

# Table of Content

S. no	Particular		Page no.
1	Introduction		2
2	Background		2
3	Data Analysis Process		3
	3.1	Ask	3
	3.2	Prepare	3
	3.3	Process	4
	3.4	Analyze and Share	5
	3.5	Act	12

# Introduction

In this case study, I will perform many real-world tasks of a junior data analyst at a fictional Multi-Category retailer, Priya Store. In order to answer the key business questions, I will follow the steps of the data process: Ask, Prepare, Process, Analyse, Share, and Act.

# Background

## Priya Store:

- Priya Store is a multi-category retail company. Priya Store sells its product all around country.
- Priya store wants to create an Deepwali sales report for 2022. So that, Priya Store can understand their customers and grow more sales upcoming festivals.

## Scenario:

I am assuming to be a junior data analyst working in the marketing analyst team at Priya Store, a Multi-Category retailer. The director of marketing believes the company's future success depends on their customers and grow more sales in festive seasons. Therefore, my team wants to understand how different variables affected sales and growth of company in last Deepwali.

From these insights, my team will design a new marketing strategy to grow more sales in upcoming festivals. But first, Priya Store executives must approve our recommendations, so they must be backed up with compelling data insights and professional data visualizations.

# Data Analysis Process

## Ask

### Business Task

Priya store wants to create an annual sales report for 2022. So that, Priya Store can understand their customers and grow more sales in 2023.

### Analysis Questions and Queries.

1. Compare men customers with women customers for number of order and total sales.
2. Compare all age groups for numbers of order and total sales.
3. Find the number of order and total sales of top 10 states.
4. Compare between married and unmarried for the numbers of order and total sales.
5. Classify number of orders and total sales according to occupations.
6. Find number of orders and total sales for each product category.

## Prepare

### Data source

I will use Priya Store sales data to analyze and identify trends for diwali 2022 which can be downloaded from GitHub website.

This is public data that can be used to explore how different customer types buying different products of Priya Store from all over India.

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

In [2]: # import csv file
df = pd.read_csv('Diwali Sales Data.csv', encoding= 'unicode_escape')
```

## Data organization

There is one CSV file named ‘Diwali Sales Data.csv’ includes information for Deewali sales data with column names such as 'User\_ID', 'Cust\_name', 'Product\_ID', 'Gender', 'Age Group', 'Age', 'Marital\_Status', 'State', 'Zone', 'Occupation', 'Product\_Category', 'Orders', 'Amount'.

In [4]: df.head()

Out[4]:

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

## Process

### Data exploration

Python Jupyter notebook is used for data exploration and to get familiarize with the data

In [5]: 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

Data cleaning and modification is done with the help of Python Jupyter notebook

```
In [6]: #drop unrelated/blank columns  
df.drop(['Status', 'unnamed1'], axis=1, inplace=True)
```

```
In [7]: #check for null values  
pd.isnull(df).sum()
```

```
Out[7]: 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
```

```
In [8]: # drop null values  
df.dropna(inplace=True)
```

```
In [9]: # change data type  
df['Amount'] = df['Amount'].astype('int')
```

```
In [10]: df['Amount'].dtypes
```

```
Out[10]: dtype('int32')
```

```
In [12]: #rename column  
df.rename(columns= {'Marital_Status':'Shaadi'})
```

