In [5]: `!pip install matplotlib`

```
Requirement already satisfied: matplotlib in c:\users\user\anaconda3\lib\site
-packages (3.7.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\user\anaconda
3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: pillow>=6.2.0 in c:\users\user\anaconda3\lib\s
ite-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\user\anaconda3\li
b\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\user\anaconda3\li
b\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\user\anaconda3\lib\si
te-packages (from matplotlib) (0.11.0)
Requirement already satisfied: numpy>=1.20 in c:\users\user\anaconda3\lib\sit
e-packages (from matplotlib) (1.23.5)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\user\anaconda3\l
ib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: packaging>=20.0 in c:\users\user\anaconda3\lib
\site-packages (from matplotlib) (22.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\user\anaconda3\l
ib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: six>=1.5 in c:\users\user\anaconda3\lib\site-p
ackages (from python-dateutil>=2.7->matplotlib) (1.16.0)
```

In [6]:
```
!pip install seaborn
```

```
Requirement already satisfied: seaborn in c:\users\user\anaconda3\lib\site-pa
ckages (0.12.2)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\user\anaco
nda3\lib\site-packages (from seaborn) (3.7.0)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\users\user\anaconda
3\lib\site-packages (from seaborn) (1.23.5)
Requirement already satisfied: pandas>=0.25 in c:\users\user\anaconda3\lib\si
te-packages (from seaborn) (1.5.3)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\user\anaconda3\l
ib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\user\anaconda3\li
b\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\user\anaconda3\li
b\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\user\anaconda
3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: packaging>=20.0 in c:\users\user\anaconda3\lib
\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (22.0)
Requirement already satisfied: cycler>=0.10 in c:\users\user\anaconda3\lib\si
te-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: pillow>=6.2.0 in c:\users\user\anaconda3\lib\s
ite-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.4.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\user\anaconda3\l
ib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: pytz>=2020.1 in c:\users\user\anaconda3\lib\si
te-packages (from pandas>=0.25->seaborn) (2022.7)
Requirement already satisfied: six>=1.5 in c:\users\user\anaconda3\lib\site-p
ackages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.
0)
```

In [29]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [110]:
```python
df = pd.read_csv ('Diwali Sales Data.csv' , encoding = "ISO-8859-1")
```

In [111]:
```python
df.shape
```

Out[111]:
```
(11251, 15)
```

In [112]: `df.head()`

Out[112]:

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

In [113]: `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 [114]: `df.drop(['Status','unnamed1'], axis=1, inplace=True)`

In [115]:
```python
pd.isnull(df).sum()
```

Out[115]:
```
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 [116]:
```python
df.dropna(inplace=True)
```

In [117]:
```python
df.shape
```

Out[117]: (11239, 13)

In [118]:
```python
df['Amount'] = df['Amount'].astype('int')
```

In [119]:
```python
df['Amount'].dtypes
```

Out[119]: dtype('int32')

In [120]:
```python
df.columns
```

Out[120]: Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')

In [121]: `df.rename(columns = {'Marital_Status' :'Shaadi'})`

Out[121]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Shaadi	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
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western
...	...	...	...	...	...	...	...	...	...
11246	1000695	Manning	P00296942	M	18-25	19	1	Maharashtra	Western
11247	1004089	Reichenbach	P00171342	M	26-35	33	0	Haryana	Northern
11248	1001209	Oshin	P00201342	F	36-45	40	0	Madhya Pradesh	Central
11249	1004023	Noonan	P00059442	M	36-45	37	0	Karnataka	Southern
11250	1002744	Brumley	P00281742	F	18-25	19	0	Maharashtra	Western

11239 rows × 13 columns

In [122]: `df.describe()`

Out[122]:

	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

In [123]:
```python
df[['Age','Orders','Amount']].describe()
```

Out[123]:

	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 [124]:
```python
ax = sns.countplot(x='Gender', data = df)

