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

In [2]:
```
data=pd.read_csv('Sales Data.csv',encoding='unicode_escape')
data.shape
```

Out[2]:
```
(11251, 15)
```

In [3]:
```
data.head()
```

Out[3]:

	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	(
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	So
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	W

In [4]:
```
#we can check or get the information about the data
data.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]:
```
#drop not related/blank columns
data.drop(['Status','unnamed1'],axis=1,inplace=True)
```

In [6]:
```
#Now we are getting the 12 columns only
data.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 [7]:
```
pd.isnull(data)
```

Out[7]:

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

11251 rows × 13 columns

In [8]:
```
#checking for the null values
pd.isnull(data).sum()
```

Out[8]:
```
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 [9]:
```
#So, we are now deleting the null values in from the data set
data.dropna(inplace=True)
```

In [10]:
```
data.shape
```

Out[10]:
```
(11239, 13)
```

In [11]:
```
data['Amount']=data['Amount'].astype('int')
```

In [12]:
```
data['Amount'].dtypes
```

Out[12]:
```
dtype('int32')
```

In [13]:
```
data.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')
```

In [14]:
```
#describe() method returns description of the data in the dataframe (i.e. c
data.describe()
```

Out[14]:

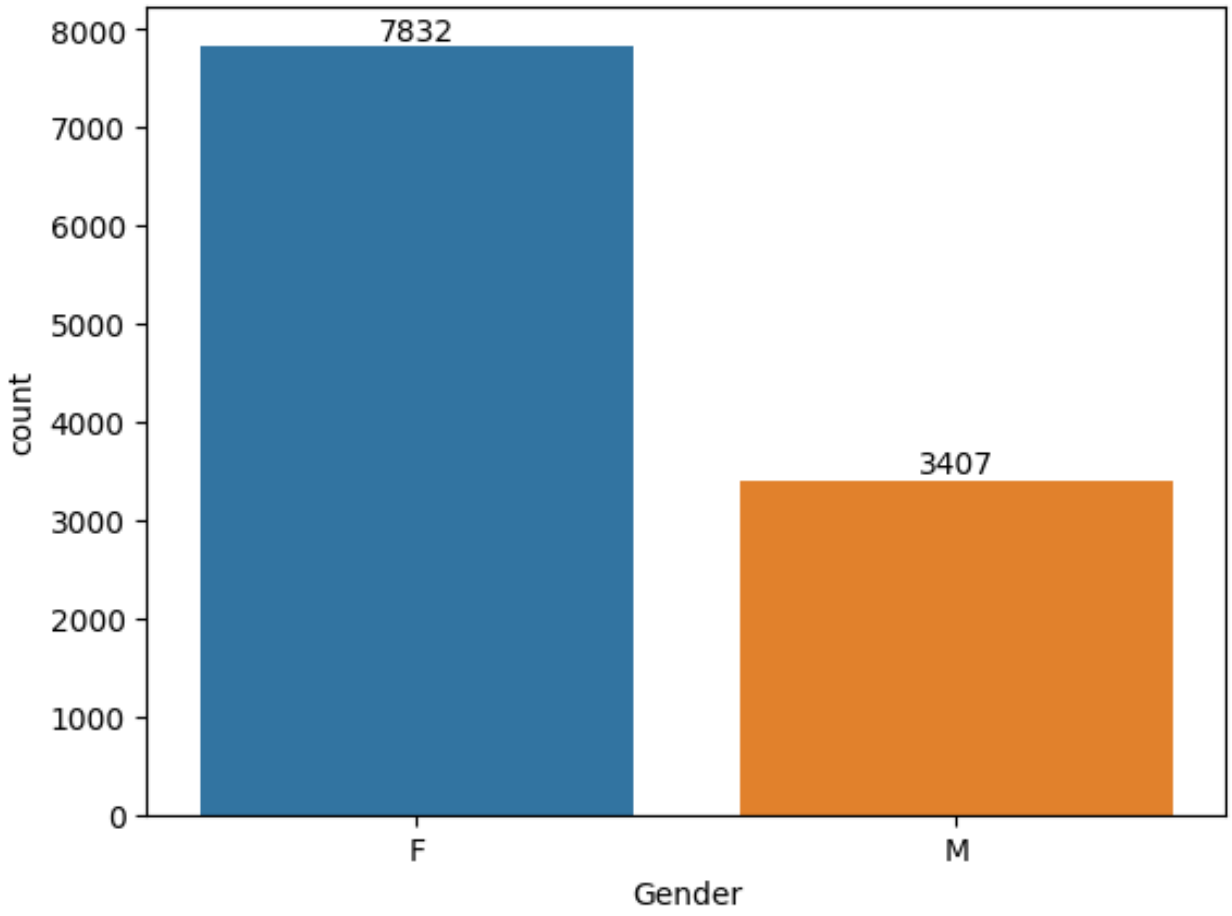
	User_ID	Age	Marital_Status	Orders	Amount
count	1.123900e+04	11239.000000	11239.000000	11239.000000	11239.000000
mean	1.003004e+06	35.410357	0.420055	2.489634	9453.610553
std	1.716039e+03	12.753866	0.493589	1.114967	5222.355168
min	1.000001e+06	12.000000	0.000000	1.000000	188.000000
25%	1.001492e+06	27.000000	0.000000	2.000000	5443.000000
50%	1.003064e+06	33.000000	0.000000	2.000000	8109.000000
75%	1.004426e+06	43.000000	1.000000	3.000000	12675.000000
max	1.006040e+06	92.000000	1.000000	4.000000	23952.000000

Analysing the Data set

Gender column Analysis

In [15]:
```
gen=sns.countplot(x='Gender',data=data)

