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

In [2]:
```python
data = pd.read_csv('C:\\Users\\kumar\\Downloads\\Python proj\\Diwali Sales Data.csv', encoding='unicode_escape')
```

In [3]:
```python
data.head(5)
```

Out[3]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	unnamed1
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.0	NaN	NaN
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0	NaN	NaN
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0	NaN	NaN
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0	NaN	NaN
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0	NaN	NaN

In [4]:
```python
df = pd.DataFrame(data)
```

In [5]:
```python
df.count()
```

Out[5]:
```
User_ID             11251
Cust_name           11251
Product_ID          11251
Gender              11251
Age Group           11251
Age                 11251
Marital_Status      11251
State               11251
Zone                11251
Occupation          11251
Product_Category    11251
Orders              11251
Amount              11239
Status                  0
unnamed1                0
dtype: int64
```

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

In [7]:
```python
dfa.count()
```

Out[7]:
```
User_ID             11251
Cust_name           11251
Product_ID          11251
Gender              11251
Age Group           11251
Age                 11251
Marital_Status      11251
State               11251
Zone                11251
Occupation          11251
Product_Category    11251
Orders              11251
Amount              11239
dtype: int64
```

In [8]:
```python
dfa.head()
```

Out[8]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.0
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0

In [9]:
```python
dfa.isnull().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
```

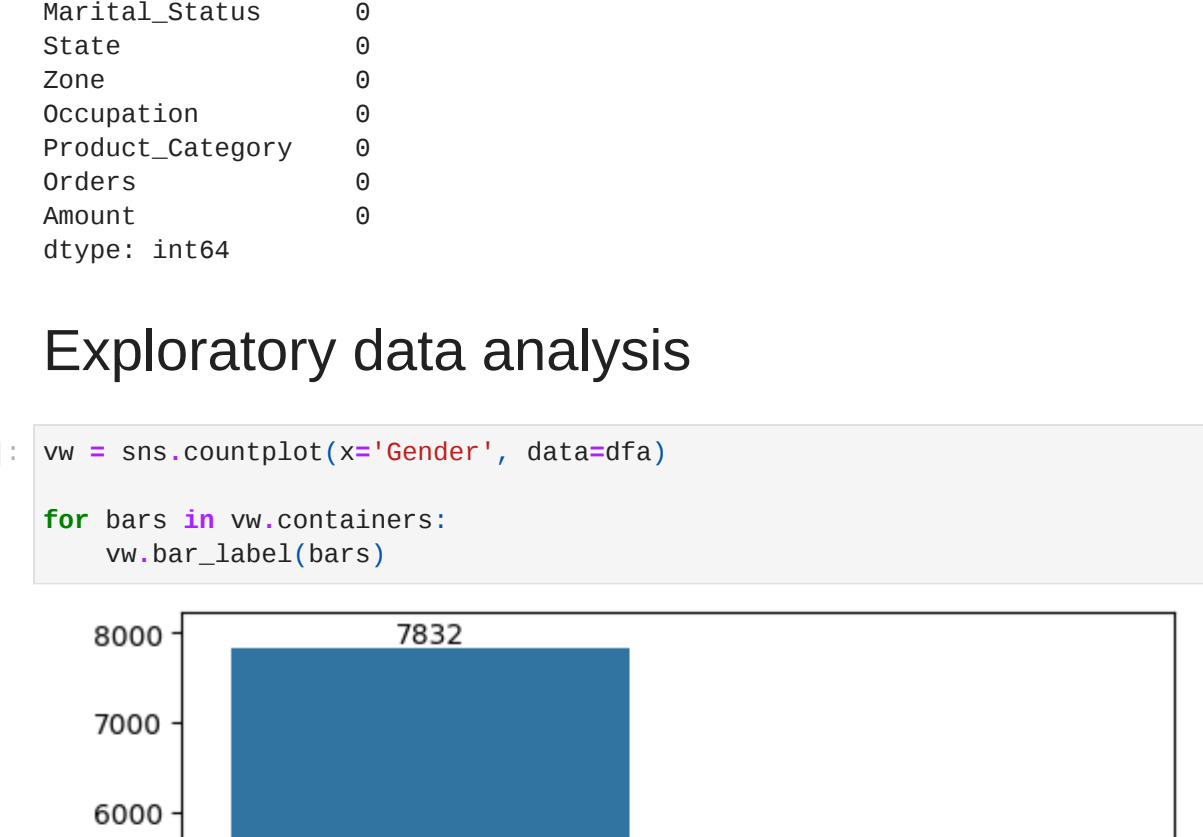
In [10]:
```python
dfa.dropna(inplace=True)
```

