DIWALI SALES ANALYSIS

Importing Libraries & Data:

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

In [2]:
```
url = "https://raw.githubusercontent.com/MAliHasnain/Sales_Analysis/main/Dat
df = pd.read_csv(url,encoding='latin1')
df.head()
```

Out[2]:

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

Data Cleaning & Pre Processing:

In [3]:
```
df.shape
```

Out[3]:
```
(11251, 15)
```

In [4]:
```
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 [5]:
```
df.drop({"Status",'unnamed1'},axis = 1,inplace=True)
```

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

Out[6]:

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

In [7]:
```
df.isna().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]:
```
df.dropna(inplace = True)
```

In [9]:
```
df.shape
```

Out[9]:
```
(11239, 13)
```

In [10]:
```
df["Amount"] = df["Amount"].astype("int")
```

In [11]:
```
df["Amount"].dtype
```

Out[11]:
```
dtype('int32')
```

In [12]:
```
df[["Age","Orders","Amount"]].describe().astype('int')
```

Out[12]:

	Age	Orders	Amount
count	11239	11239	11239
mean	35	2	9453
std	12	1	5222
min	12	1	188
25%	27	2	5443
50%	33	2	8109
75%	43	3	12675
max	92	4	23952

Exploratory Data Analysis:

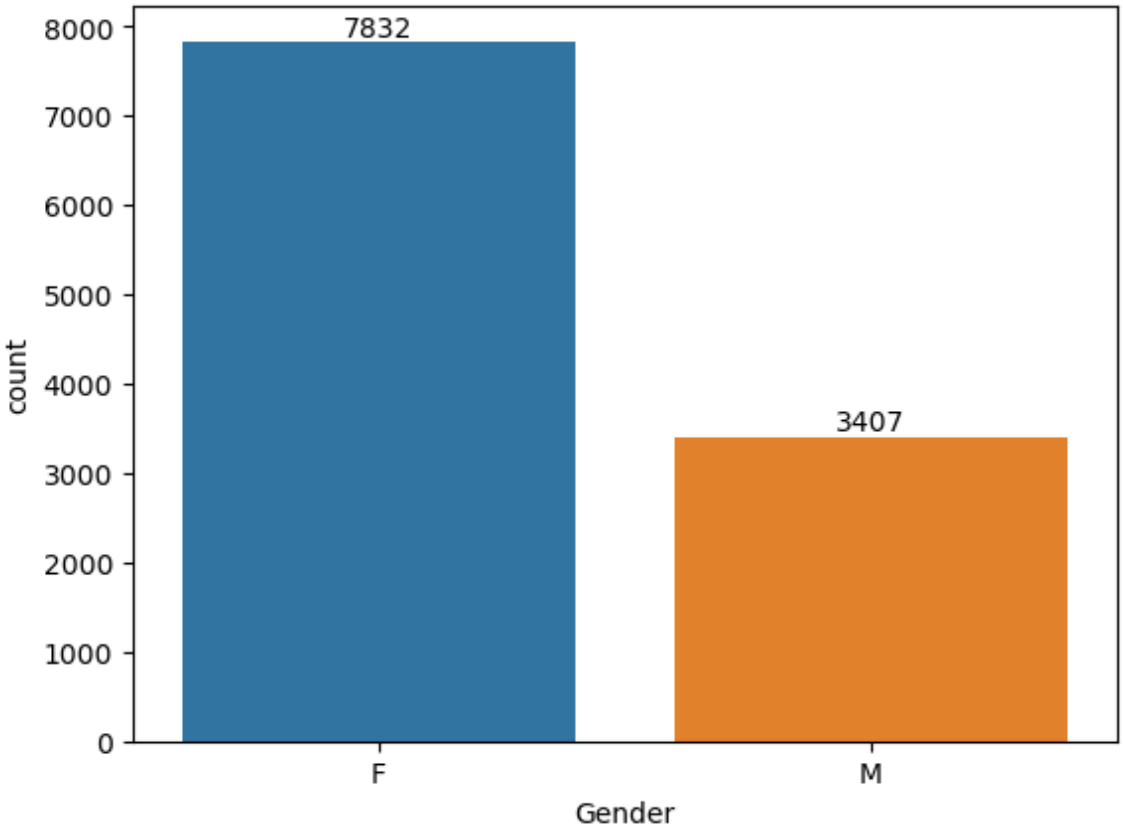
In [13]:
```
df.columns
```

Out[13]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Categor
y',
       'Orders', 'Amount'],
      dtype='object')
```

Gender:

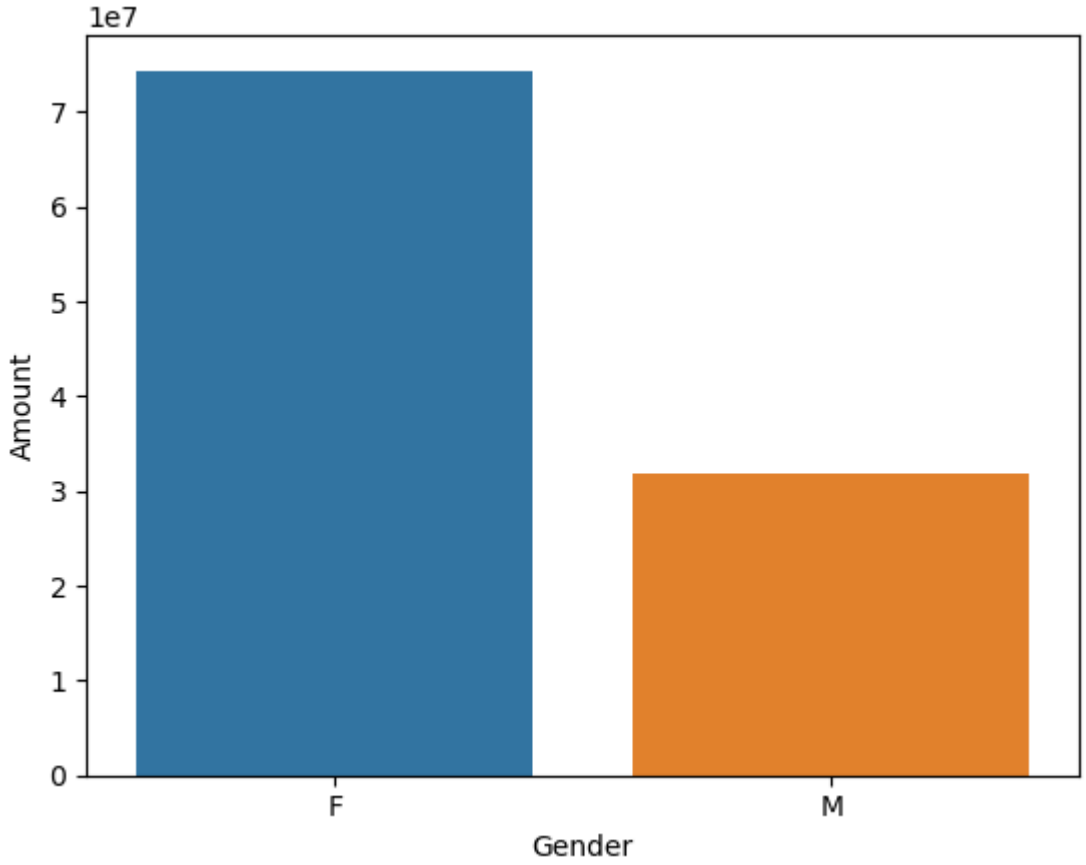
In [14]:
```
ax = sns.countplot(x = "Gender",data=df)

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



In [15]:
```
df1 = df.groupby(["Gender"], as_index=False)["Amount"].sum().sort_values(by=
sns.barplot(x = "Gender", y = "Amount",data = df1)
```