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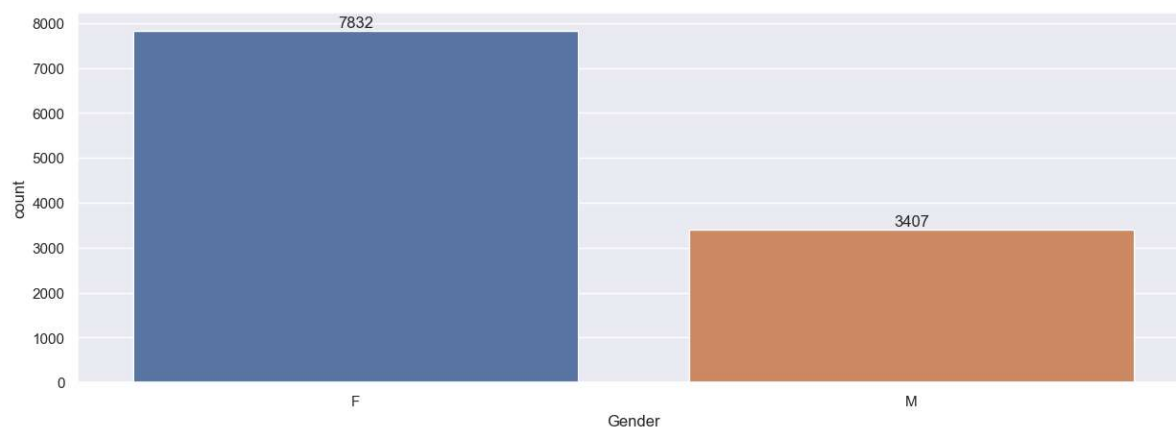

In [125]:
```python
Sales_gen = df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values
sns.barplot(x = 'Gender',y = 'Amount', data =Sales_gen )
```

Out[125]: <Axes: xlabel='Gender', ylabel='Amount'>
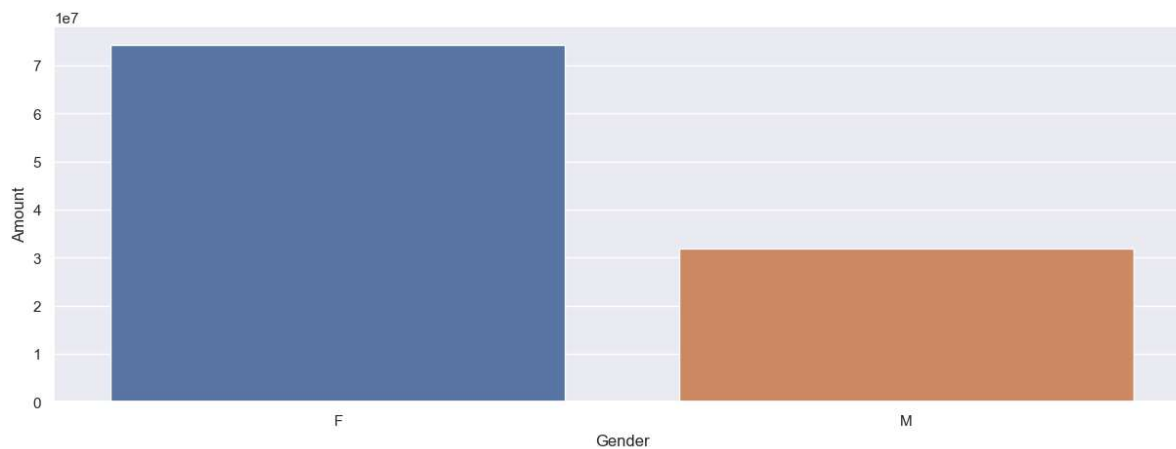
from above graphs we can see that most of the buyers are females and even
the purchasing power of females are grater than man.