In [11]:
```python
dfa.isnull().sum()
```

Out[11]:
```
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
```

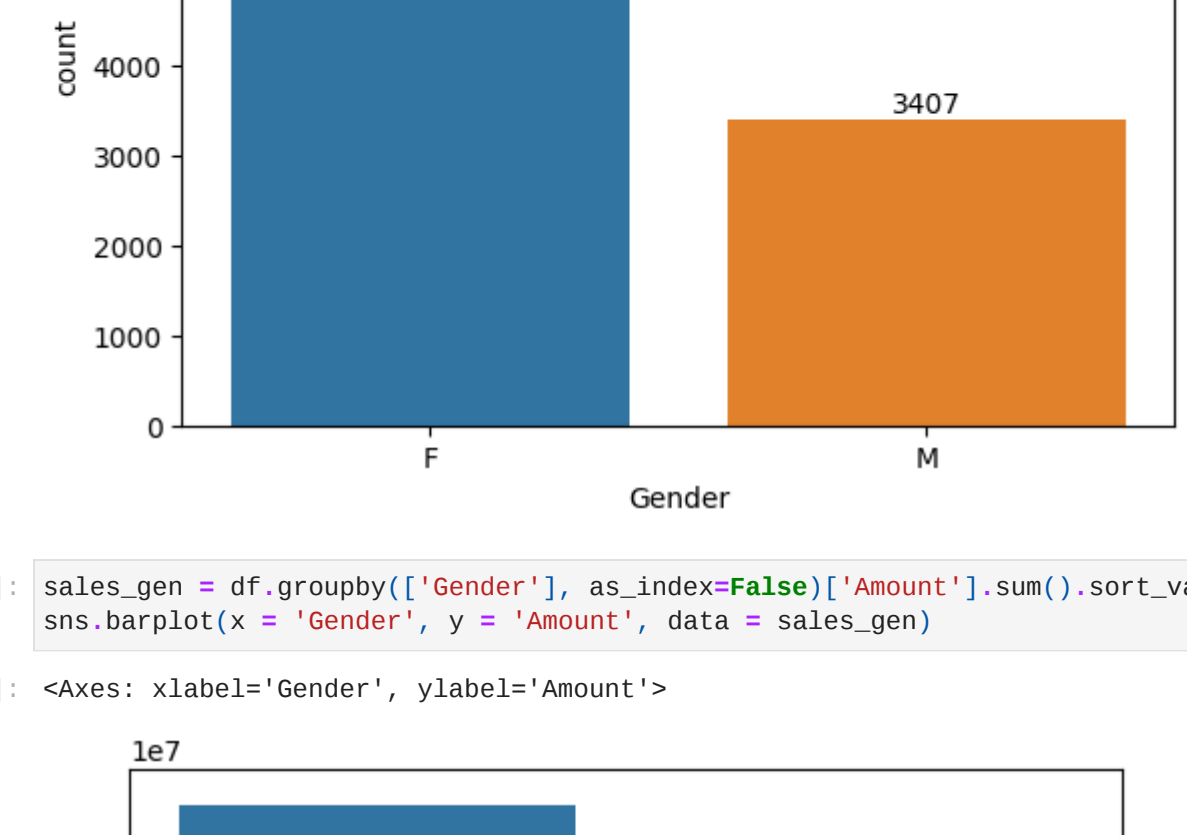
Exploratory data analysis

In [12]:
```python
vw = sns.countplot(x='Gender', data=dfa)

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

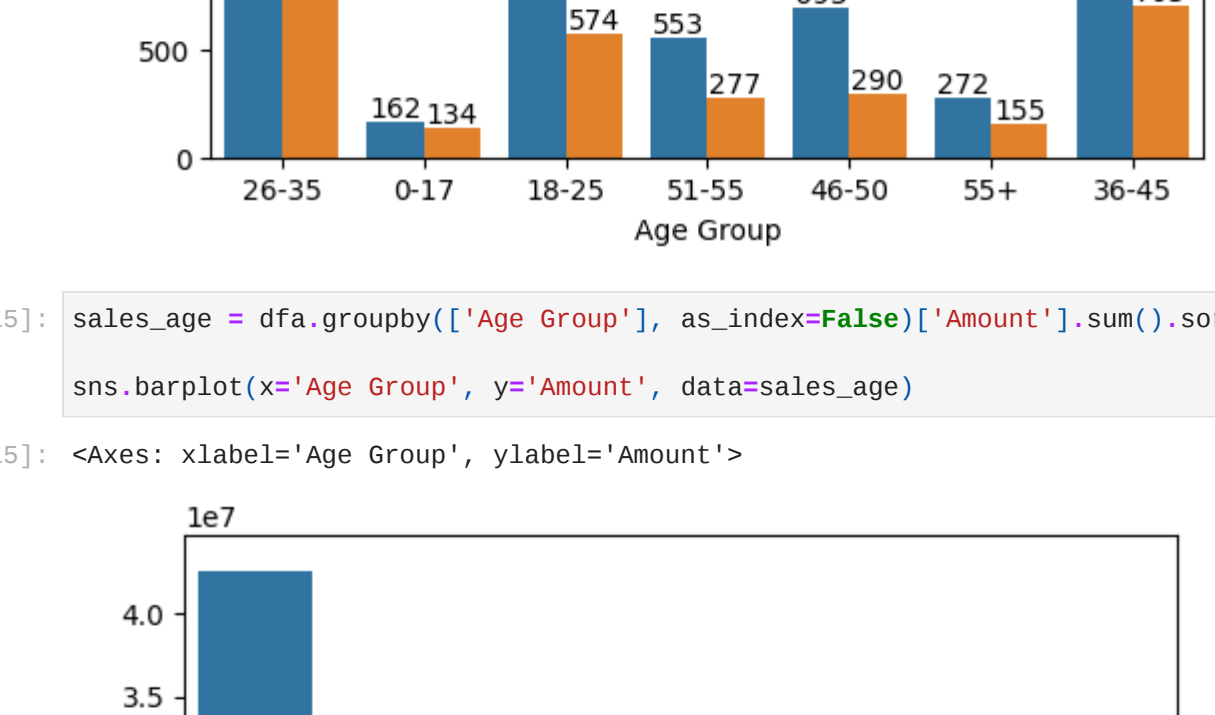


In [13]:
```python
sales_gen = df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(x = 'Gender', y = 'Amount', data = sales_gen)