Out[15]:
```
<Axes: xlabel='Gender', ylabel='Amount'>
```

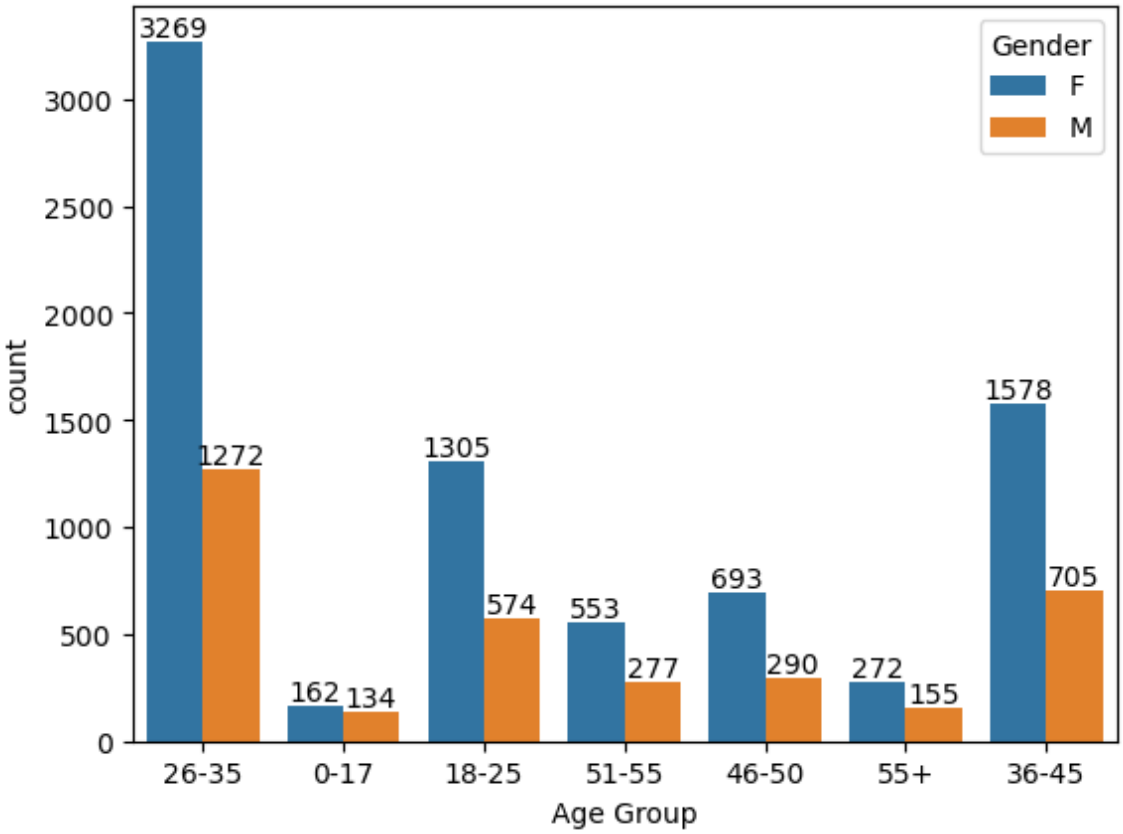


From above visualization we can extract that Females are most buyers and their purchasing power is also greater than man.

Age Group:

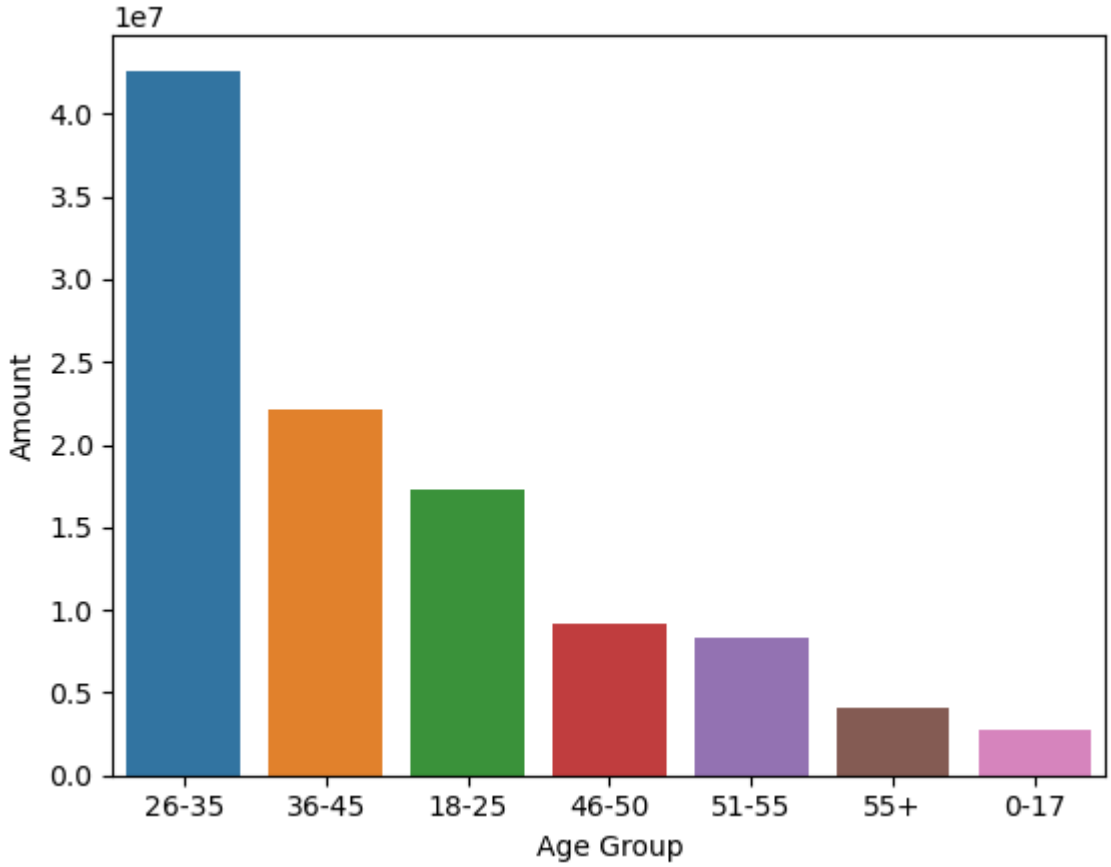
In [16]:
```
ax1 = sns.countplot(x = "Age Group",data=df,hue = "Gender")

for bars in ax1.containers:
    ax1.bar_label(bars)
```



In [17]:
```
df2 = df.groupby(["Age Group"], as_index=False)["Amount"].sum().sort_values(
sns.barplot(x = "Age Group", y = "Amount",data = df2)
```

Out[17]:
```
<Axes: xlabel='Age Group', ylabel='Amount'>
```



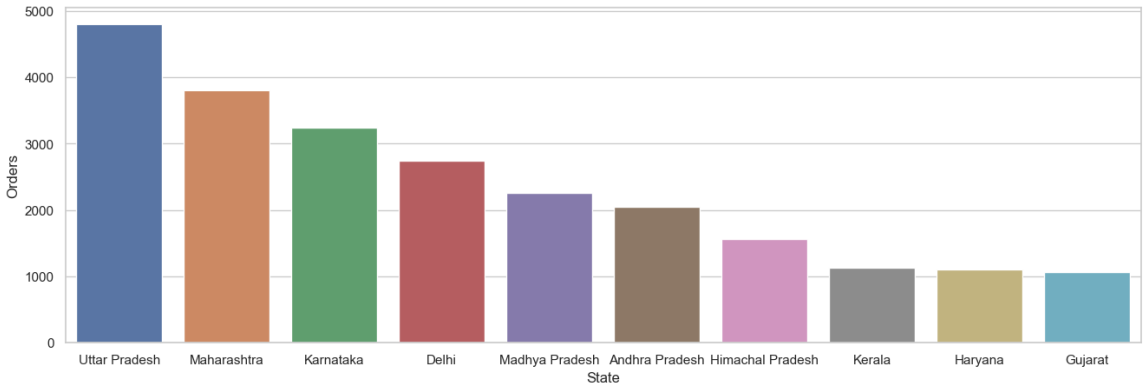
From Above visualization we can infere that the Female buyers of age group 26-35 are most buyers.