for i in gen.containers:
    gen.bar_label(i)
```



In [16]:
```
#analysing the Sales by Gender
gender_sales=data.groupby(['Gender'],as_index=False)['Amount'].sum().sort_v
```

In [17]:
```
gender_sales
```

Out[17]:

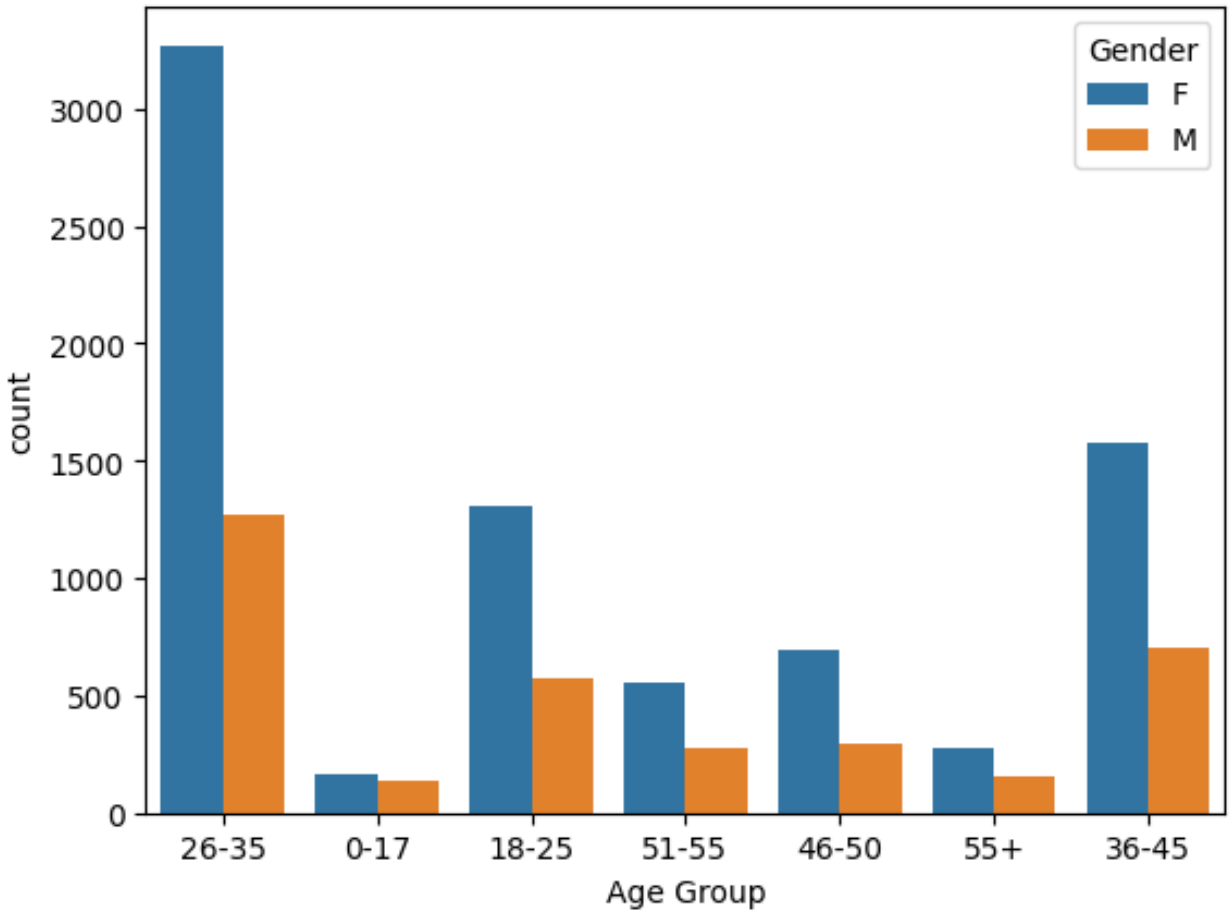
	Gender	Amount
0	F	74335853
1	M	31913276

So, From here we are getting that females are doing more shoping than males.

Ages

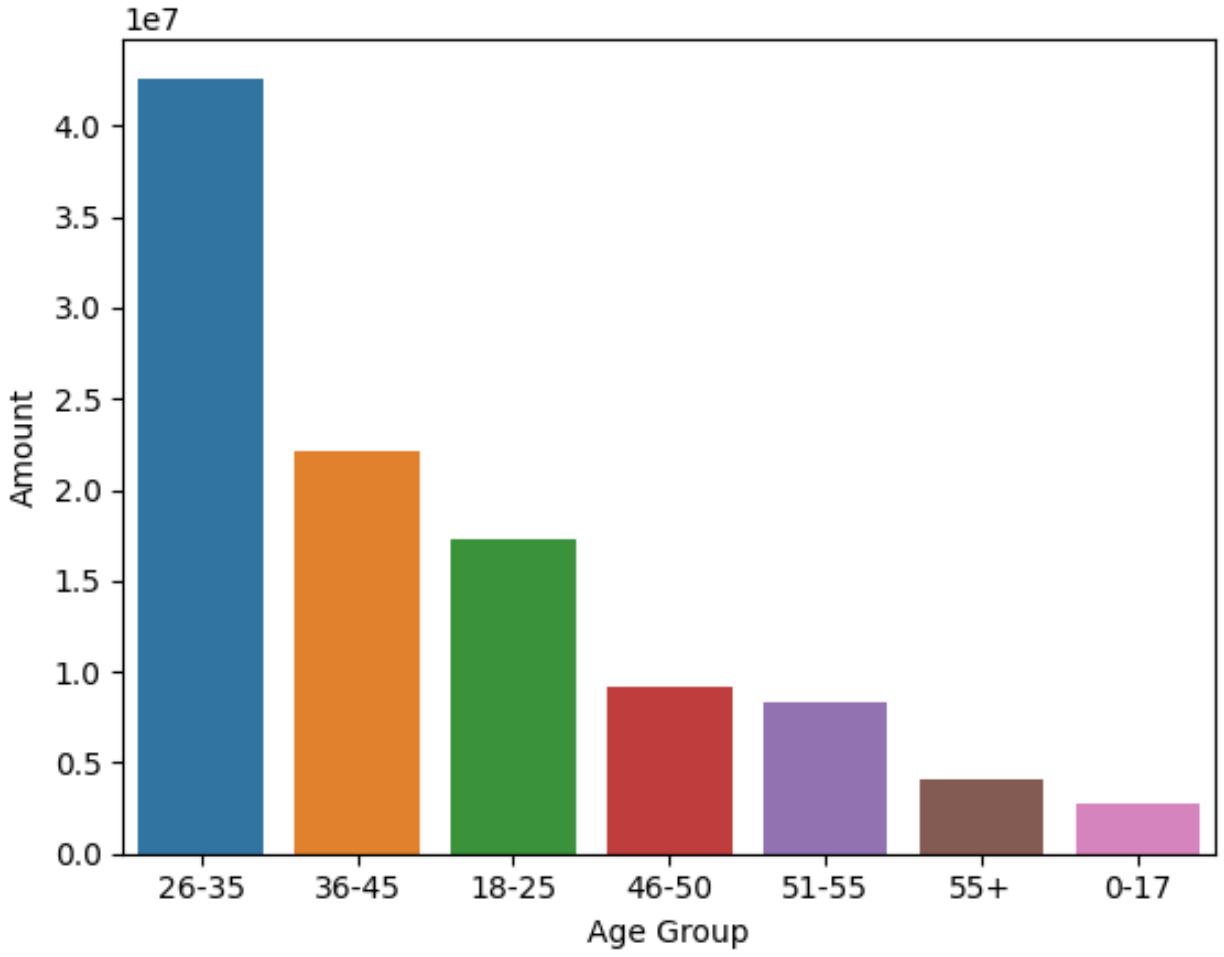
In [18]:
```
sns.countplot(data=data, x='Age Group',hue='Gender')
```

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



In [19]:
```
#Sales by Age Groups
sales_Age=data.groupby(['Age Group'],as_index=False)['Amount'].sum().sort_v
sns.barplot(x='Age Group',y='Amount',data=sales_Age)
```

Out[19]: <Axes: xlabel='Age Group', ylabel='Amount'>

