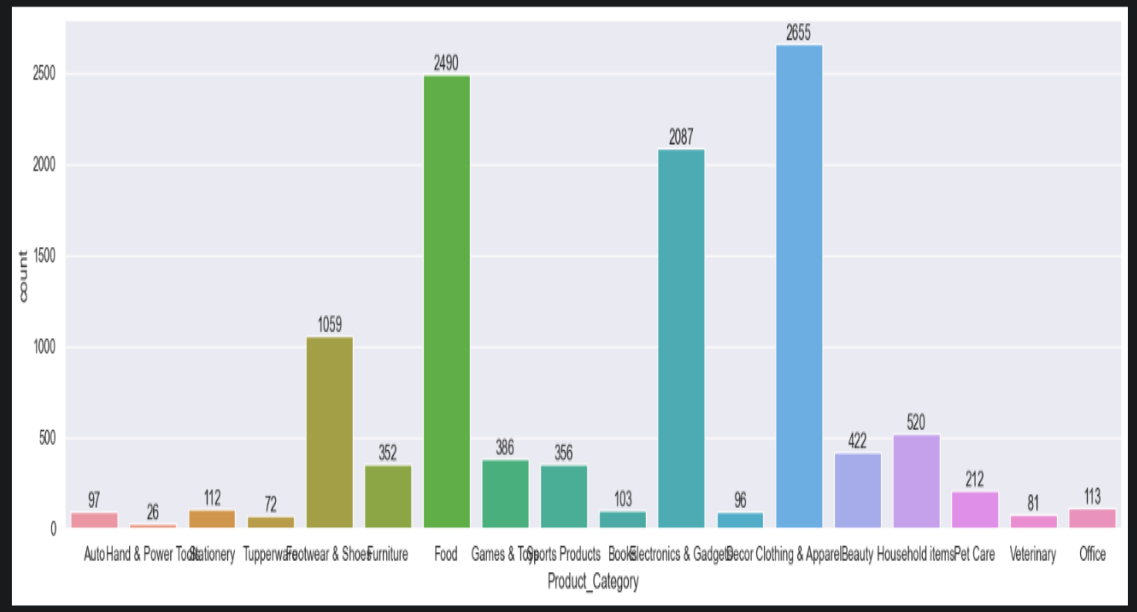
```
<Axes: xlabel='Occupation', ylabel='Amount'>
```



Amount of sales generated are highest by IT Sector Employees

Orders a/c to Product Category

```
In [29]: sns.set(rc={'figure.figsize':(20,5)})
         ax=sns.countplot(data=df,x='Product_Category')

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



Orders are highest by clothing and apparel followed by food then Electronics and Gadgets

Conclusion

Married woman age from 26-35 years from U.P, Maharashtra and Karnataka working in IT, Healthcare and aviation are more likely to buy products from food, clothing and electronics category.

THANK YOU