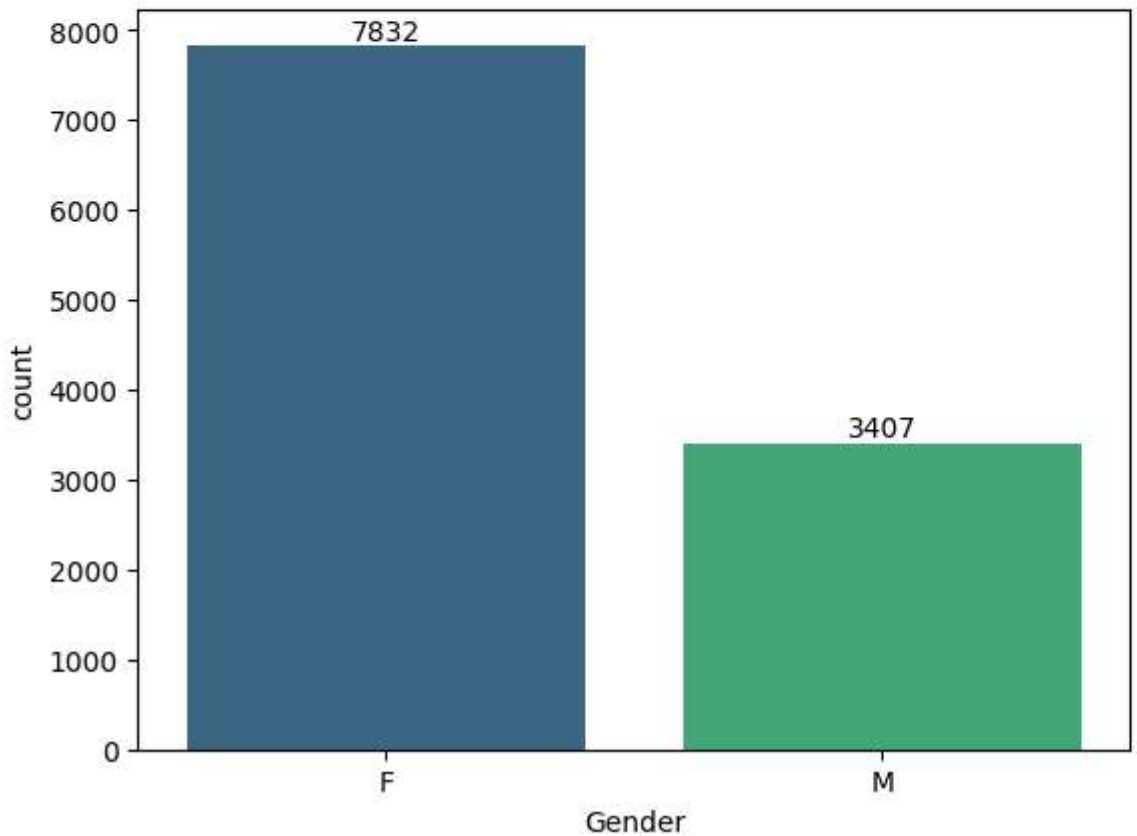
Exploratory Data Analysis (EDA)

Gender

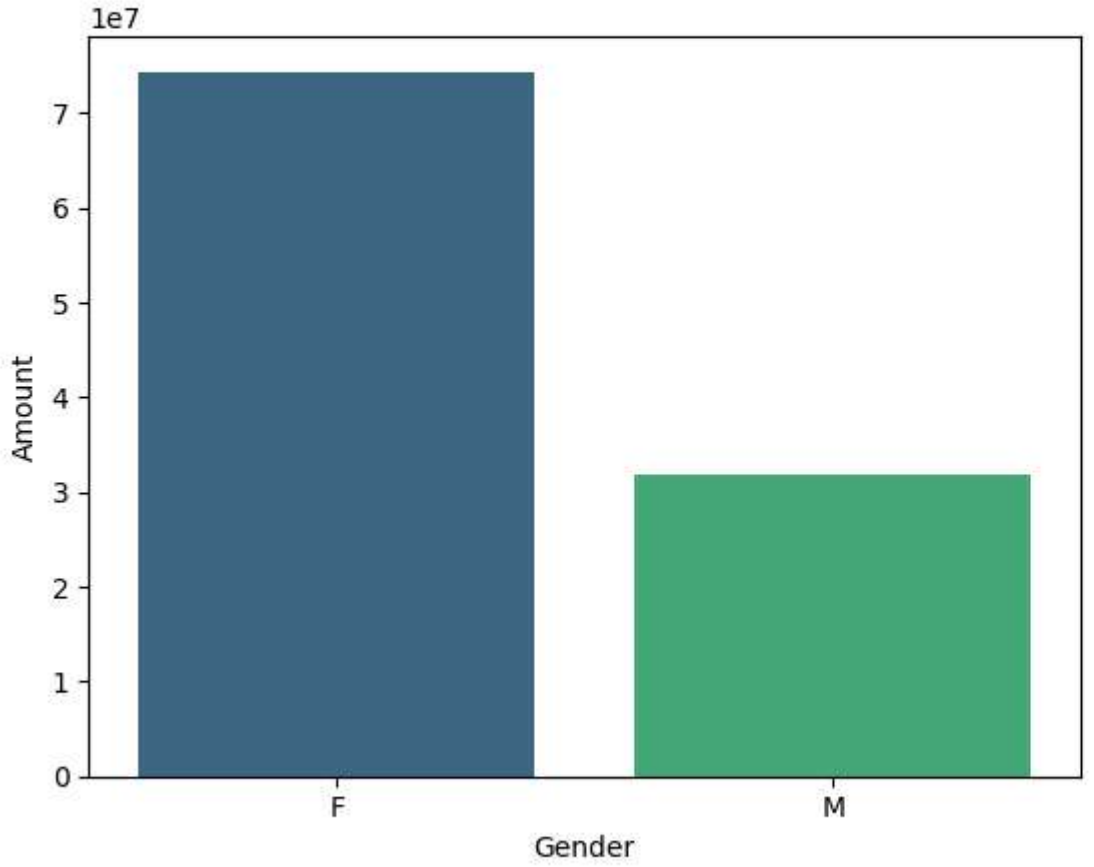
Plotting a bar chart for gender and it's count

[13]:
```
ax = sns.countplot(x = "Gender", data = df, hue = "Gender", palette = "viridis")
for bars in ax.containers:
    ax.bar_label(bars)
plt.show()
```



Plotting a bar chart for gender vs total amount

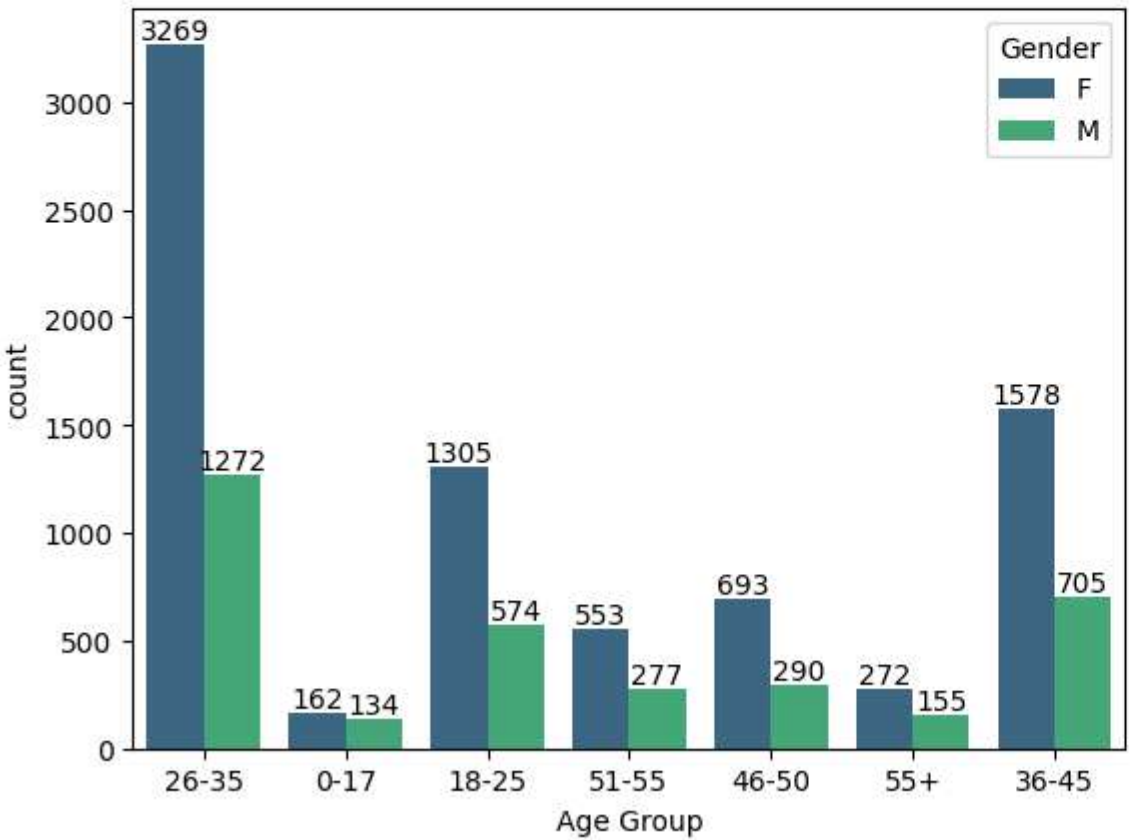
[14]:
```
sales_gen = df.groupby(['Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(x = "Gender", y = "Amount", data = sales_gen, hue = "Gender", palette = "viridis")
plt.show()
```



From above graphs we can see that most of the buyers are females and even the purchasing power of females are greater than men.