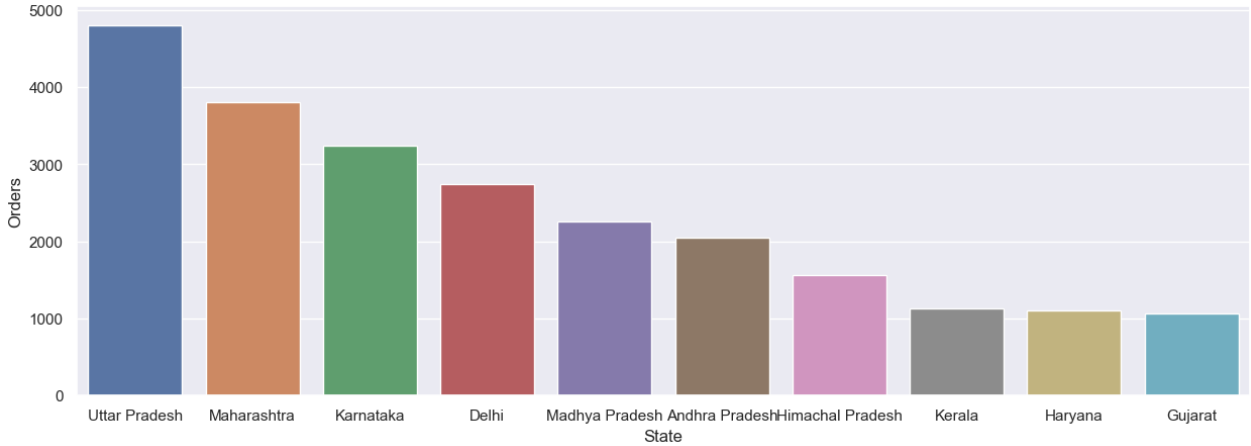


So, Here we can see that most of the buyers are between the age of 26-35 Age Group

Sate wise analysis

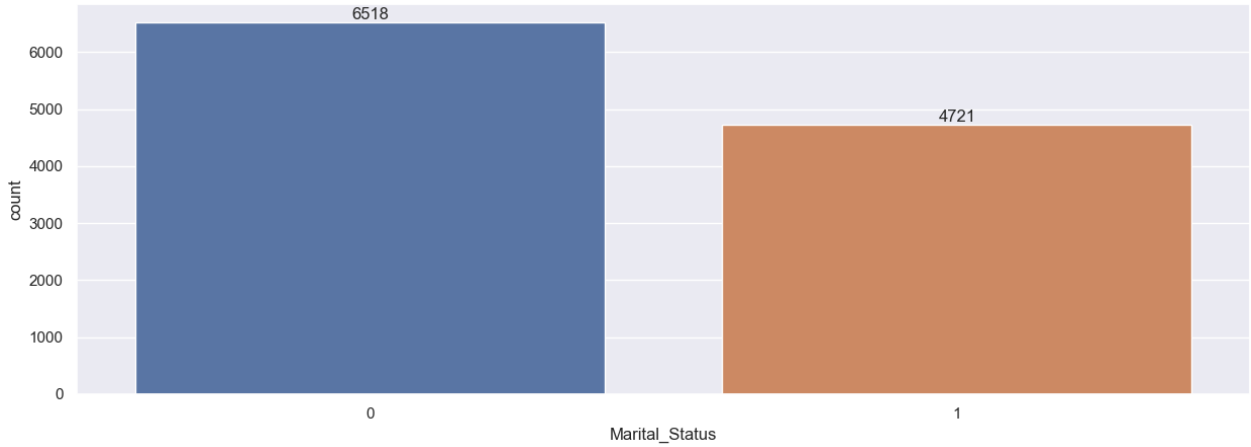
In [20]:
```
sate_sales=data.groupby(['State'],as_index=False)['Orders'].sum().sort_valu
sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data=sate_sales,x='State',y='Orders')
```

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



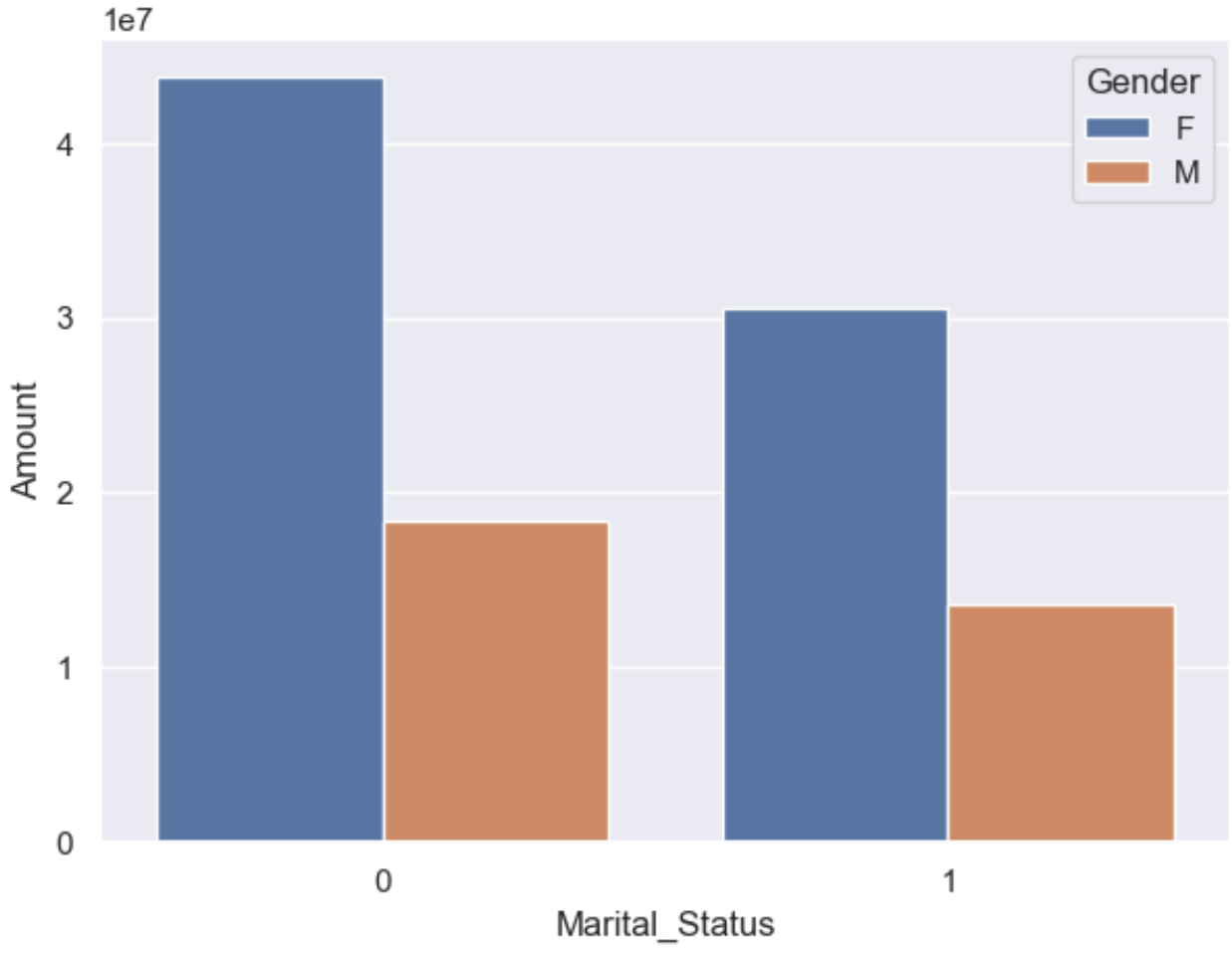
Marital Status

In [21]:
```
ma=sns.countplot(data=data,x='Marital_Status')
sns.set(rc={'figure.figsize':(7,5)})
for i in ma.containers:
    ma.bar_label(i)
```



In [22]:
```
ma=data.groupby(['Marital_Status','Gender'],as_index=False)['Amount'].sum()
sns.set(rc={'figure.figsize':(7,5)})
sns.barplot(data=ma,x='Marital_Status',y='Amount',hue='Gender')
```

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

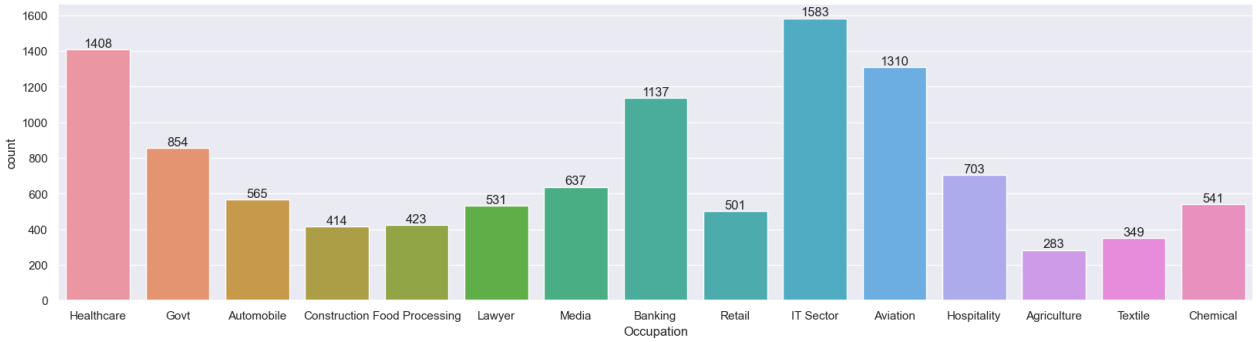


from above graph we can see that most of thr buyers are married (women) and the have High Purchasing power.

Occupations Analysis

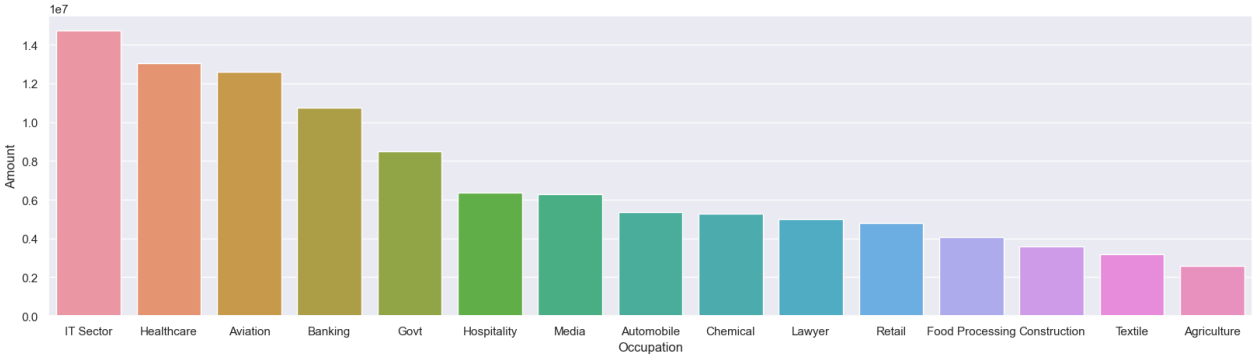
In [23]:
```
sns.set(rc={'figure.figsize':(20,5)})
ac=sns.countplot(data=data,x='Occupation')

for i in ac.containers:
    ac.bar_label(i)
```



In [24]:
```
sate_sales=data.groupby(['Occupation'],as_index=False)['Amount'].sum().sort_
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data=sate_sales,x='Occupation',y='Amount')
```

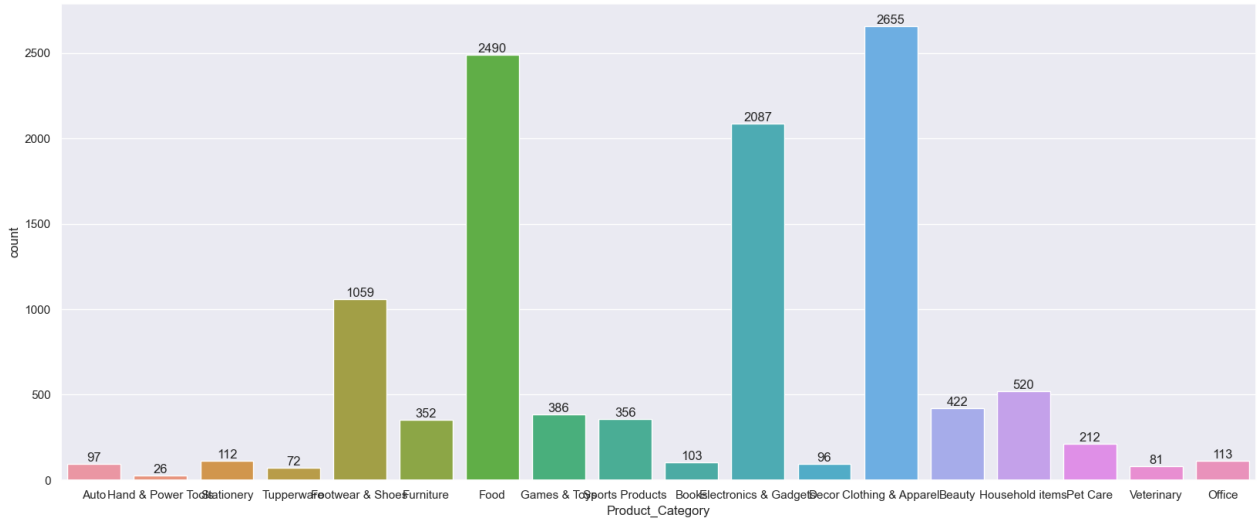
Out[24]: <Axes: xlabel='Occupation', ylabel='Amount'>



from the above graphwe can say that most of the buyers are from Healthcare and IT Sector.

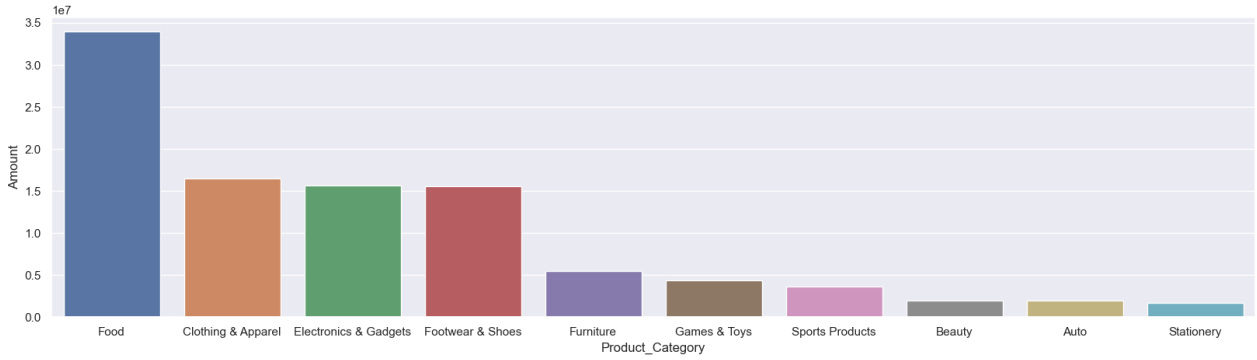
Product Category Analysis

In [25]:
```
sns.set(rc={'figure.figsize':(20,8)})
ac=sns.countplot(data=data,x='Product_Category')
for i in ac.containers:
    ac.bar_label(i)
```



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

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



from the above graphwe can say that most of the buyers are from Food and Clothing & Apparel.

Conclusion-

As we can see in the above analysis that most of the married womens in age group of 26-35 years are purchasing the High amount of Products and most of the purchase is done from the U.P., Maharastra, Karnataka in Healthcare, IT and Aviation Sector. The most purchased products are Food,Clothing and Electronics items.

By-

Rajat Gupta