In [4]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [10]:
```
df=pd.read_csv(r"C:\Users\Admin\Desktop\Diwali Sales Data.csv",encoding="un
```

In [11]:
```
df.shape
```

Out[11]: (11251, 15)

In [12]:
```
df.head()
```

Out[12]:

	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	Sc
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Sc
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	W

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

In [21]:
```
df.drop(["Status","unnamed1"], axis=1, inplace=True)
```

In [22]:
```
pd.isnull(df)
```

Out[22]:

	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 [23]:
```
pd.isnull(df).sum()
```

Out[23]:
```
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 [26]:
```
df.dropna(inplace=True)
```

In [27]:
```
pd.isnull(df.sum())
```

Out[27]:
```
User_ID             False
Cust_name           False
Product_ID          False
Gender              False
Age Group           False
Age                 False
Marital_Status      False
State               False
Zone                False
Occupation          False
Product_Category    False
Orders              False
Amount              False
dtype: bool
```

In [28]:
```
df["Amount"]=df["Amount"].astype("int")
```

In [30]:
```
df["Amount"].dtypes
```

Out[30]:
```
dtype('int32')
```

In [31]:
```
df.describe()
```

Out[31]:

	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 [35]:
```
df[["Age","Orders","Amount"]].describe()
```

Out[35]:

	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

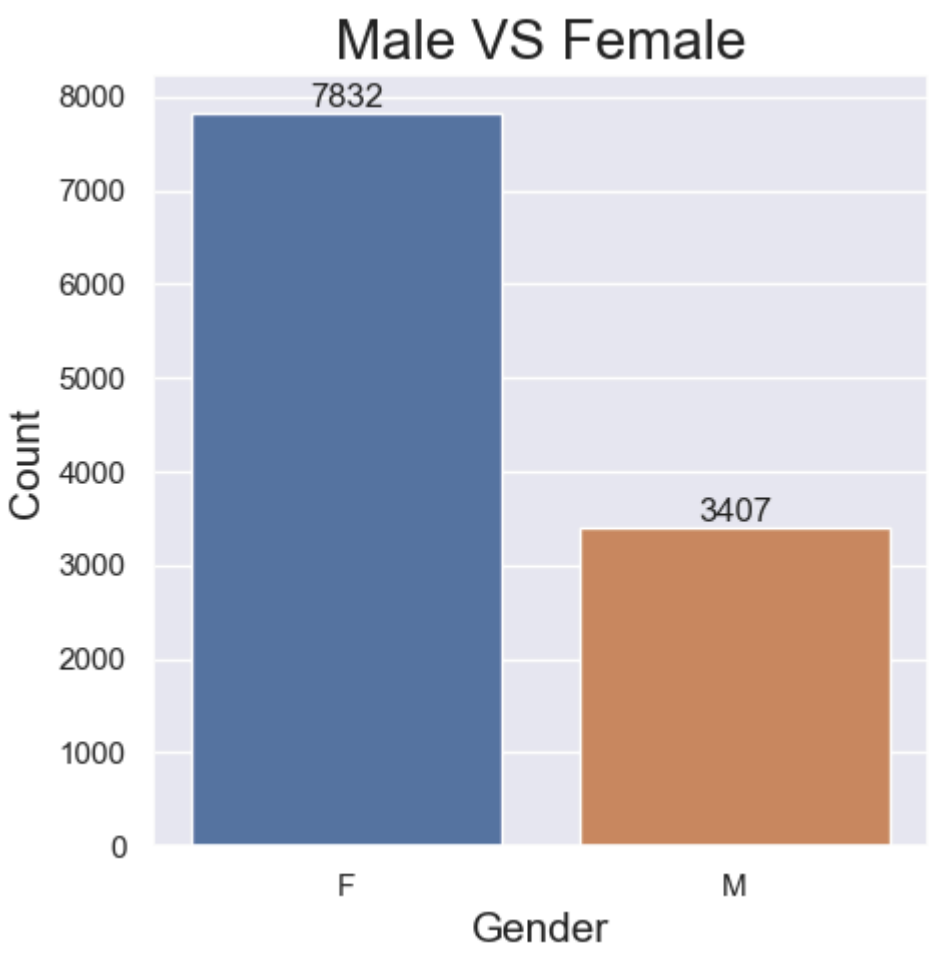
Exploratory Data Analysis

In [39]:
```
df.columns
```

Out[39]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Categor
y',
       'Orders', 'Amount'],
      dtype='object')
```

In [137]:
```
ax=sns.countplot(x="Gender", data=df)
sns.set(rc={"figure.figsize":(5,5)})
for bars in ax.containers:
    ax.bar_label(bars)

plt.title("Male VS Female",fontsize=20)
plt.xlabel("Gender",fontsize=15)
plt.ylabel("Count",fontsize=15)
```

Out[137]:
```
Text(0, 0.5, 'Count')
```



In [47]:
```
df.groupby(["Gender"],as_index=False)["Amount"].sum().sort_values(by="Amoun
```

Out[47]:

	Gender	Amount
0	F	74335853
1	M	31913276

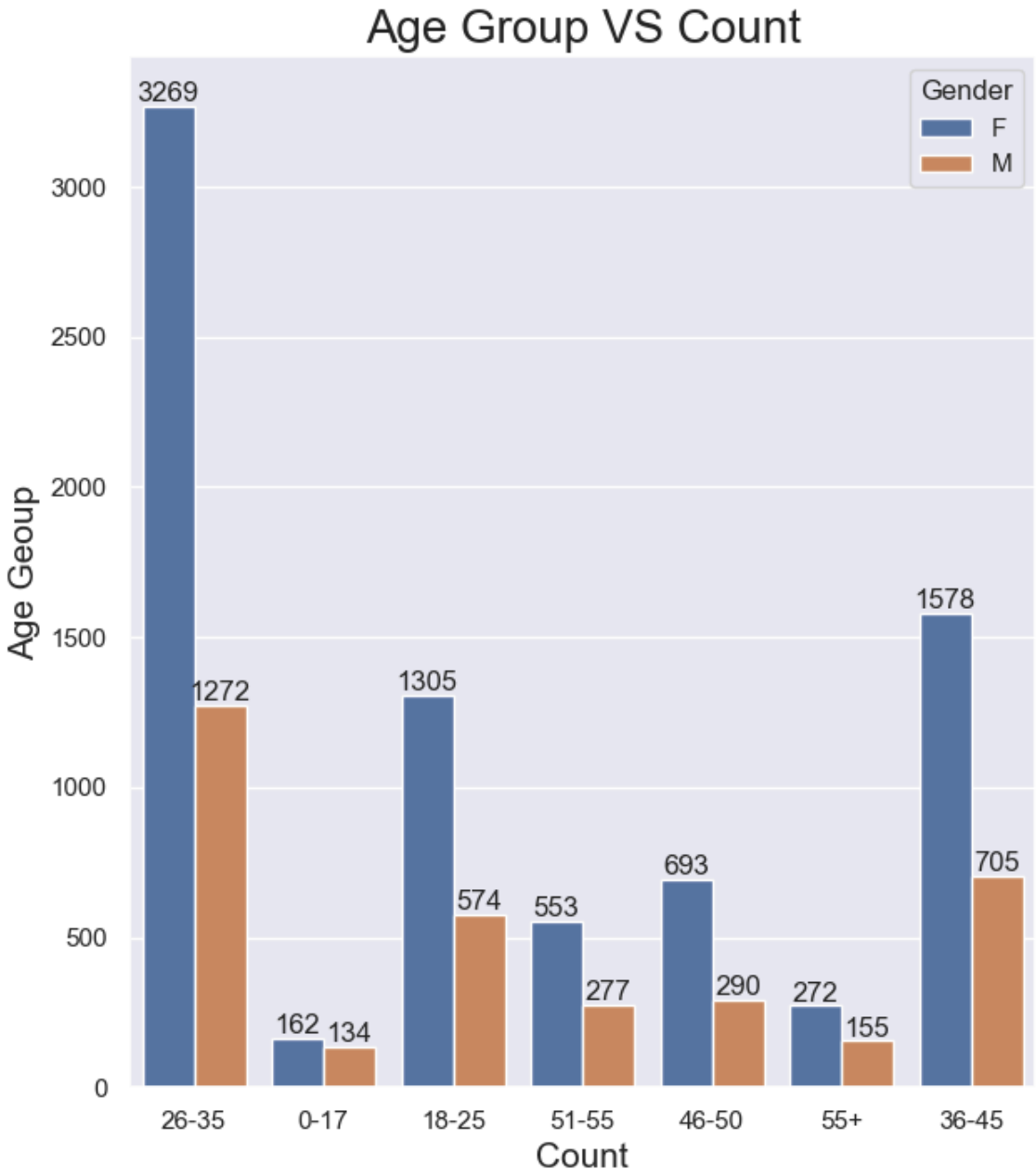
from above graph we can see that most of the buyer are females and even the ourchasing power of the female are greater than men

In [48]:
```
df.columns
```

Out[48]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Categor
y',
       'Orders', 'Amount'],
      dtype='object')
```

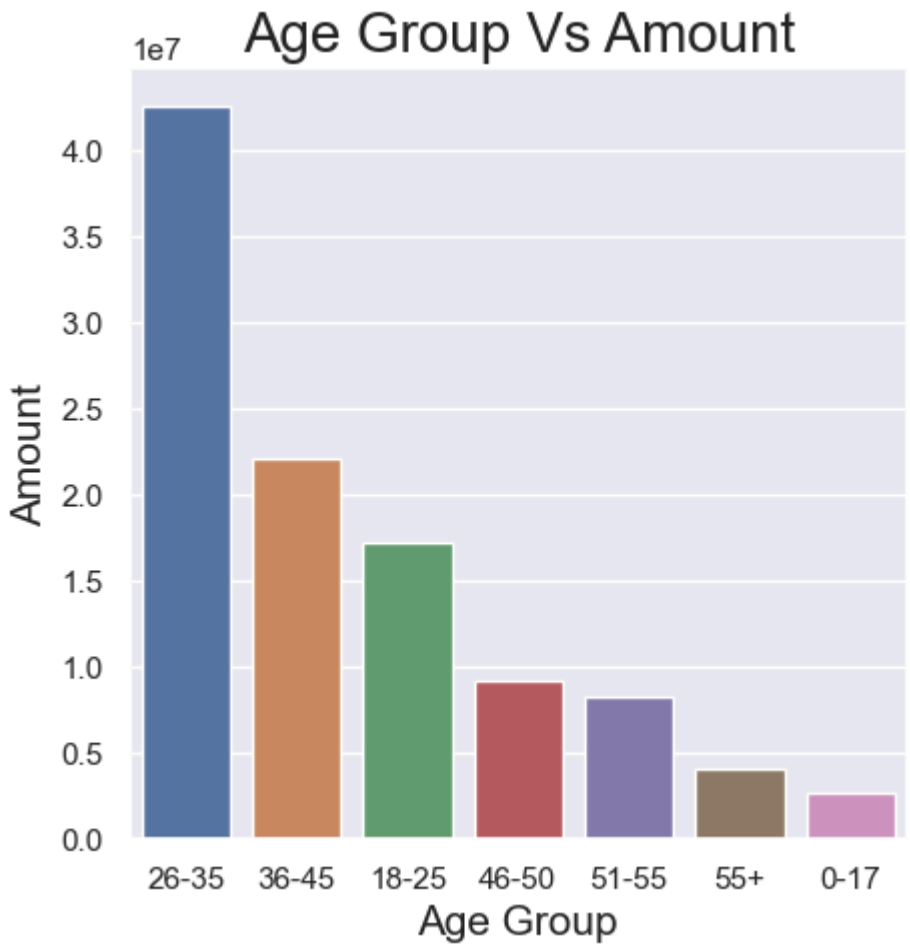
In [143]:
```
ax=sns.countplot(data=df ,x="Age Group", hue="Gender")
sns.set(rc={"figure.figsize":(7,8)})
plt.title("Age Group VS Count",fontsize=20)
for bars in ax.containers:
        ax.bar_label(bars)