```

Out[13]: <Axes: xlabel='Gender', ylabel='Amount'>

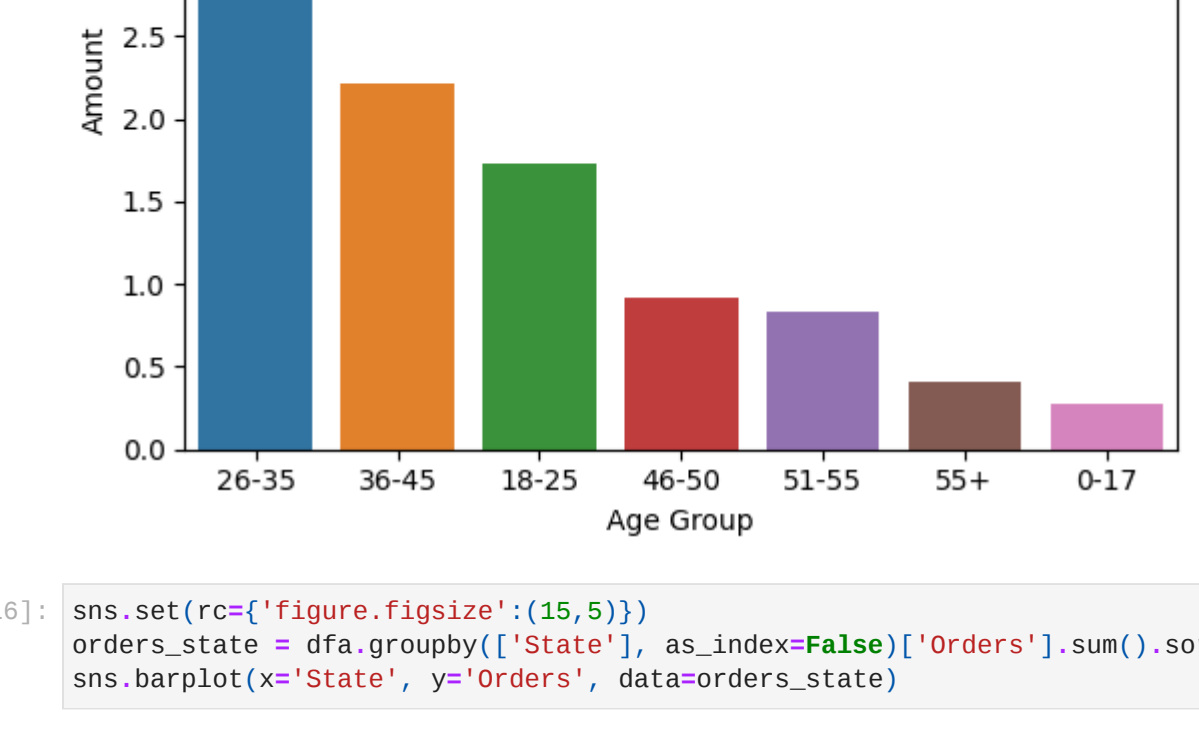


In [14]:
```python
ax=sns.countplot(x='Age Group', data=dfa, hue='Gender')
for bars in ax.containers:
    ax.bar_label(bars)
```



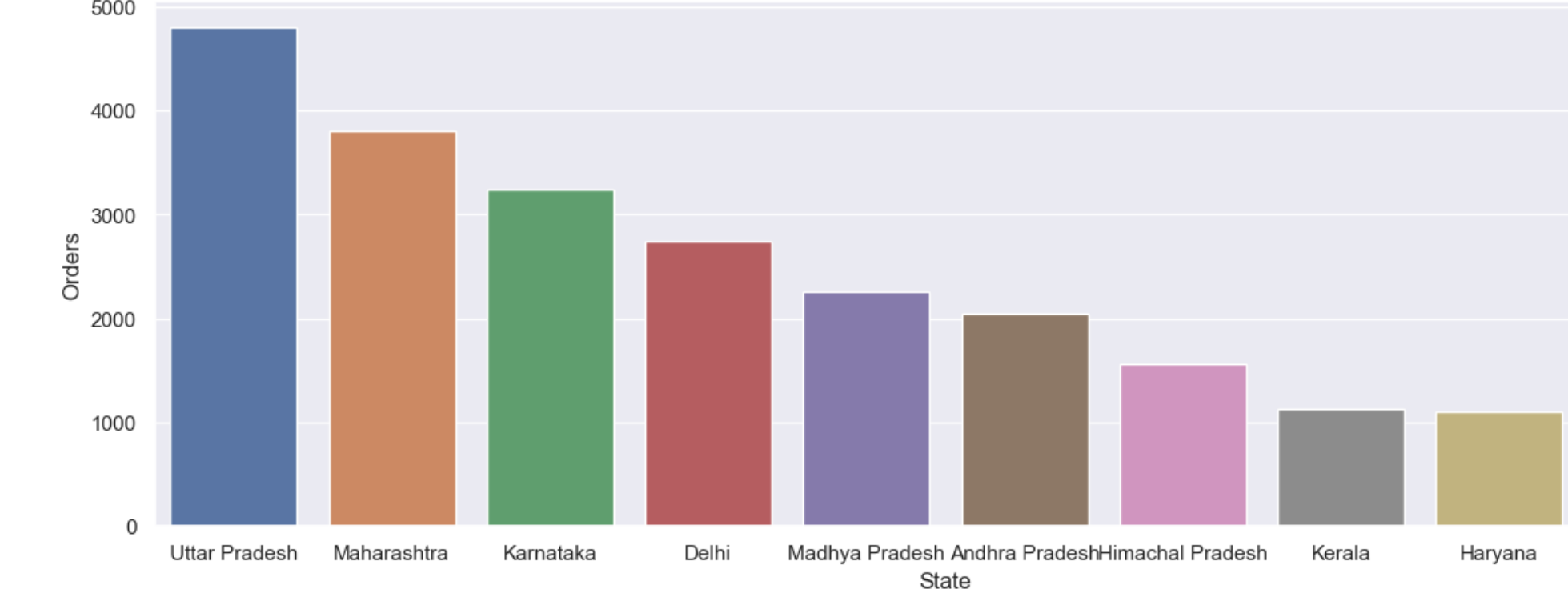
In [15]:
```python
sales_age = dfa.groupby(['Age Group'], as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False)