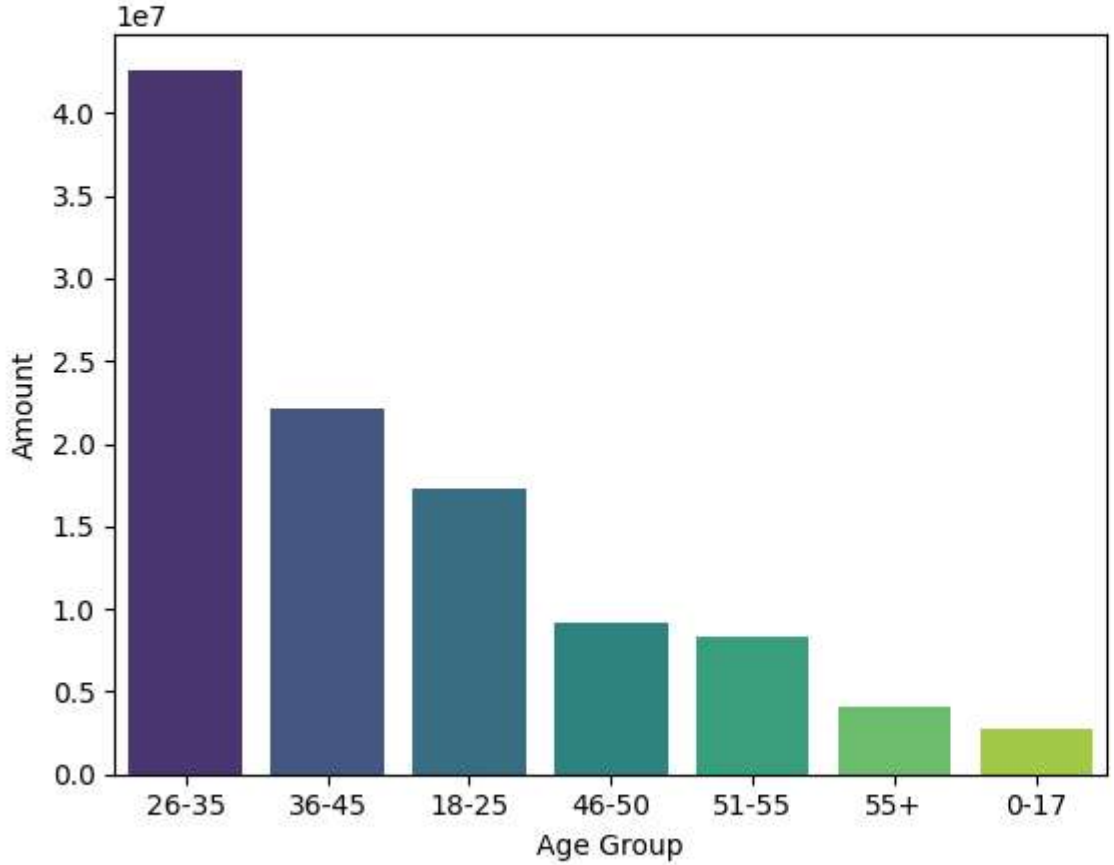
Age

[15]:
```
ax = sns.countplot(data = df, x = "Age Group", hue = "Gender", palette = "viridis")

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



Total amount vs Age Group

[16]:
```
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, hue = "Age Group", palette = "viridis")
plt.show()
```

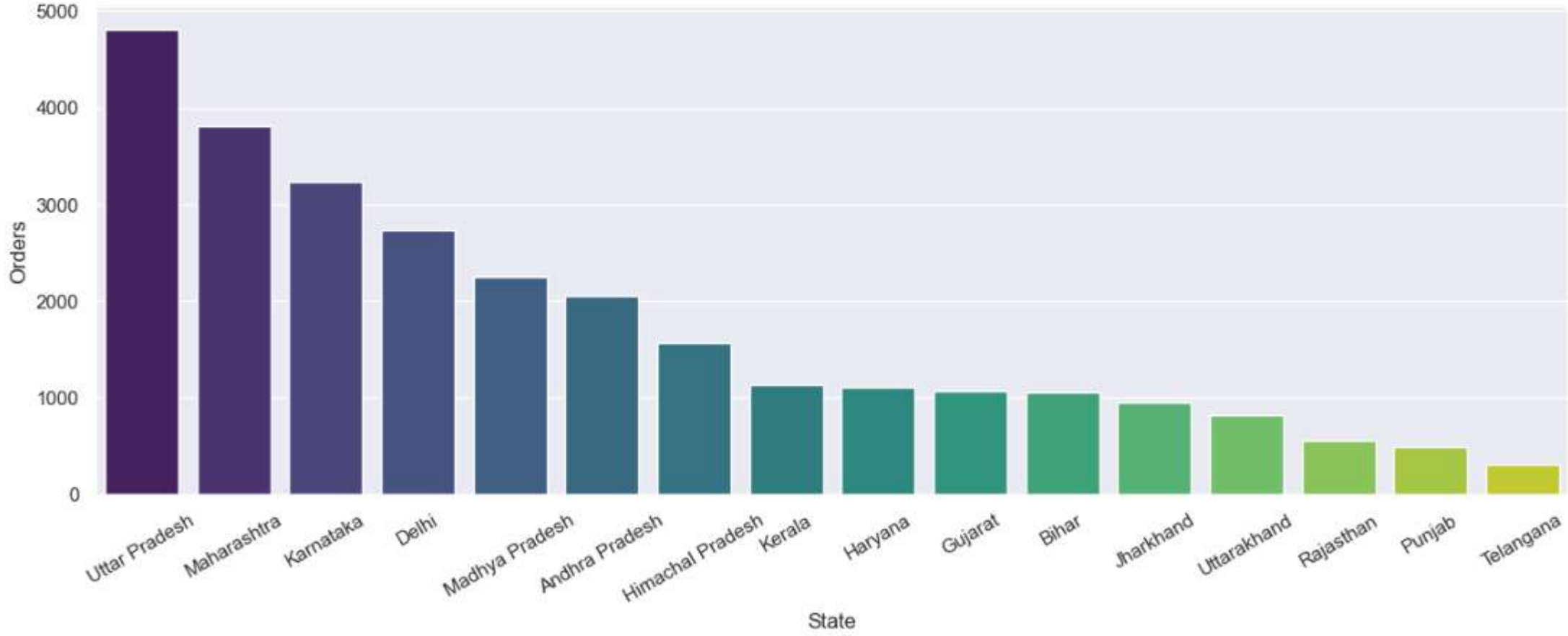


From above graphs we can see that most of the buyers are of age group between 26-35 yrs. female.

State

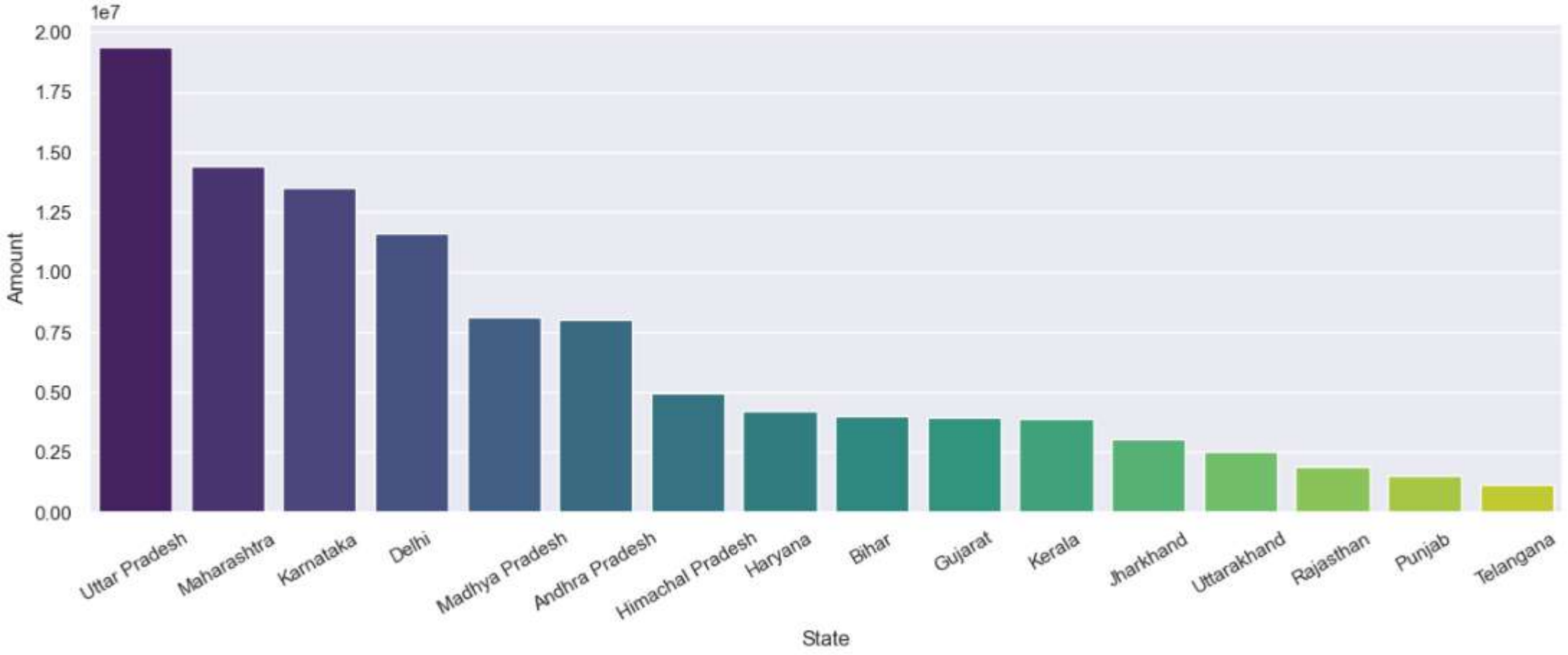
Total number of orders from top 10 states

[17]:
```
sales_state = df.groupby(["State"], as_index = False)["Orders"].sum().sort_values(by = "Orders", ascending = False)

sns.set(rc = {"figure.figsize":(15,5)})
sns.barplot(data = sales_state, x = "State", y = "Orders", hue = "State", palette = "viridis")
plt.xticks(rotation = 30)
plt.show()
```



Total amount / sales from top 10 states

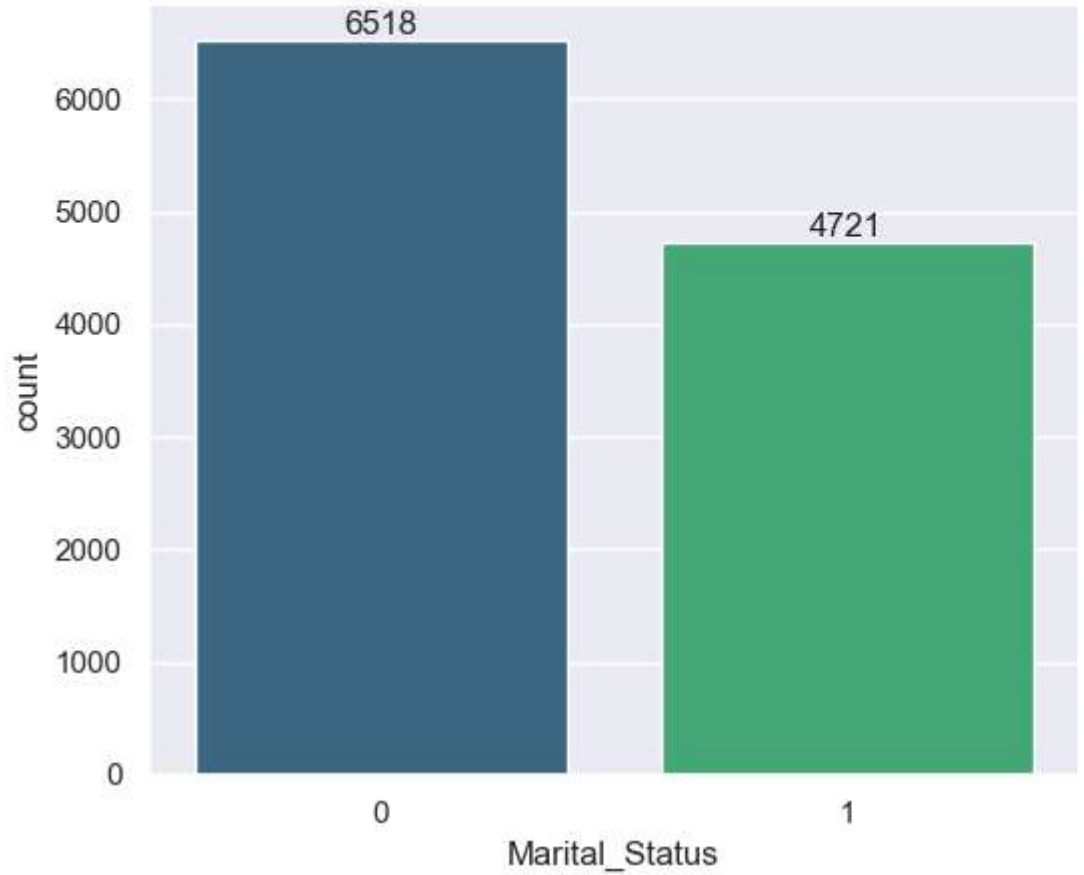
[18]:
```
sales_state = df.groupby(["State"], as_index = False)["Amount"].sum().sort_values(by = "Amount", ascending = False)

sns.set(rc = {"figure.figsize":(15,5)})
sns.barplot(data = sales_state, x = "State", y = "Amount", hue = "State", palette = "viridis")
plt.xticks(rotation = 30)
plt.show()
```



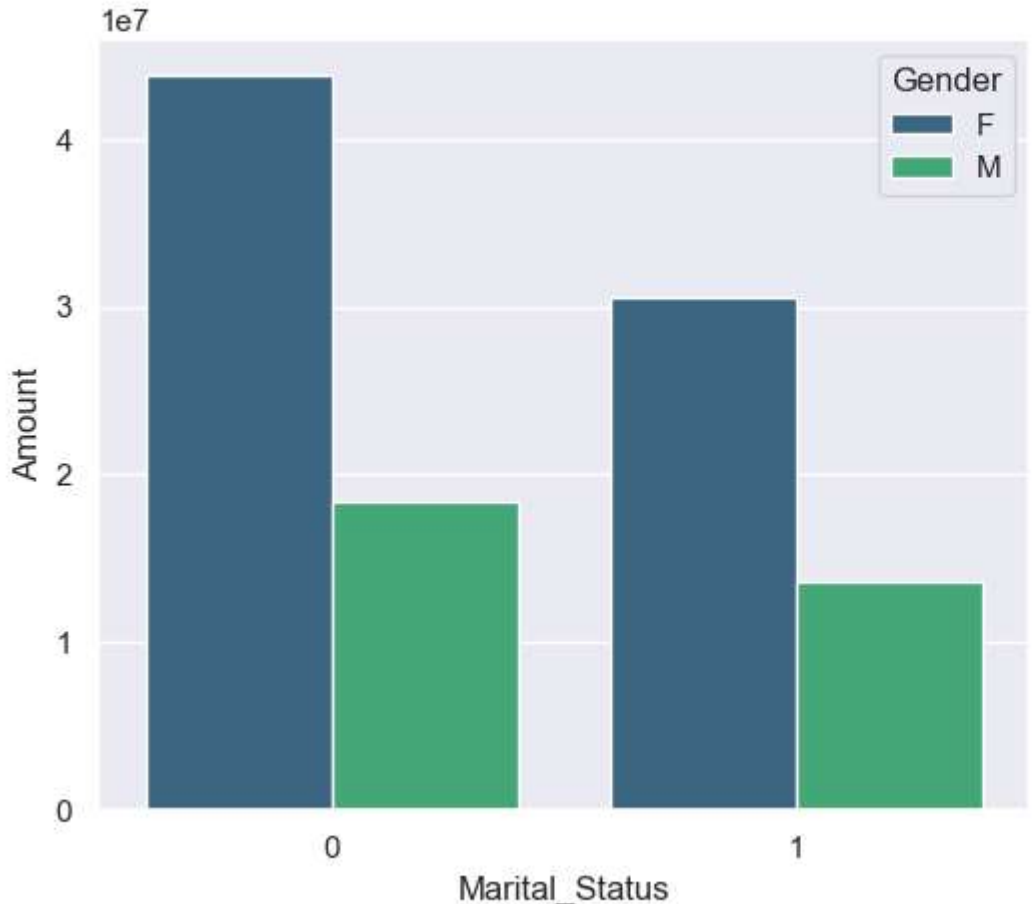
From above graphs we can see that most of the orders & total sales / amount are from Uttar Pradesh, Maharashtra and Karnataka respectively.

Marital Status

[23]:
```
ax = sns.countplot(data = df, x = "Marital_Status", hue = "Marital_Status", legend = False, palette = "viridis")

sns.set(rc = {"figure.figsize":(0,5)})
for bars in ax.containers:
    ax.bar_label(bars)
plt.show()
```



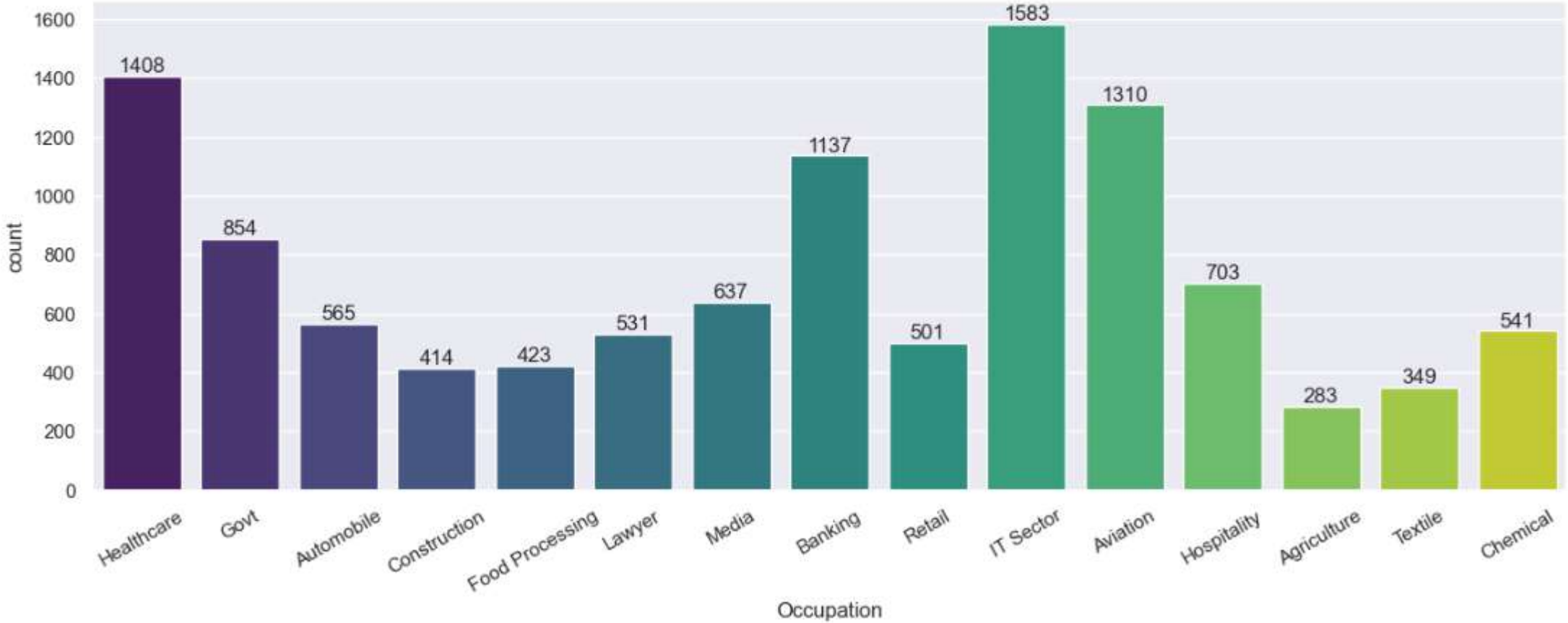
[24]:
```
sales_state = df.groupby(["Marital_Status", "Gender"], as_index = False)["Amount"].sum().sort_values(by = "Amount", ascending = False)

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



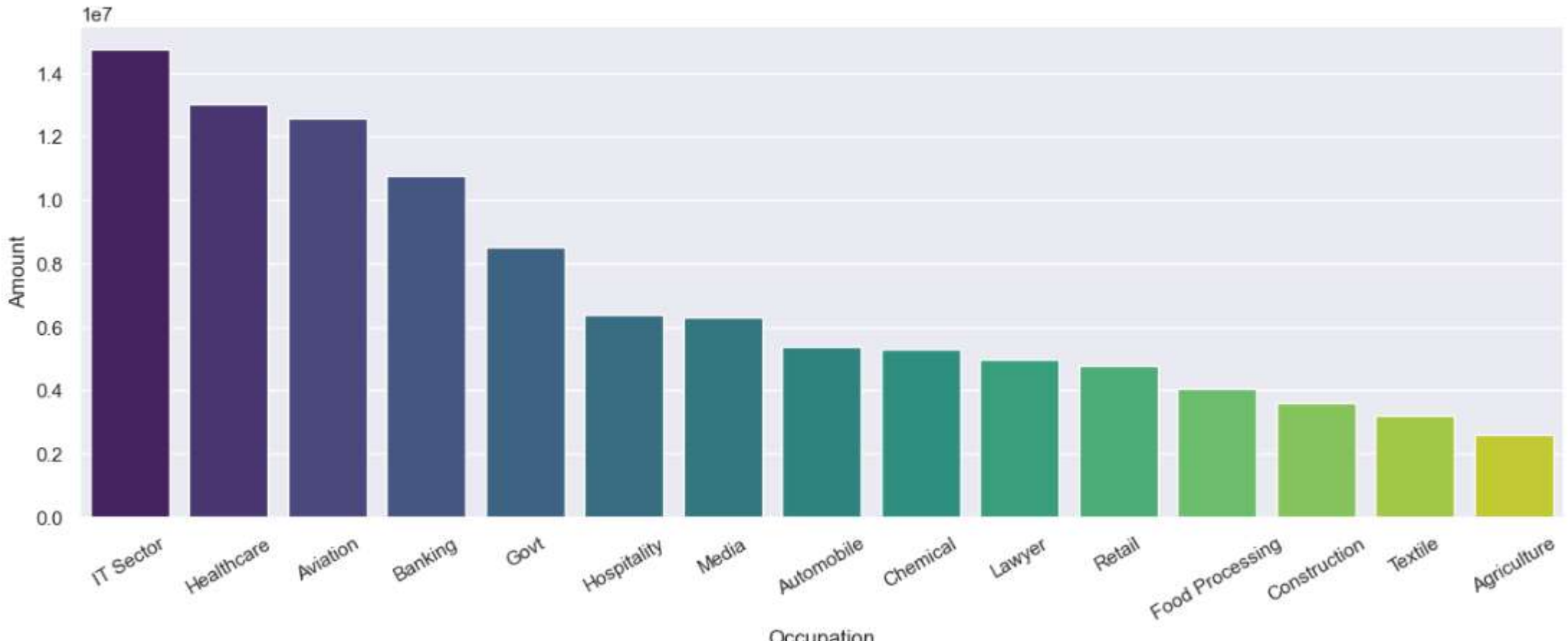
From above graphs we can see that most of the buyers are married (women) and they have high purchasing power.

Occupation

[27]:
```
ax = sns.countplot(data = df, x = "Occupation", hue = "Occupation", palette = "viridis")

sns.set(rc = {"figure.figsize":(15,5)})
for bars in ax.containers:
    ax.bar_label(bars)
plt.xticks(rotation = 30)
plt.show()
```



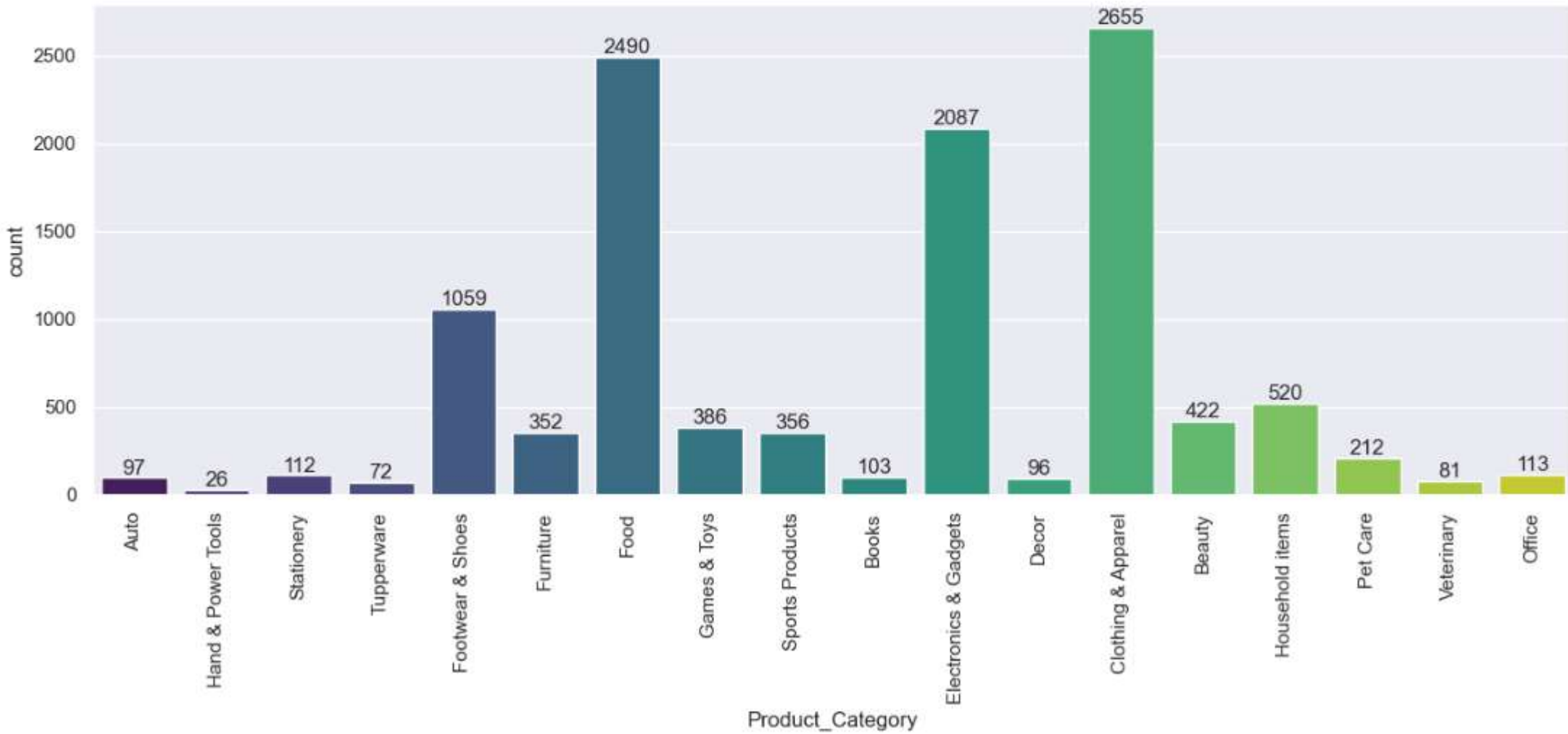
[28]:
```
sales_state = df.groupby(["Occupation"], as_index = False)["Amount"].sum().sort_values(by = "Amount", ascending = False)

sns.set(rc = {"figure.figsize":(15,5)})
sns.barplot(data = sales_state, x = "Occupation", y = "Amount", hue = "Occupation", palette = "viridis")
plt.xticks(rotation = 30)
plt.show()
```



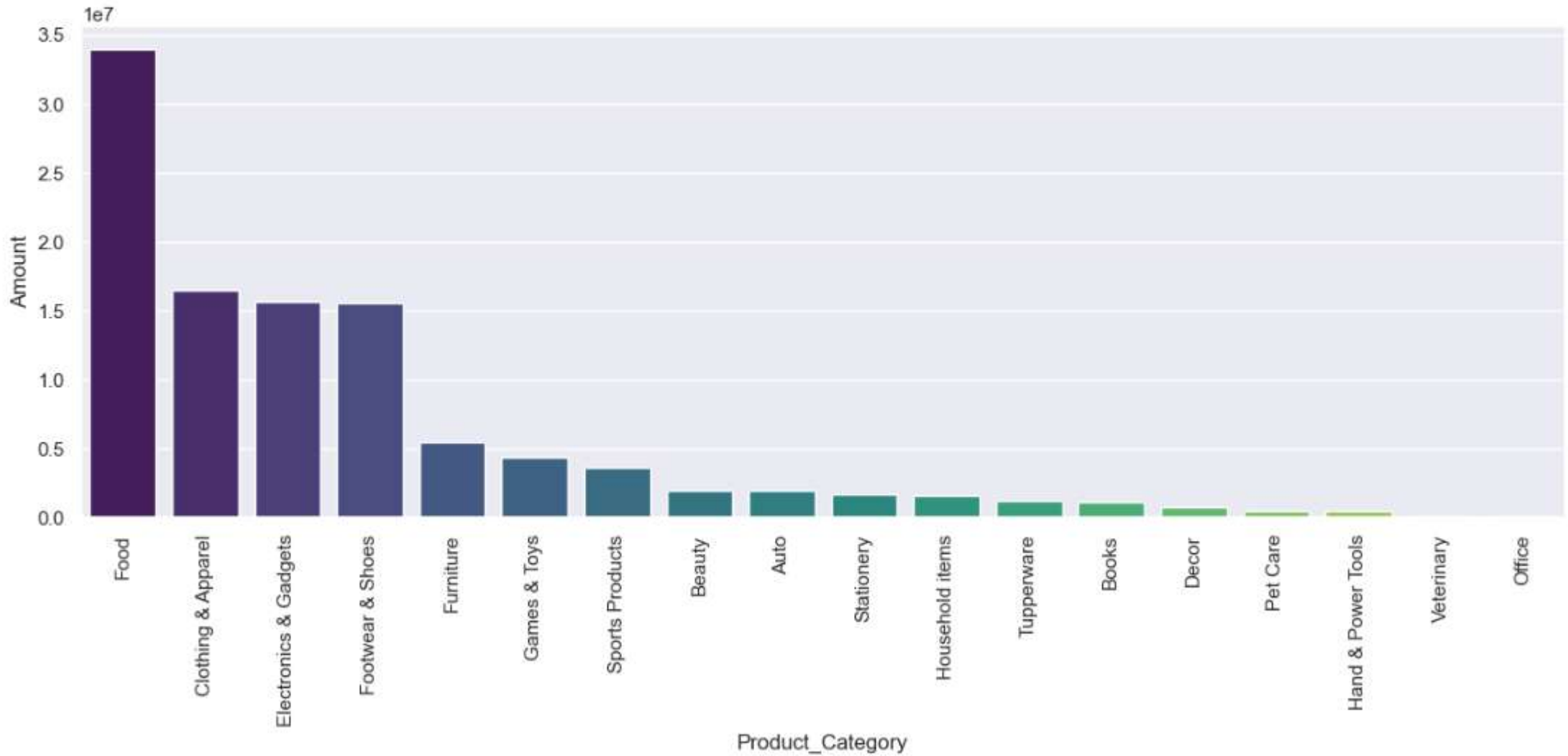
From above graphs we can see that most of the buyers are working in IT, Healthcare and Aviation sector.

Product Category

[29]:
```
ax = sns.countplot(data = df, x = "Product_Category", hue = "Product_Category", palette = "viridis")
sns.set(rc = {"figure.figsize":(15,5)})

for bars in ax.containers:
    ax.bar_label(bars)
plt.xticks(rotation = 90)
plt.show()
```



[30]:
```
sales_state = df.groupby(["Product_Category"], as_index = False)["Amount"].sum().sort_values(by = "Amount", ascending = False)

sns.set(rc = {"figure.figsize":(15,5)})
sns.barplot(data = sales_state, x = "Product_Category", y = "Amount", hue = "Product_Category", palette = "viridis")
plt.xticks(rotation = 90)
plt.show()
```



From above graphs we can see that most of the sold products are from Food, Clothing and Electronics category.

Top 10 most sold products (same thing as above)

[46]:
```
sales_state = df.groupby(["Product_ID"], as_index = False)["Orders"].sum().sort_values(by = "Orders", ascending = False).head(10)

sns.set(rc = {"figure.figsize":(15,5)})
sns.barplot(data = sales_state, x = "Product_ID", y = "Orders", hue = "Product_ID", palette = "viridis")
plt.show()
```