

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

```
In [2]: df = pd.read_csv(r'C:\Users\DEHI MARKAZ MALL\Desktop\Jupyter Notebook\Python_Diwali_Sa
```

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

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

```
In [4]: df.head(10)
```

```
Out[4]:
```

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	C
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	
5	1000588	Joni	P00057942	M	26-35	28	1	Himachal Pradesh	Northern	
6	1001132	Balk	P00018042	F	18-25	25	1	Uttar Pradesh	Central	
7	1002092	Shivangi	P00273442	F	55+	61	0	Maharashtra	Western	
8	1003224	Kushal	P00205642	M	26-35	35	0	Uttar Pradesh	Central	
9	1003650	Ginny	P00031142	F	26-35	26	1	Andhra Pradesh	Southern	

```
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
```

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

```
In [7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 13 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
dtypes: float64(1), int64(4), object(8)
memory usage: 1.1+ MB
```

```
In [8]: pd.isnull(df)
```

Out[8]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...
11246	False	False	False	False	False	False	False	False	False	False
11247	False	False	False	False	False	False	False	False	False	False
11248	False	False	False	False	False	False	False	False	False	False
11249	False	False	False	False	False	False	False	False	False	False
11250	False	False	False	False	False	False	False	False	False	False

11251 rows × 13 columns

In [9]: `pd.isnull(df).sum()`

Out[9]:

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 [10]: `df.dropna(inplace=True)`In [11]: `df.shape`

Out[11]: (11239, 13)

In [12]: `pd.isnull(df).sum()`

```
Out[12]: 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
```

```
In [13]: df['Amount'] = df['Amount'].astype('int')
```

```
In [14]: df['Amount'].dtypes
```

```
Out[14]: dtype('int32')
```

```
In [15]: df[['Age', 'Orders', 'Amount']].describe()
```