sns.barplot(x='Age Group', y='Amount', data=sales_age)
```

Out[15]: <Axes: xlabel='Age Group', ylabel='Amount'>



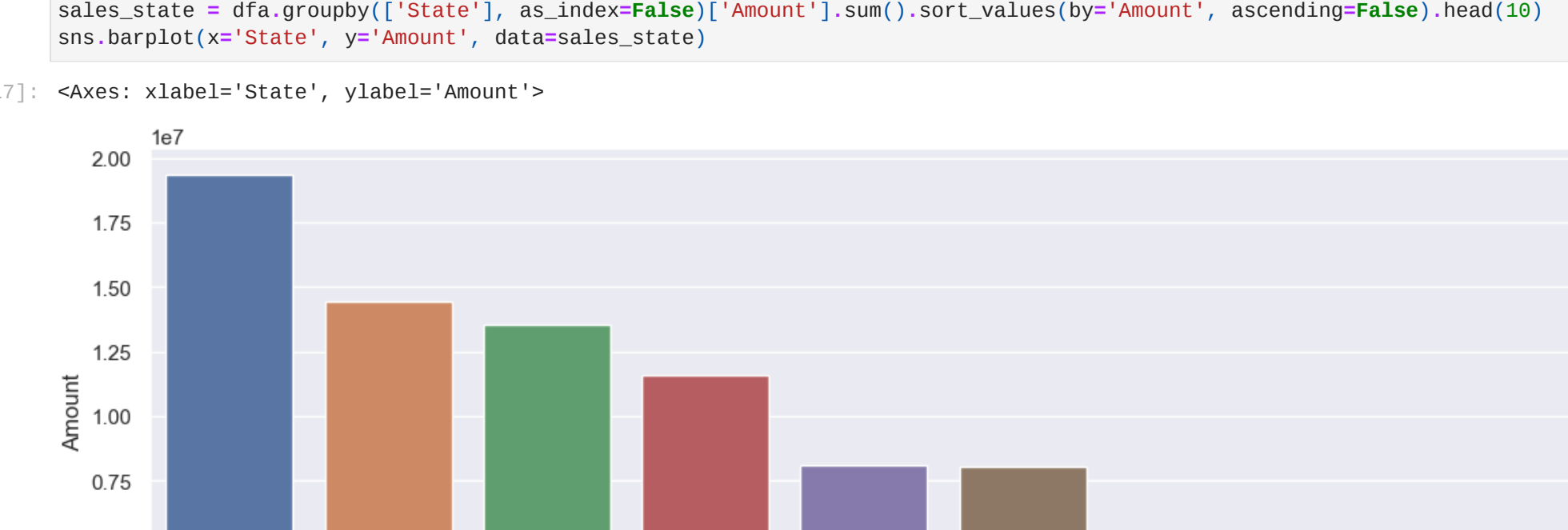
In [16]:
```python
sns.set(rc={'figure.figsize':(15,5)})
orders_state = dfa.groupby(['State'], as_index=False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)
sns.barplot(x='State', y='Orders', data=orders_state)
```

Out[16]: <Axes: xlabel='State', ylabel='Orders'>

