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

df = pd.read_csv(r"C:\Users\RAJNI PAL\Downloads\Diwali Sales
Data.csv",encoding='latin1')

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

```
df
```

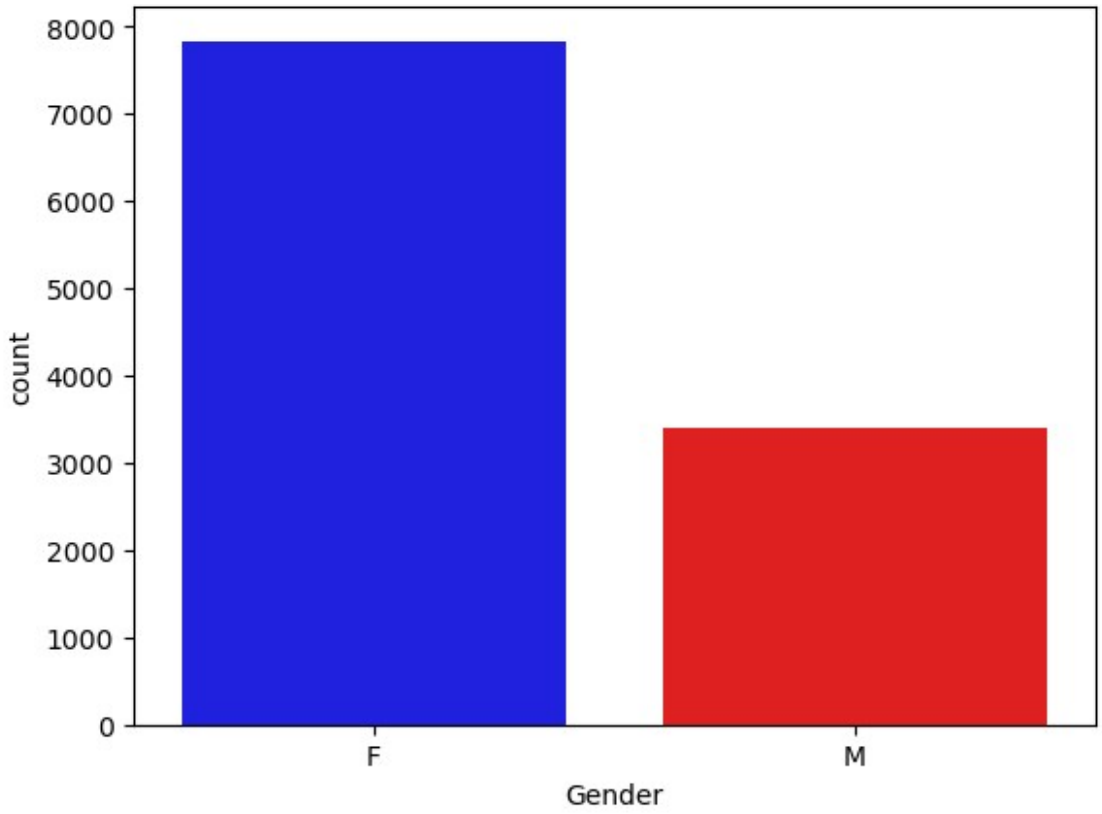
```
       User_ID    Cust_name Product_ID Gender Age Group  Age
Marital_Status  \
0      1002903    Sanskriti  P00125942      F     26-35   28
0
1      1000732       Kartik  P00110942      F     26-35   35
1
2      1001990        Bindu  P00118542      F     26-35   35
1
3      1001425       Sudevi  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
0       23952.0
1       23934.0
2       23924.0
3       23912.0
4       23877.0
...         ...
11246     370.0
11247     367.0
11248     213.0
11249     206.0
11250     188.0

[11251 rows x 13 columns]
```

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

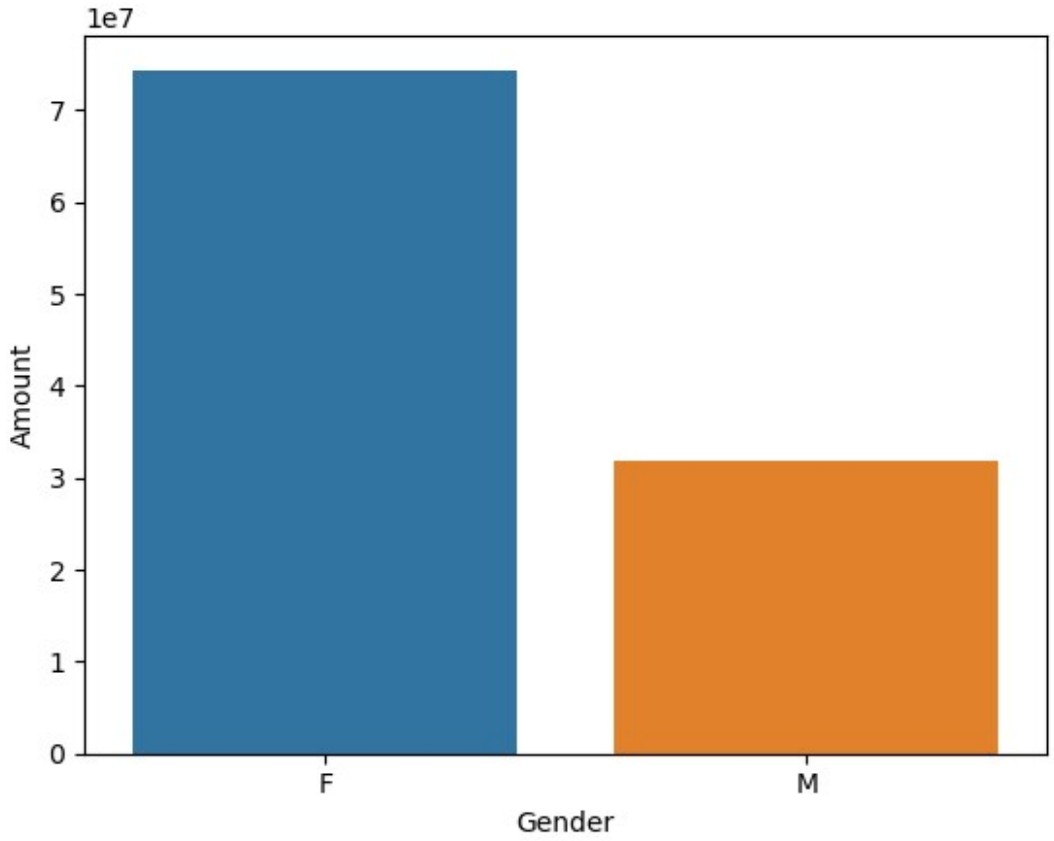
```
df["Amount"] = df["Amount"].astype("int")
```

```
ax = sns.countplot(x = "Gender",data = df,palette = ["blue","red"])
```



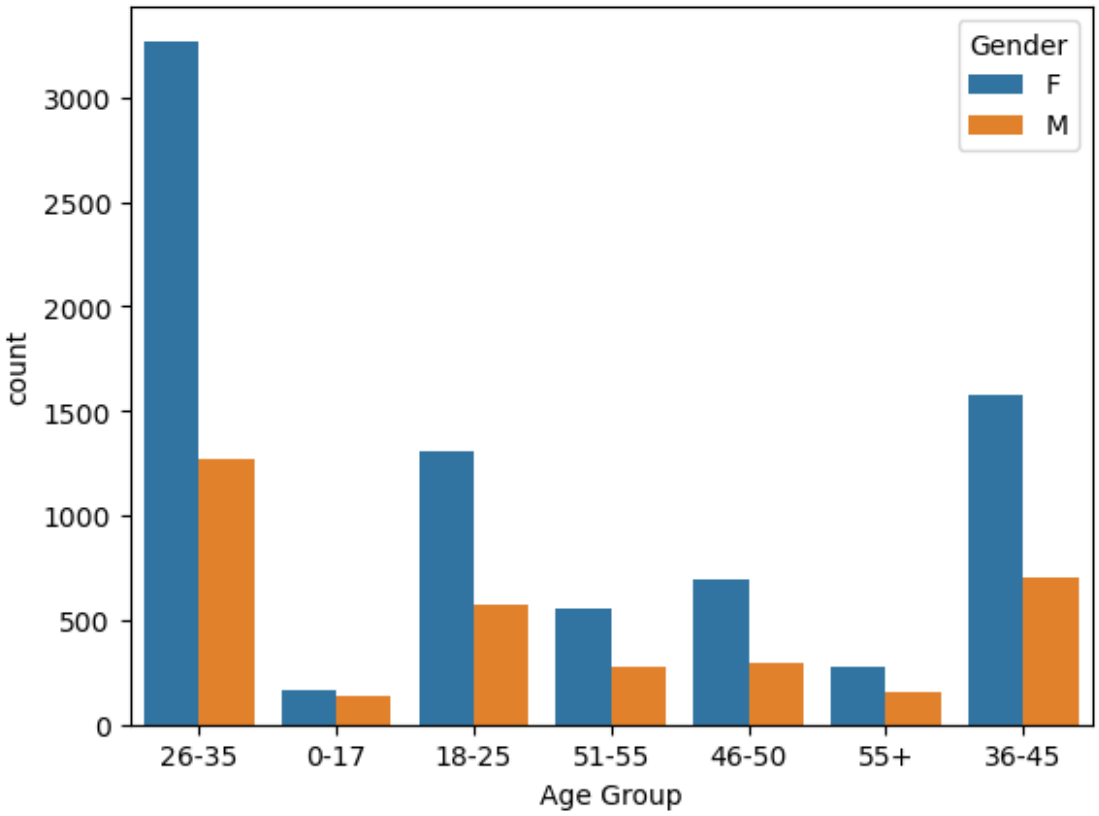

```
sales_gen  = df.groupby(["Gender"],as_index = False)
["Amount"].sum().sort_values(by = "Amount",ascending  = False)


sns.barplot(x = "Gender", y = "Amount", data = sales_gen)
```

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



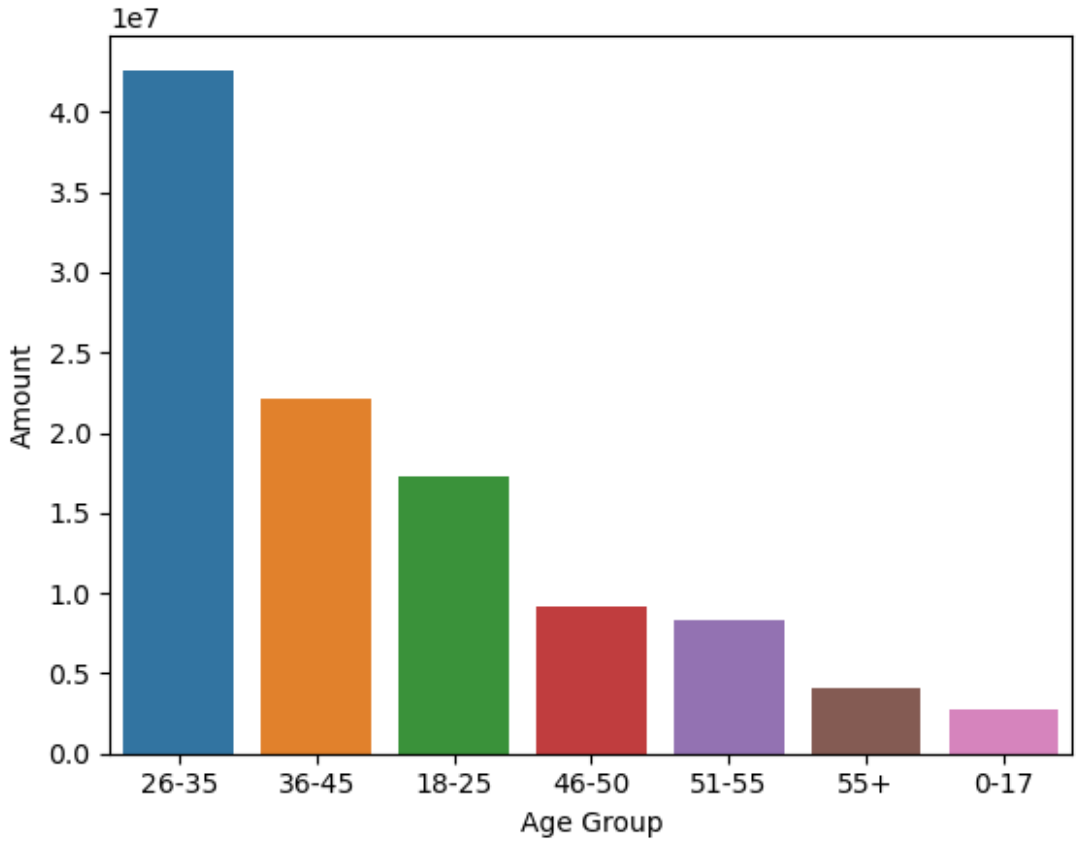
From above graphs we can conclude that female order more than men and have high purchasing power than men

```
count_age = sns.countplot(x = "Age Group", data = df,hue = "Gender")
```



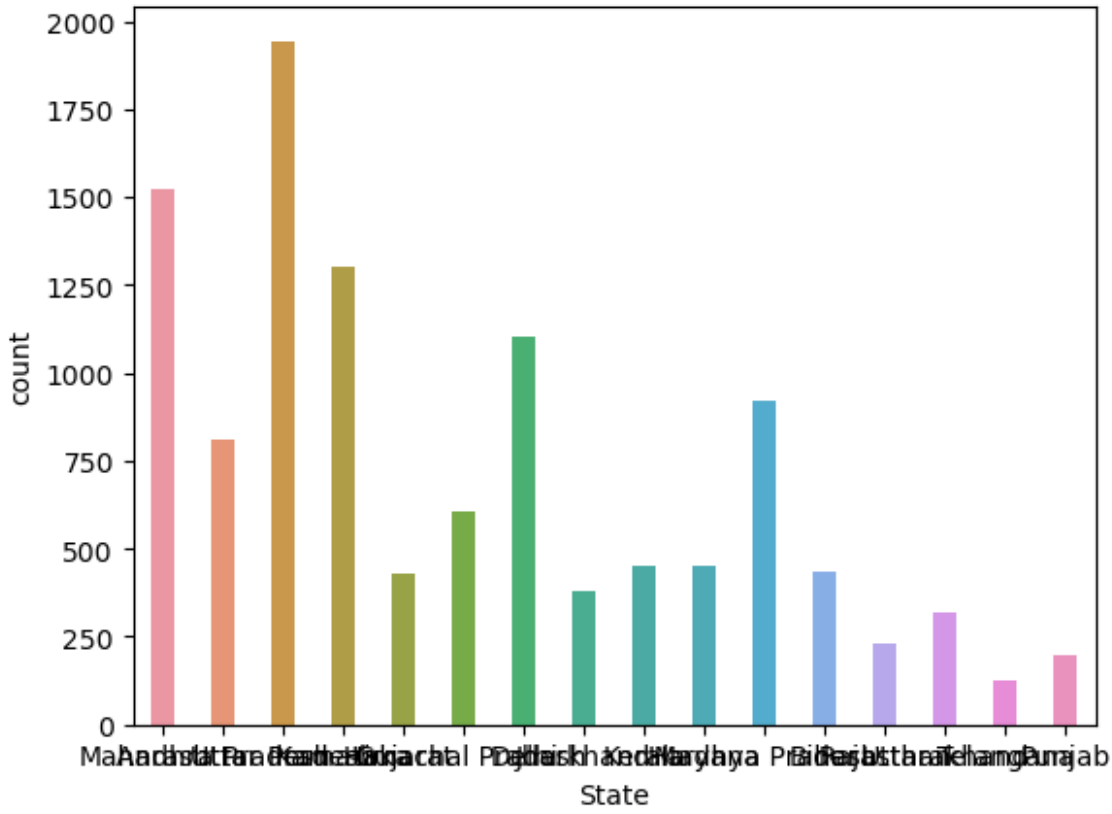

```
sales_age = df.groupby(["Age Group"],as_index = False)
["Amount"].sum().sort_values(by = "Amount",ascending = False)

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

```
<Axes: xlabel='Age Group', ylabel='Amount'>
```



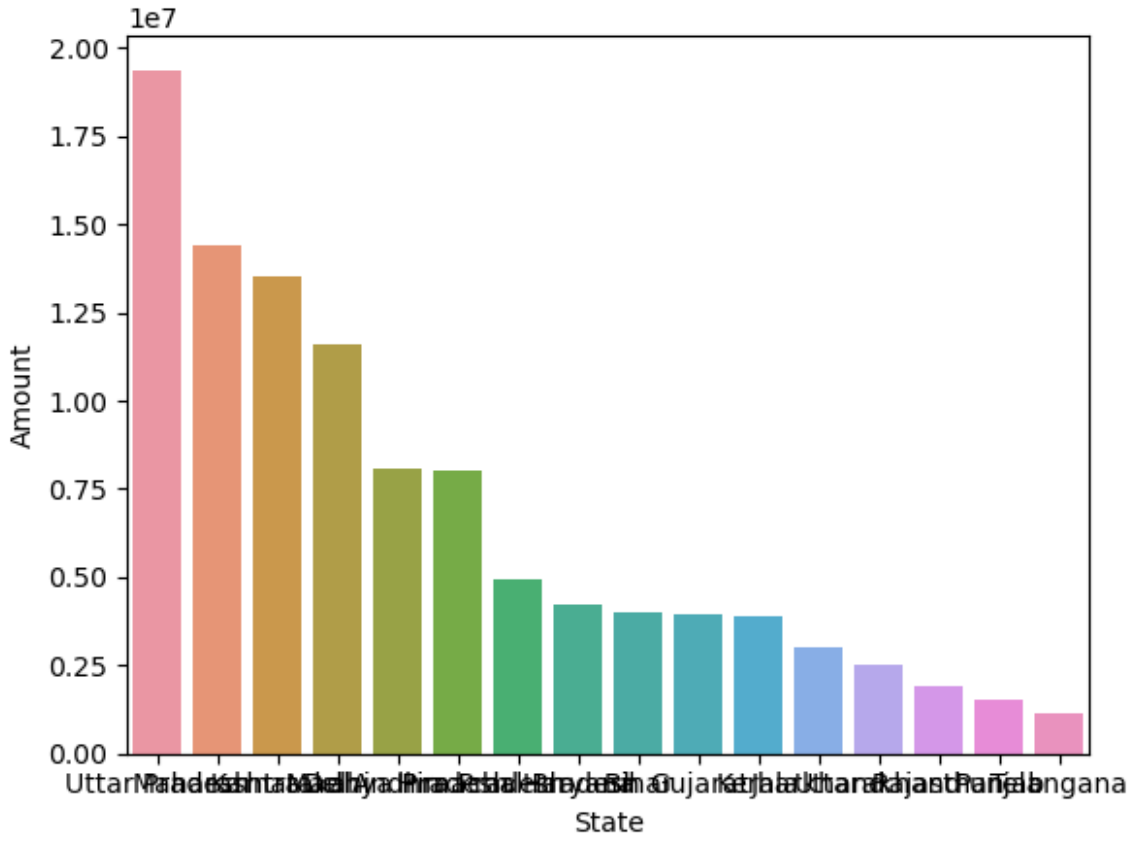
From above graph we can conlude that female under age bracket of 26-35 order more than other age brackets

```
count_states = sns.countplot(x = "State", data = df, width = 0.4)
```




```
sales_state = df.groupby(["State"],as_index = False)
["Amount"].sum().sort_values(by = "Amount",ascending = False)

sns.barplot(x = "State",y = "Amount", data = sales_state)
```

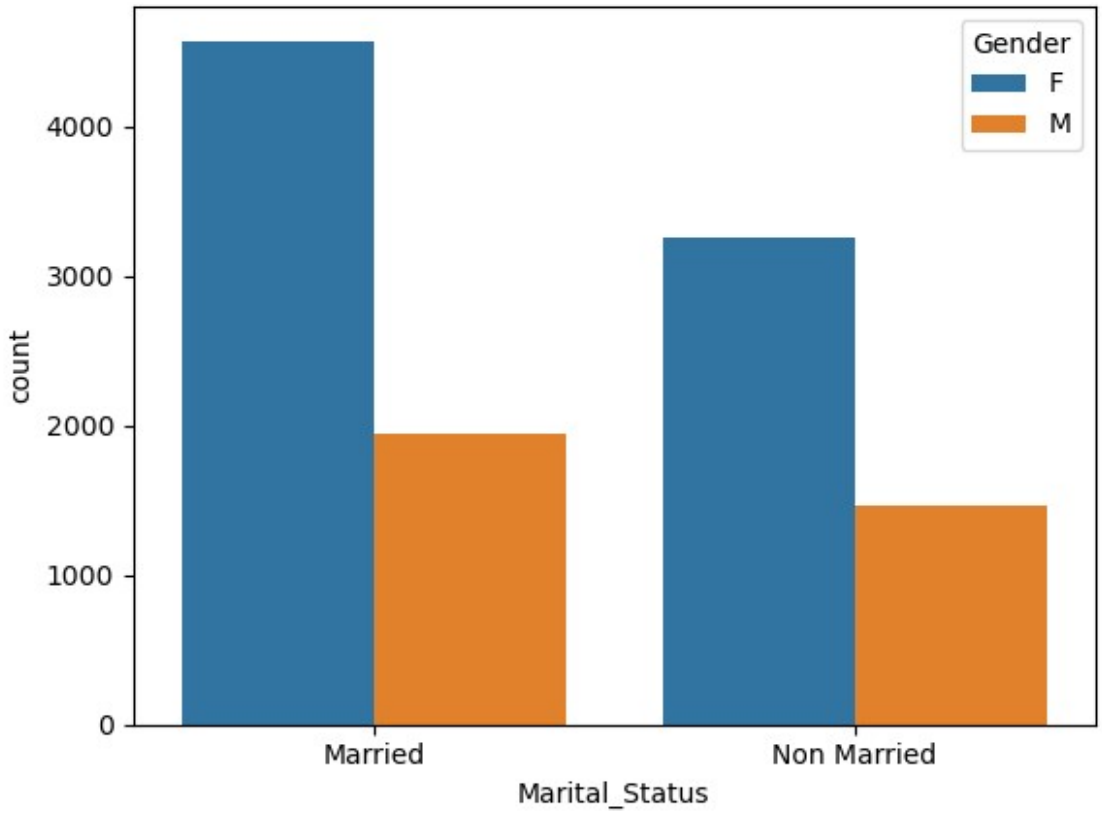
```
<Axes: xlabel='State', ylabel='Amount'>
```



From above graph we can conclude that Maharastra order more than other states but Uttar Pradesh spend more money

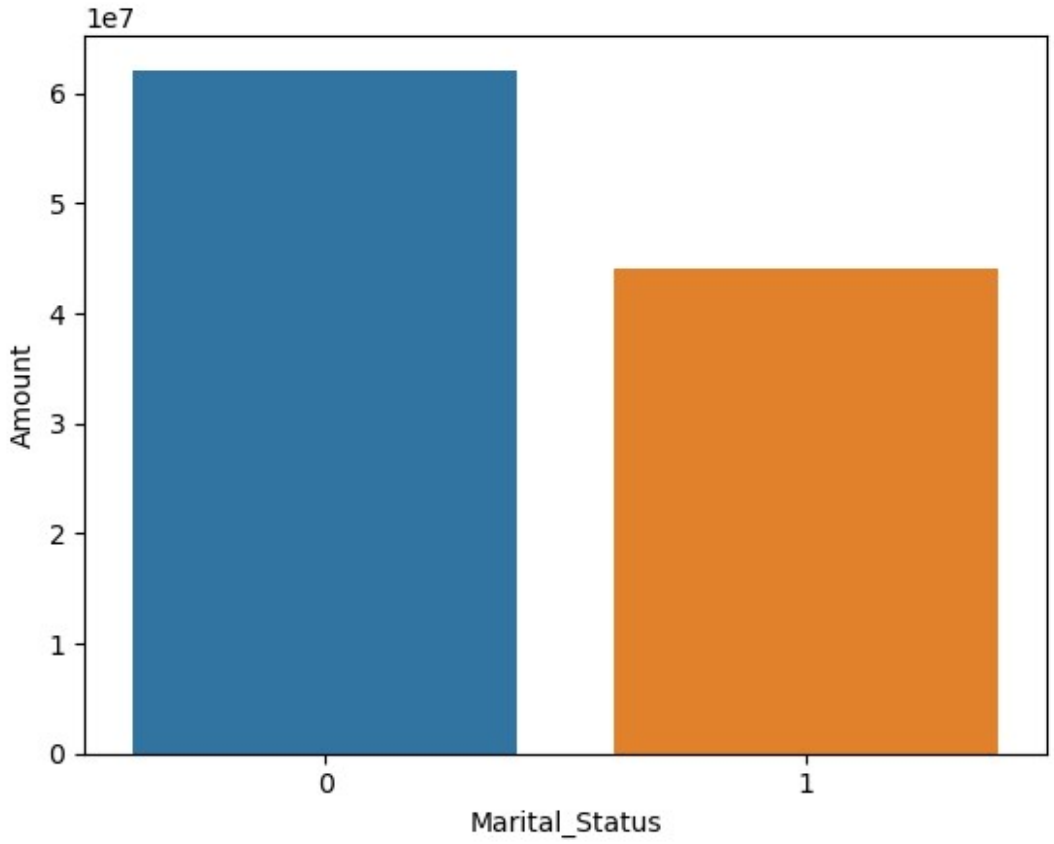
```
count_marital = sns.countplot(x = "Marital_Status",data = df,hue=
"Gender")
count_marital.set_xticklabels(["Married", "Non Married"])
#count_marital.set_xticklabels = (["Married","Non Married"])
```

```
[Text(0, 0, 'Married'), Text(1, 0, 'Non Married')]
```



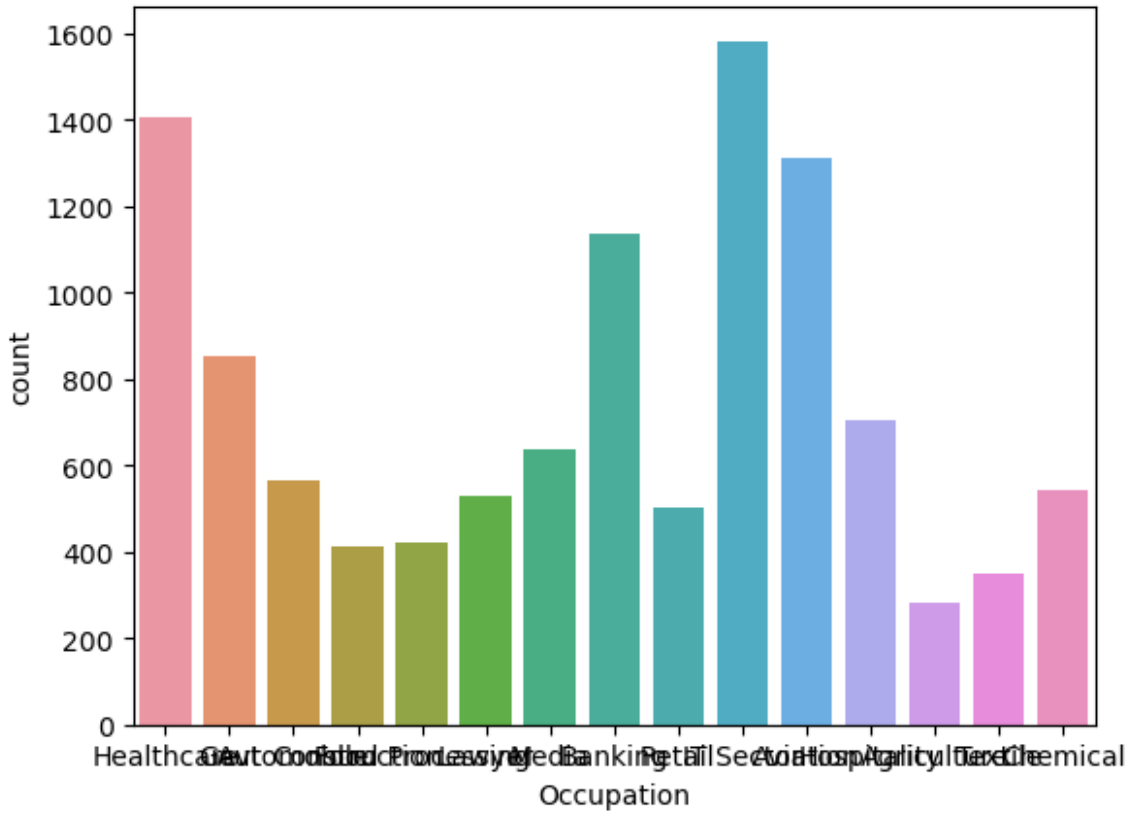

```
sales_marital = df.groupby(["Marital_Status"],as_index = False)
["Amount"].sum().sort_values(by = "Amount",ascending = False)
sns.barplot(x = "Marital_Status",y = "Amount", data = sales_marital)
```

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



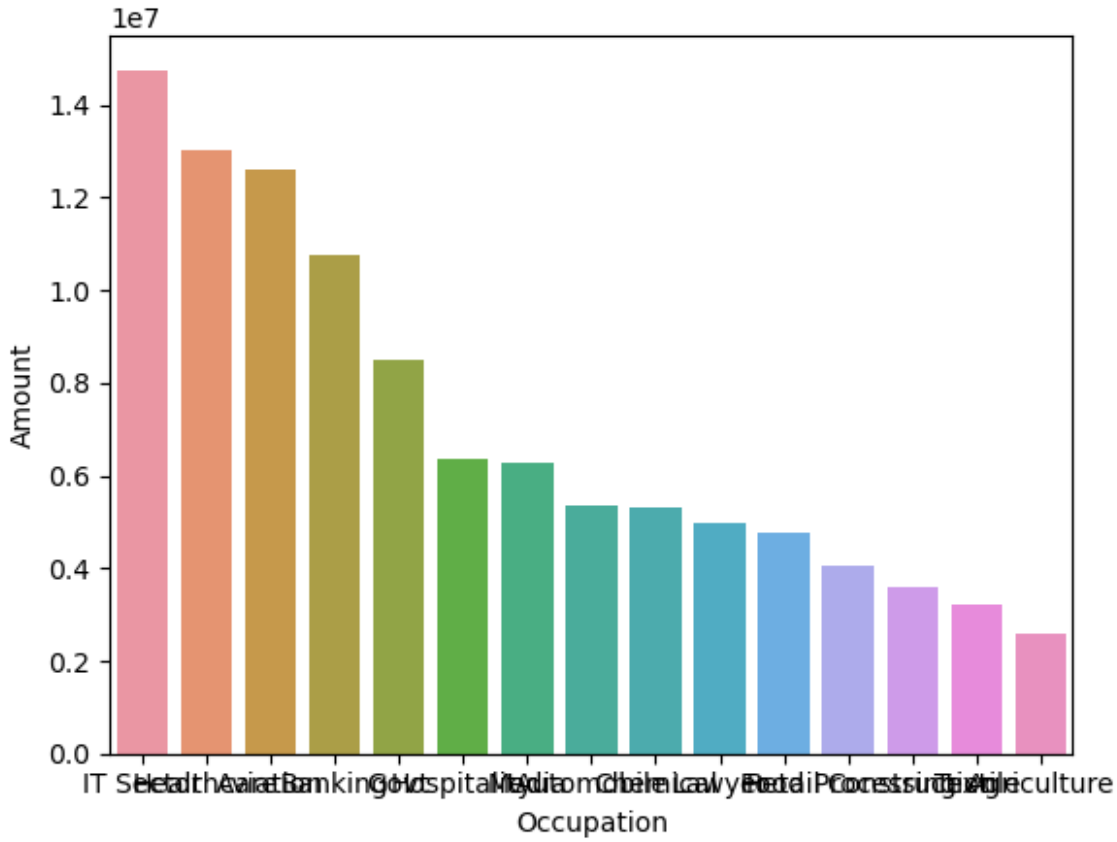
From above graph we can conlude that married people order more than unmarried people

```
count_occ = sns.countplot(x = "Occupation", data = df)
```



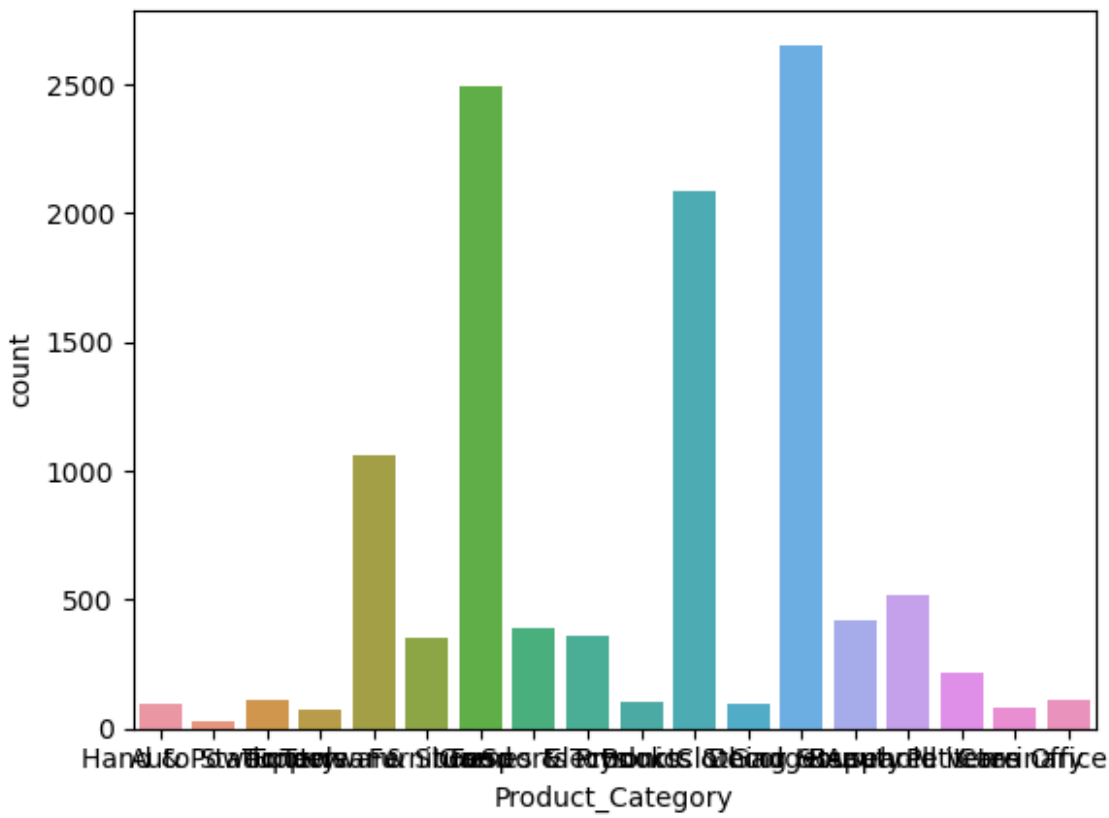

```
sales_occ = df.groupby(["Occupation"],as_index = False)
["Amount"].sum().sort_values(by = "Amount", ascending = False)
sns.barplot(x = "Occupation",y = "Amount", data = sales_occ)
```

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



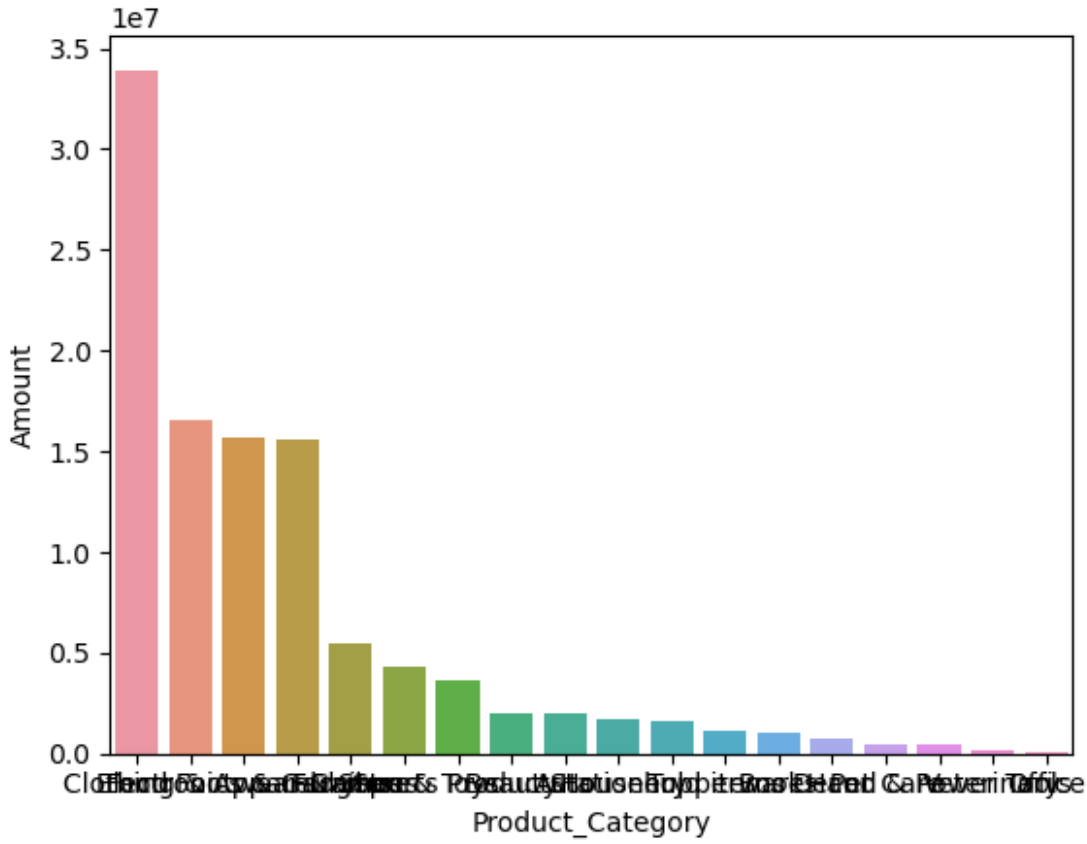
From above graph we can conclude that most of the customer comes from IT, Health and Aviation Industry

```
count_pro = sns.countplot(x = "Product_Category" , data = df)
```




```
sales_pro = df.groupby(["Product_Category"], as_index = False)
["Amount"].sum().sort_values(by = "Amount", ascending = False)
sns.barplot(x = "Product_Category",y = "Amount",data = sales_pro)
```

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



From above graph we can conclude that customers order more cloths but spend more on food items

Conclusion

We can conclude that female married customers from Uttar Pradesh and Maharastra who works in IT, Health and Aviation Industry is our largest buyers