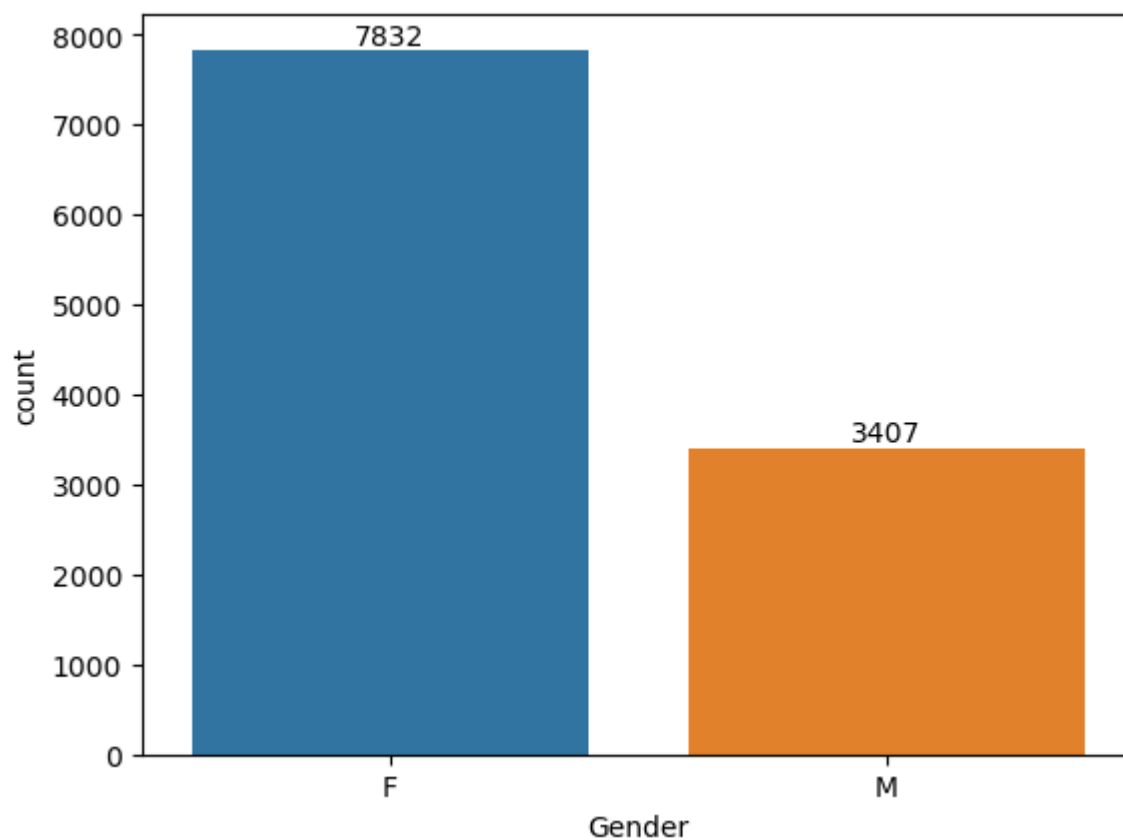
```
Out[15]:
```

	Age	Orders	Amount
count	11239.000000	11239.000000	11239.000000
mean	35.410357	2.489634	9453.610553
std	12.753866	1.114967	5222.355168
min	12.000000	1.000000	188.000000
25%	27.000000	2.000000	5443.000000
50%	33.000000	2.000000	8109.000000
75%	43.000000	3.000000	12675.000000
max	92.000000	4.000000	23952.000000

```
In [16]: df.columns
```

```
Out[16]: Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
              'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
              'Orders', 'Amount'],
              dtype='object')
```

```
In [17]: ax=sns.countplot(x = 'Gender', data = df)
         for bars in ax.containers:
             ax.bar_label(bars)
```



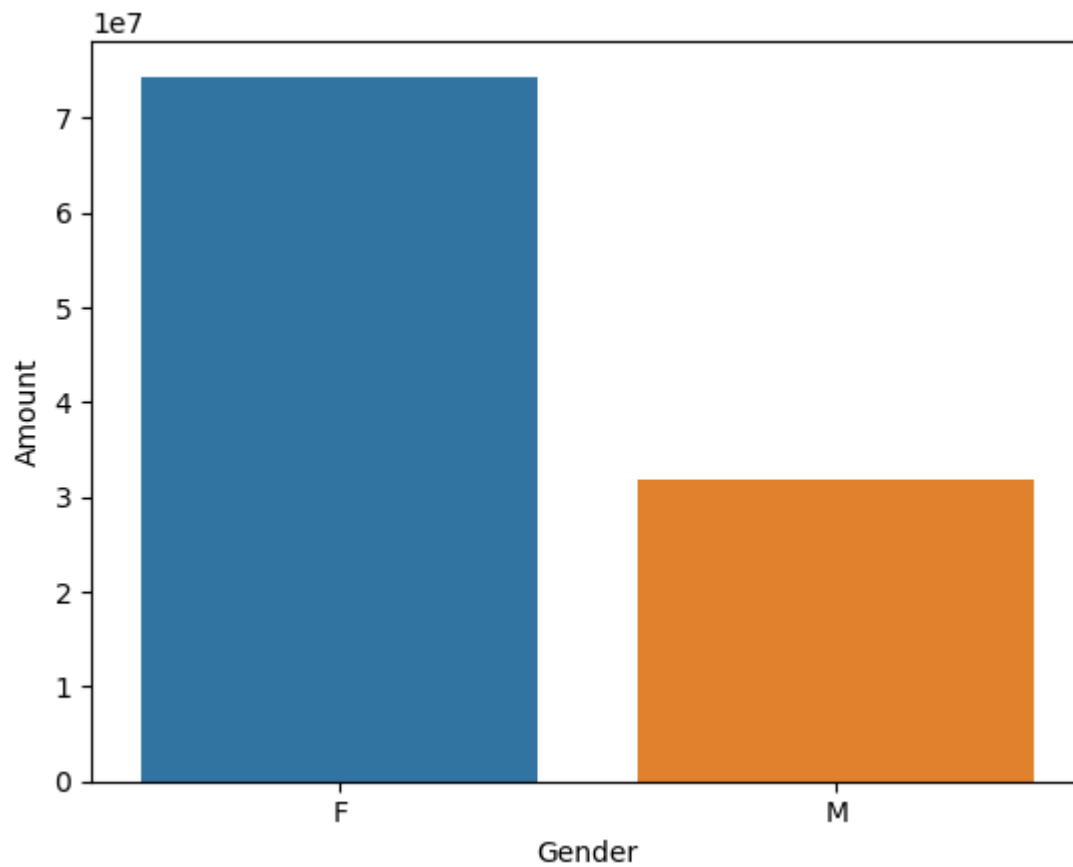
```
In [18]: df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=True)
```

```
Out[18]:
```

	Gender	Amount
0	F	74335853
1	M	31913276

```
In [19]: sales_gen = df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=True)
sns.barplot(x = 'Gender' ,y= 'Amount' ,data=sales_gen)
```

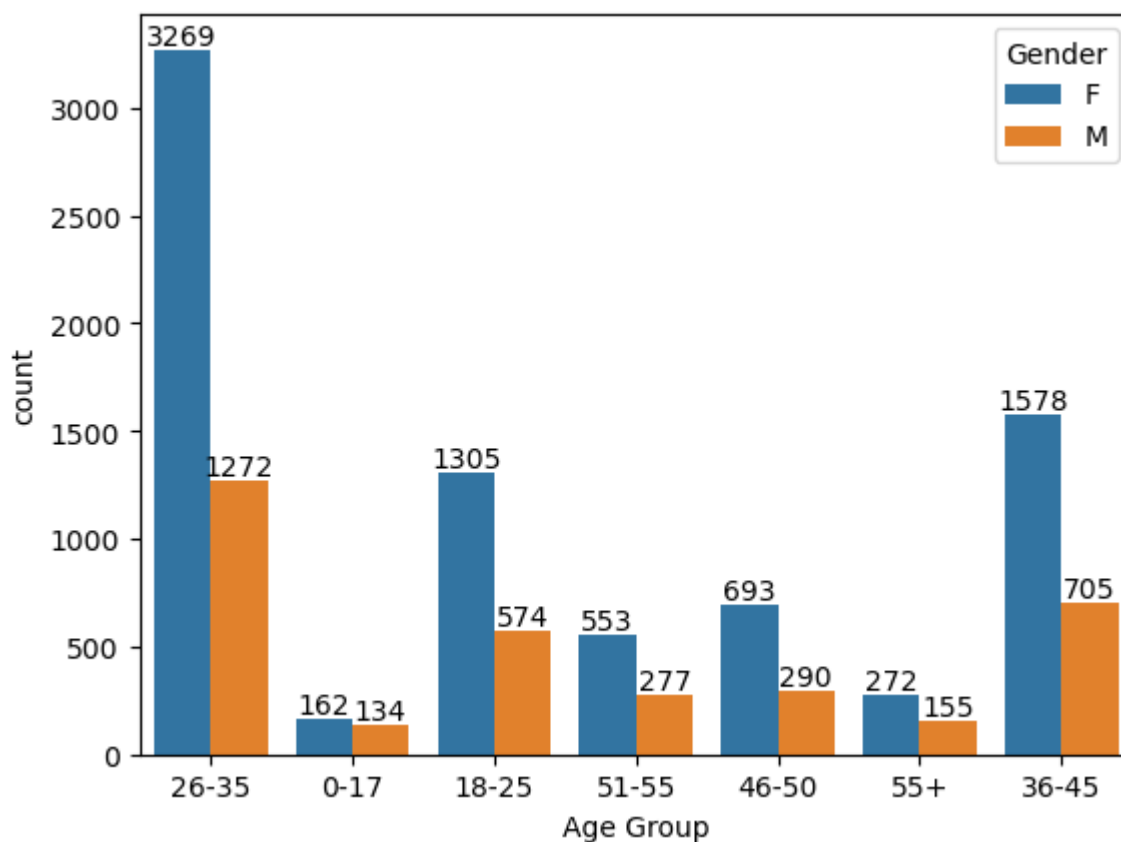
```
Out[19]: <Axes: xlabel='Gender', ylabel='Amount'>
```



```
In [20]: df.columns
```

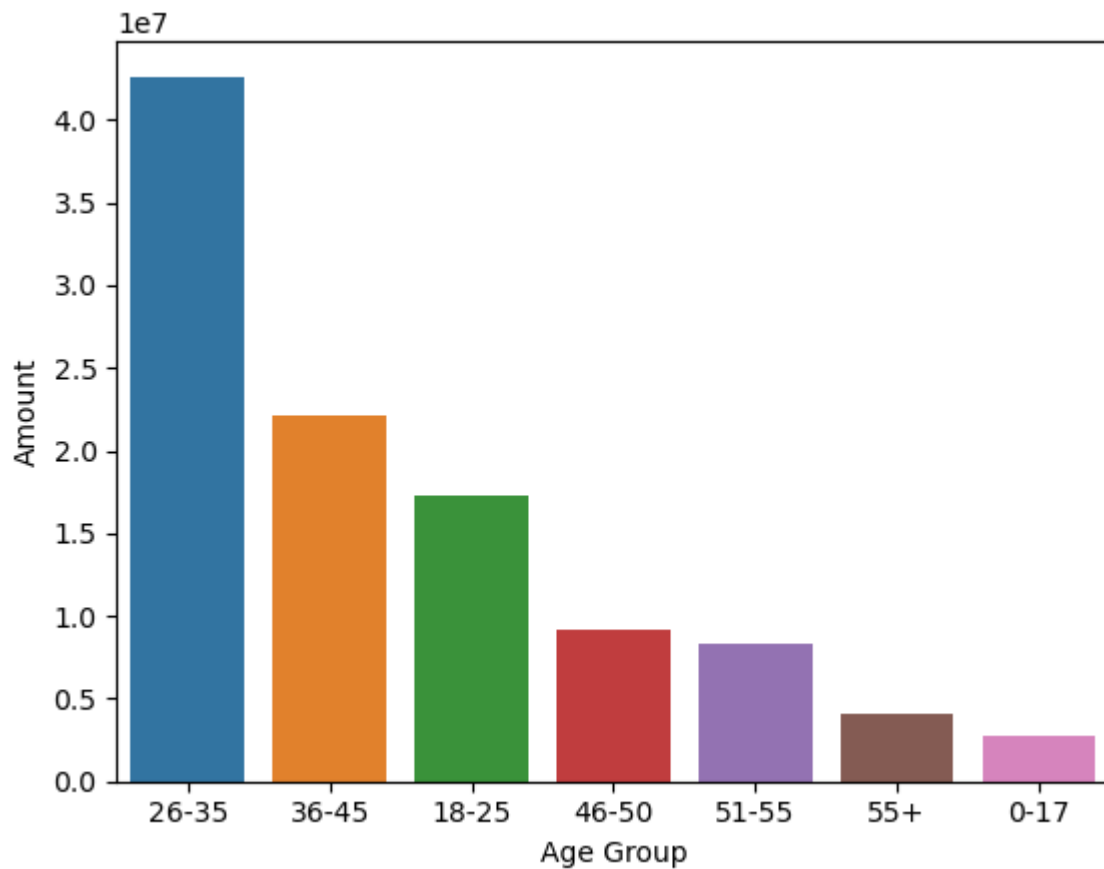
```
Out[20]: Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',  
            'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',  
            'Orders', 'Amount'],  
            dtype='object')
```

```
In [21]: ax = sns.countplot(data=df, x='Age Group', hue='Gender')  
         for bars in ax.containers:  
             ax.bar_label(bars)
```



```
In [22]: sales_age = df.groupby('Age Group')['Amount'].sum().reset_index().sort_values(by='Amount')
sns.barplot(x='Age Group', y='Amount', data=sales_age)
```

```
Out[22]: <Axes: xlabel='Age Group', ylabel='Amount'>
```

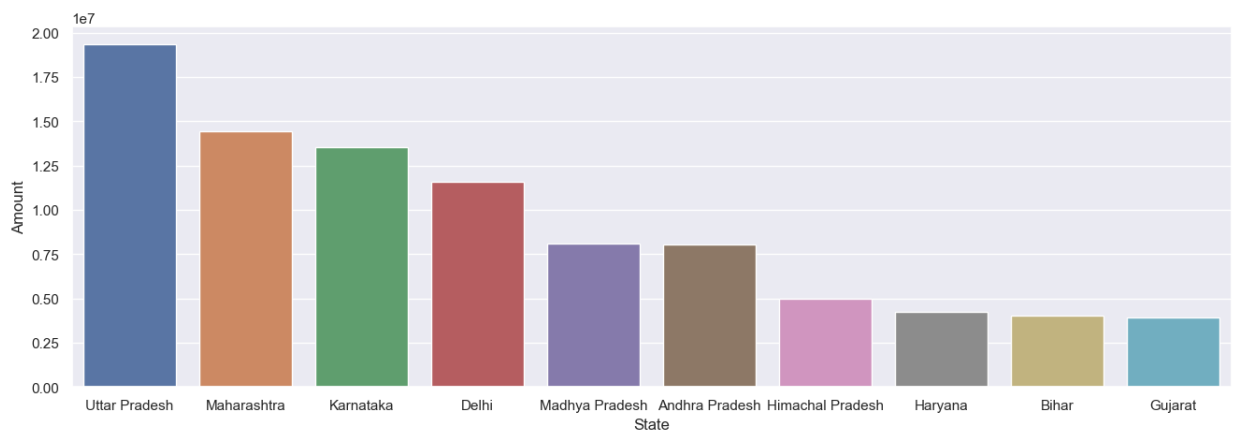


In [23]: `df.columns`

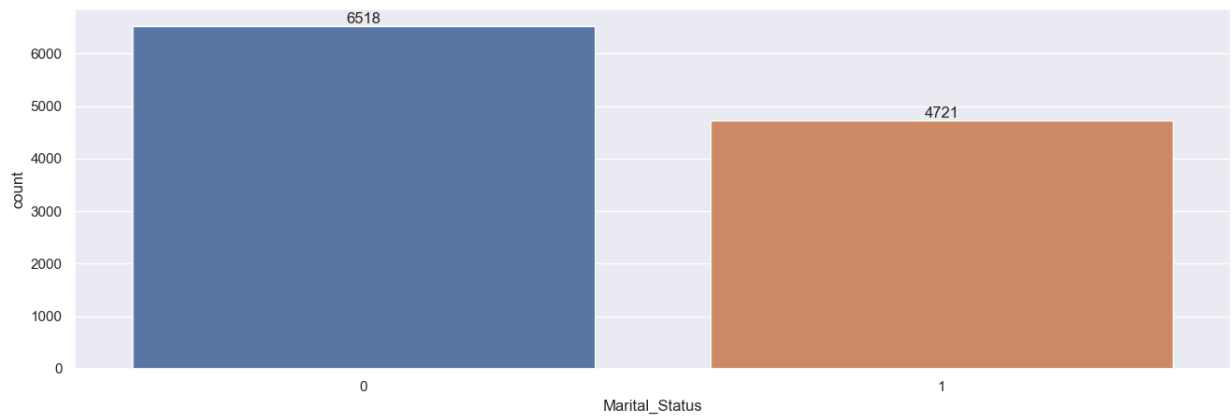
Out[23]: Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age', 'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category', 'Orders', 'Amount'], dtype='object')

In [24]: `sales_state = df.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)`
`sns.set(rc={'figure.figsize':(16,5)})`
`sns.barplot(data = sales_state, x = 'State', y='Amount')`

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



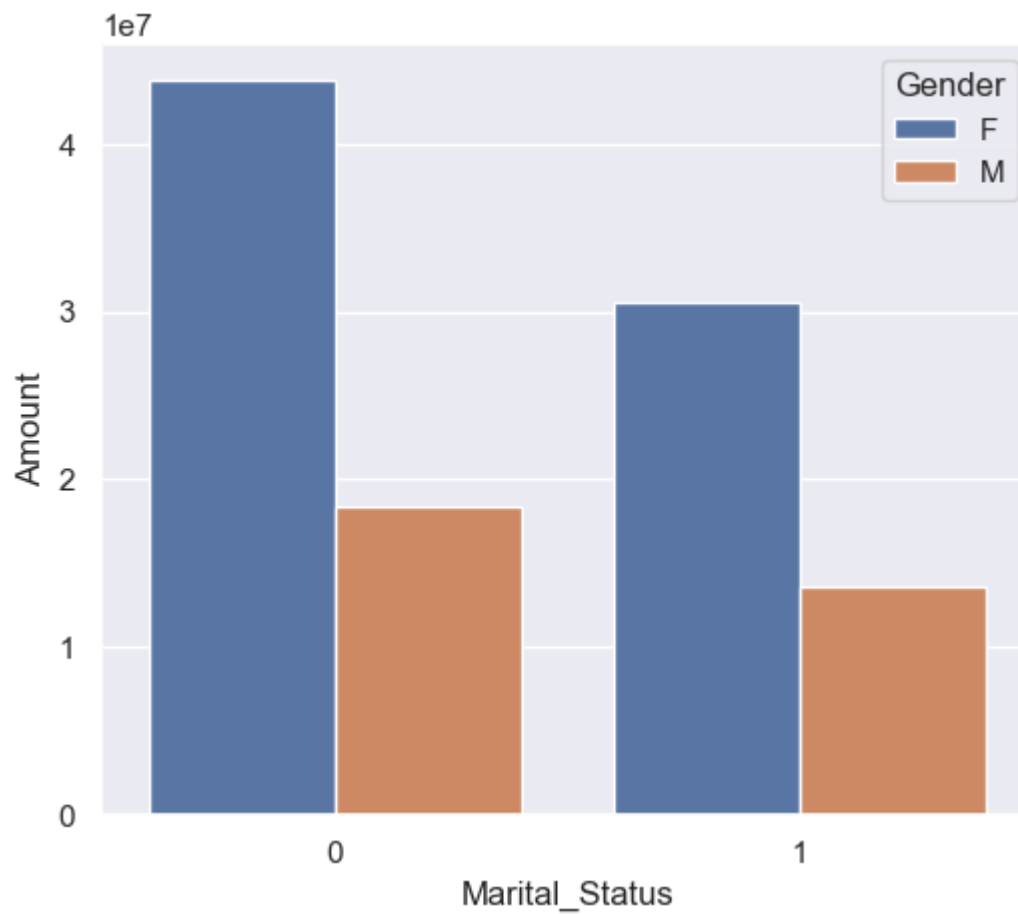
In [25]: `ax = sns.countplot(data = df, x= 'Marital_Status')`
`sns.set(rc={'figure.figsize':(7,5)})`
`for bars in ax.containers:`
`ax.bar_label(bars)`



```
In [26]: sales_state = df.groupby(['Marital_Status', 'Gender'], as_index=False)['Amount'].sum()

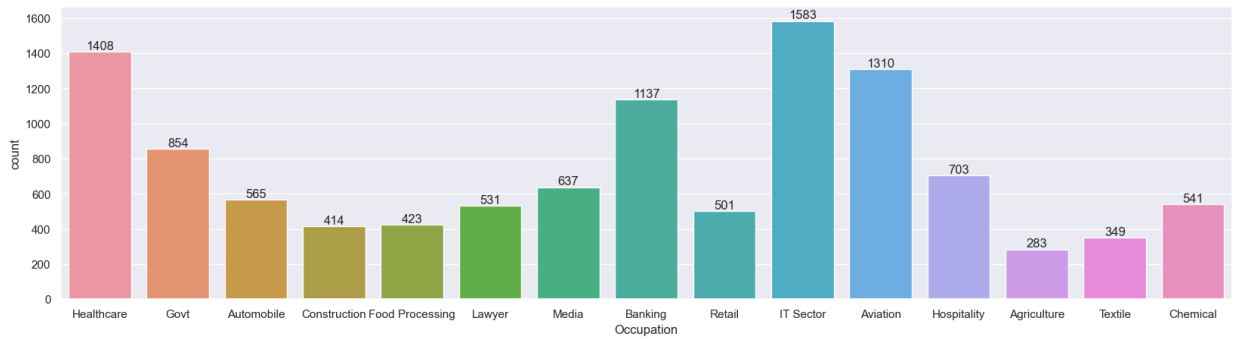
sns.set(rc={'figure.figsize':(6,5)})
sns.barplot(data = sales_state, x = 'Marital_Status', y = 'Amount', hue='Gender')
```

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



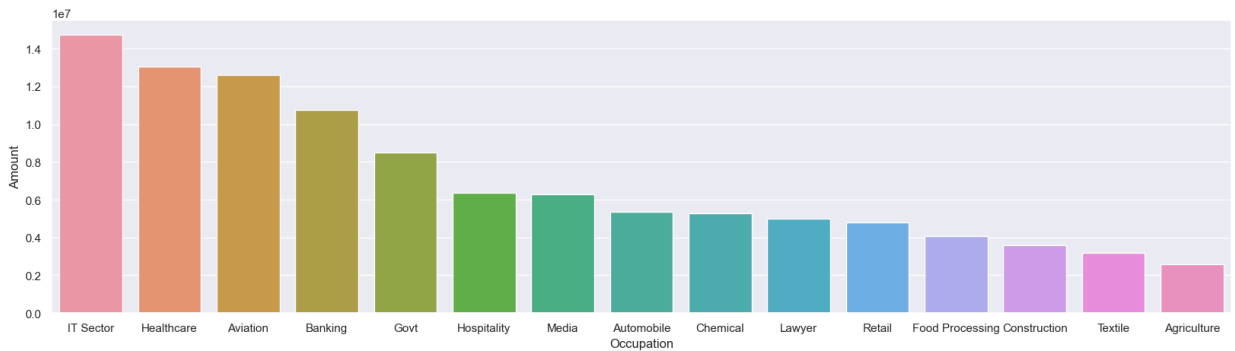
```
In [27]: sns.set(rc={'figure.figsize':(20,5)})
ax = sns.countplot(data =df, x= 'Occupation')

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



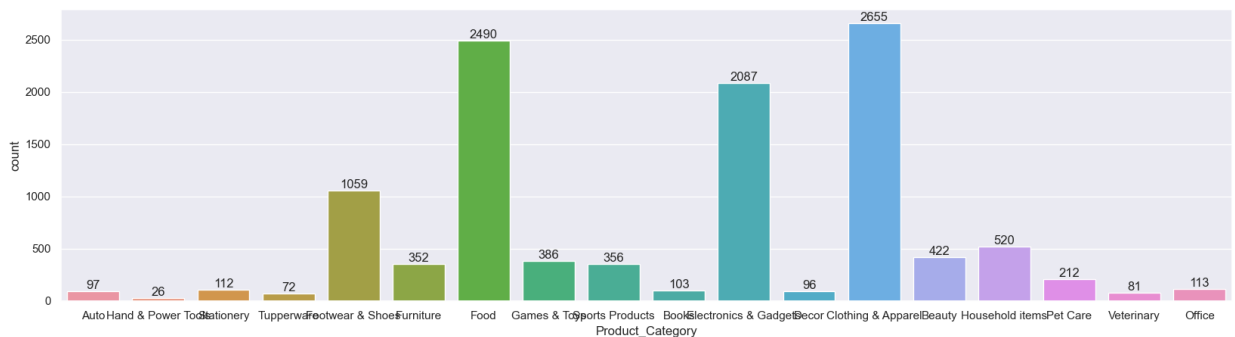
```
In [28]: sales_state = df.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_values(
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Occupation', y= 'Amount')
```

Out[28]: <Axes: xlabel='Occupation', ylabel='Amount'>



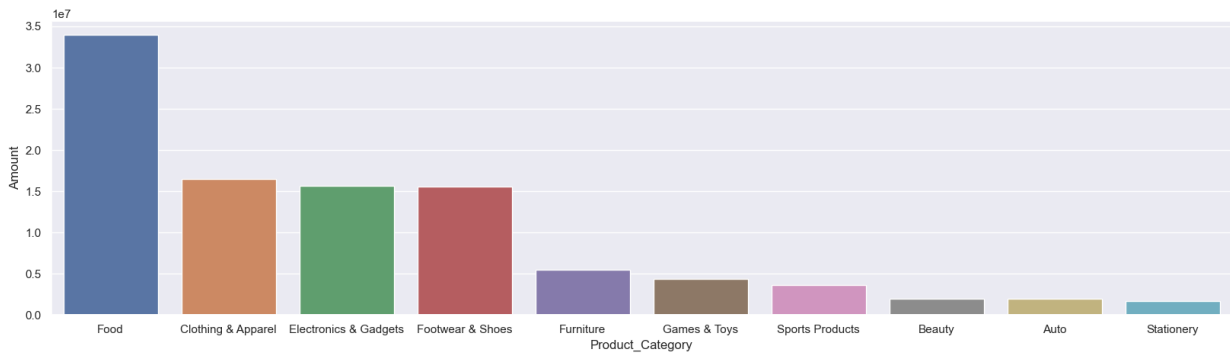
```
In [29]: sns.set(rc={'figure.figsize':(20,5)})
ax = sns.countplot(data =df, x= 'Product_Category')

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



```
In [30]: sales_state = df.groupby(['Product_Category'], as_index=False)['Amount'].sum().sort_val
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Product_Category', y= 'Amount')
```

Out[30]: <Axes: xlabel='Product_Category', ylabel='Amount'>



```
In [31]: sales_state = df.groupby(['Product_ID'], as_index=False)['Orders'].sum().sort_values(
sns.set(rc={'figure.figsize':(25,5)})
sns.barplot(data = sales_state, x = 'Product_ID', y= 'Orders')
```

Out[31]: <Axes: xlabel='Product_ID', ylabel='Orders'>