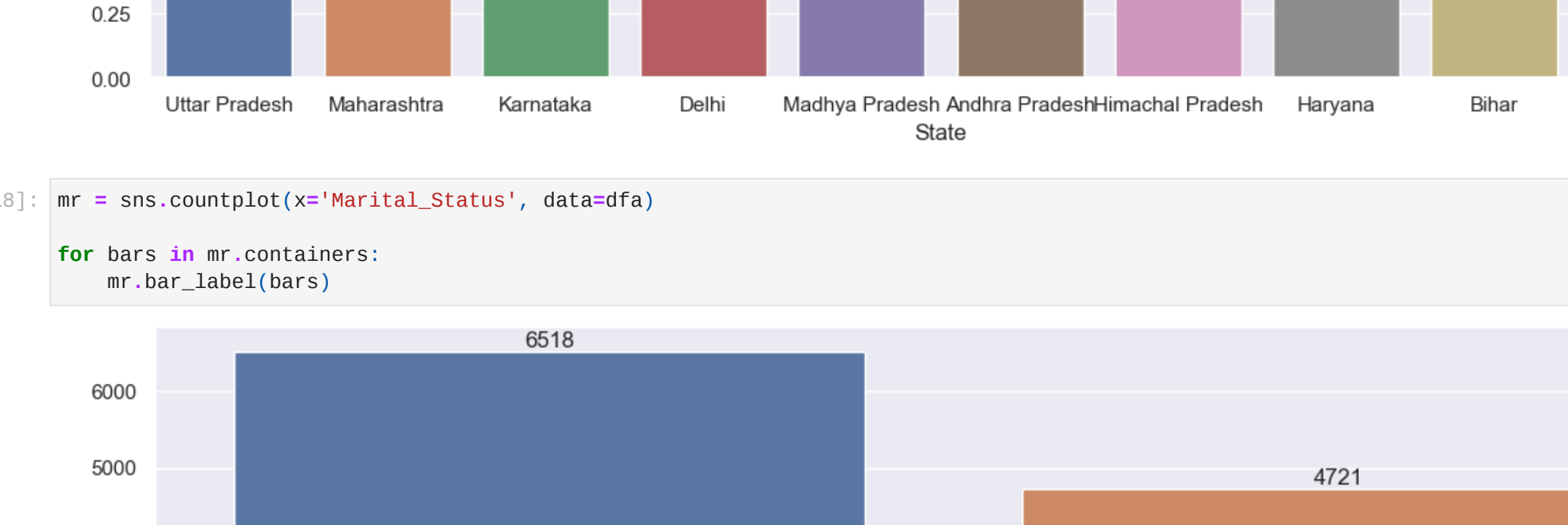


In [17]:
```python
sns.set(rc={'figure.figsize':(15,5)})
sales_state = dfa.groupby(['State'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sns.barplot(x='State', y='Amount', data=sales_state)
```

Out[17]: <Axes: xlabel='State', ylabel='Amount'>

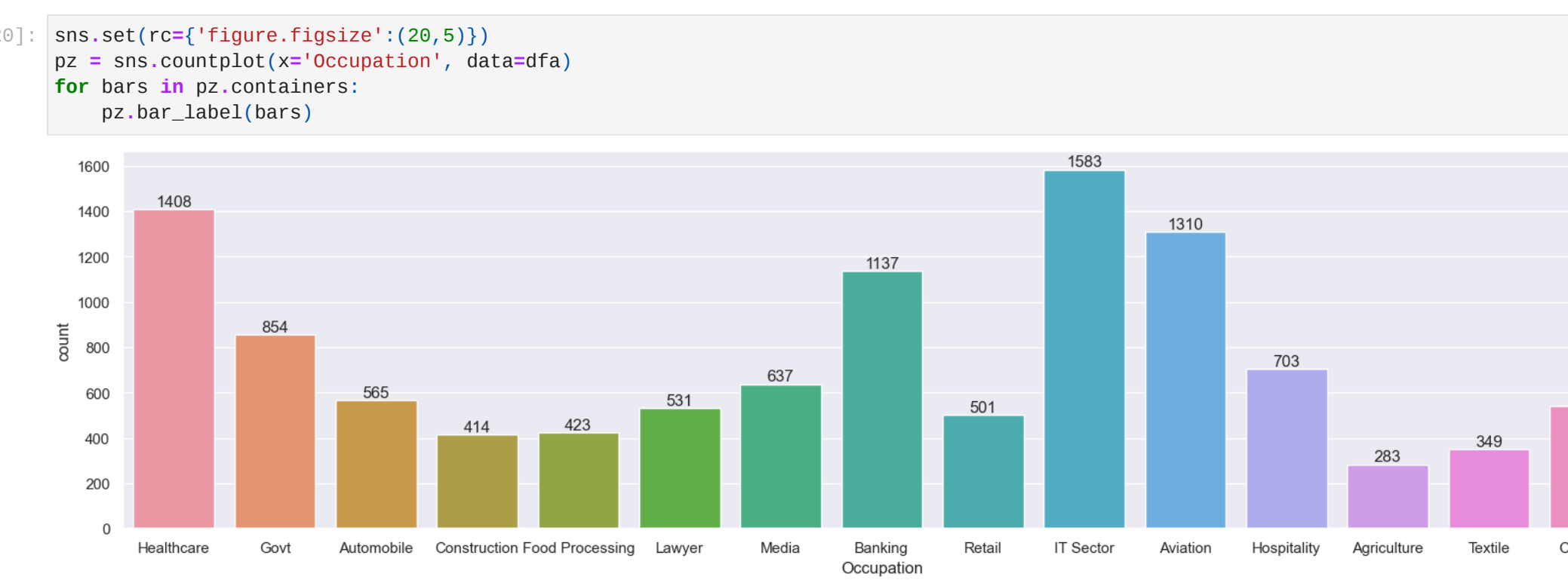


In [18]:
```python
mr = sns.countplot(x='Marital_Status', data=dfa)

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

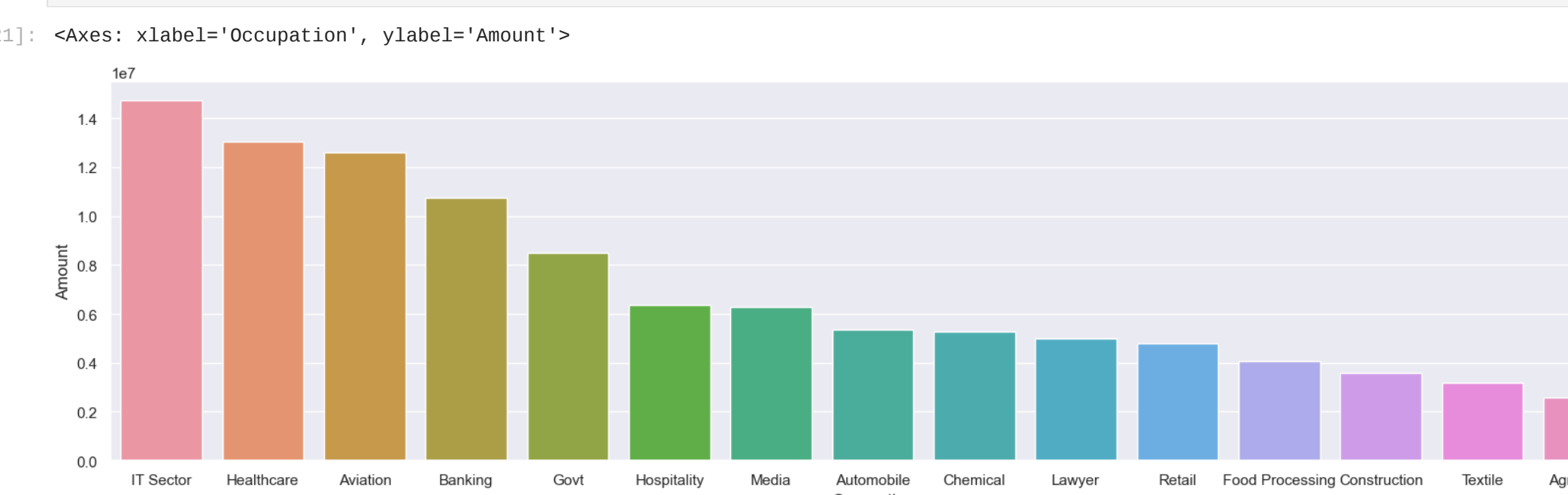


In [19]:
```python
mar_sales = dfa.groupby(['Marital_Status', 'Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(x='Marital_Status', y='Amount', data= mar_sales, hue='Gender')
```

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

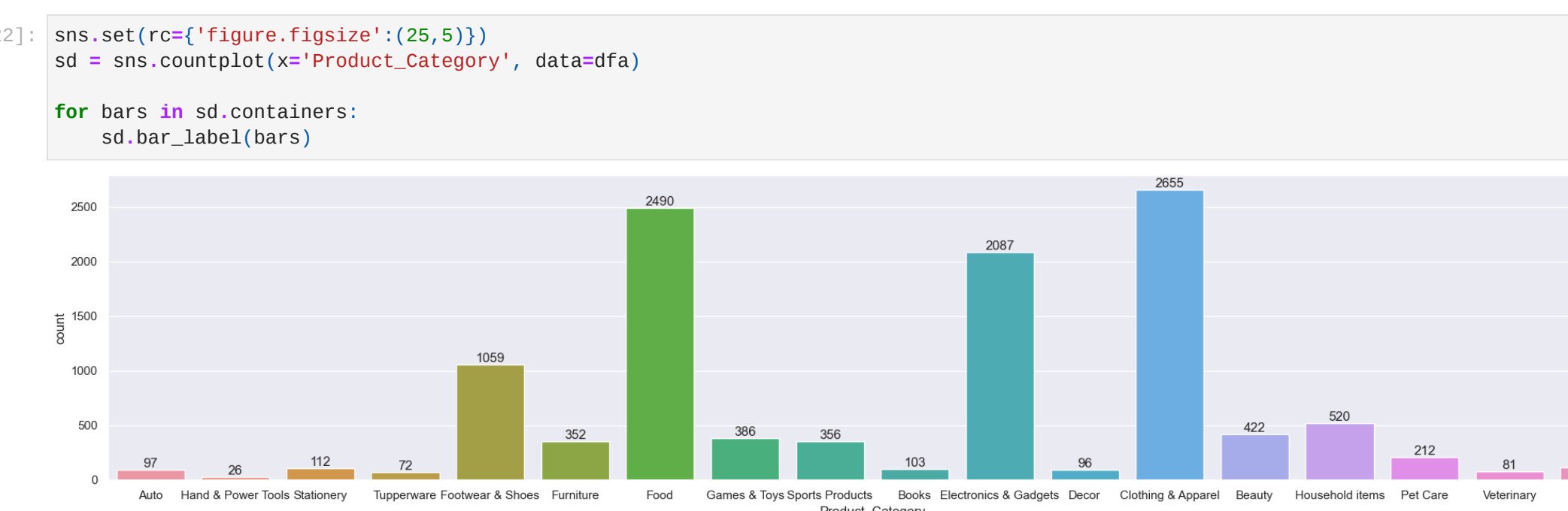


In [20]:
```python
sns.set(rc={'figure.figsize':(20,5)})
pz = sns.countplot(x='Occupation', data=dfa)
for bars in pz.containers:
    pz.bar_label(bars)
```

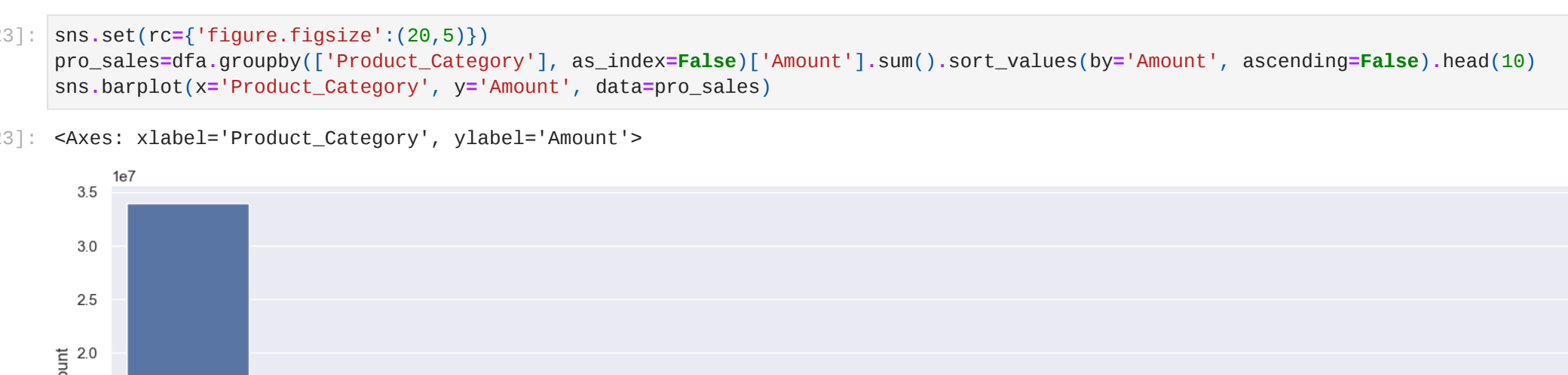


In [21]:
```python
sns.set(rc={'figure.figsize':(20,5)})
occ_sales=dfa.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(x='Occupation', y='Amount', data=occ_sales)
```

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



In [22]:
```python
sns.set(rc={'figure.figsize':(25,5)})
sd = sns.countplot(x='Product_Category', data=dfa)

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



In [23]:
```python
sns.set(rc={'figure.figsize':(20,5)})
pro_sales=dfa.groupby(['Product_Category'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sns.barplot(x='Product_Category', y='Amount', data=pro_sales)
```

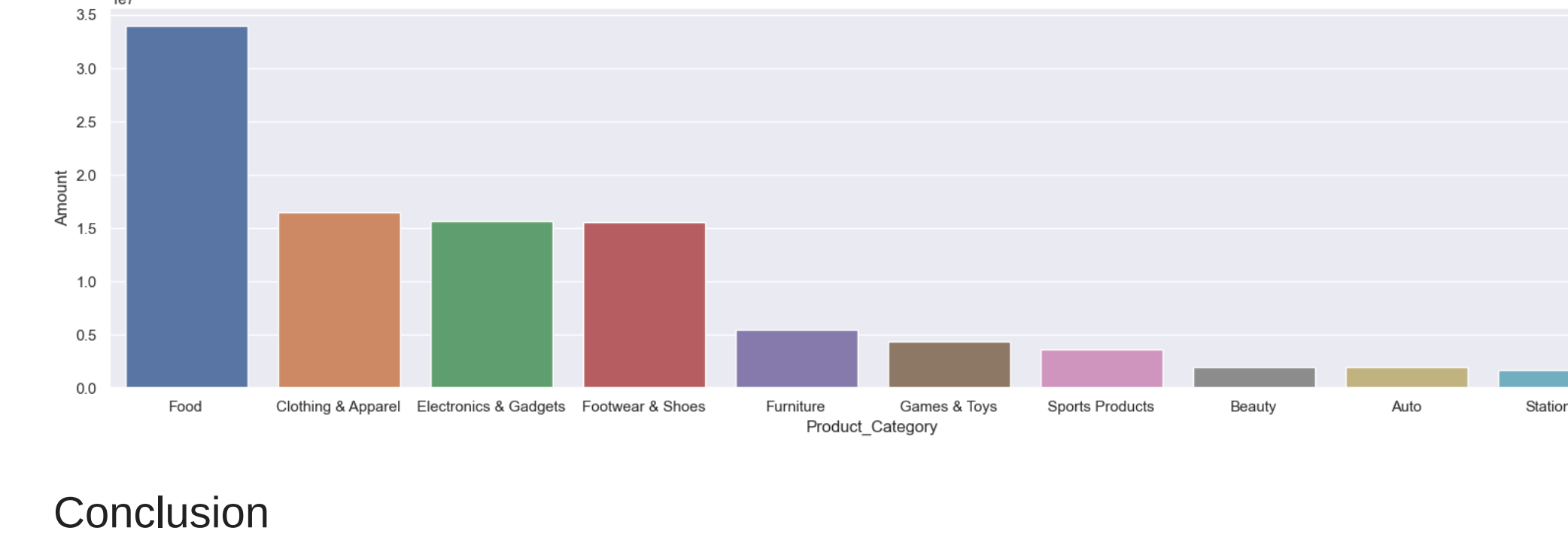
Out[23]: <Axes: xlabel='Product_Category', ylabel='Amount'>

In [24]:
```python
pro_ord = 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 = pro_ord, x = 'Product_ID',y= 'Orders')
```

Out[24]: <Axes: xlabel='Product_ID', ylabel='Orders'>

In [25]:
```python
pro_amn = 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 = pro_amn, x = 'Product_Category',y= 'Amount')
```

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



Conclusion

Married women of age group 26-35 yrs from UP, Maharastra and Karnataka working in IT, Healthcare and Aviation are more likely to buy products from Food, Clothing and Electronics category

In [ ]:

In [ ]: