```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

df = pd.read_csv("C:\\Users\\Sony\\Desktop\\pandas practice\\DIWALI
SALES\\Diwali Sales Data.csv",encoding = 'unicode_escape' )
df
```

```
       User_ID    Cust_name Product_ID Gender Age Group  Age
Marital_Status  \
0      1002903    Sanskriti  P00125942      F     26-35   28
0
1      1000732       Kartik  P00110942      F     26-35   35
1
2      1001990        Bindu  P00118542      F     26-35   35
1
3      1001425       Sudevi  P00237842      M      0-17   16
0
4      1000588         Joni  P00057942      M     26-35   28
1
...        ...          ...        ...    ...       ...  ...
...
11246  1000695      Manning  P00296942      M     18-25   19
1
11247  1004089  Reichenbach  P00171342      M     26-35   33
0
11248  1001209        Oshin  P00201342      F     36-45   40
0
11249  1004023       Noonan  P00059442      M     36-45   37
0
11250  1002744      Brumley  P00281742      F     18-25   19
0

                State      Zone       Occupation Product_Category
Orders  \
0         Maharashtra   Western       Healthcare             Auto
1
1      Andhra Pradesh  Southern             Govt             Auto
3
2       Uttar Pradesh   Central       Automobile             Auto
3
3           Karnataka  Southern     Construction             Auto
2
4             Gujarat   Western  Food Processing             Auto
2
...               ...       ...              ...              ...
...
11246     Maharashtra   Western         Chemical           Office
4
```

```
11247         Haryana  Northern       Healthcare       Veterinary
3
11248  Madhya Pradesh   Central          Textile           Office
4
11249       Karnataka  Southern      Agriculture           Office
3
11250     Maharashtra   Western       Healthcare           Office
3

        Amount  Status  unnamed1
0      23952.0     NaN       NaN
1      23934.0     NaN       NaN
2      23924.0     NaN       NaN
3      23912.0     NaN       NaN
4      23877.0     NaN       NaN
...        ...     ...       ...
11246    370.0     NaN       NaN
11247    367.0     NaN       NaN
11248    213.0     NaN       NaN
11249    206.0     NaN       NaN
11250    188.0     NaN       NaN

[11251 rows x 15 columns]
```

```
df.shape
```

```
(11251, 15)
```

```
df.head(2)
```

```
   User_ID  Cust_name Product_ID Gender Age Group  Age  Marital_Status
\
0  1002903  Sanskriti  P00125942      F     26-35   28               0

1  1000732     Kartik  P00110942      F     26-35   35               1


            State      Zone  Occupation Product_Category  Orders
Amount  \
0     Maharashtra   Western  Healthcare             Auto       1
23952.0
1  Andhra Pradesh  Southern        Govt             Auto       3
23934.0

   Status  unnamed1
0     NaN       NaN
1     NaN       NaN
```

```
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
```

```
df.drop(['Status','unnamed1' ],axis = 1, inplace = True)
df.head(2)
```

```
   User_ID  Cust_name Product_ID Gender Age Group  Age  Marital_Status
\
0  1002903  Sanskriti  P00125942      F     26-35   28               0

1  1000732     Kartik  P00110942      F     26-35   35               1


            State      Zone  Occupation Product_Category  Orders  
Amount  
0     Maharashtra   Western  Healthcare             Auto       1  
23952.0  
1  Andhra Pradesh  Southern        Govt             Auto       3  
23934.0  
```

```
#check null values
pd.isnull(df).sum()
```

```
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
```

```
df.dropna(inplace =True)
df.shape
```

```
(11239, 13)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 11239 entries, 0 to 11250
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   User_ID           11239 non-null  int64  
 1   Cust_name         11239 non-null  object 
 2   Product_ID        11239 non-null  object 
 3   Gender            11239 non-null  object 
 4   Age Group         11239 non-null  object 
 5   Age               11239 non-null  int64  
 6   Marital_Status    11239 non-null  int64  
 7   State             11239 non-null  object 
 8   Zone              11239 non-null  object 
 9   Occupation        11239 non-null  object 
 10  Product_Category  11239 non-null  object 
 11  Orders            11239 non-null  int64  
 12  Amount            11239 non-null  float64
dtypes: float64(1), int64(4), object(8)
memory usage: 1.2+ MB
```

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

```
df['Amount'].dtype
```

```
dtype('int32')
```

```
df.describe()
```

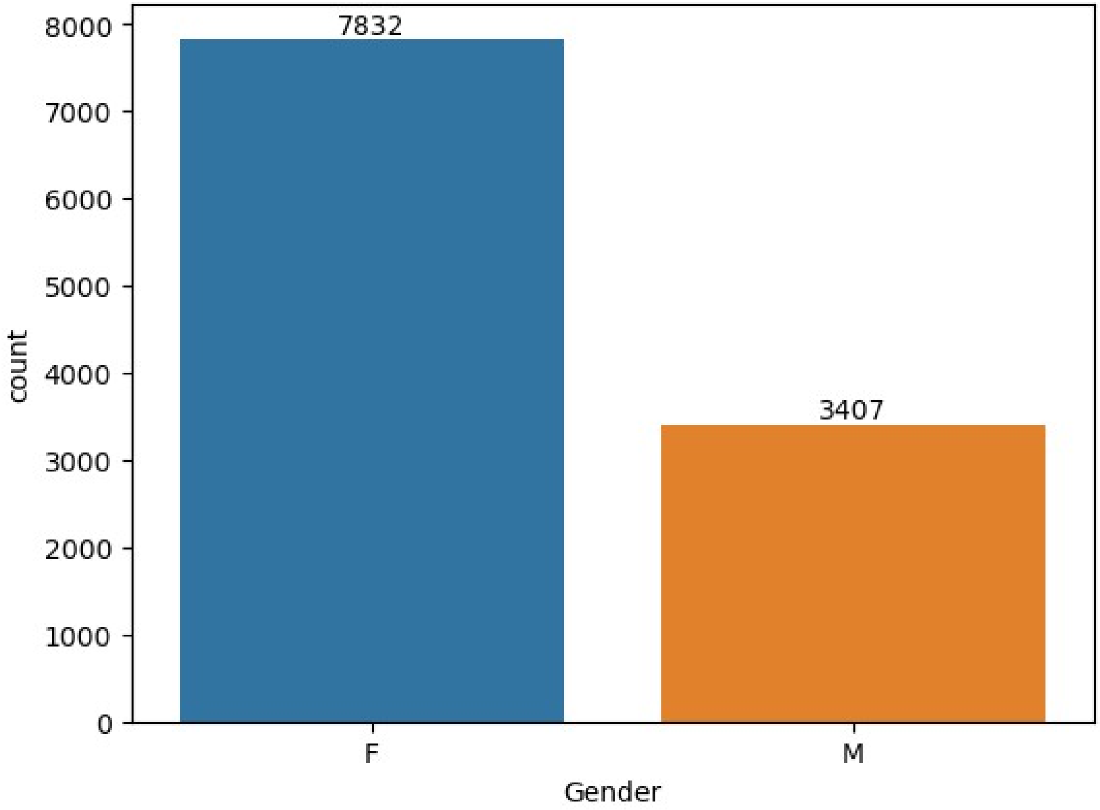
```
            User_ID           Age  Marital_Status        Orders
Amount
count  1.123900e+04  11239.000000    11239.000000  11239.000000
11239.000000
mean   1.003004e+06     35.410357        0.420055      2.489634
9453.610553
std    1.716039e+03     12.753866        0.493589      1.114967
```

```
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
```

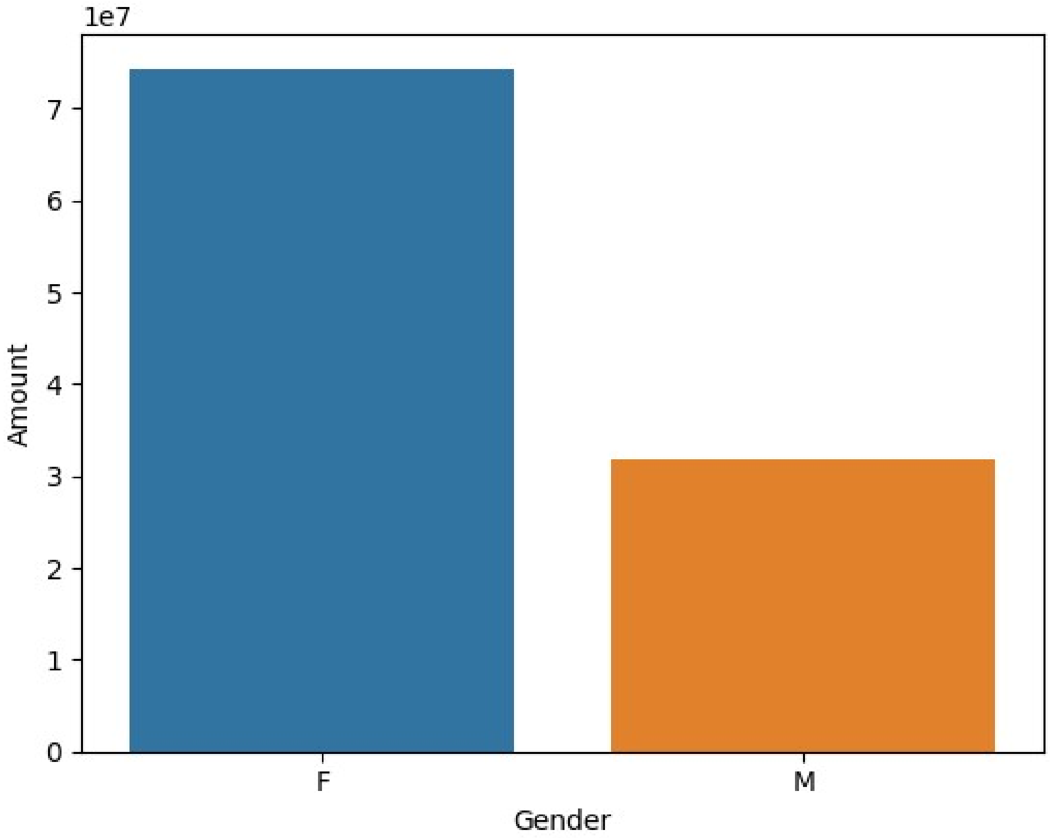
```
df.columns
```

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

```
ax = sns.countplot(x = "Gender", data = df)

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

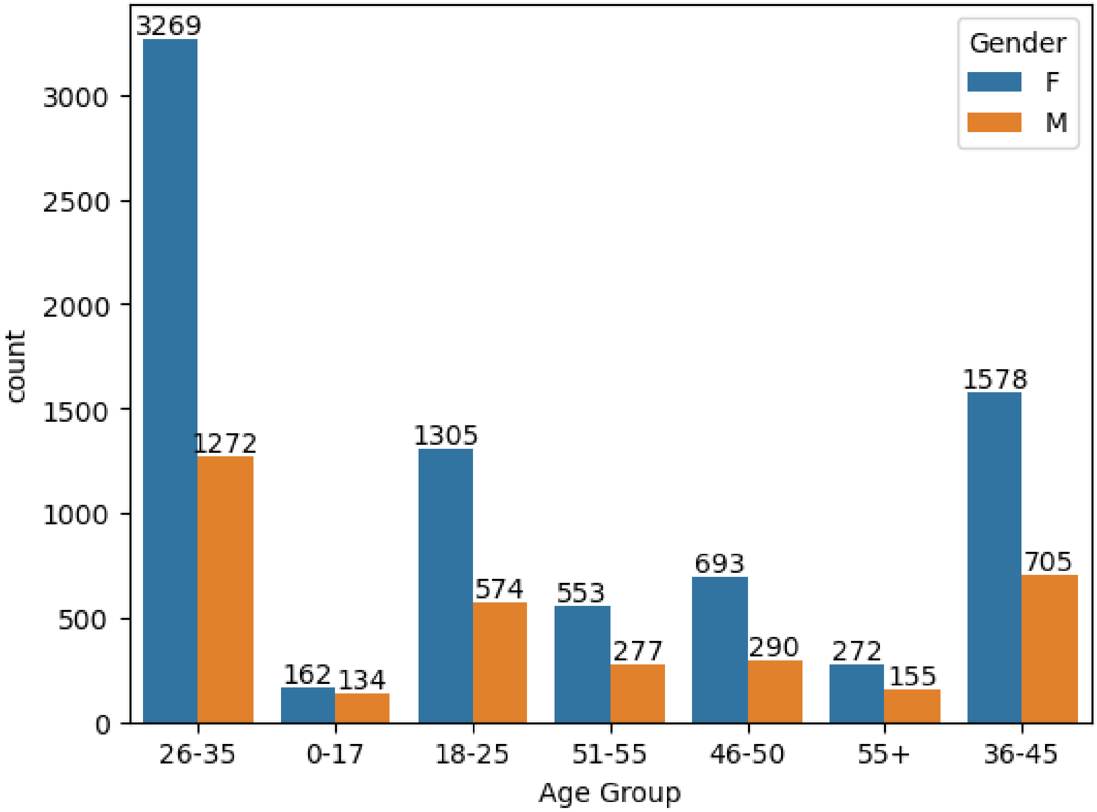


```
sales_gen = df.groupby(['Gender'], as_index =False)
['Amount'].sum().sort_values(by='Amount', ascending =False)
az = sns.barplot(x = "Gender", y = "Amount", data = sales_gen)
```

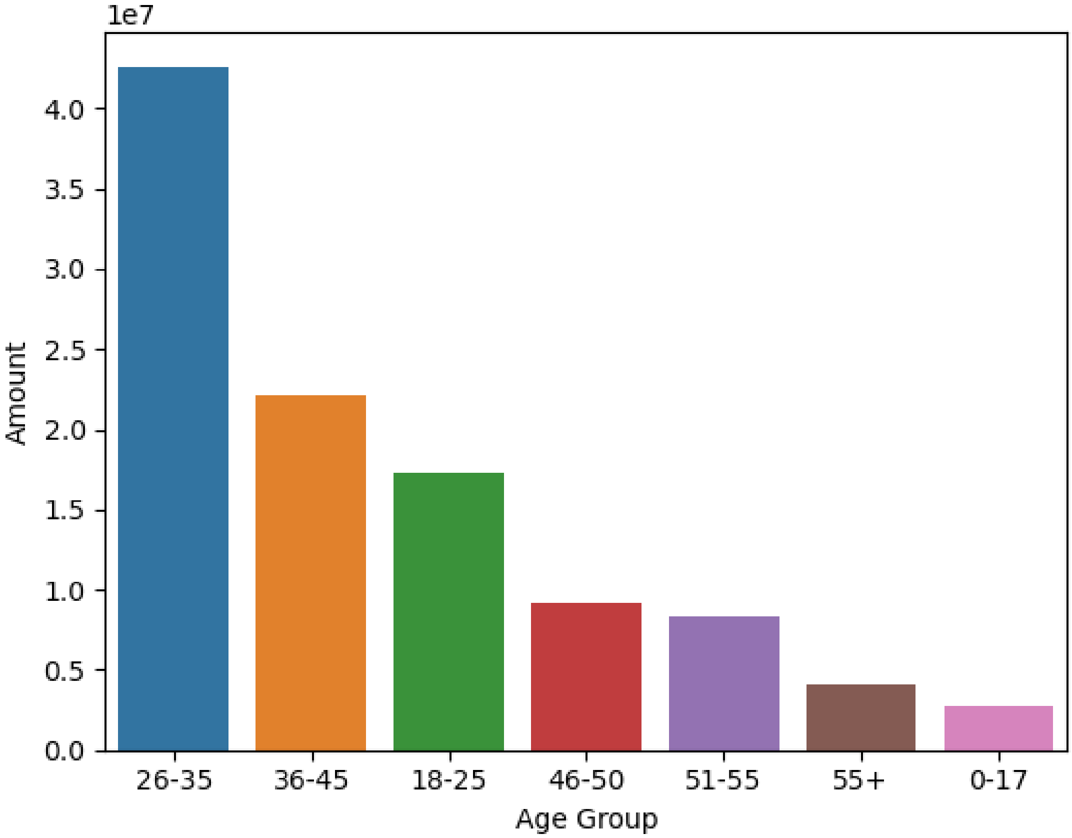


From the above data we can see that most of the buyers are female and even the purchasing power of females is more than men.

```
df.columns

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

ax = sns.countplot(x = 'Age Group' , data = df ,hue ="Gender" )

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

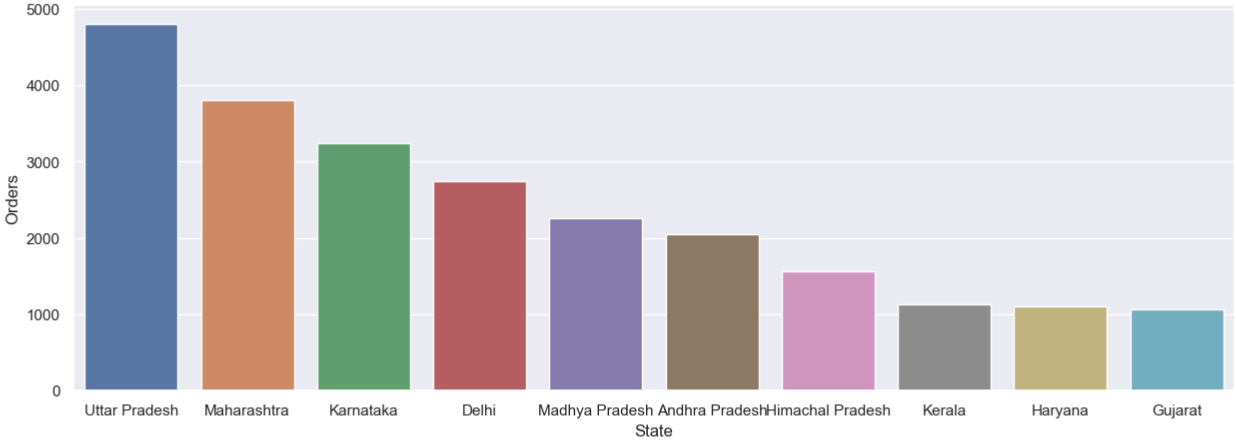


```
sales_age = df.groupby(['Age Group'] ,as_index = False)
['Amount'].sum().sort_values(by = 'Amount',ascending = False)

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



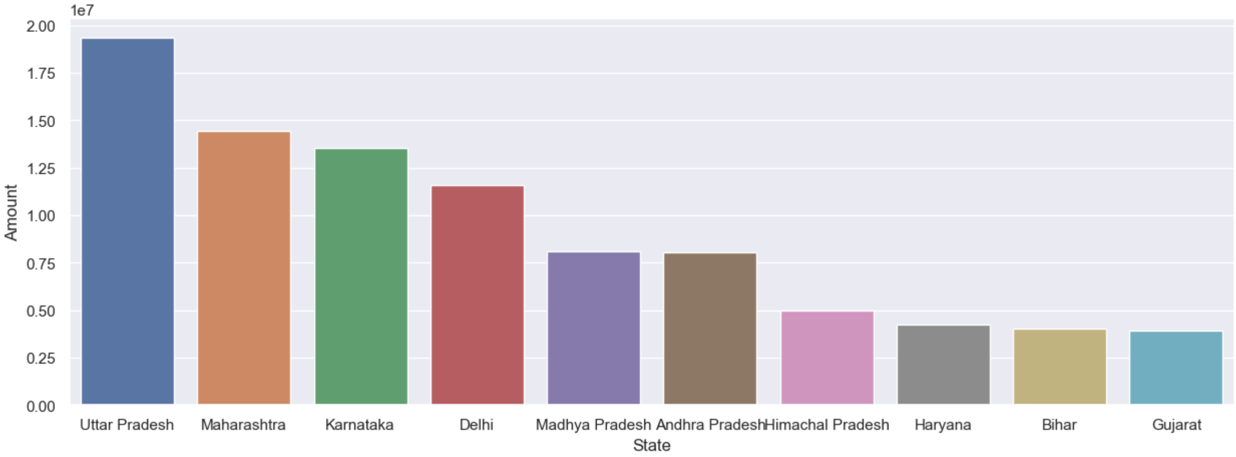
From the above we can see thet most of the buyers are from ager group 26-35 and are females.

```
df.columns

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

State_sale = df.groupby(['State'] , as_index = False)
['Orders'].sum().sort_values(by='Orders' , ascending = False).head(10)

ax = sns.set(rc ={'figure.figsize':(15,5)})
sns.barplot(x = 'State', y = "Orders", data =State_sale )



<Axes: xlabel='State', ylabel='Orders'>
```



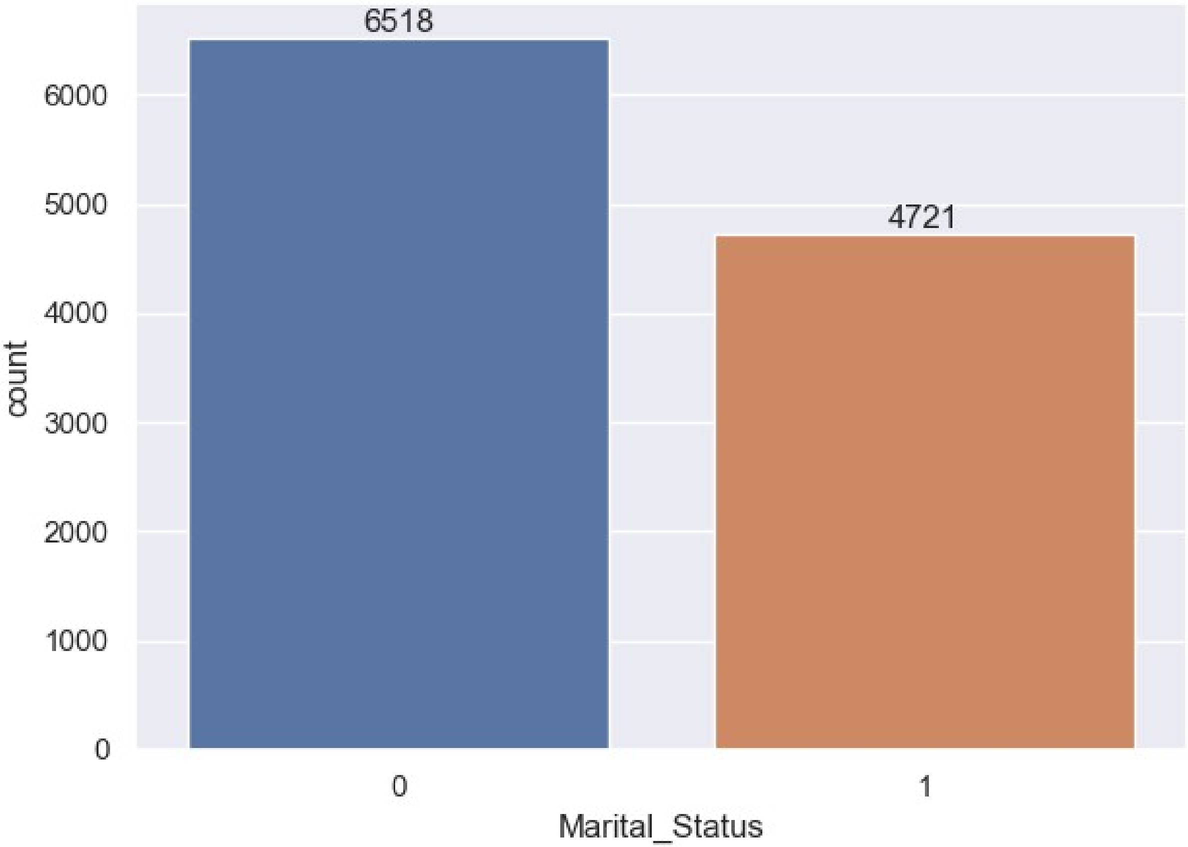
```
State_sale = df.groupby(['State'] , as_index = False)
['Amount'].sum().sort_values(by='Amount' , ascending = False).head(10)

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

```
<Axes: xlabel='State', ylabel='Amount'>
```



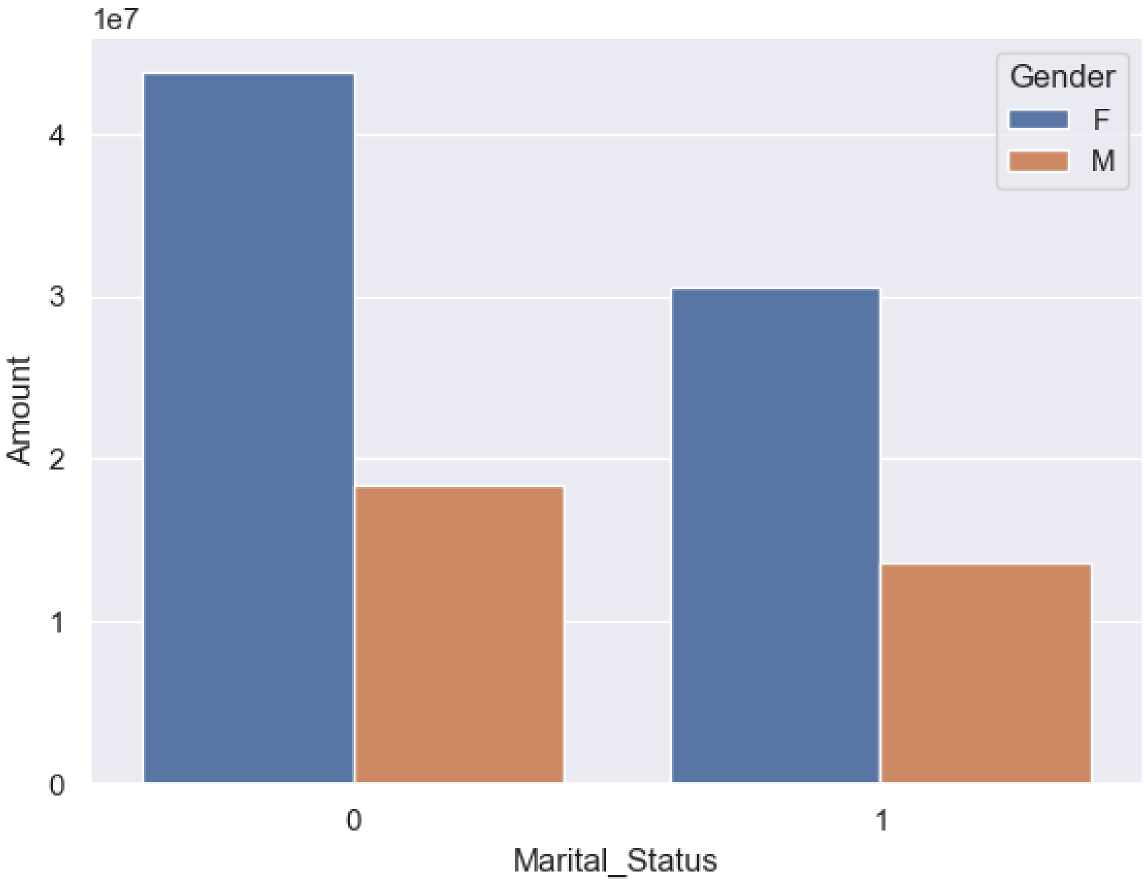
We can see that most of the orders are from UP , Maharashtra and karnatka and their purchasing pwer is also maximum

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



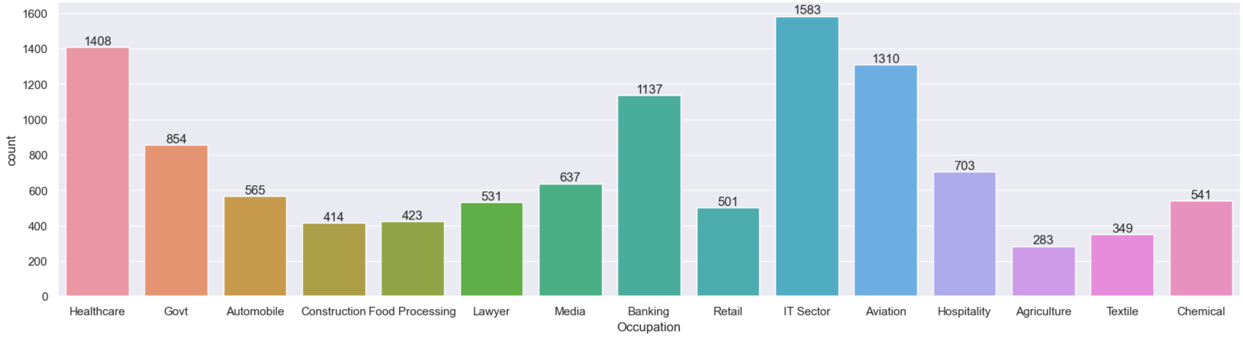
```
m_sale= df.groupby(['Marital_Status','Gender'] , as_index = False)
['Amount'].sum().sort_values(by='Amount' , ascending = False).head(10)

ax = sns.set(rc={'figure.figsize':(7,5)})
sns.barplot(x = 'Marital_Status', y = "Amount", data =m_sale, hue
='Gender')
```

```
<Axes: xlabel='Marital_Status', ylabel='Amount'>
```



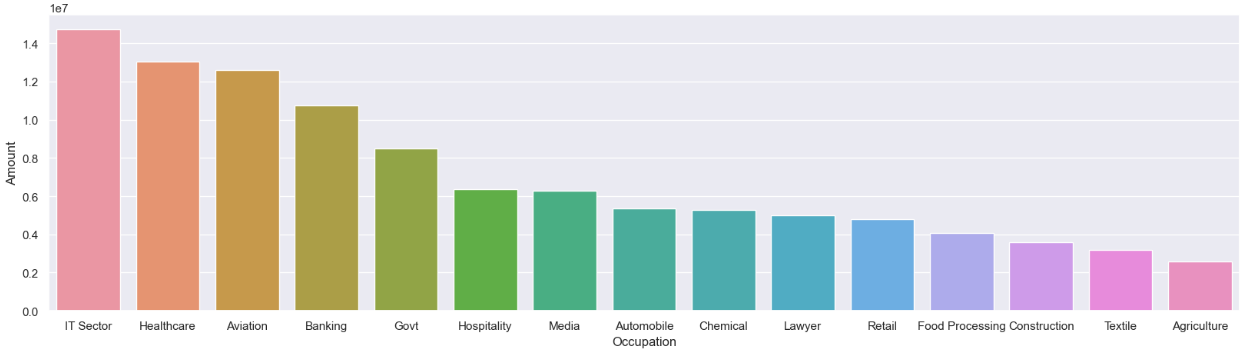
From the above data we can see that women who are married have the maximum purchasing power.

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



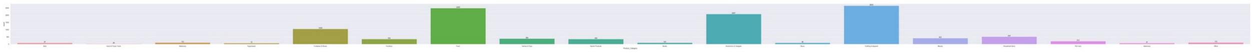
```
o_sale= df.groupby(['Occupation'] , as_index = False)
['Amount'].sum().sort_values(by='Amount' , ascending = False)
```

```
ax = sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(x = 'Occupation', y = "Amount", data =o_sale)

<Axes: xlabel='Occupation', ylabel='Amount'>
```

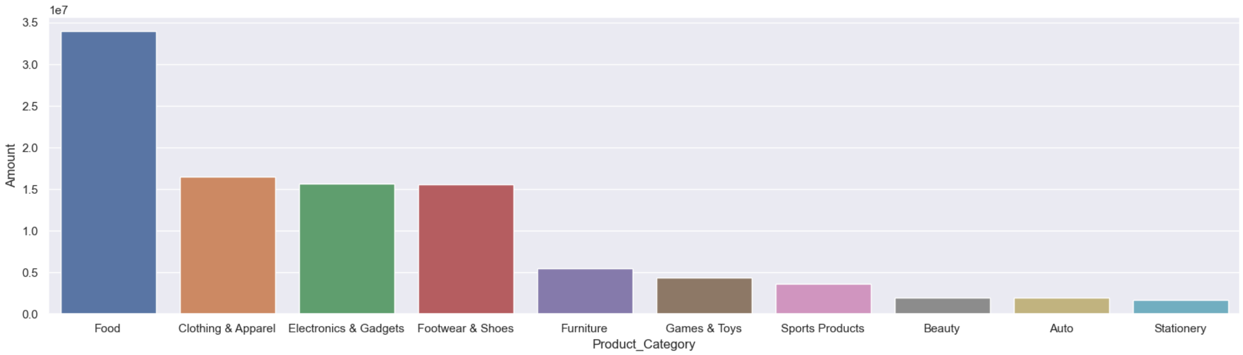


From the above we can see that most of the buyers are working in IT, Aviation and Healthcare sector.

```
df.columns

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

ax = sns.countplot(x =  'Product_Category', data = df)
sns.set(rc={'figure.figsize':(15,5)})
for bars in ax.containers:
    ax.bar_label(bars)
```



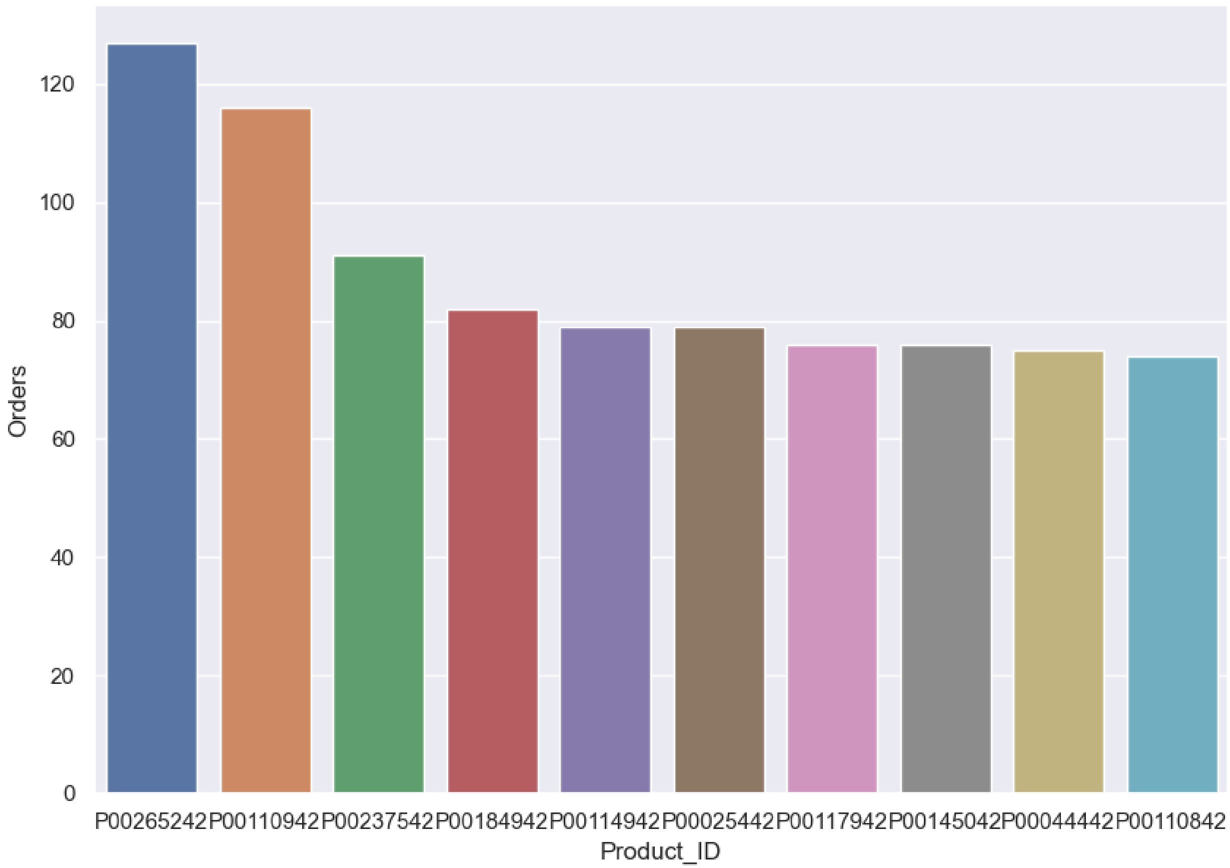
```
p_sale= df.groupby(['Product_Category'] , as_index = False)
['Amount'].sum().sort_values(by='Amount' , ascending = False).head(10)

ax = sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(x = 'Product_Category', y = "Amount", data =p_sale)

<Axes: xlabel='Product_Category', ylabel='Amount'>
```



From the above we can see that mos of the revenue is generated from food, clothing , Electronic and footwears.

```
p_sale= df.groupby(['Product_ID'] , as_index = False)
['Orders'].sum().sort_values(by='Orders' , ascending = False).head(10)

ax = sns.set(rc={'figure.figsize':(10,7)})
sns.barplot(x = 'Product_ID', y = "Orders", data =p_sale)
```

```
<Axes: xlabel='Product_ID', ylabel='Orders'>
```



Conclusion

1. women especially who are married are the biggest buyers with maximum purchasing power.
2. UP, maharashtra and karnataka are the biggest markets with maximum purchasing power and maximum orders.
3. People wirking in IT, healthcare and Aviation holds maximum purchasing power.
4. food, clothing. electronic and foothwear are the biggest revenue generator