plt.xlabel("Count",fontsize=15)
plt.ylabel("Age Geoup",fontsize=15)
```

Out[143]: Text(0, 0.5, 'Age Geoup')



In [145]:
```
#total Amount vs Age Group
sales_age=df.groupby(["Age Group"],as_index=False)["Amount"].sum().sort_val
sns.set(rc={"figure.figsize":(5,5)})
sns.barplot(x="Age Group",y="Amount",data=sales_age)
plt.title("Age Group Vs Amount",fontsize=20)
plt.xlabel("Age Group",fontsize=15)
plt.ylabel("Amount",fontsize=15)
```

Out[145]:
```
Text(0, 0.5, 'Amount')
```



From above graph we can see that most of the buyers are of age group between 26-35 yrs Female

In [66]:
```
df.columns
```

Out[66]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Categor
y',
       'Orders', 'Amount'],
      dtype='object')
```

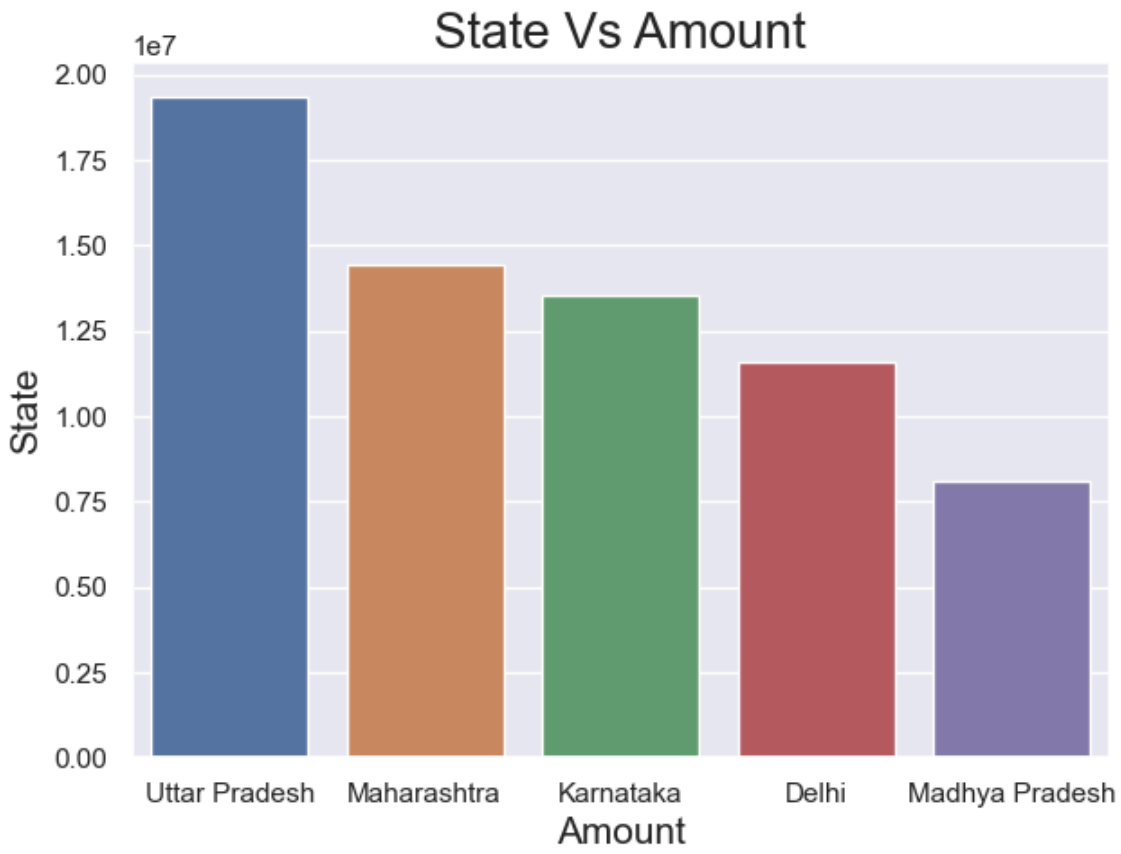
In [148]:

```
sales_State = df.groupby(["State"], as_index=False)["Orders"].sum().sort_va
sns.set(rc={"figure.figsize":(8,5)})
sns.barplot(data=sales_State, x="State", y="Orders")
plt.title("State Vs Orders",fontsize=20)
plt.xlabel("Orders",fontsize=15)
plt.ylabel("State",fontsize=15)
```

Out[148]: Text(0, 0.5, 'State')

In [149]:
```
sales_State = df.groupby(["State"], as_index=False)["Amount"].sum().sort_va
sns.set(rc={"figure.figsize":(7,5)})
sns.barplot(data=sales_State, x="State", y="Amount")
plt.title("State Vs Amount",fontsize=20)
plt.xlabel("Amount",fontsize=15)
plt.ylabel("State",fontsize=15)
```

Out[149]:
```
Text(0, 0.5, 'State')
```



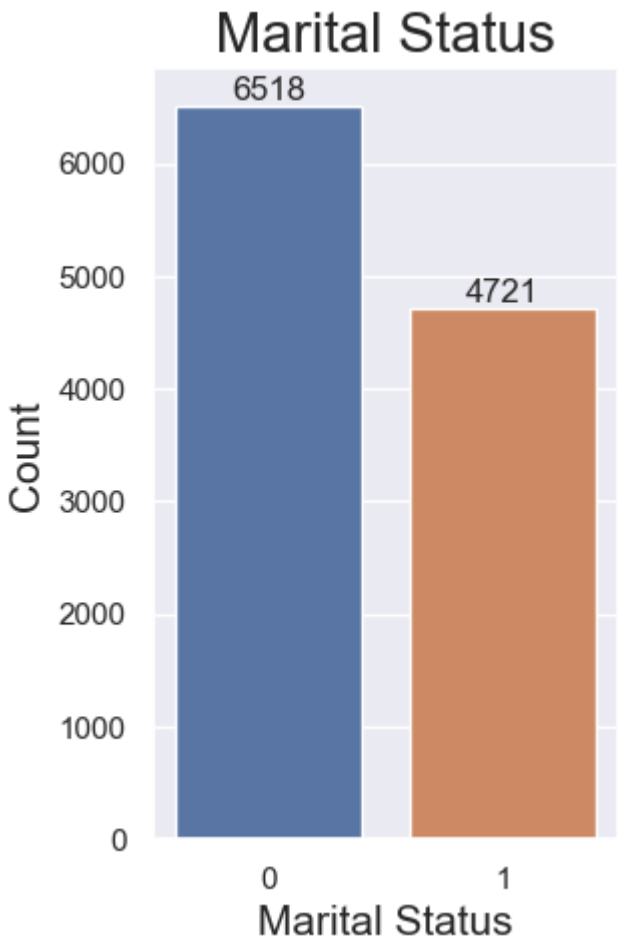
from above graph we can see that most of the orders & total amount are from Uttar Pradesh ,Maharashtra and Karnataka respectively

In [89]:
```
df.columns
```

Out[89]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Categor
y',
       'Orders', 'Amount'],
      dtype='object')
```

In [105]:
```
ax = sns.countplot(data=df,x="Marital_Status")
sns.set(rc={"figure.figsize":(7,5)})
for bars in ax.containers:
    ax.bar_label(bars)
plt.title("Marital Status",fontsize=20)
plt.xlabel("Marital Status",fontsize=15)
plt.ylabel("Count",fontsize=15)
```

Out[105]:
```
Text(0, 0.5, 'Count')
```

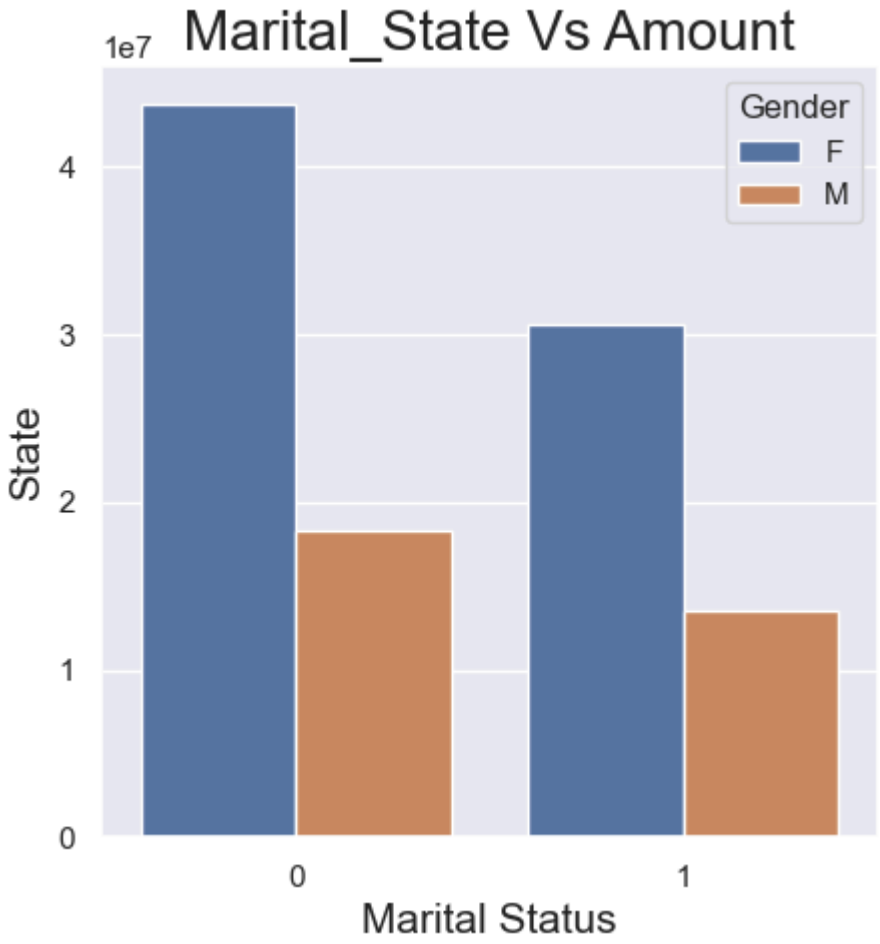


In [108]:
```
df.columns
```

Out[108]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Categor
y',
       'Orders', 'Amount'],
      dtype='object')
```

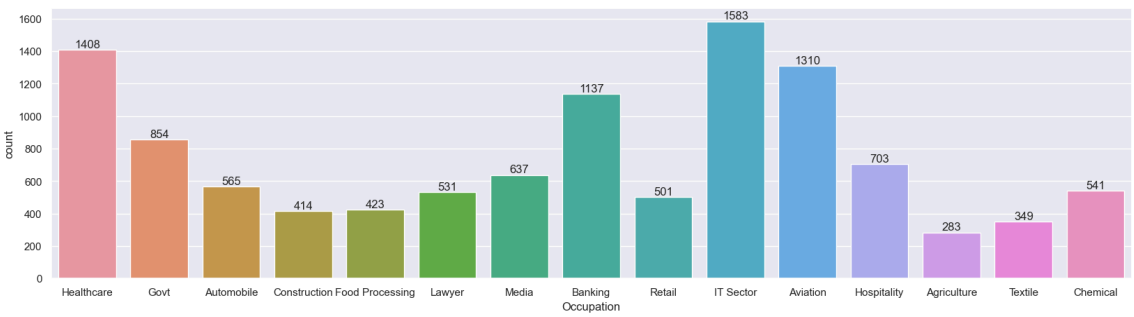
In [151]:
```
sales_State = df.groupby(["Marital_Status","Gender"], as_index=False)["Amou
sns.set(rc={"figure.figsize":(5,5)})
sns.barplot(data=sales_State, x="Marital_Status", y="Amount",hue="Gender")

plt.title("Marital_State Vs Amount",fontsize=20)
plt.xlabel("Marital Status",fontsize=15)
plt.ylabel("State",fontsize=15)
```

Out[151]: Text(0, 0.5, 'State')



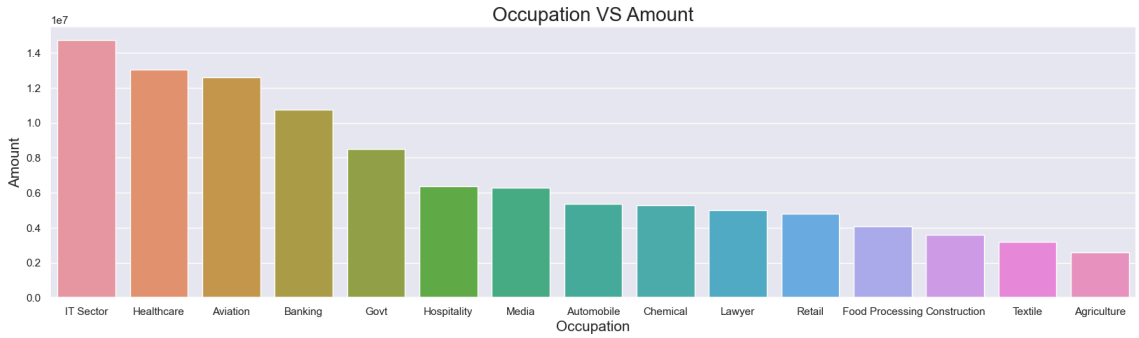
from above graph we can see that most of the buyers are married women they have high pirchasing power

In [115]:
```
sns.set(rc={"figure.figsize":(20,5)})
ax=sns.countplot(data=df,x="Occupation")
for bars in ax.containers:
    ax.bar_label(bars)
```



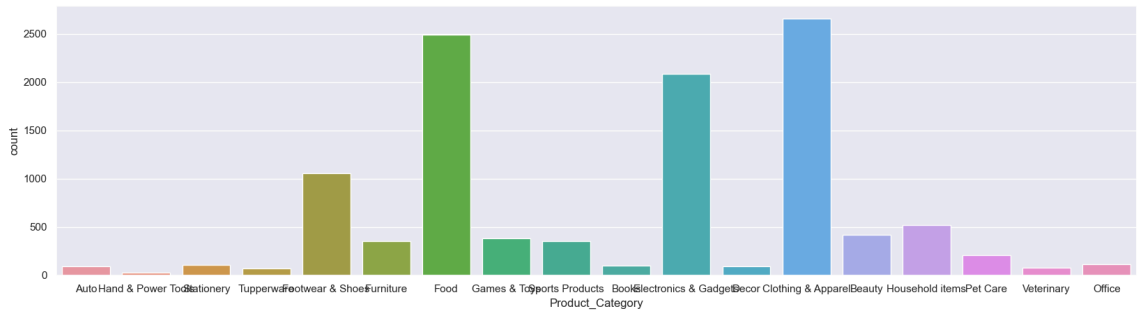
In [118]:
```
sales_State = df.groupby(["Occupation"], as_index=False)["Amount"].sum().so
sns.set(rc={"figure.figsize":(20,5)})
sns.barplot(data=sales_State, x="Occupation", y="Amount")
plt.xlabel('Occupation',fontsize=15)
plt.ylabel('Amount',fontsize=15)
plt.title("Occupation VS Amount",fontsize=20)
```

Out[118]:
```
Text(0.5, 1.0, 'Occupation VS Amount')
```



from above graph we can see that most of the buyers are working in IT , Healthcare and Aviation Secctor

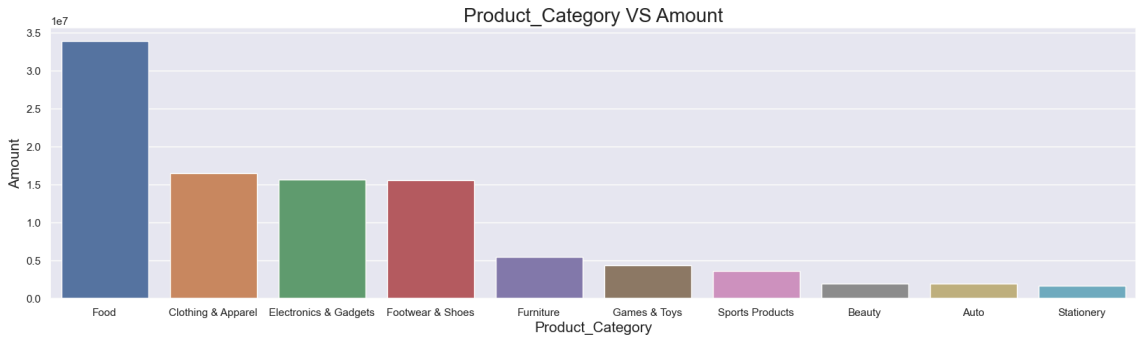
In [124]:
```
sns.set(rc={"figure.figsize":(20,5)})
sns.countplot(data=df, x="Product_Category")
for bars in ax.containers:
    ax.bar_label(bars)
```



In [122]:
```
df.columns
```

Out[122]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Categor
y',
       'Orders', 'Amount'],
      dtype='object')
```

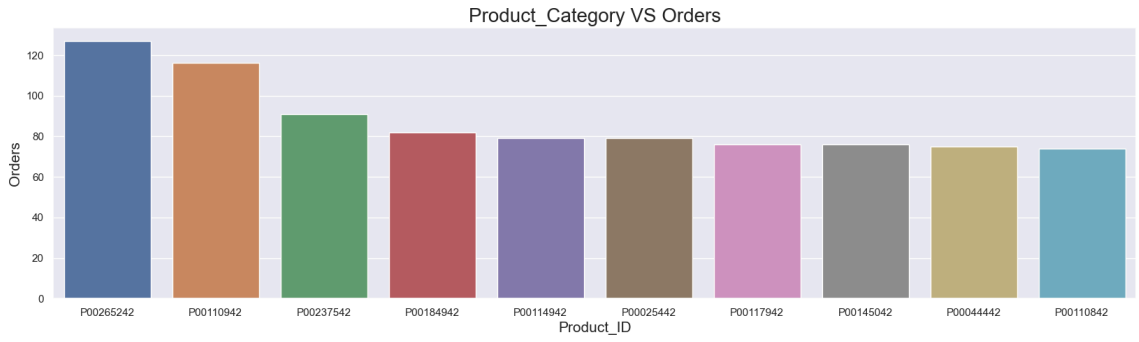
In [129]:
```
sales_State = df.groupby(["Product_Category"], as_index=False)["Amount"].su
sns.set(rc={"figure.figsize":(20,5)})
sns.barplot(data=sales_State, x="Product_Category",y="Amount")
plt.xlabel('Product_Category',fontsize=15)
plt.ylabel('Amount',fontsize=15)
plt.title("Product_Category VS Amount",fontsize=20)
for bars in ax.containers:
    ax.bar_label(bars)
```



from above graph we can see that most of the sold product are from Food,Clothing,Footwear and Electronics category

In [132]:
```
sales_State = df.groupby(["Product_ID"], as_index=False)["Orders"].sum().so
sns.set(rc={"figure.figsize":(20,5)})
sns.barplot(data=sales_State, x="Product_ID",y="Orders")
plt.xlabel('Product_ID',fontsize=15)
plt.ylabel('Orders',fontsize=15)
plt.title("Product_Category VS Orders",fontsize=20)
```

Out[132]:
```
Text(0.5, 1.0, 'Product_Category VS Orders')
```



Conclusion

Married women age group 26-35 yers from up,Maharasthra and karnataka working in IT,Healthcare and Aviation are more likely to buy products from food,clothing and Electronics category.

In [ ]:
```
```