DIWALI SALES ANALYSIS

Data Analysis using Python (Pandas, Matplotlib, Seaborn, Numpy)

INTRODUCTION

This analysis aims to understand customer behavior, sales trends, and top-performing products during Diwali season.

DATASET OVERVIEW

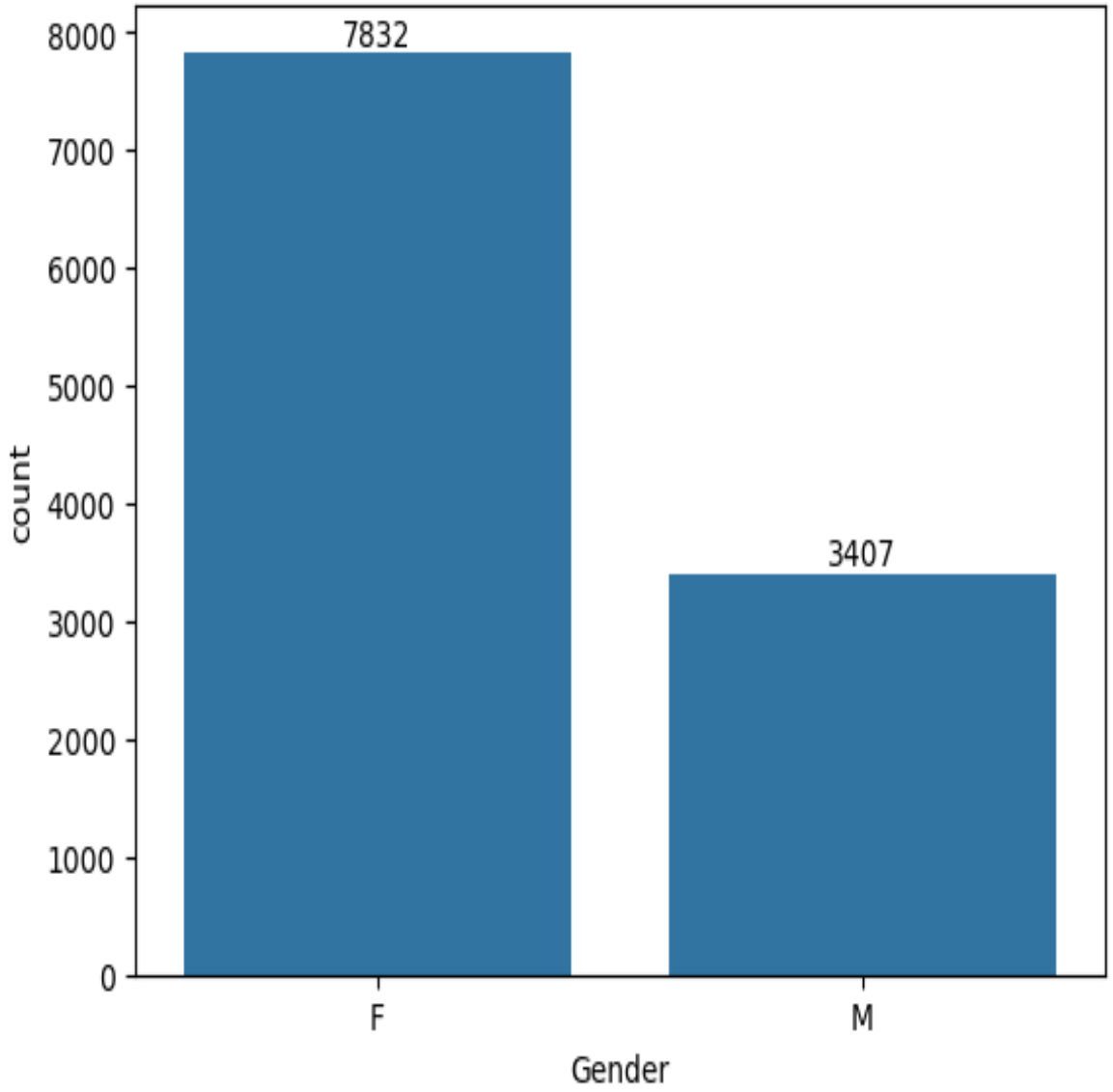
```
df.shape
```

```
(11239, 13)
```

```
df.columns
```

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

The dataset contains more Females (7832) than males (3407)



CUSTOMER DISTRIBUTION BY GENDER

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

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

plt.show()
```

PURCHASE AMOUNT BY GENDER

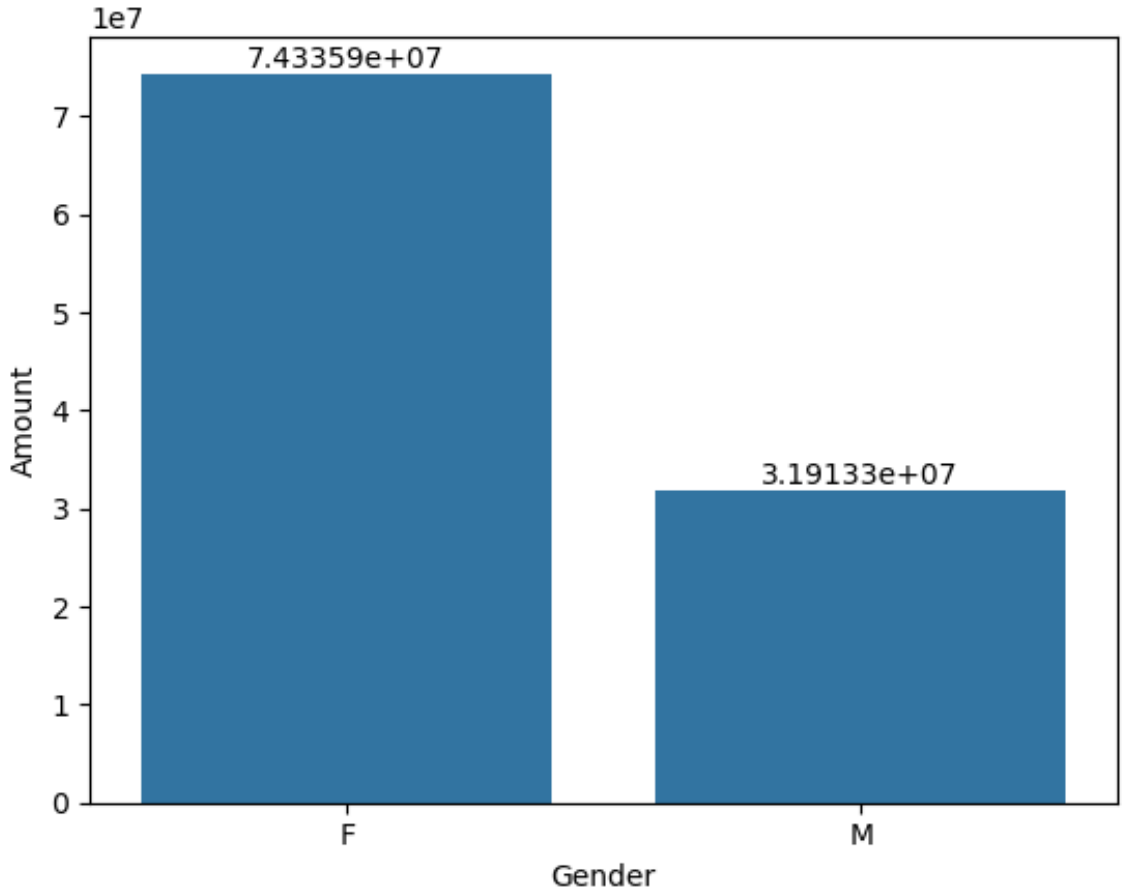
The total purchase amount by Females is higher than that of Males.

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

ax = sns.barplot(x = 'Gender', y = 'Amount', data= sales_gen)

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

plt.show()
```

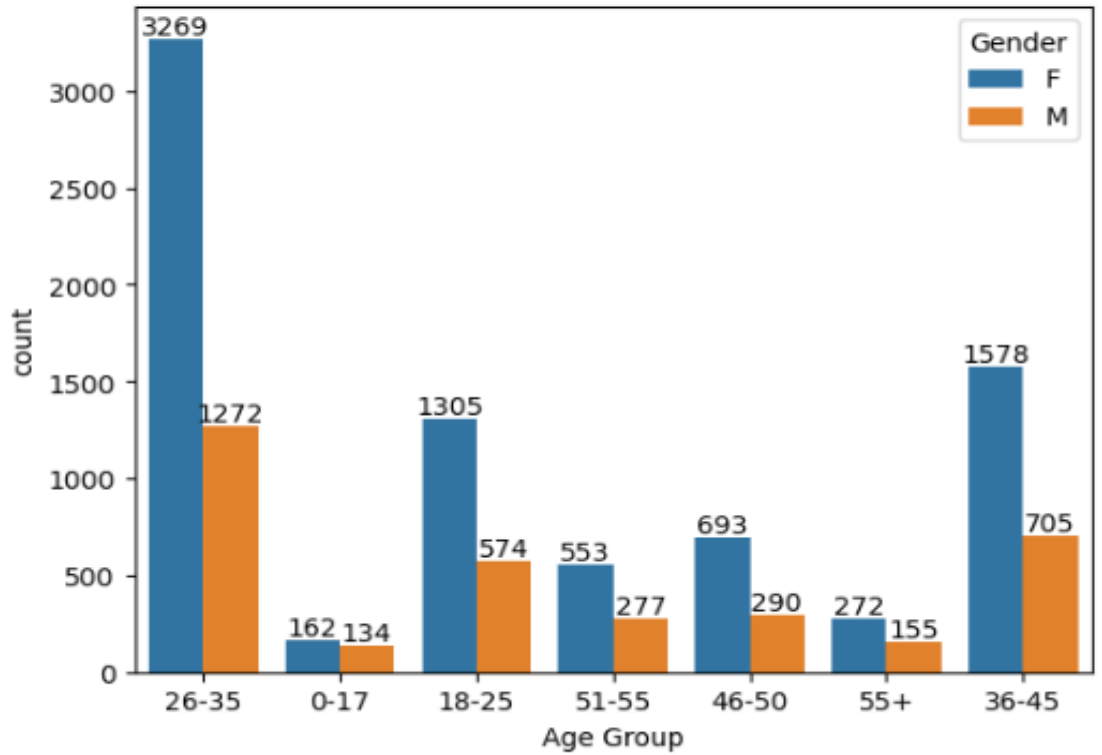


ORDERS BY AGE GROUP AND GENDER

```
age_order = df['Age Group'].value_counts().sort_values(ascending=False).index
ax = sns.countplot(data = df, x = 'Age Group', hue='Gender')

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

plt.show()
```



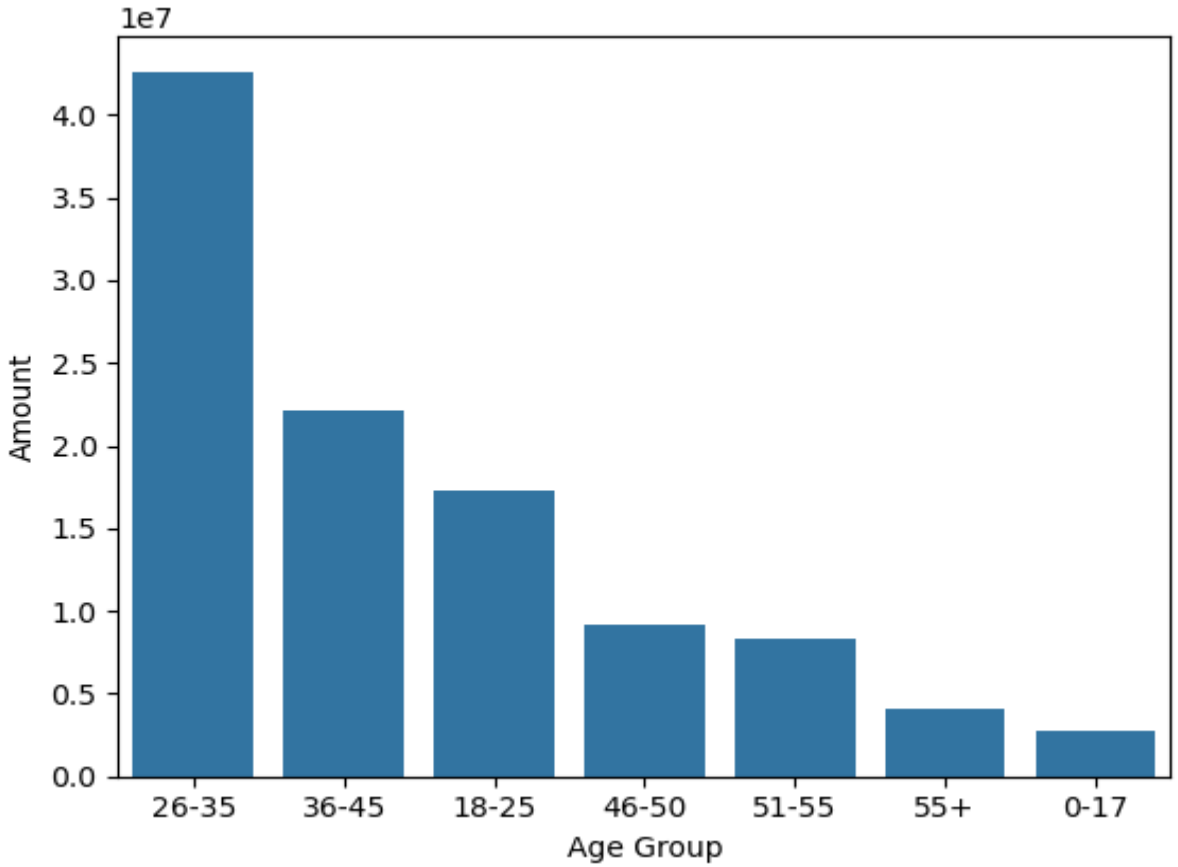
*The chart shows that the **26-35 age group** has the highest number of customers, while the **0-17 age group** has the lowest. Across most age groups, Females outnumber Males.*

TOTAL AMOUNT BY AGE GROUP

Adults aged 26–35 years contributed the highest total purchase amount, followed by the 36–45 age group.

```
#Total amount by Age group
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)
plt.show()
```

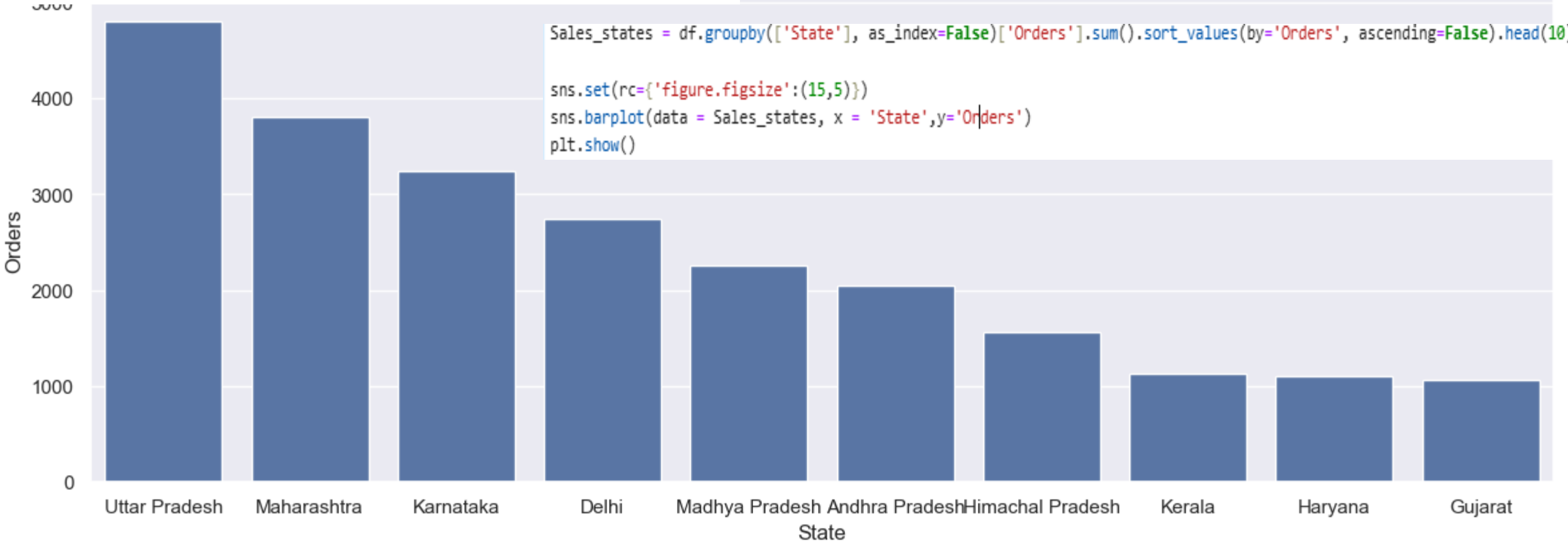


TOP 10 STATES BY TOTAL NUMBER OF ORDERS

Uttar Pradesh, Maharashtra, and Karnataka have the highest number of orders among all states.

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

sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = Sales_states, x = 'State',y='Orders')
plt.show()
```

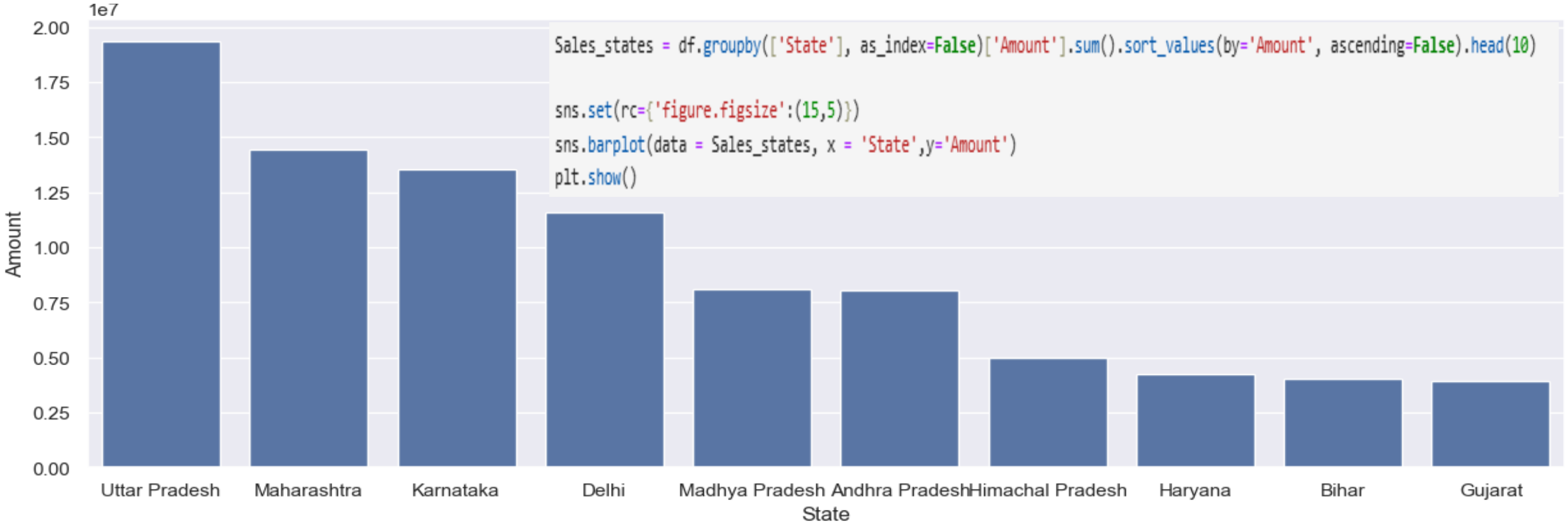


TOP 10 STATES BY TOTAL NUMBER OF AMOUNT

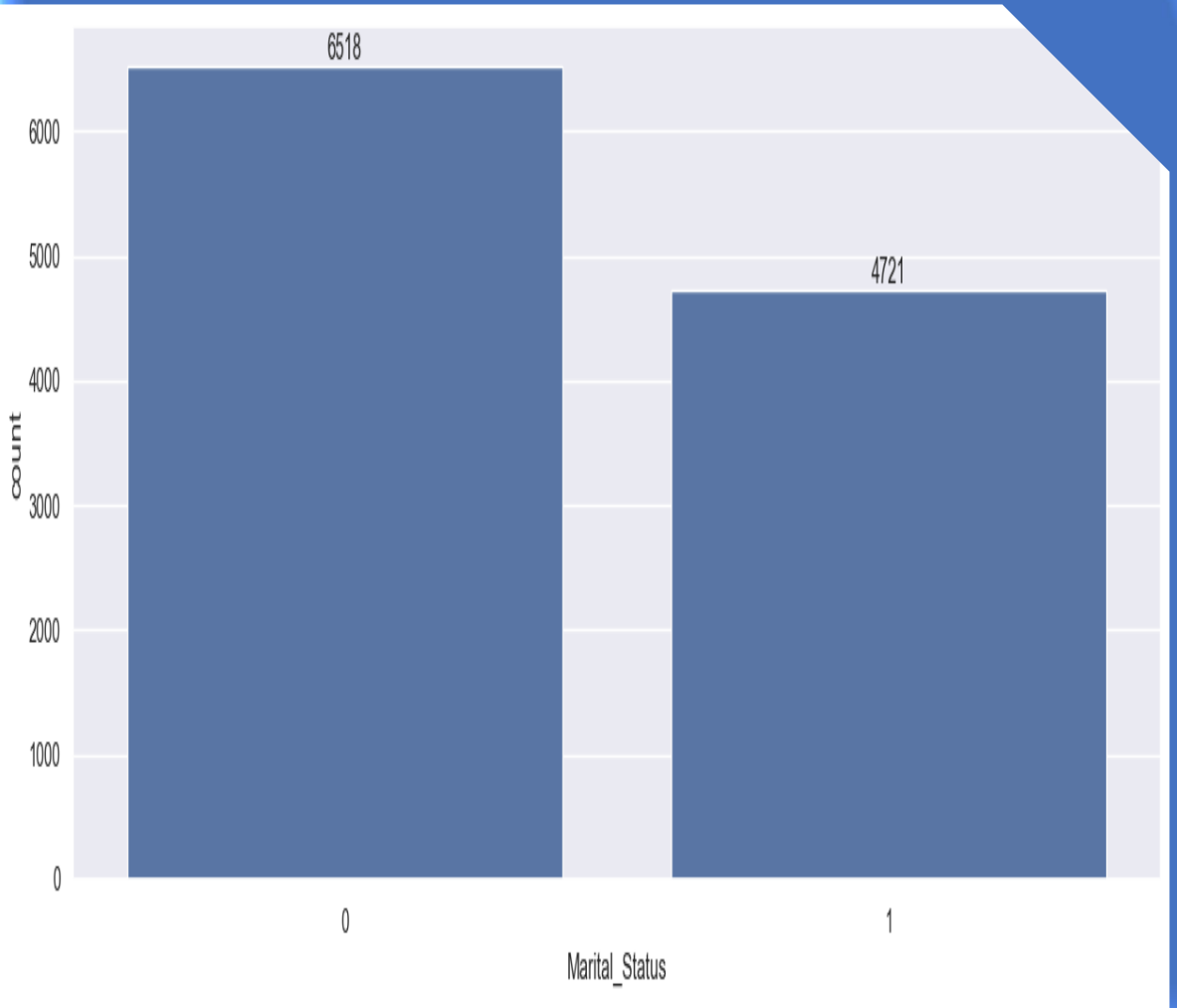
Uttar Pradesh, Maharashtra, and Karnataka generated the highest sales amount among all states.

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

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



ORDERS BY MARITAL STATUS

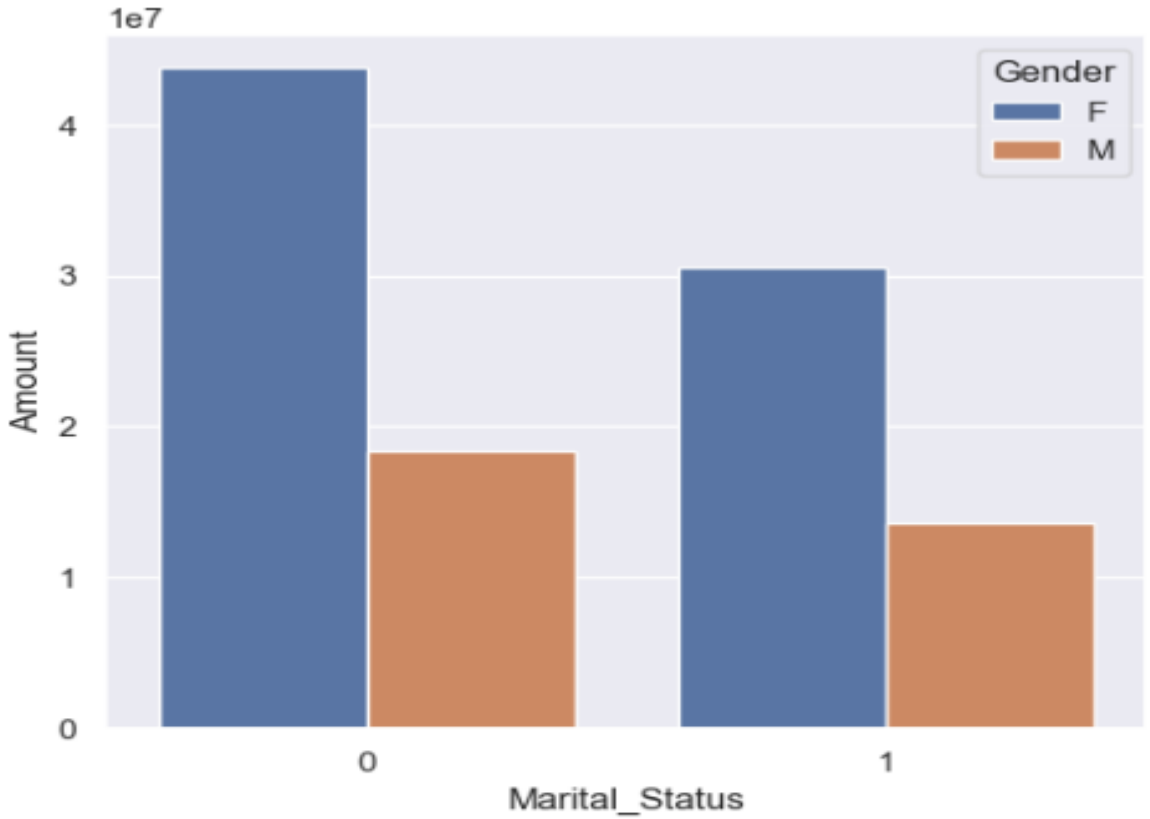


Unmarried customers are higher in number compared to married customers

```
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.show()
```

Total Purchase Amount by Marital Status and Gender




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

sns.set(rc={'figure.figsize':(6,5)})
sns.barplot(data = sales_state, x='Marital_Status', y='Amount', hue='Gender')
plt.show()
print(sales_state)
```

	Marital_Status	Gender	Amount
0	0	F	43786646
2	1	F	30549207
1	0	M	18338738
3	1	M	13574538

Unmarried women contribute(43786646) the highest purchase amount compared to all other groups.

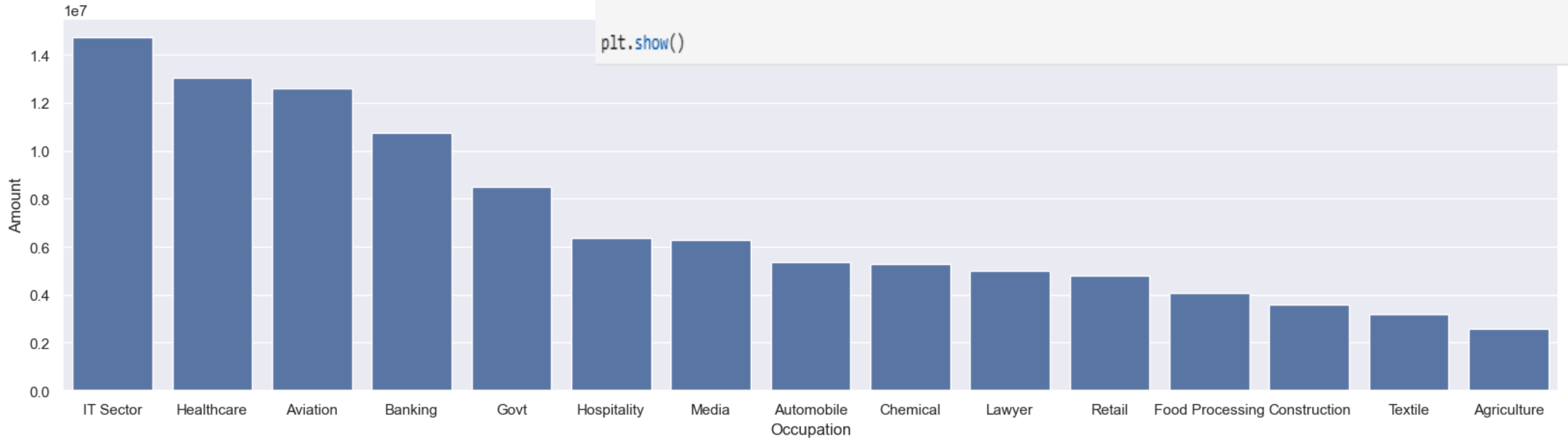
PURCHASE TRENDS ACROSS OCCUPATIONS

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

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

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

plt.show()
```

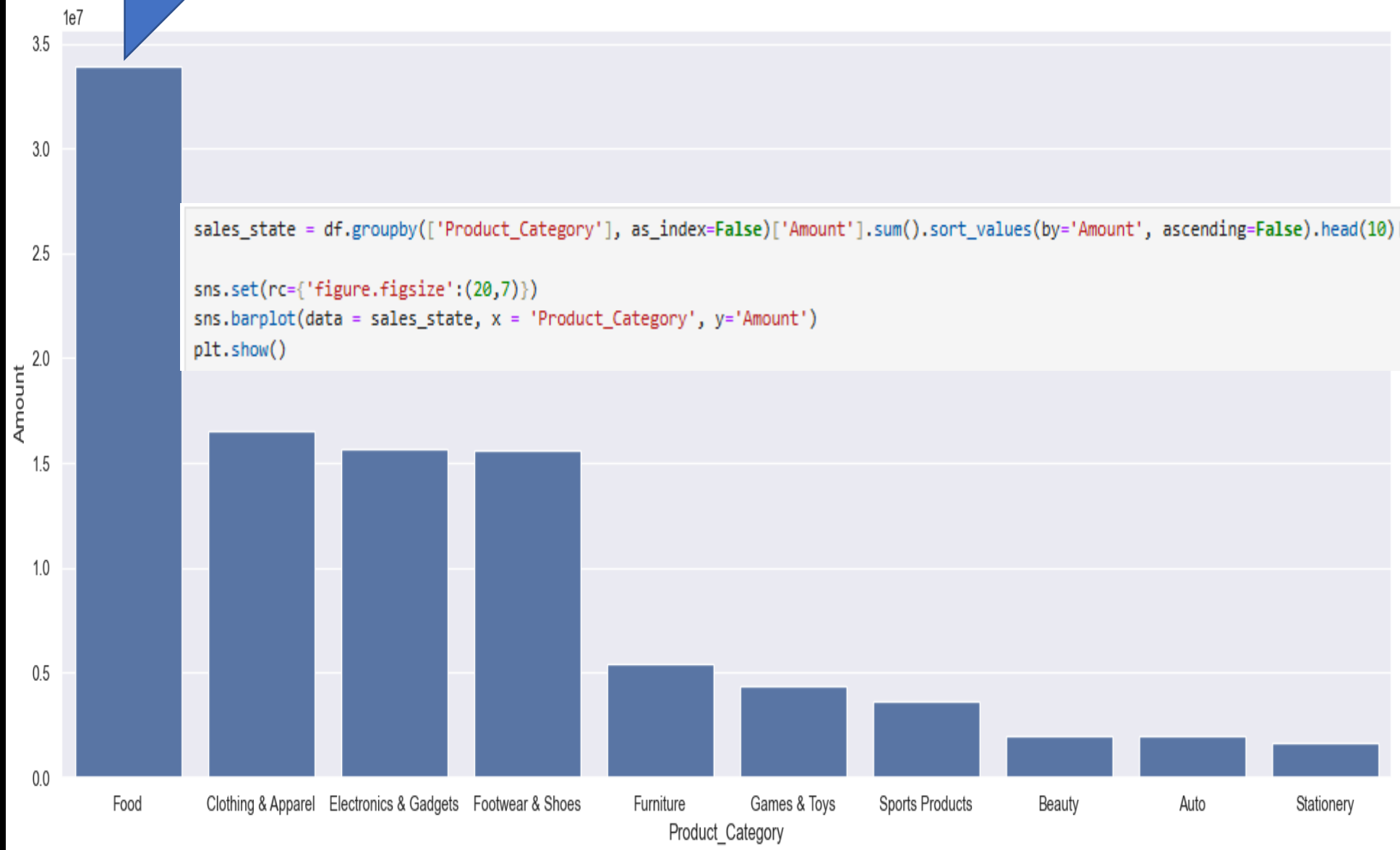


TOP 10 PRODUCT CATEGORIES
(BY SALES AMOUNT)

We can see that most of the sold products are from Food, Clothing and Electronics category

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

sns.set(rc={'figure.figsize':(20,7)})
sns.barplot(data = sales_state, x = 'Product_Category', y='Amount')
plt.show()
```



KEY INSIGHTS

- ***Females*** *lead in both customer count and total purchase amount.*
- *The* ***26–35 age group*** *contributes the highest number of orders and total sales.*
- ***Uttar Pradesh, Maharashtra, and Karnataka*** *rank as the top three states in both sales and orders.*
- ***Unmarried females*** *account for the highest purchase amount.*
- *Majority of buyers are from* ***IT, Healthcare, and Aviation*** *sectors.*
- ***Clothing, Food, and Electronics*** *are the top-selling product categories.*
- ***Product ID P00265242*** *is the most frequently ordered product.*