```
Out[12]:   User_ID Cust_name Product_ID Gender Age Group Age Shaadi State Zone Occupation Product_Category Orders Amount  
0  1002903  Sanskriti P00125942      F  26-35  28    0 Maharashtra Western Healthcare       Auto    1  23952  
1  1000732      Kartik P00110942      F  26-35  35    1 Andhra Pradesh Southern Govt          Auto    3  23934  
2  1001990      Bindu P00118542      F  26-35  35    1 Uttar Pradesh Central Automobile       Auto    3  23924  
3  1001425     Sudevi P00237842      M  0-17   16    0 Karnataka Southern Construction      Auto    2  23912  
4  1000588      Joni P00057942      M  26-35  28    1 Gujarat Western Food Processing      Auto    2  23877  
... ... ... ... ... ... ... ... ... ... ... ...  
11246 1000695      Manning P00296942      M  18-25  19    1 Maharashtra Western Chemical       Office   4   370  
11247 1004089  Reichenbach P00171342      M  26-35  33    0 Haryana Northern Healthcare      Veterinary   3   367  
11248 1001209      Oshin P00201342      F  36-45  40    0 Madhya Pradesh Central Textile          Office   4   213  
11249 1004023      Noonan P00059442      M  36-45  37    0 Karnataka Southern Agriculture      Office   3   206  
11250 1002744     Brumley P00281742      F  18-25  19    0 Maharashtra Western Healthcare       Office   3   188
```

11239 rows × 13 columns

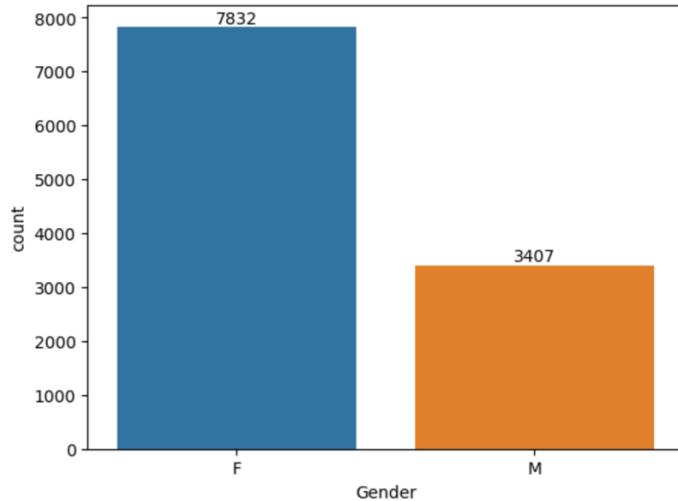
## Analyse & Share

After cleaning the data, it is ready for analysis. Python Jupyter notebook is used for Data Analysis.

In data analysis we will find the answers to the Data Analysis Questions and Queries.

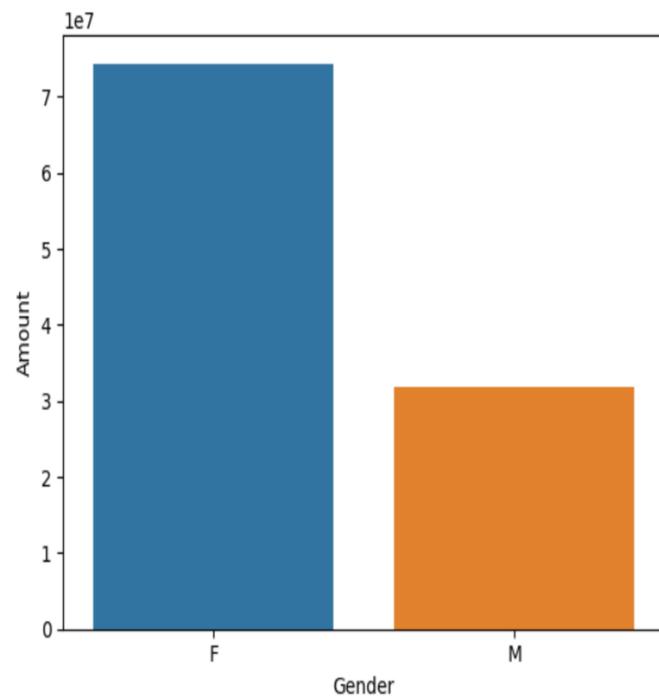
# 1. Compare men customers with women customers for number of order and total sales.

```
In [15]: # plotting a bar chart for Gender and it's count  
ax = sns.countplot(x = 'Gender', data = df)  
  
for bars in ax.containers:  
    ax.bar_label(bars)
```



```
In [16]: # plotting a bar chart for gender vs total amount  
  
sales_gen = df.groupby(['Gender'], as_index=False)[['Amount']].sum().sort_values(by='Amount', ascending=False)  
  
sns.barplot(x = 'Gender', y = 'Amount', data = sales_gen)
```

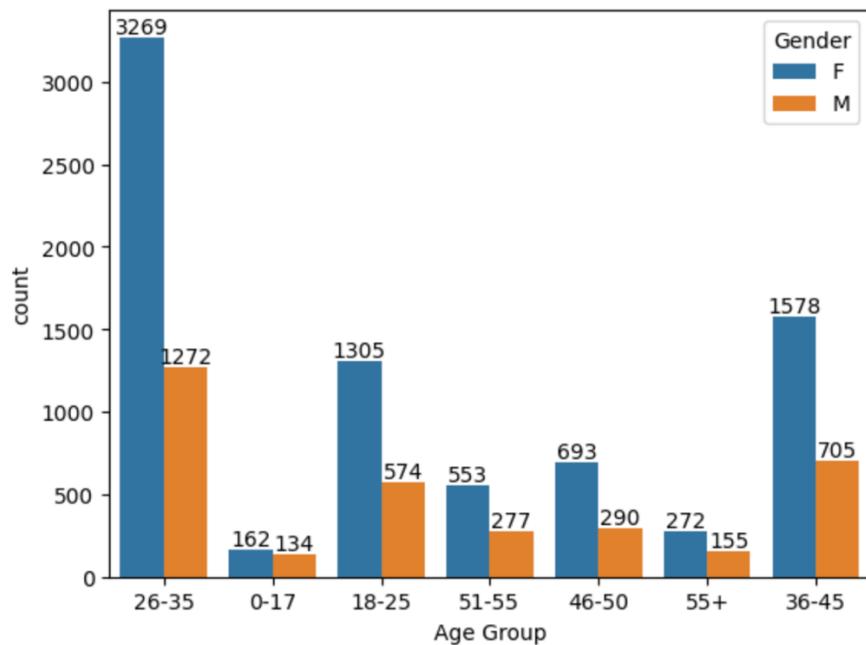
Out[16]: <Axes: xlabel='Gender', ylabel='Amount'>



## 2. Compare all age groups for numbers of order and total sales.

```
In [17]: ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender')

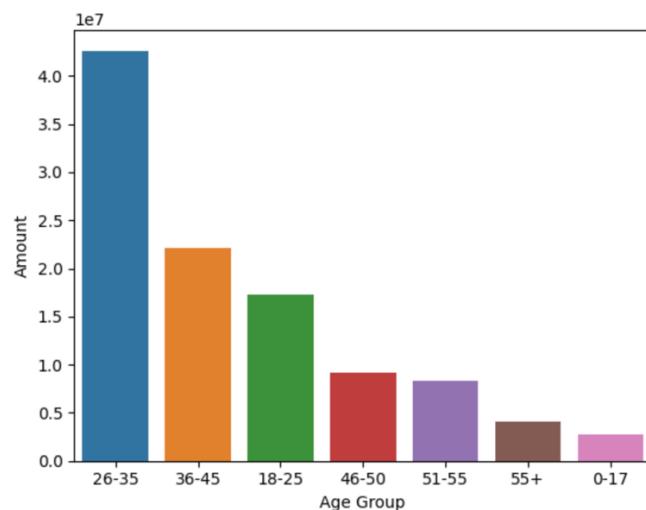
for bars in ax.containers:
    ax.bar_label(bars)
```



```
In [18]: # Total Amount vs Age Group
sales_age = df.groupby(['Age Group'], as_index=False)[['Amount']].sum().sort_values(by='Amount', ascending=False)

sns.barplot(x = 'Age Group',y= 'Amount' ,data = sales_age)
```

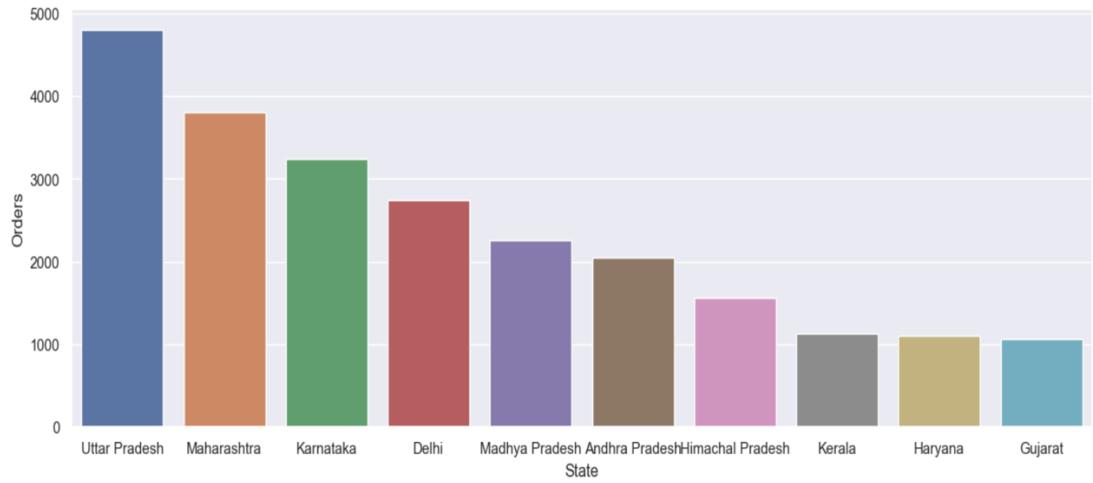
Out[18]: <Axes: xlabel='Age Group', ylabel='Amount'>



### 3. Find the number of order and total sales of top 10 states.

```
In [19]: # 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':(15,5)})  
sns.barplot(data = sales_state, x = 'State',y= 'Orders')
```

```
Out[19]: <Axes: xlabel='State', ylabel='Orders'>
```



```
In [20]: # total amount/sales from top 10 states
```

```
sales_state = df.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)  
  
sns.set(rc={'figure.figsize':(15,5)})  
sns.barplot(data = sales_state, x = 'State',y= 'Amount')
```

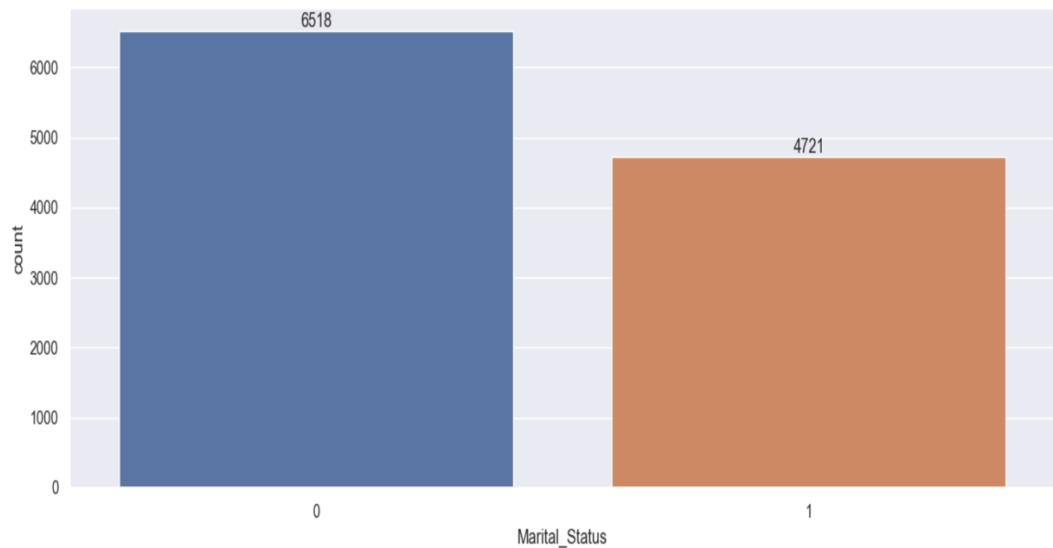
```
Out[20]: <Axes: xlabel='State', ylabel='Amount'>
```



4. Compare between married and unmarried for the numbers of order and total sales.

```
In [21]: ax = sns.countplot(data = df, x = 'Marital_Status')

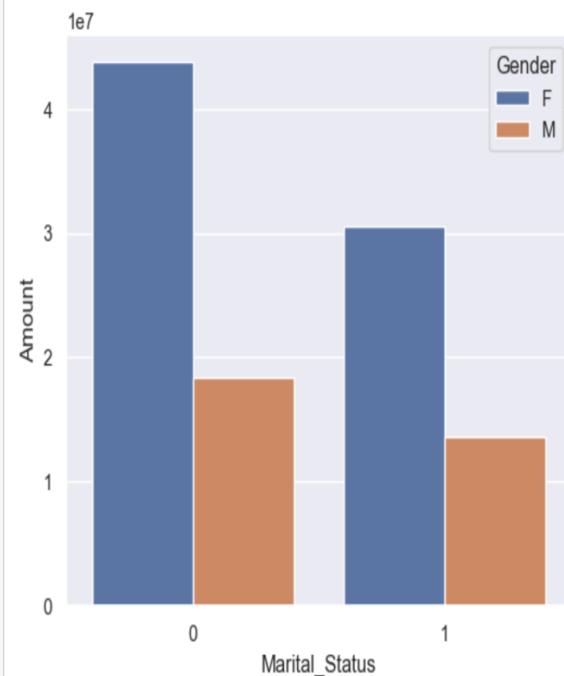
sns.set(rc={'figure.figsize':(7,5)})
for bars in ax.containers:
    ax.bar_label(bars)
```



```
In [22]: sales_state = df.groupby(['Marital_Status', 'Gender'], as_index=False)[['Amount']].sum().sort_values(by='Amount', ascending=False)

sns.set(rc={'figure.figsize':(6,5)})
sns.barplot(data = sales_state, x = 'Marital_Status', y= 'Amount', hue='Gender')
```

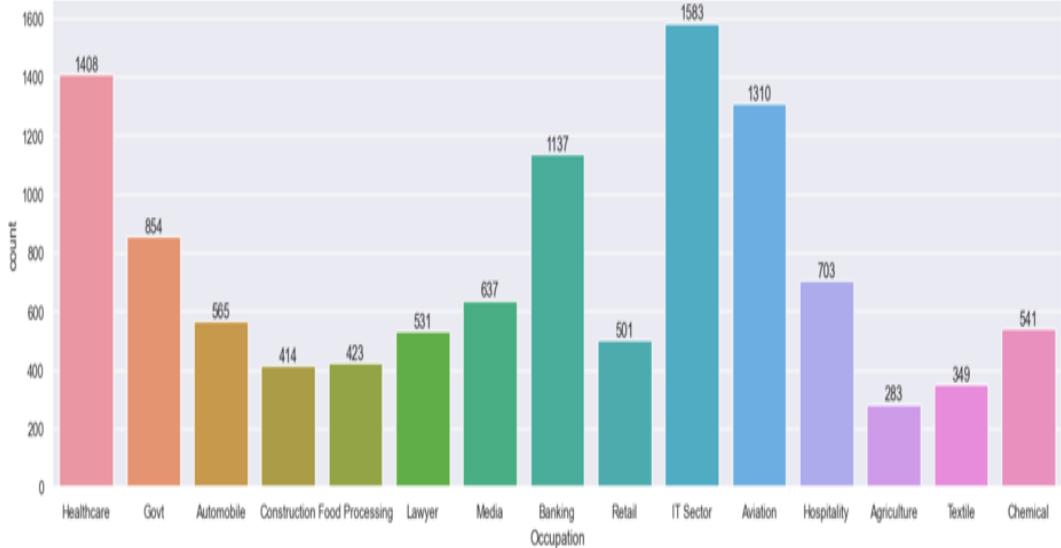
Out[22]: <Axes: xlabel='Marital\_Status', ylabel='Amount'>



## 5. Classify number of orders and total sales according to occupations.

```
In [23]: sns.set(rc={'figure.figsize':(20,5)})
ax = sns.countplot(data = df, x = 'Occupation')

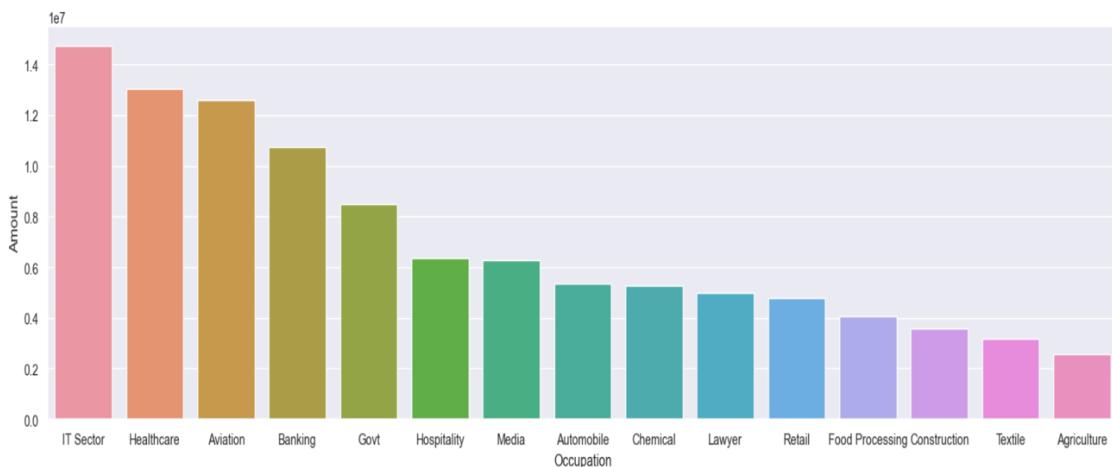
for bars in ax.containers:
    ax.bar_label(bars)
```



```
In [24]: sales_state = df.groupby(['Occupation'], as_index=False)[['Amount']].sum().sort_values(by='Amount', ascending=False)

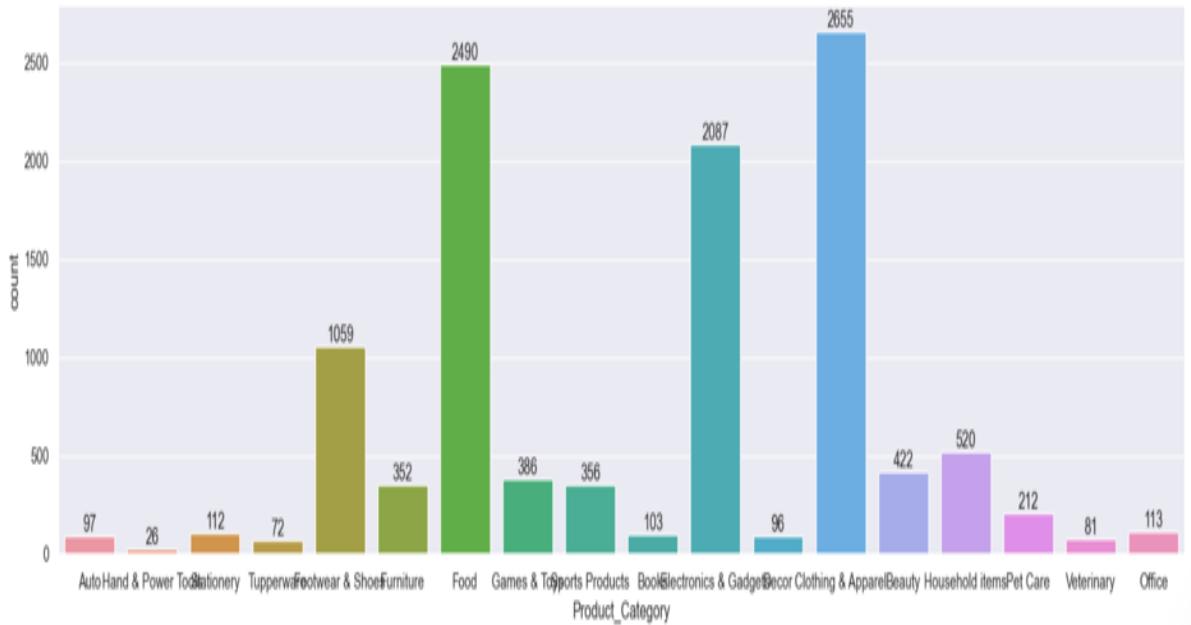
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Occupation', y = 'Amount')
```

Out[24]: <Axes: xlabel='Occupation', ylabel='Amount'>



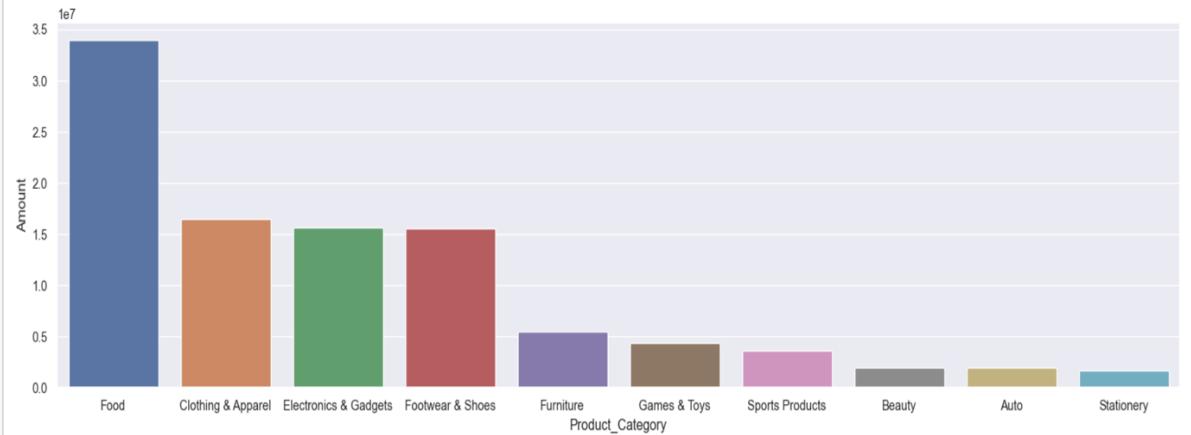
6. Find number of orders and total sales for each product category.

```
In [25]: sns.set(rc={'figure.figsize':(20,5)})  
ax = sns.countplot(data = df, x = 'Product_Category')  
  
for bars in ax.containers:  
    ax.bar_label(bars)
```



```
In [26]: sales_state = df.groupby(['Product_Category'], as_index=False)[['Amount']].sum().sort_values(by='Amount', ascending=False)  
  
sns.set(rc={'figure.figsize':(20,5)})  
sns.barplot(data = sales_state, x = 'Product_Category',y= 'Amount')
```

```
Out[26]: <Axes: xlabel='Product_Category', ylabel='Amount'>
```



## **Summary of analysis**

- Most of the buyers are females and even the purchasing power of females are greater than men.
- Most of the buyers are of age group between 26-35 years female.
- Most of the buyers are married (women) and they have high purchasing power.
- Most of the buyers are working in IT, Healthcare and Aviation sector.
- Most of the sold products are from Food, Clothing and Electronics category.

## **Act**

After thorough review of analysis of sales data of Deewali 2022 following points have been concluded.

- Women customers of the store have more purchasing power so in festive season store should focus more on the demand of women customer base.
- Store should introduce new product line focusing on women of age between 26-35 years.
- Store should exploit the demand of married women centric products as this category has the high potential of more purchase.
- Demand among the people working in IT, Healthcare and Aviation sector are very high, more analysis should be conducted in this area.
- More product lines can be ventured into within the category of Food, clothing and Electronics.