State:

In [18]:
```
df3 = df.groupby(["State"], as_index=False)["Orders"].sum().sort_values(by='

sns.set(style="whitegrid")
plt.figure(figsize=(16, 5))
sns.barplot(x = "State", y = "Orders",data = df3)
```

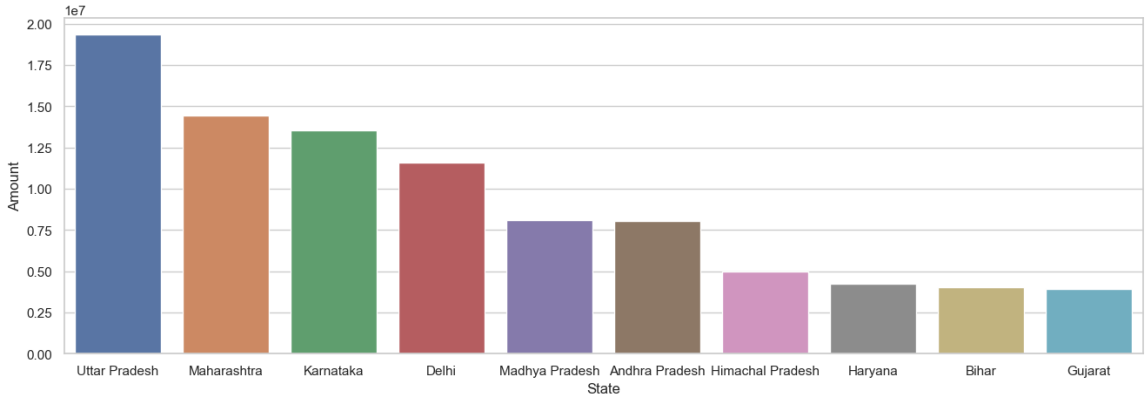
Out[18]:
```
<Axes: xlabel='State', ylabel='Orders'>
```



In [19]:
```
df4 = df.groupby(["State"], as_index=False)["Amount"].sum().sort_values(by='

sns.set(style="whitegrid")
plt.figure(figsize=(16, 5))
sns.barplot(x = "State", y = "Amount",data = df4)
```

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

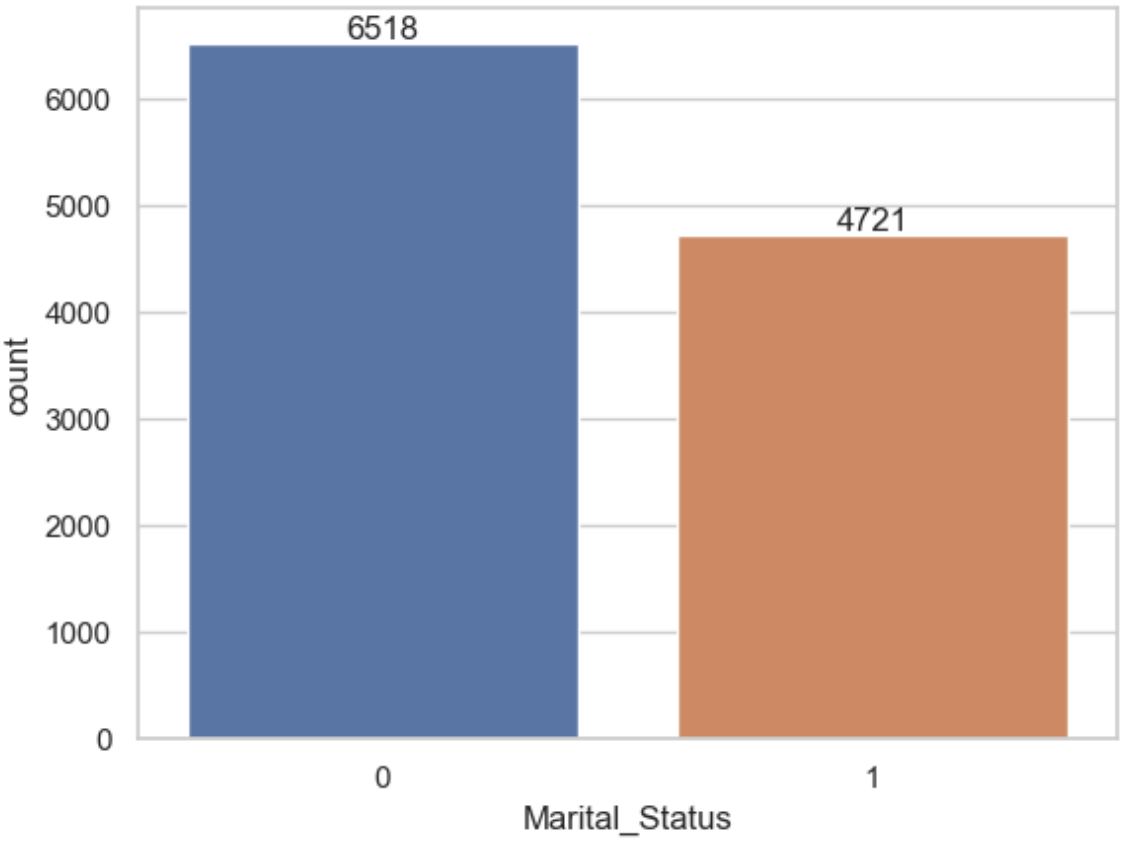


Above Visualizations shows that Uttar Pradesh State, Maharashtra, Karnatka have more orders and Buying power respectively.

Marital_Status:

In [20]:
```
ax3 = sns.countplot(x = "Marital_Status",data=df)

plt.figure(figsize=(7, 5))
for bars in ax3.containers:
    ax3.bar_label(bars)
```

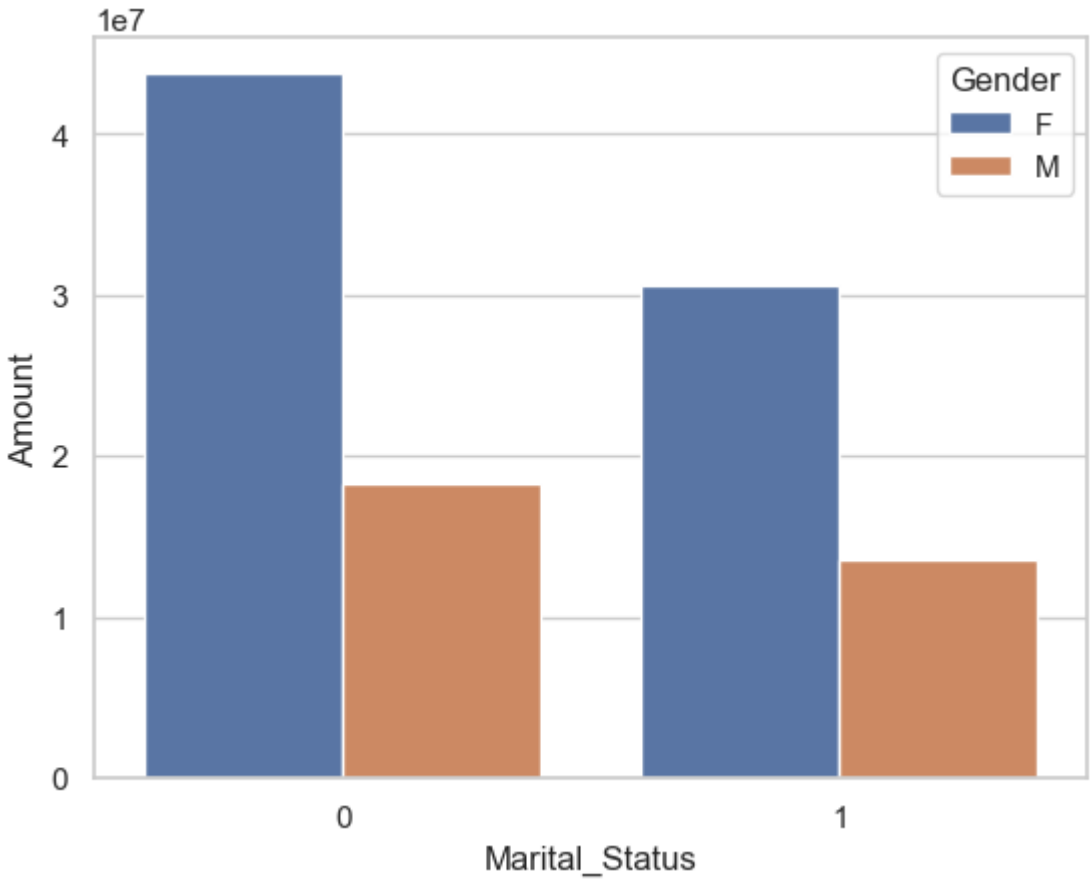



```
<Figure size 700x500 with 0 Axes>
```