In [126]: `df.columns`

Out[126]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')
```

In [127]:
```python
ax = sns.countplot(data= df , x ='Age Group' , hue = 'Gender')

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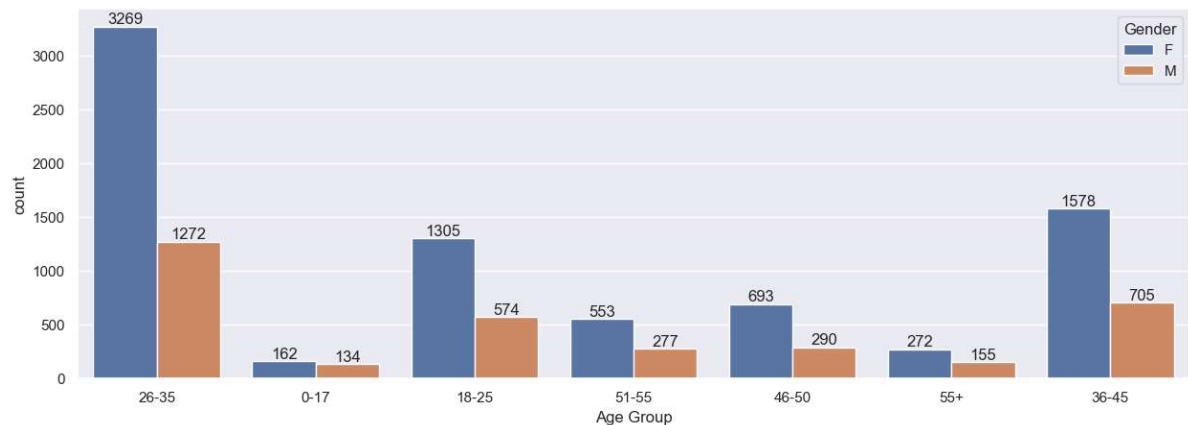

In [128]:
```python
Sales_Age = df.groupby(['Age Group'], as_index=False)['Amount'].sum().sort_val
sns.barplot(x = 'Age Group',y = 'Amount', data =Sales_Age )
```

Out[128]: `<Axes: xlabel='Age Group', ylabel='Amount'>`
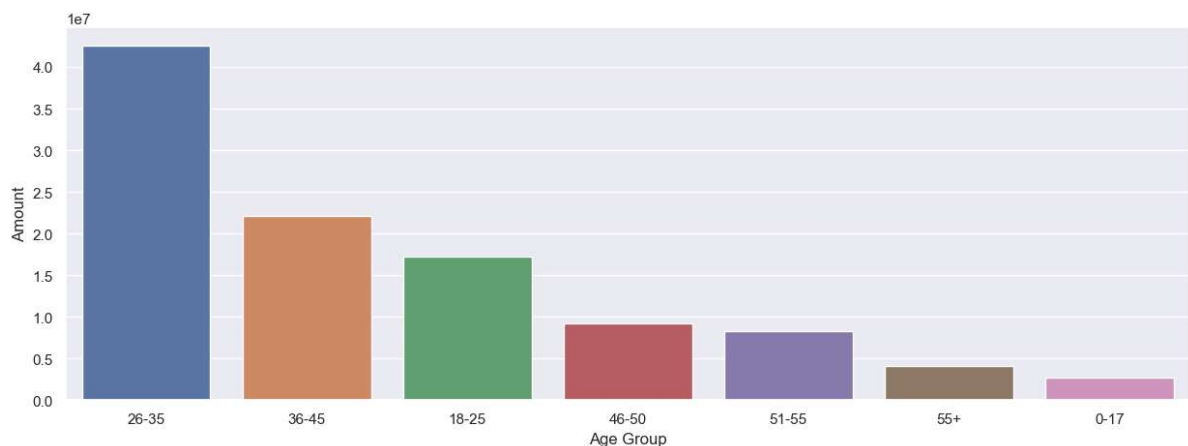
#from above graph we can see that most of the buyers are of the age group
between 26-35yrs female

