```
#Importing packages:
import numpy as np
import pandas as pd
import matplotlib as pyplot
import seaborn as sns


#Importing CSV file:
df=pd.read_csv("/content/Diwali Sales Data.csv",encoding='unicode_escape')
df
```

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	Sta
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharash
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Prade
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Prade
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnata
4	1000588	Joni	P00057942	M	26-35	28	1	Guja
...	...	...	...	...	...	...	...	
11246	1000695	Manning	P00296942	M	18-25	19	1	Maharash
11247	1004089	Reichenbach	P00171342	M	26-35	33	0	Harya
11248	1001209	Oshin	P00201342	F	36-45	40	0	Madh Prade
11249	1004023	Noonan	P00059442	M	36-45	37	0	Karnata
11250	1002744	Brumley	P00281742	F	18-25	19	0	Maharash

11251 rows × 15 columns

```
df.head()
```

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupat
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthc
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	G
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automo
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construct
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Fo Process

```
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
```

```
df.columns
```

```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount', 'Status', 'unnamed1'],
      dtype='object')
```

```
#dropping unnecessary columns:
df.drop(['Status','unnamed1'],axis=1,inplace=True)
```

```
#check for null values:
df.isna().sum()
```

```
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
```

```
df.shape
```

```
(11251, 13)
```

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

```
df.isna().sum()
```

```
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              0
dtype: int64
```

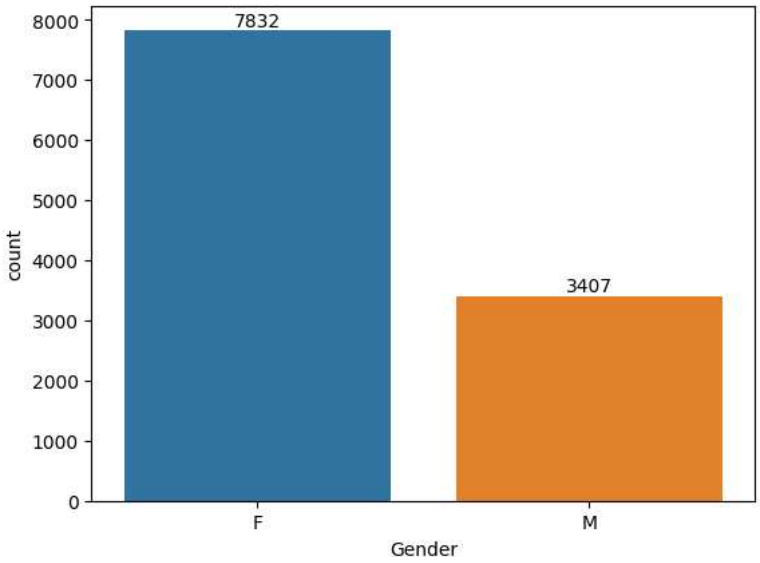
```
df['Amount']=df['Amount'].astype('int')
```

```
df['Amount'].dtype
```

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

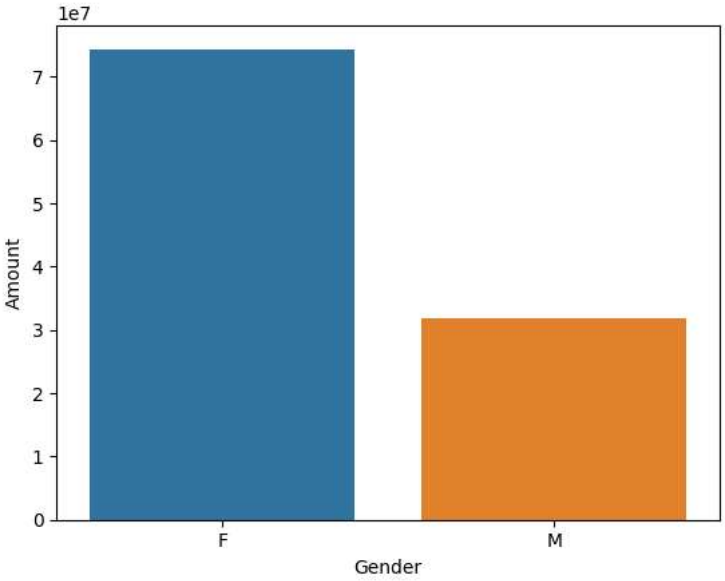
```
#Exploratory Data Analysis:
```

```
# plotting a bar chart for gender and its count:
ax=sns.countplot(x='Gender',data=df)
for bars in ax.containers:
    ax.bar_label(bars)
```




```
# 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)
```

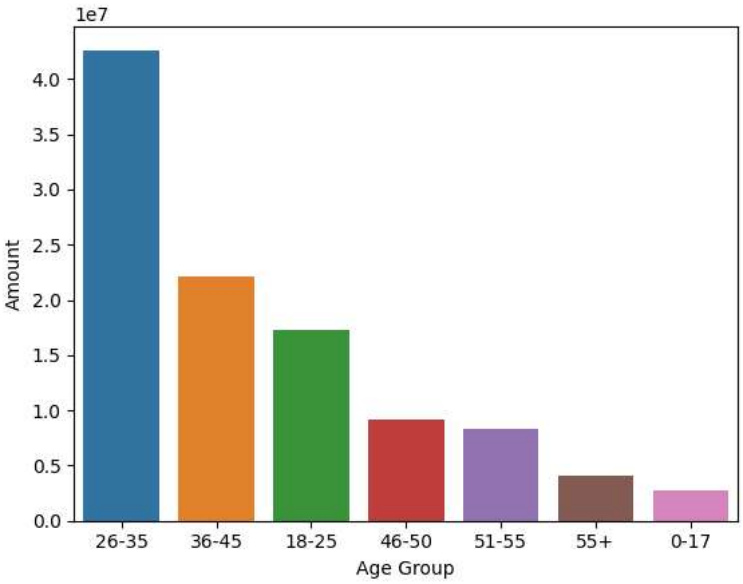
```
<Axes: xlabel='Gender', ylabel='Amount'>
```




```
#From above graphs we can see that most of the buyers are females


#Age Group:
#plotting graph between age groups and its count:
ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender')
for bars in ax.containers:
    ax.bar_label(bars)
```


```
# 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)
```

```
<Axes: xlabel='Age Group', ylabel='Amount'>
```




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

```
df.columns
```

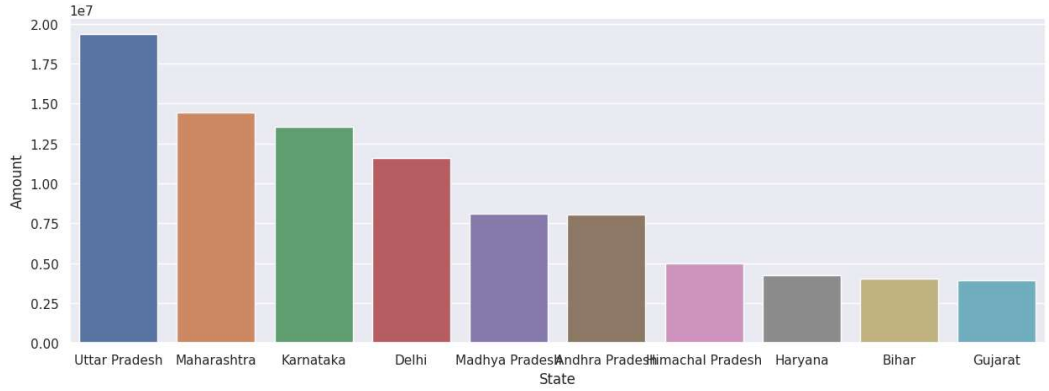
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')
```

```
#State:
# 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')
```

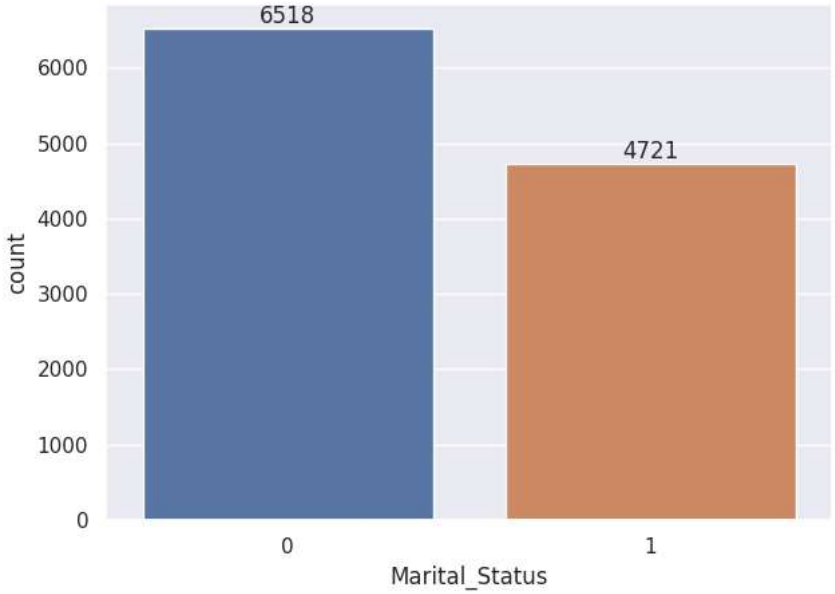
```
<Axes: xlabel='State', ylabel='Orders'>
```


```
# Total amount spend 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')
```

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

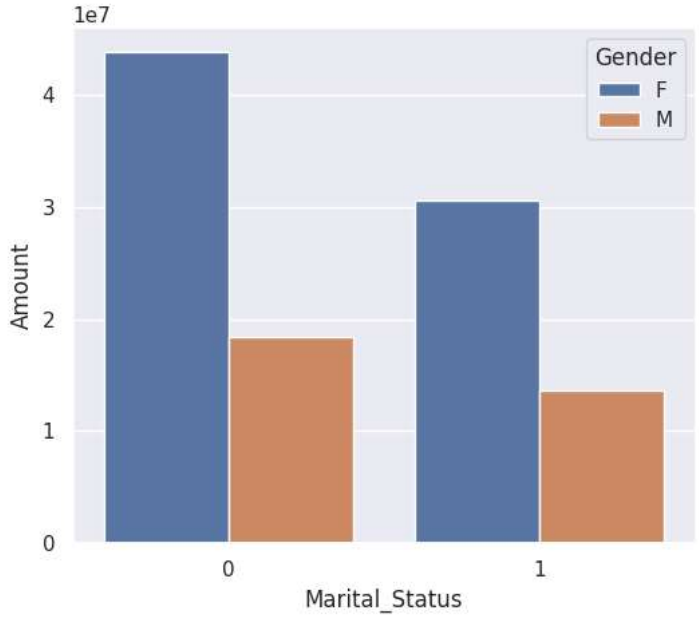



```
#From above graphs we can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively


#Martial Status:
#Plotting bar chart between martial status and total counts
ax = sns.countplot(x = 'Marital_Status',data=df)
sns.set(rc={'figure.figsize':(5,5)})
for bars in ax.containers:
    ax.bar_label(bars)
```

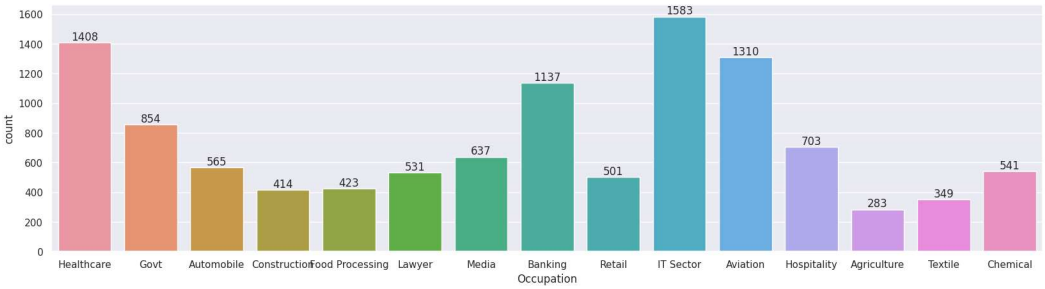



```
#graph between martial status and total amount spend
sales_status = df.groupby(['Marital_Status', 'Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.set(rc={'figure.figsize':(6,5)})
sns.barplot(data = sales_status, x = 'Marital_Status',y= 'Amount', hue='Gender')
```

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



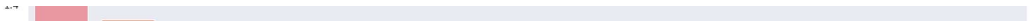

```
#From above graphs we can see that most of the buyers are married women


#Occupation:
#Plotting a Graph between Occupation and its count
sns.set(rc={'figure.figsize':(20,5)})
ax = sns.countplot(x = 'Occupation',data=df)
for bars in ax.containers:
    ax.bar_label(bars)
```



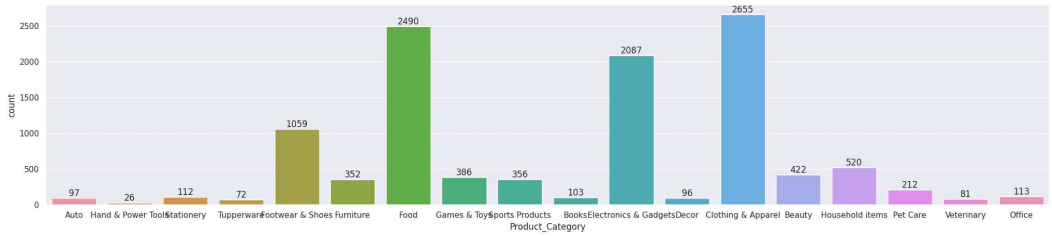

```
#Plotting a chart between Occupation and the total amount in each
sales_state = df.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Occupation',y= 'Amount')
```

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

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

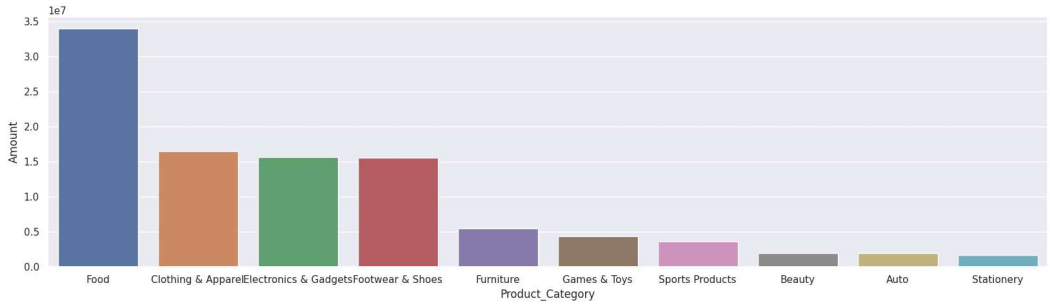


```
#Product Category:
#plotting a graph between product categories and its count:
sns.set(rc={'figure.figsize':(25,5)})
ax = sns.countplot(x = 'Product_Category',data=df)
for bars in ax.containers:
    ax.bar_label(bars)
```



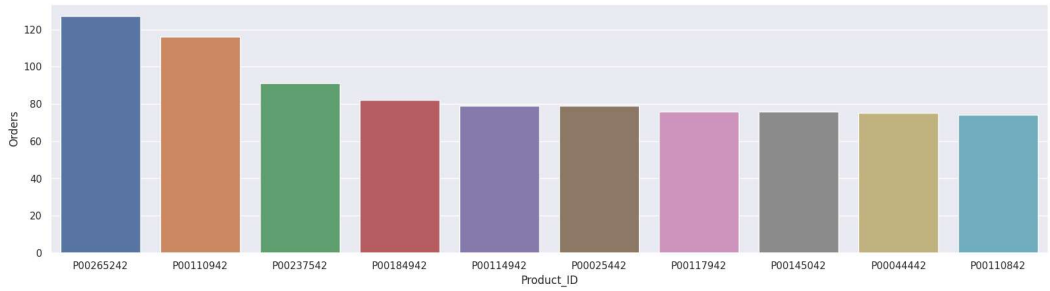

```
#Plotting a chart between product category and total amount:
sales_state = df.groupby(['Product_Category'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Product_Category',y= 'Amount')
```

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




```
#checking the top selling products based on product_id
sales_state = df.groupby(['Product_ID'], as_index=False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Product_ID',y= 'Orders')
```

<Axes: xlabel='Product_ID', ylabel='Orders'>



```
#Insights from EDA:
#The maximum sales are done by married women aged between 26-35 years from States UP, Maharastra and Karnataka working in IT, Healthcare and
#and are more likely to buy products from food,clothing and Electronics category.
```

✓ 0s completed at 4:29 PM