In [21]:
```
df5 = df.groupby(["Marital_Status","Gender"], as_index=False)["Amount"].sum(

# sns.set(style="whitegrid")
# plt.figure(figsize=(16, 5))
sns.barplot(x = "Marital_Status", y = "Amount",hue = "Gender",data = df5)
```

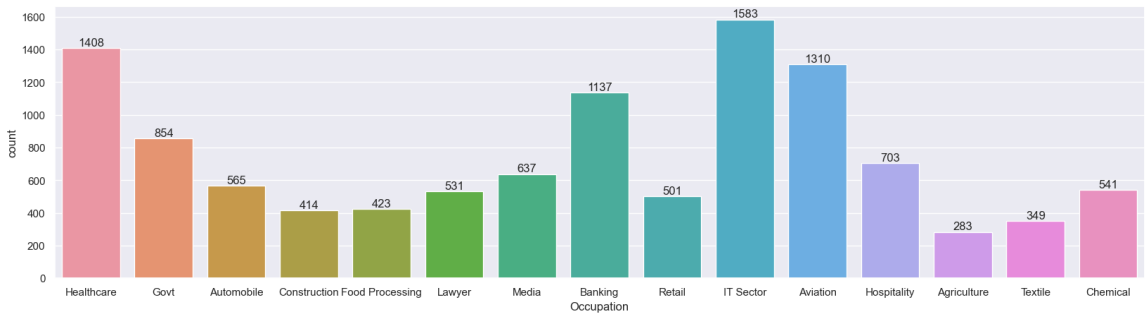
Out[21]: <Axes: xlabel='Marital_Status', ylabel='Amount'>



Above Visualization infere that most Married womens are the buyers in diwali sales.

Occupation:

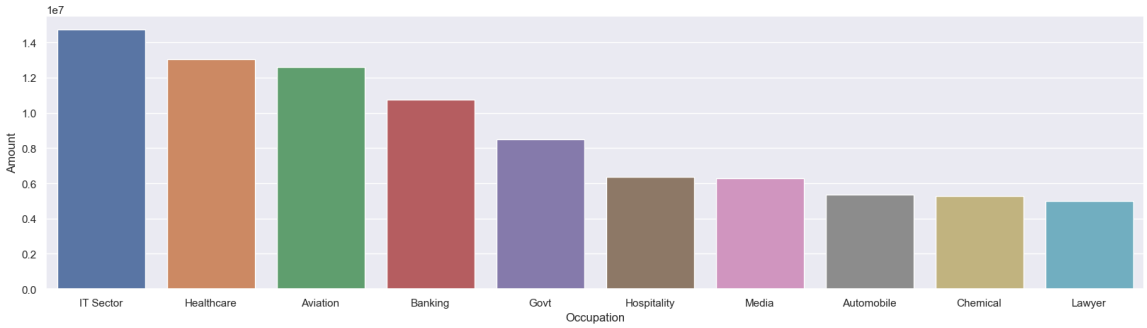
In [22]:
```
sns.set(rc={"figure.figsize":(20,5)})
ax4 = sns.countplot(x = "Occupation",data=df)
for bars in ax4.containers:
    ax4.bar_label(bars)
```



In [23]:
```
df6 = df.groupby(["Occupation"], as_index=False)["Amount"].sum().sort_values

sns.barplot(x = "Occupation", y = "Amount",data = df6)
```

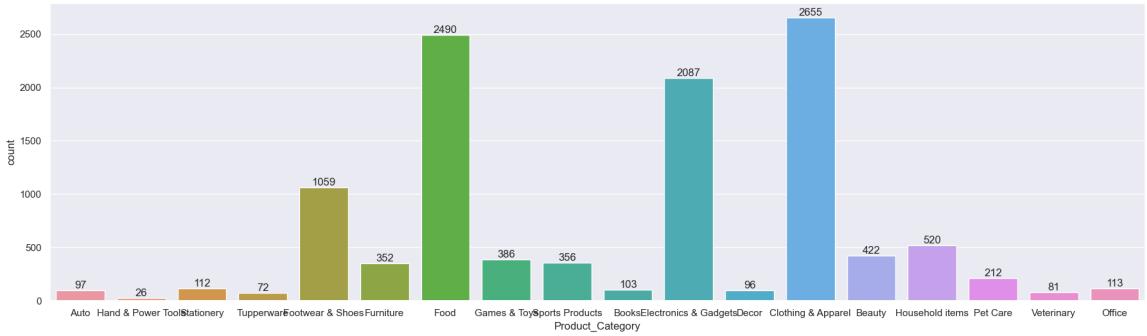
Out[23]: <Axes: xlabel='Occupation', ylabel='Amount'>



This above visualization shows that the IT sector, Health Care and Aviation employees have more diwali shopping respectively than other occupation people.

Product Category:

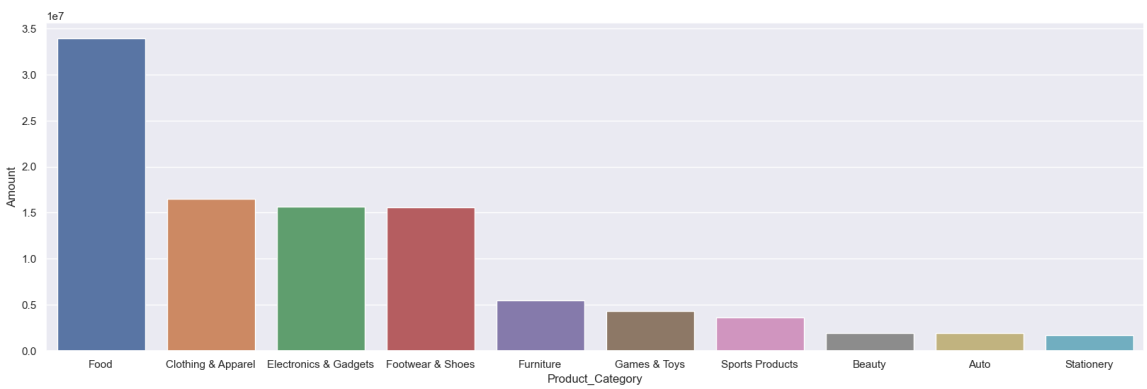
In [24]:
```
plt.figure(figsize=(22, 6))
ax4 = sns.countplot(x = "Product_Category",data=df)
for bars in ax4.containers:
    ax4.bar_label(bars)
```



In [25]:
```
plt.figure(figsize=(20, 6))
df7 = df.groupby(["Product_Category"], as_index=False)["Amount"].sum().sort_

sns.barplot(x = "Product_Category", y = "Amount",data = df7)
```

Out[25]: <Axes: xlabel='Product_Category', ylabel='Amount'>



The most product item purchased in diwali are Food items.