In [129]:
```python
Sales_State = df.groupby(['State'], as_index=False)['Orders'].sum().sort_value

sns.set( rc= {'figure.figsize':(15,5)})
sns.barplot(x = 'State',y = 'Orders', data =Sales_State )
```
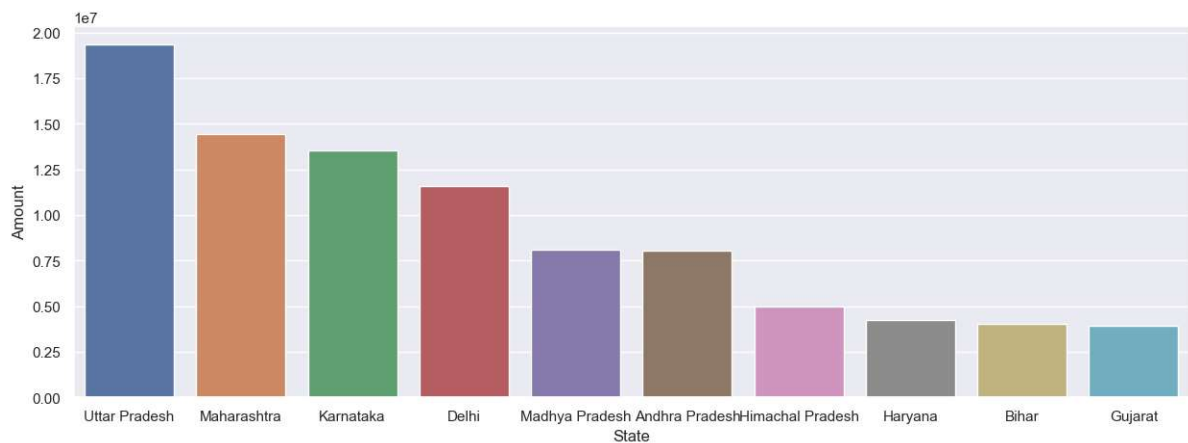
Out[129]: <Axes: xlabel='State', ylabel='Orders'>
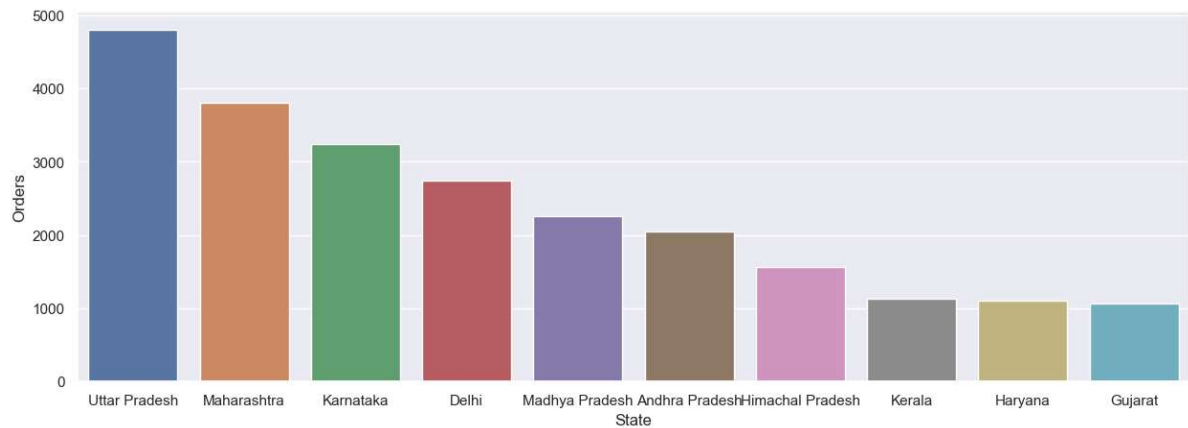
In [130]:
```python
Sales_State = df.groupby(['State'], as_index=False)['Amount'].sum().sort_value

sns.set( rc= {'figure.figsize':(15,5)})
sns.barplot(x = 'State',y = 'Amount', data =Sales_State )
```

Out[130]: <Axes: xlabel='State', ylabel='Amount'>


```
#from above graph we can see that most of the orders are from uttar padesh,
Maharastraand karnataka respectively
```

In [131]:
```python
ax = sns.countplot(x='Marital_Status', data = df)

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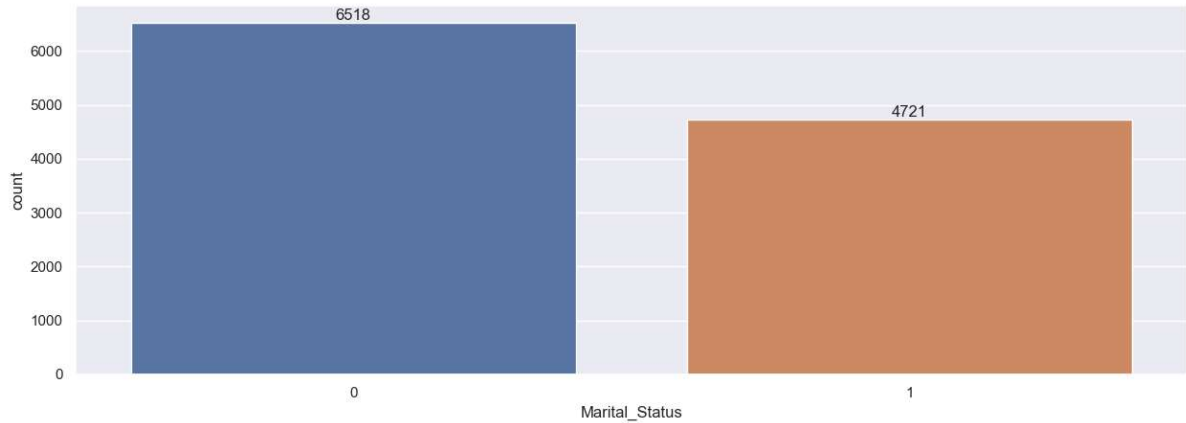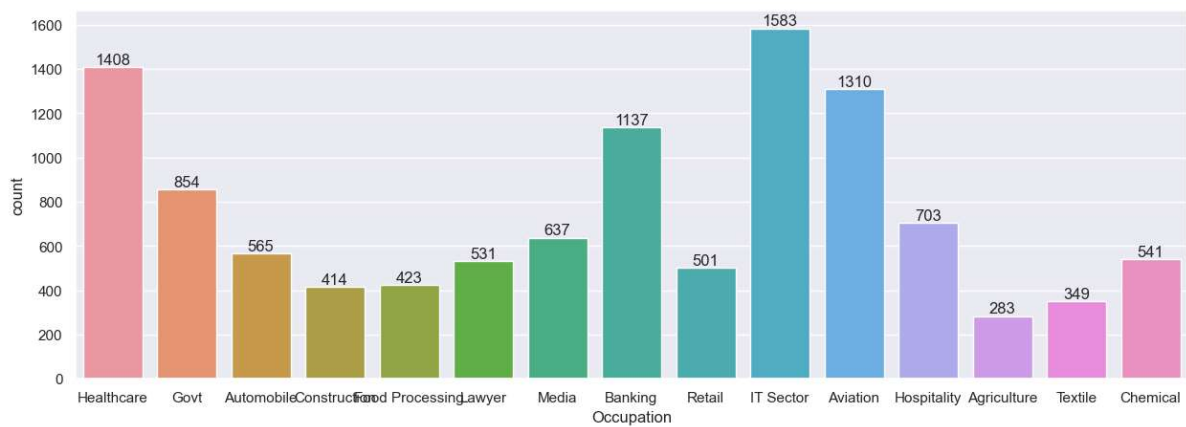

#from above graph we can see that most of the buyers are married (Women) and they have high purchasing power.

In [132]:
```python
ax = sns.countplot(x='Occupation', data = df)

for bars in ax.containers:
    ax.bar_label(bars)
```
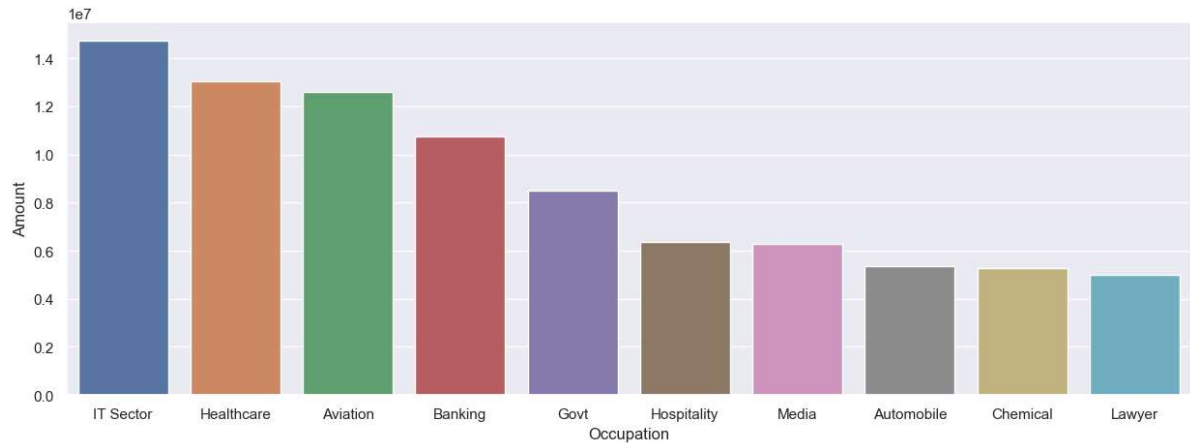
In [133]:
```python
Sales_Occ = df.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_va

sns.set( rc= {'figure.figsize':(15,5)})
sns.barplot(x = 'Occupation',y = 'Amount', data =Sales_Occ )
```
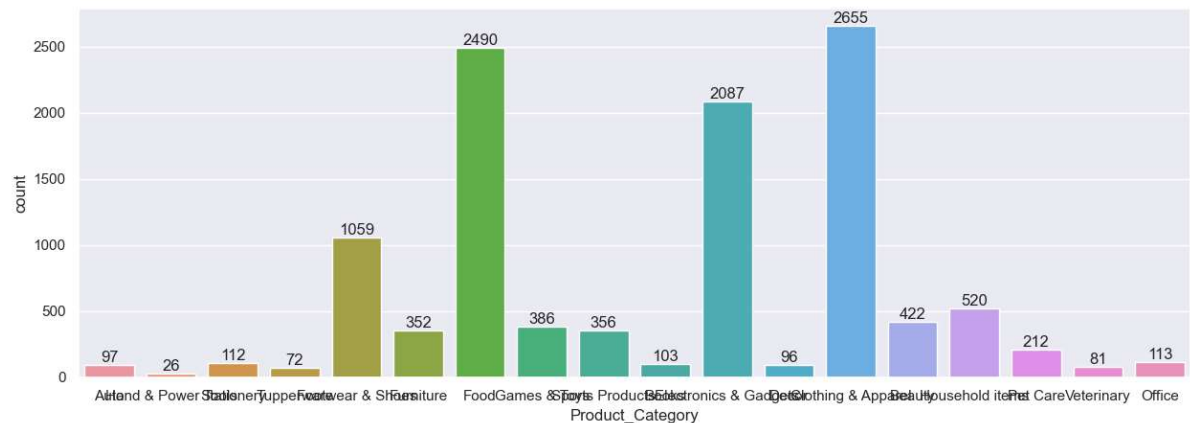
Out[133]: `<Axes: xlabel='Occupation', ylabel='Amount'>`

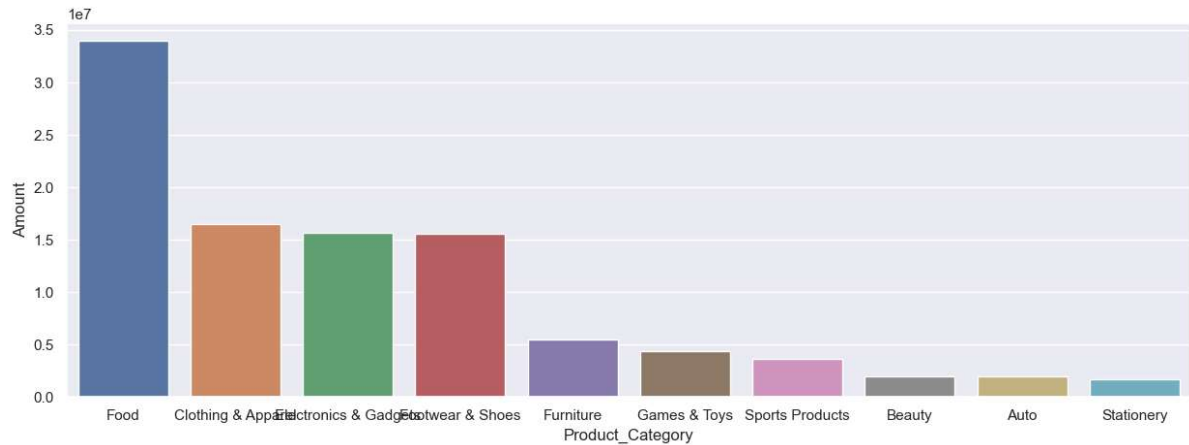#Above graph we can see that mpst of the buyers are workin in It. Healthcare and Aviation

In [134]:
```python
ax = sns.countplot(x='Product_Category', data = df)

