



1 Setting The Working Environment

In [1]:

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 %matplotlib inline
5 import seaborn as sns
```

2 Importing The Dataset

In [2]:

```
1 df = pd.read_csv('Diwali Sales Data.csv', encoding= 'unicode_escape')
```

3 Exploring Dataset

In [3]:

```
1 df.shape
```

Out[3]:

```
(11251, 15)
```

In [4]:

```
1 df.head()
```

Out[4]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	V
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Sc
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Sc
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	V



In [5]:

```
1 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
```

In [6]:

```
1 df.drop(['Status', 'unnamed1'], axis=1, inplace=True)
```

In [7]:

```
1 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]:

```
1 df.dropna(inplace=True)
```

In [9]:

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

In [10]:

```
1 df['Amount'].dtypes
```

Out[10]:

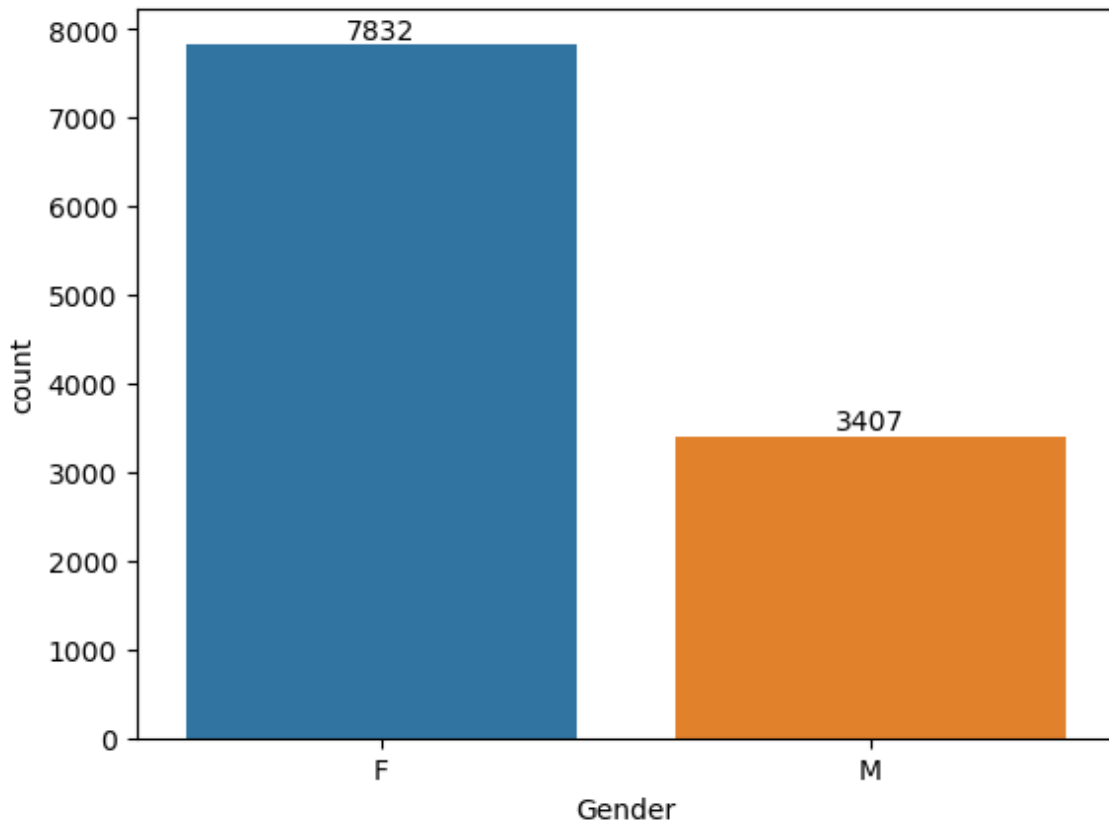
```
dtype('int32')
```

Gender



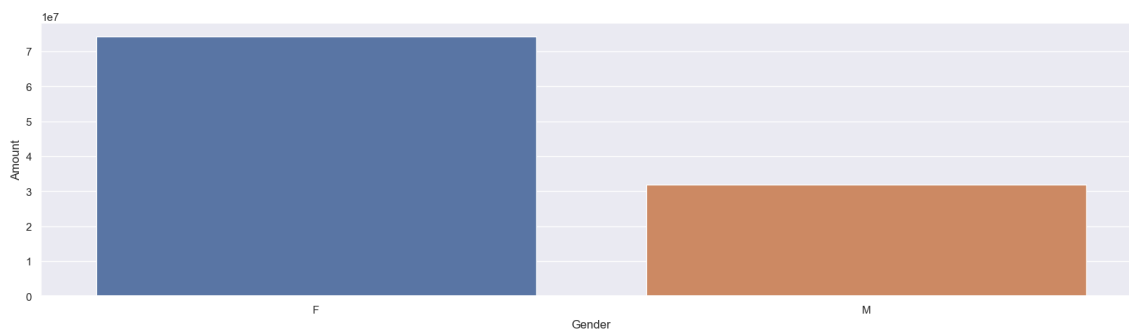
In [11]:

```
1 ax = sns.countplot(x = 'Gender',data = df)
2
3 for bars in ax.containers:
4     ax.bar_label(bars)
```



In [49]:

```
1 sales_gen = df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='A
2
3 sns.barplot(x = 'Gender',y= 'Amount' ,data = sales_gen)
4 plt.show()
```



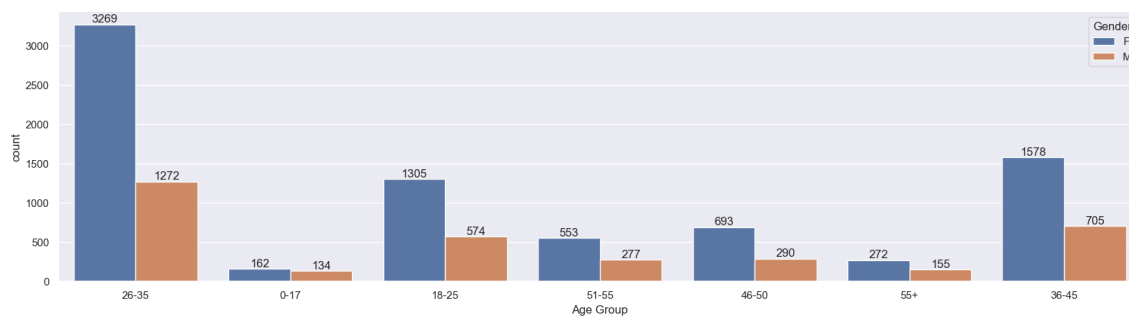
From above graphs we can see that most of the buyers are females and even the purchasing power of females are greater than men

Age



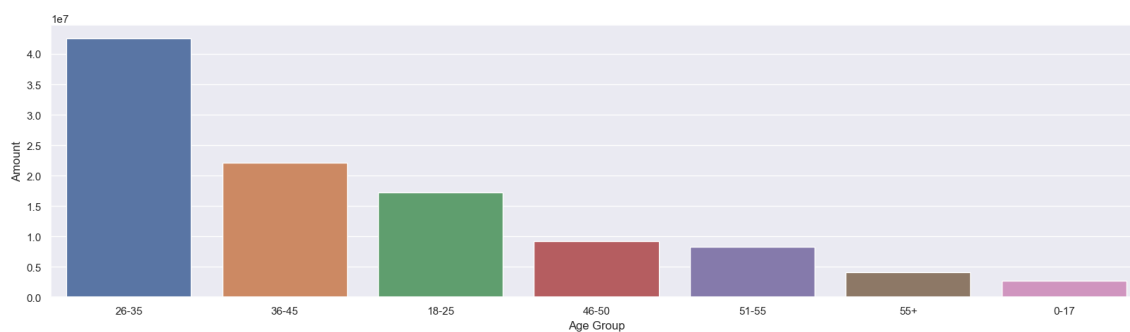
In [42]:

```
1 ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender')
2
3 for bars in ax.containers:
4     ax.bar_label(bars)
```



In [48]:

```
1 sales_age = df.groupby(['Age Group'], as_index=False)['Amount'].sum().sort_values(by
2
3 sns.barplot(x = 'Age Group', y = 'Amount', data = sales_age)
4 plt.show()
```



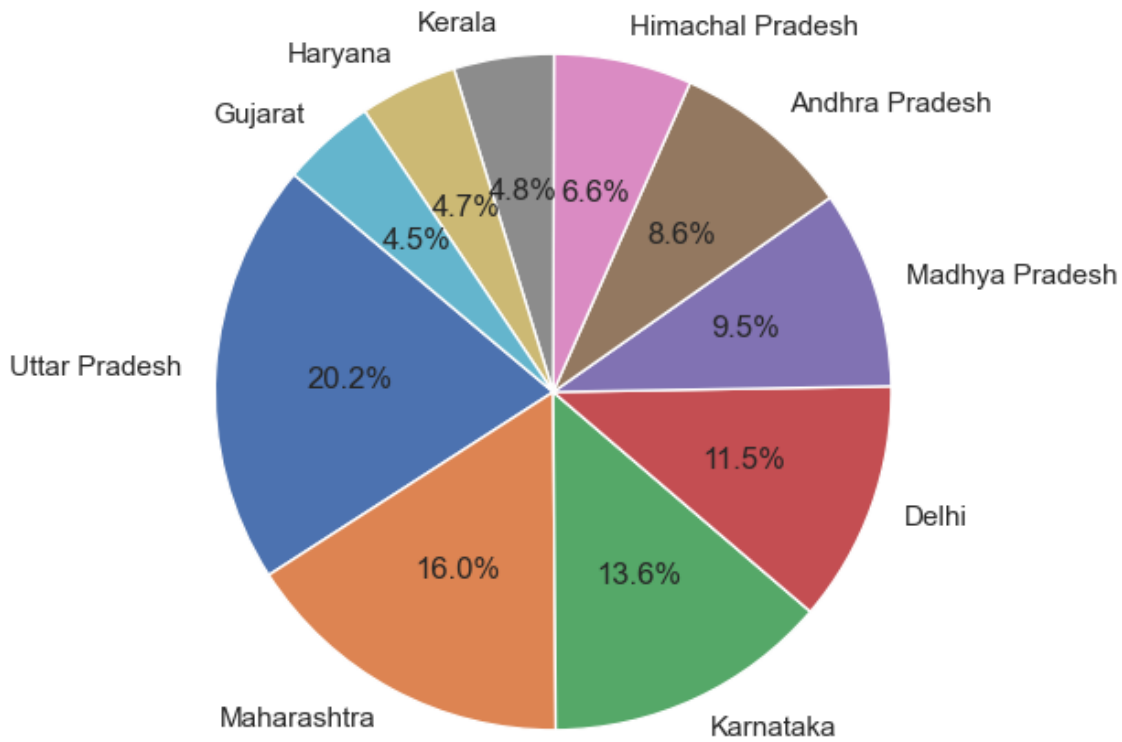
From above graphs we can see that most of the buyers are of age group between 26-35 yrs female

State



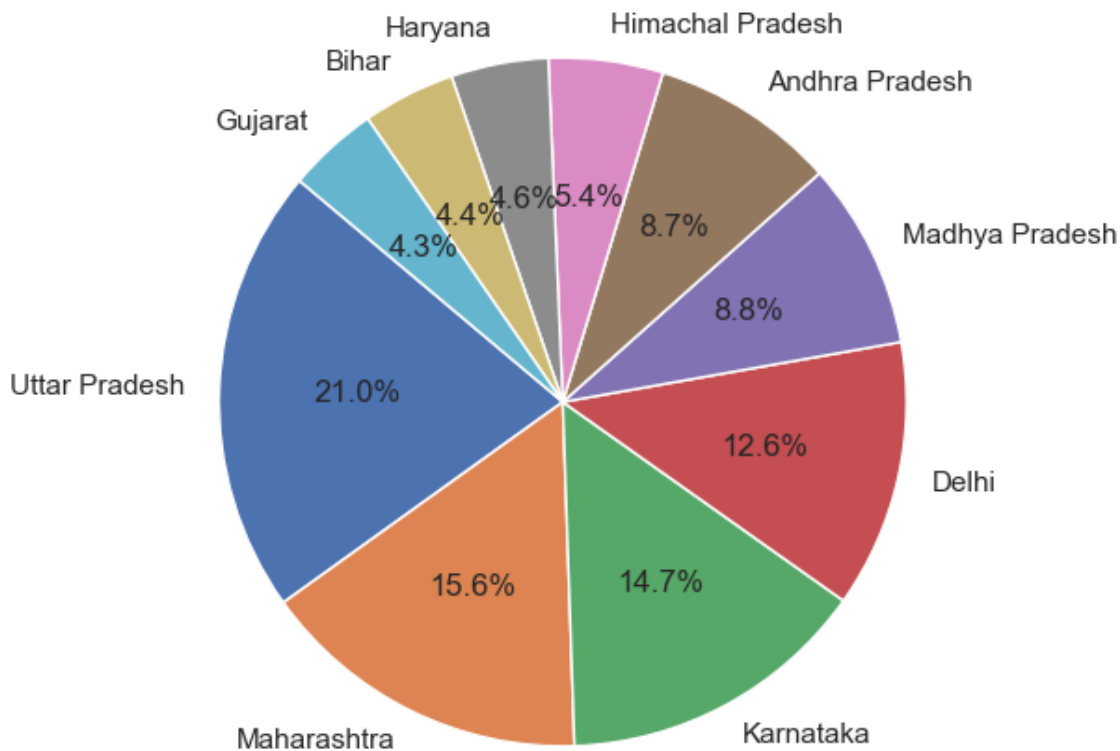
In [45]:

```
1 sales_state = df.groupby(['State'], as_index=False)['Orders'].sum().sort_values(by='Orders')
2
3 plt.figure(figsize=(10, 6))
4
5 plt.pie(sales_state['Orders'], labels=sales_state['State'], autopct='%1.1f%%', startangle=90)
6 plt.show()
```



In [46]:

```
1 sales_state = df.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount')
2
3 plt.figure(figsize=(10, 6))
4
5 plt.pie(sales_state['Amount'], labels=sales_state['State'], autopct='%1.1f%%', startangle=90)
6 plt.show()
```



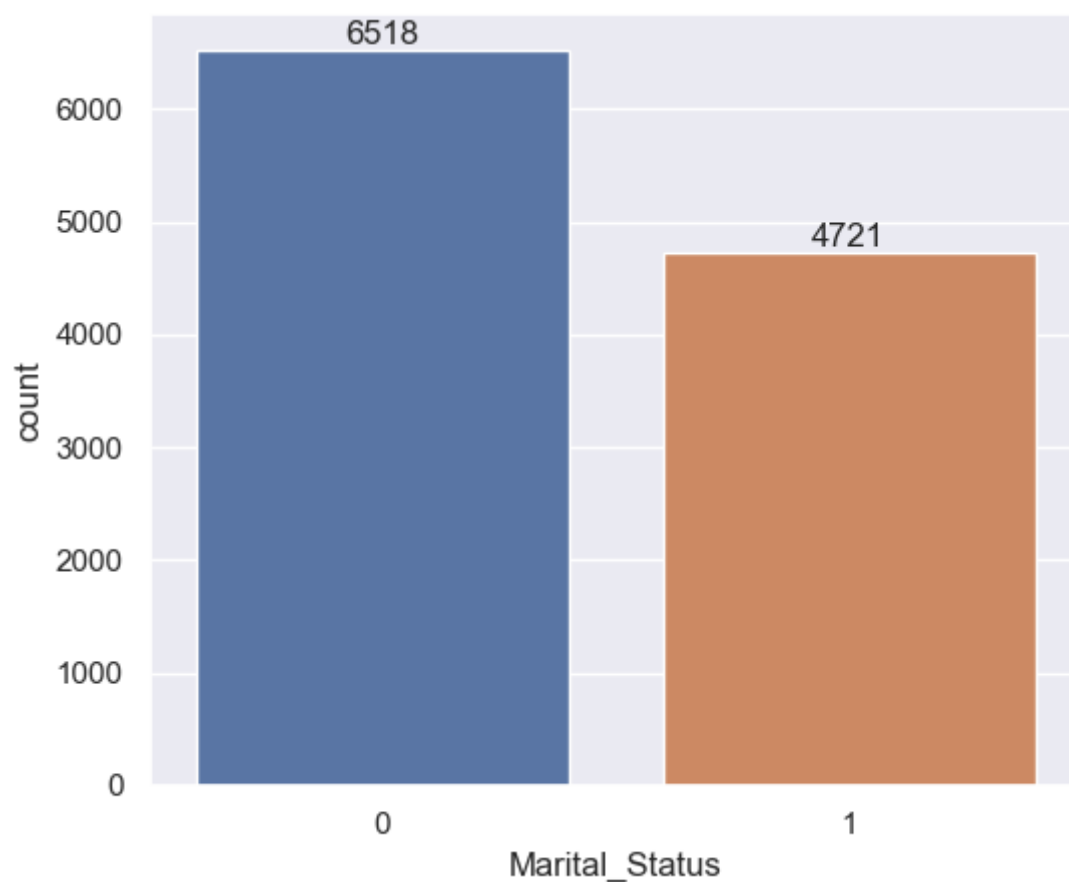
From above graphs we can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively

Marital Status



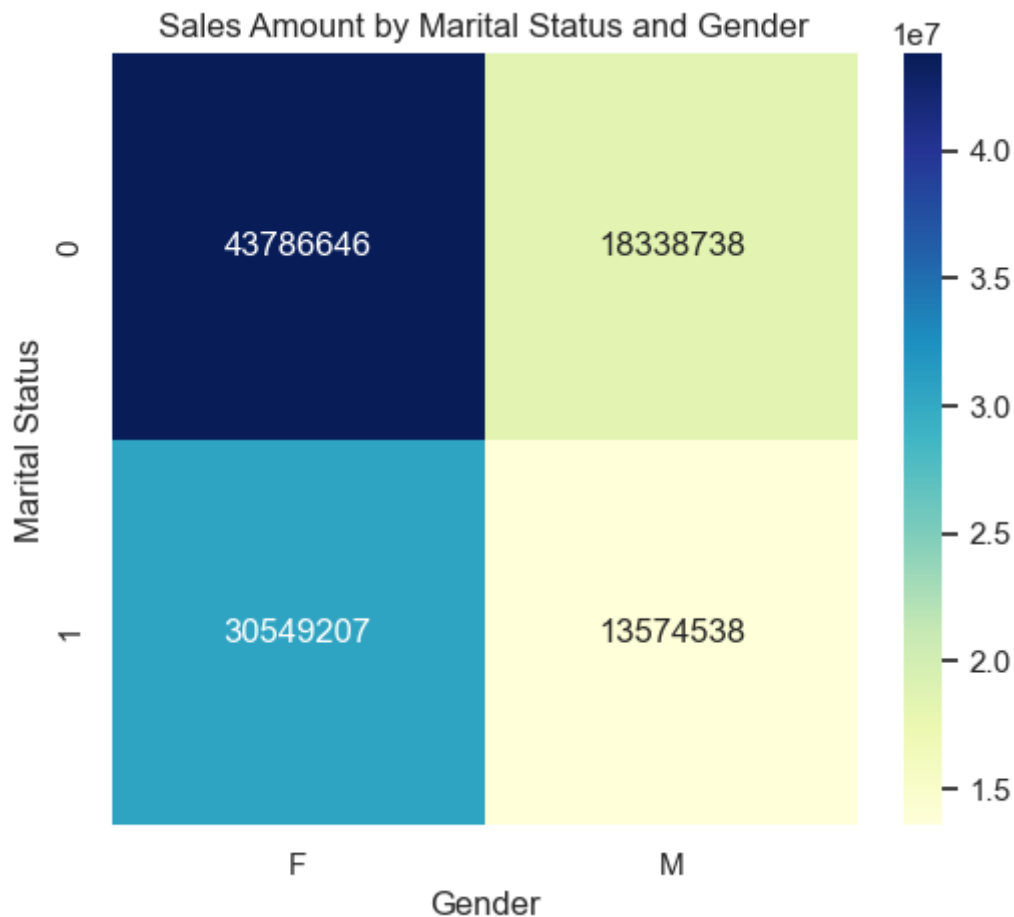
In [63]:

```
1 ax = sns.countplot(data = df, x = 'Marital_Status')
2
3 sns.set(rc={'figure.figsize':(7,5)})
4 for bars in ax.containers:
5     ax.bar_label(bars)
```



In [61]:

```
1 sales_state_pivot = df.pivot_table(index='Marital_Status', columns='Gender', values=
2
3 sns.set(rc={'figure.figsize': (6, 5)})
4 sns.heatmap(sales_state_pivot, annot=True, fmt='.0f', cmap='YlGnBu')
5
6 plt.xlabel('Gender')
7 plt.ylabel('Marital_Status')
8 plt.title('Sales Amount by Marital Status and Gender')
9
10 plt.show()
11
```



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

Occupation

teacher



businessman



waitress



doctor



cook



police



judge

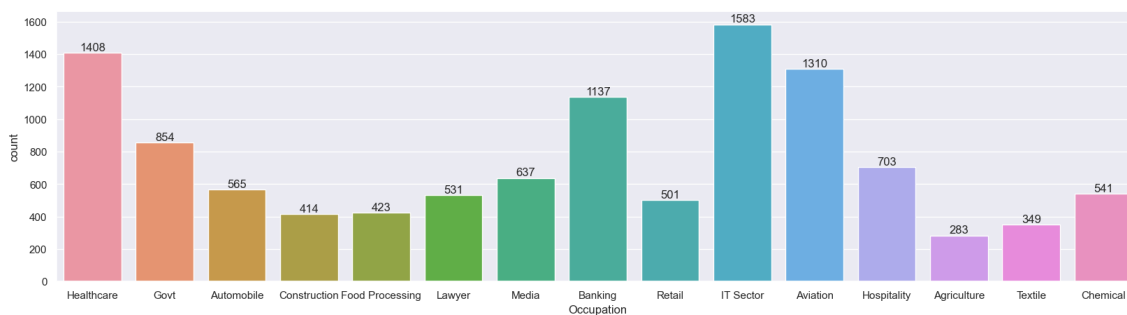


repairman



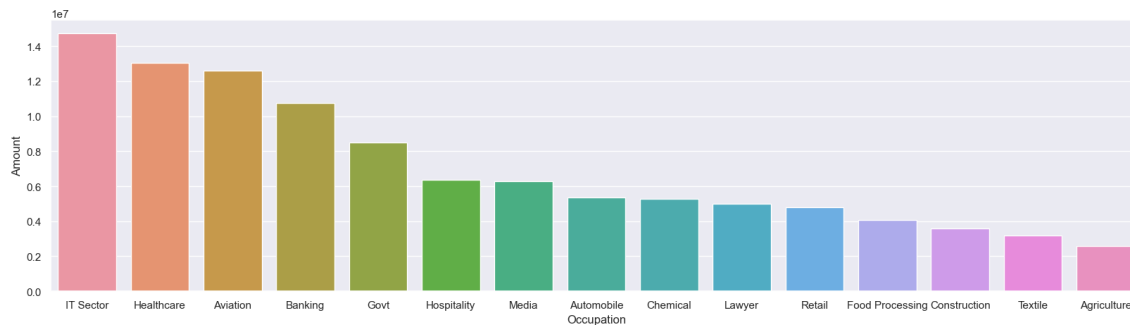
In [57]:

```
1 sns.set(rc={'figure.figsize':(20,5)})
2 ax = sns.countplot(data = df, x = 'Occupation')
3
4 for bars in ax.containers:
5     ax.bar_label(bars)
```



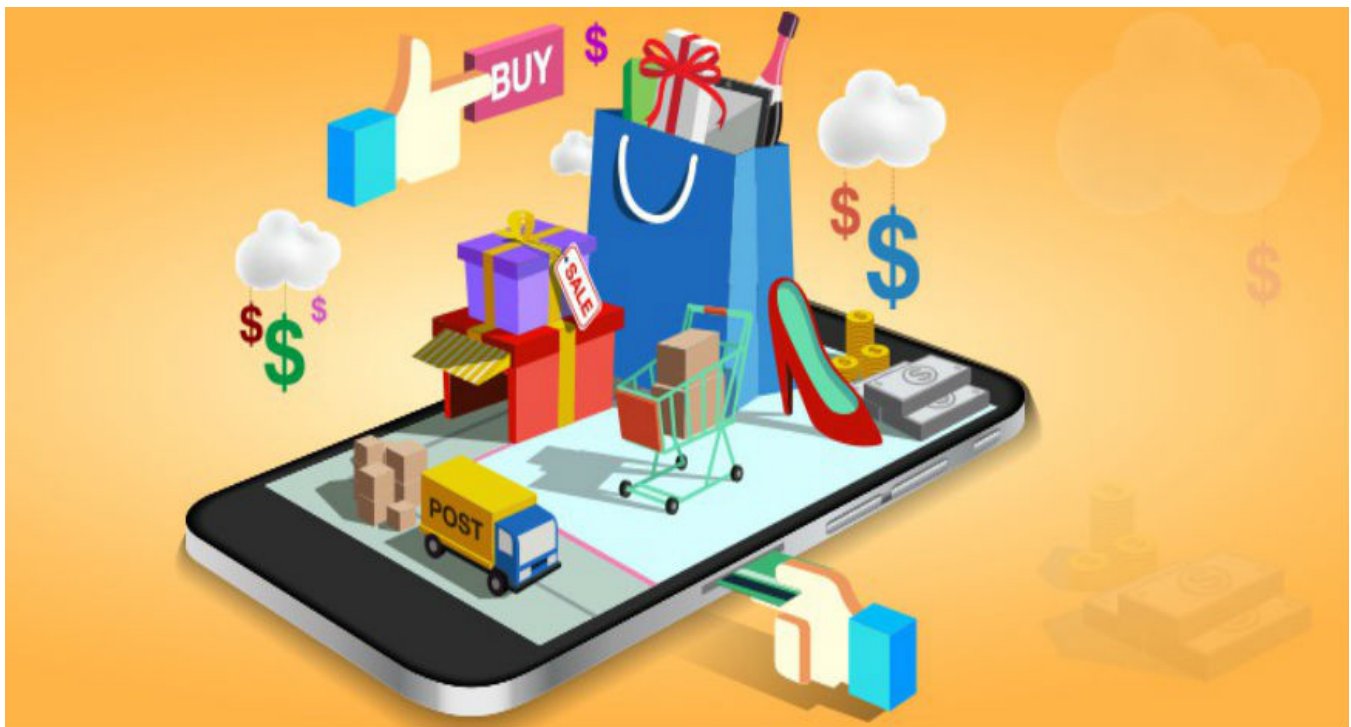
In [65]:

```
1 sales_state = df.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_values
2
3 sns.set(rc={'figure.figsize':(20,5)})
4 sns.barplot(data = sales_state, x = 'Occupation',y= 'Amount')
5 plt.show()
```



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

Product Category

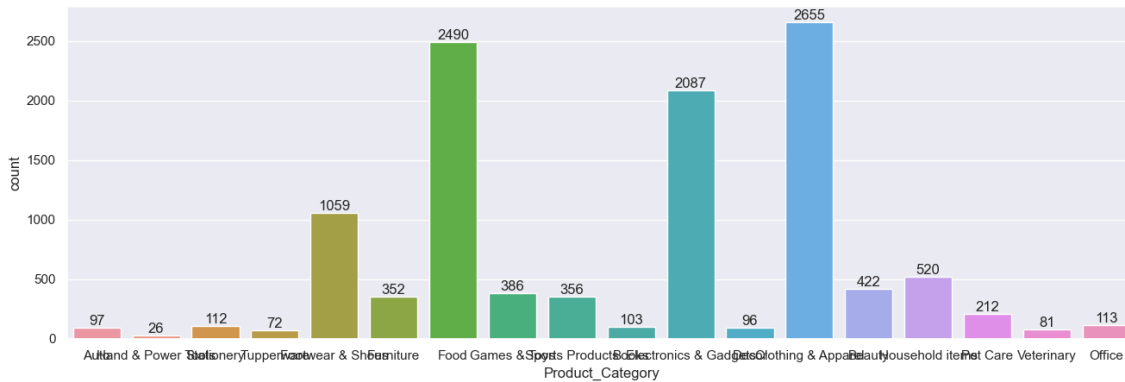


In [67]:

```

1 sns.set(rc={'figure.figsize':(16,5)})
2 ax = sns.countplot(data = df, x = 'Product_Category')
3
4 for bars in ax.containers:
5     ax.bar_label(bars)

```

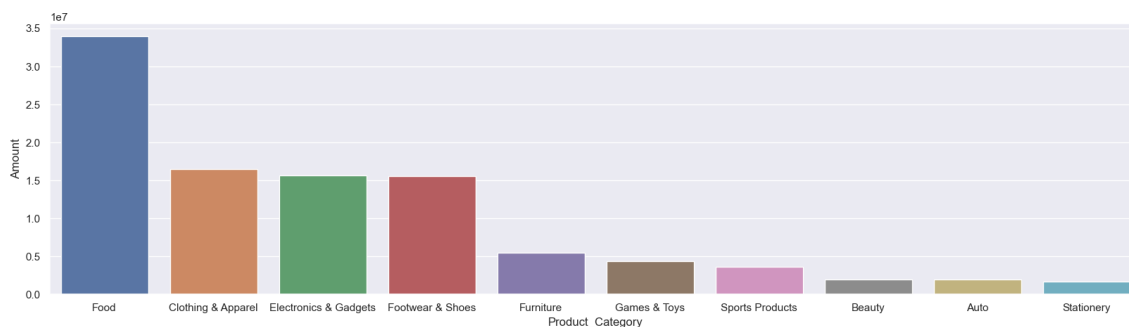


In [64]:

```

1 sales_state = df.groupby(['Product_Category'], as_index=False)['Amount'].sum().sort_
2
3 sns.set(rc={'figure.figsize':(20,5)})
4 sns.barplot(data = sales_state, x = 'Product_Category',y= 'Amount')
5 plt.show()

```

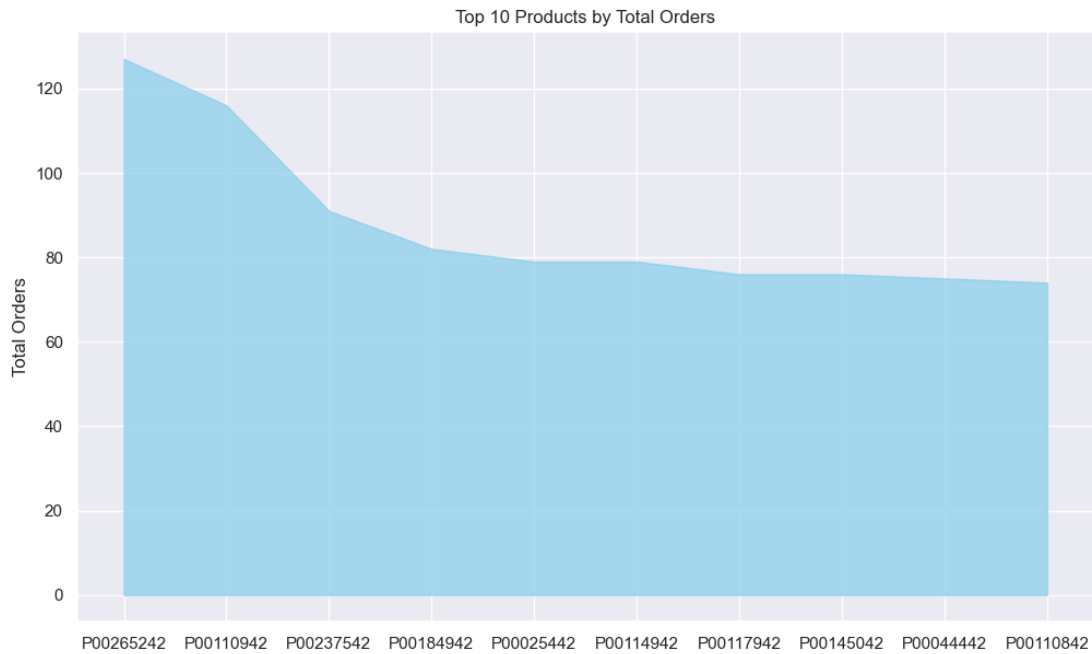


From above graphs we can see that most of the sold products are from Food, Clothing and Electronics category

Product ID

In [55]:

```
1 top_products = df.groupby('Product_ID')['Orders'].sum().nlargest(10).sort_values(asc
2
3 fig1, ax1 = plt.subplots(figsize=(12, 7))
4 ax1.fill_between(top_products.index, top_products.values, color='skyblue', alpha=0.7
5 ax1.set_xlabel('Product ID')
6 ax1.set_ylabel('Total Orders')
7 ax1.set_title('Top 10 Products by Total Orders')
8 plt.show()
```



4 Conclusion:

Married women age group 26-35 yrs from UP, Maharastra and Karnataka working in IT, Healthcare and Aviation are more likely to buy products from Food, Clothing and Electronics category

Thank you!