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

In [135]:
```python
Sales_Prod= df.groupby(['Product_Category'], as_index=False)['Amount'].sum().s
sns.barplot(x = 'Product_Category',y = 'Amount', data =Sales_Prod )
```
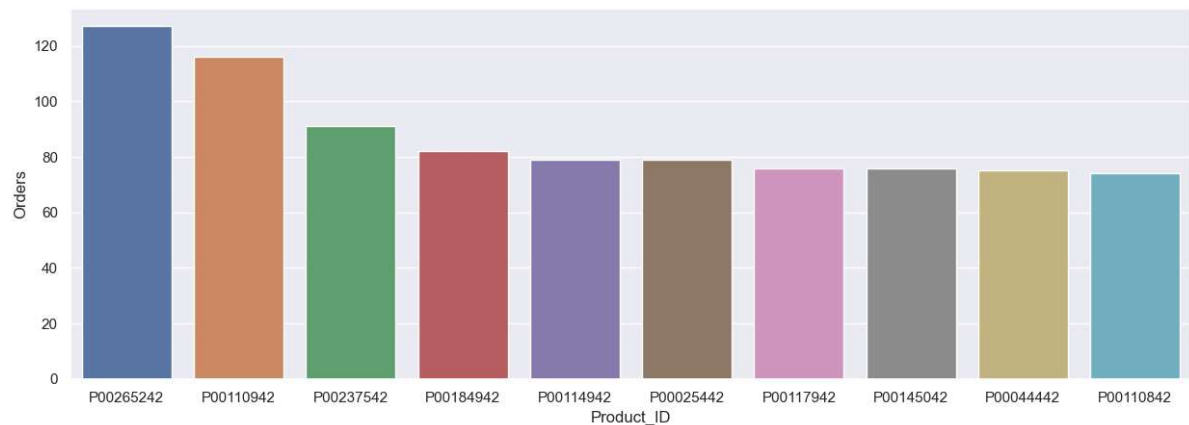
Out[135]: <Axes: xlabel='Product_Category', ylabel='Amount'>


```python
# from above graph we can see that most of the sold product are from food,
clothing and Electronic categroy
```

In [136]:
```python
Sales_Prod= df.groupby(['Product_ID'], as_index=False)['Orders'].sum().sort_va
sns.barplot(x = 'Product_ID',y = 'Orders', data =Sales_Prod )
```

Out[136]: <Axes: xlabel='Product_ID', ylabel='Orders'>

#Conclusion : Married Women group 26-35 years from Up, Maharastraand Karnatka working in It, Healthcare and Aviation are more likely to buy products from food, clothings and Electronic category.

In [ ]: