Home | Untitled

localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 26 minutes ago

File Edit View Run Kernel Settings Help | Trusted

Code | JupyterLab | Python [conda env:base] *

[1]:
```
# Let's import python libraries First

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt # visualizing data
%matplotlib inline
import seaborn as sns
```

[4]:
```
# import csv file
df = pd.read_csv('Sales_Data.csv', encoding= 'unicode_escape')
```

[5]:
```
#check number of row and column
df.shape
```

[5]:
```
(11257, 13)
```

[6]:
```
#top see the imported data
df.head()#show top5 rows
```

[6]:

	User_ID	Cust_name	Product_ID	Age	Age Group	Gender	State	Zone	Zipcode	Profession	Product_Category	Orders	Amount
0	1002903.0	Anvi	P00125942	27.0	26-35	Female	Maharashtra	West	NaN	Healthcare	Sports	4.0	20500.0
1	1000732.0	Shanta	P00110942	34.0	26-35	Female	Andhra Pradesh	South	NaN	Govt	Sports	2.0	25360.0
2	1001990.0	Sheetal	P00118542	16.0	0-17	Female	Uttar Pradesh	Central	NaN	Automobile	Health	4.0	29350.0
3	1001425.0	Virendra	P00237842	16.0	0-17	M	Karnataka	South	NaN	Construction	Clothing	6.0	23500.0
4	1000588.0	Vishal	P00057942	28.0	26-35	M	Gujarat	West	NaN	Food Processing	Electronics	4.0	23870.0

[7]:
```
#field details and data type
```

Home | Untitled

localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 26 minutes ago

File Edit View Run Kernel Settings Help | Trusted

Code | JupyterLab | Python [conda env:base] *

[7]:
```
#field details and data type
df.info ()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11257 entries, 0 to 11256
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   User_ID           11245 non-null  float64
 1   Cust_name         11245 non-null  object 
 2   Product_ID        11245 non-null  object 
 3   Age               11245 non-null  float64
 4   Age Group         11245 non-null  object 
 5   Gender            11245 non-null  object 
 6   State             11245 non-null  object 
 7   Zone              11245 non-null  object 
 8   Zipcode           0 non-null      float64
 9   Profession        11245 non-null  object 
 10  Product_Category  11245 non-null  object 
 11  Orders            11245 non-null  float64
 12  Amount            11245 non-null  float64
dtypes: float64(5), object(8)
memory usage: 1.1+ MB
```

data cleaning markdown

[40]:
```
df.drop(['Zipcode'], axis=1, inplace=True, errors='ignore')
```

[41]:
```
#list of columns avilable
df.columns
```

[41]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Age', 'Age Group', 'Gender',
```

jupyter **Untitled** Last Checkpoint: 27 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python [conda env:base] *

[41]:
```
#list of columns avilable
df.columns
```

[41]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Age', 'Age Group', 'Gender',
       'State', 'Zone', 'Profession', 'Product_Category', 'Orders', 'Amount'],
      dtype='object')
```

[42]:
```
#check for null values
pd.isnull(df).sum()
```

[42]:
```
User_ID             12
Cust_name           12
Product_ID          12
Age                 12
Age Group           12
Gender              12
State               12
Zone                12
Profession          12
Product_Category    12
Orders              12
Amount              12
dtype: int64
```

[45]:
```
#drop null value
df.dropna(how='all',inplace=True)
```

[46]:
```
df.shape
```

[46]:
```
(11245, 12)
```

[50]:
```
#replace value of gender column
df['Gender']= df['Gender'].replace('M','Male')
```

[53]:
```
#view only male gender data
df[df['Gender']=='Male']
```

[53]:

	User_ID	Cust_name	Product_ID	Age	Age Group	Gender	State	Zone	Profession	Product_Category	Orders	Amount
3	1001425.0	Virendra	P00237842	16.0	0-17	Male	Karnataka	South	Construction	Clothing	6.0	23500.0
4	1000588.0	Vishal	P00057942	28.0	26-35	Male	Gujarat	West	Food Processing	Electronics	4.0	23870.0
5	1000588.0	Suuraj	P00057942	28.0	26-35	Male	Himachal Pradesh	Northern	Food Processing	Electronics	3.0	23860.0
8	1003224.0	Kushal	P00205642	35.0	26-35	Male	Uttar Pradesh	Central	Govt	Beauty	4.0	23809.0
11	1003829.0	Harsh	P00200842	34.0	26-35	Male	Delhi	Central	Banking	Health	2.0	23770.0
...	...	...	...	...	...	...	...	...	...	...	...	...
11249	1005446.0	Sheetal	P00297742	53.0	51-55	Male	Gujarat	West	Healthcare	Health	3.0	382.0
11250	1005446.0	Sheetal	P00297742	53.0	51-55	Male	Madhya Pradesh	Central	Healthcare	Health	2.0	382.0
11252	1000695.0	Manning	P00296942	19.0	18-25	Male	Maharashtra	West	Chemical	Health	1.0	370.0
11253	1004089.0	Reichenbach	P00171342	33.0	26-35	Male	Haryana	Northern	Healthcare	Health	5.0	367.0
11255	1004023.0	Noonan	P00059442	37.0	36-45	Male	Karnataka	South	Agriculture	Clothing	4.0	206.0

3408 rows × 12 columns

-->now lets learn EDA Exploratory data analysis

[54]:
```
#describe() method returns description of the data in the data frame (i.e count,mean,std,etc
df.describe()
```

[54]:

	User_ID	Age	Orders	Amount
count	1.124500e+04	11245.000000	11245.000000	11245.000000

Home | Untitled

localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 27 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab | Python [conda env:base] *

[54]:

	User_ID	Age	Orders	Amount
count	1.124500e+04	11245.000000	11245.000000	11245.000000
mean	1.003004e+06	35.415651	3.500311	9461.934237
std	1.716207e+03	12.756369	1.713706	5234.426634
min	1.000001e+06	12.000000	1.000000	188.000000
25%	1.001492e+06	27.000000	2.000000	5443.000000
50%	1.003065e+06	33.000000	4.000000	8109.000000
75%	1.004429e+06	43.000000	5.000000	12683.000000
max	1.006040e+06	92.000000	6.000000	29350.000000

[56]:
```
#use describe for specific column
df[['Age','Orders','Amount']].describe()
```

[56]:

	Age	Orders	Amount
count	11245.000000	11245.000000	11245.000000
mean	35.415651	3.500311	9461.934237
std	12.756369	1.713706	5234.426634
min	12.000000	1.000000	188.000000
25%	27.000000	2.000000	5443.000000
50%	33.000000	4.000000	8109.000000
75%	43.000000	5.000000	12683.000000
max	92.000000	6.000000	29350.000000

Home | Untitled

localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 27 minutes ago

File Edit View Run Kernel Settings Help

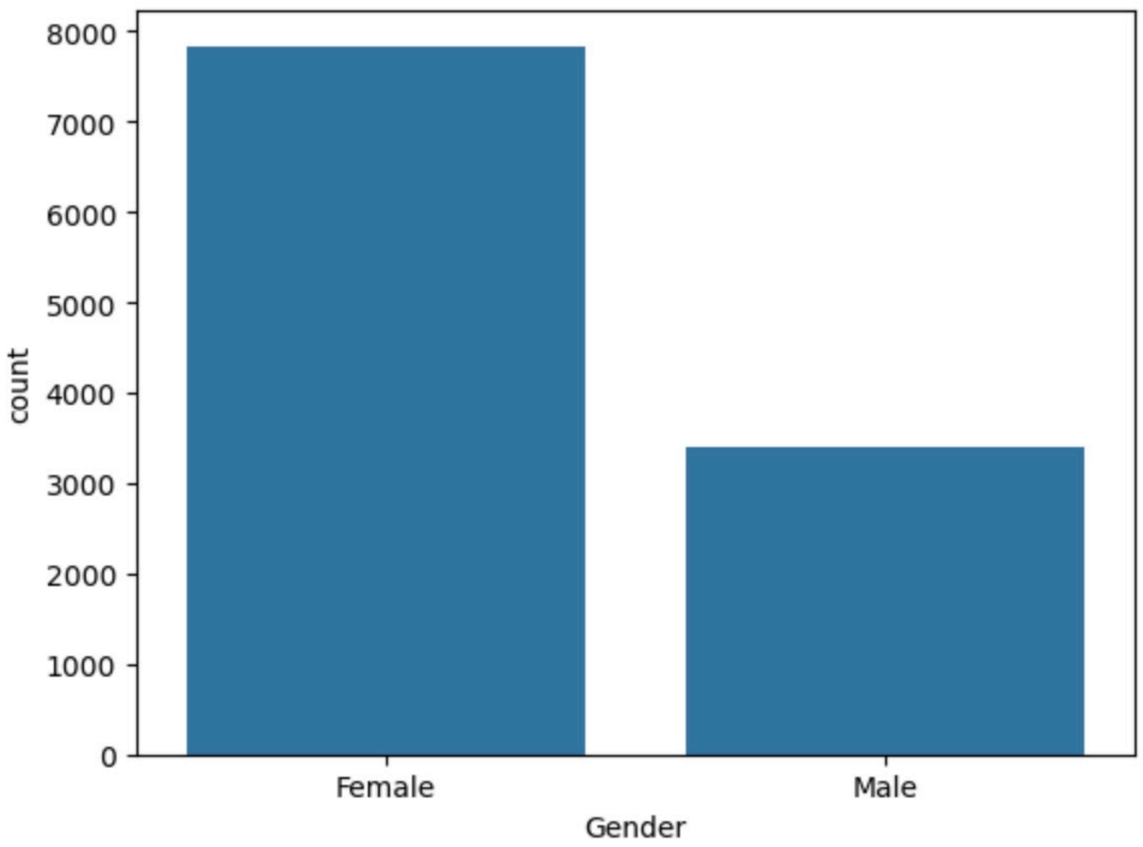
Trusted

Code

JupyterLab

Python [conda env:base] *

[59]:
```
sns.countplot(x='Gender',data=df)
plt.show()
```



[66]:
```
#1 total transaction counts by gender wise show in bar charts
ax=sns.countplot(x='Gender',data =df)
#set title and label with font size
for bars in ax.containers:
    ax.bar_label(bars)
```

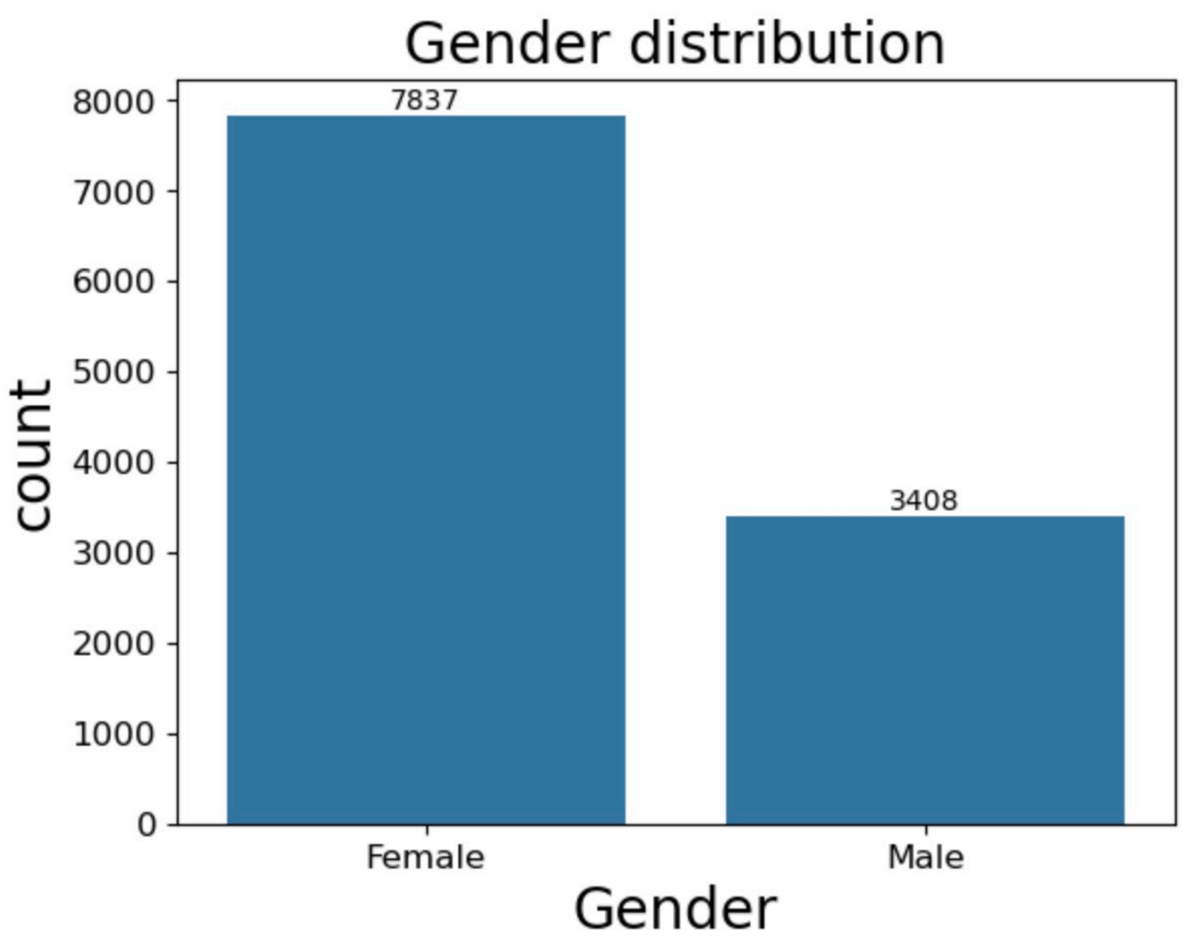
Home | Untitled | localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 27 minutes ago

File Edit View Run Kernel Settings Help | Trusted

Code | JupyterLab | Python [conda env:base] *

```
for bars in ax.containers:
    ax.bar_label(bars)
    ax.set_title ('Gender distribution',fontsize=20)#chart title
    ax.set_xlabel('Gender',fontsize=20)#x axis label
    ax.set_ylabel('count',fontsize=20)#y axis label
    ax.tick_params(axis='both',labelsize=12)#axis parameter
plt.show()
```



Home | Untitled

localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 27 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

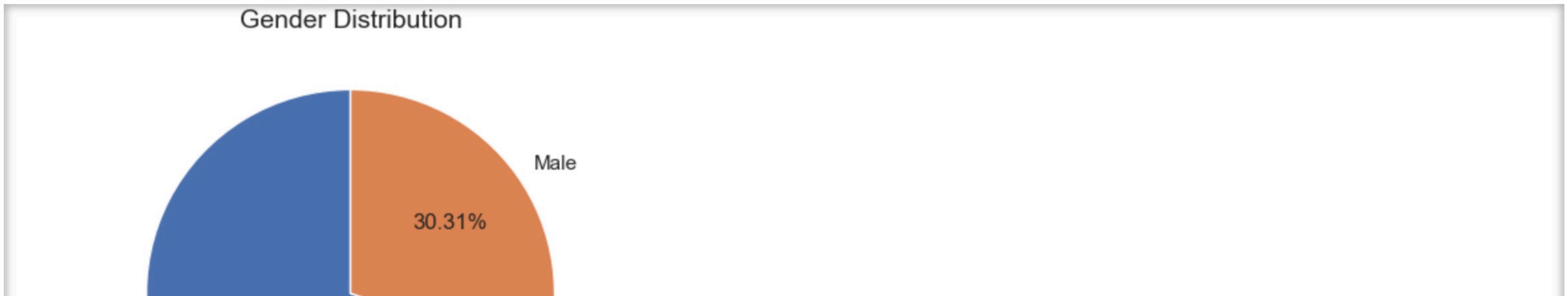
JupyterLab

Python [conda env:base] *

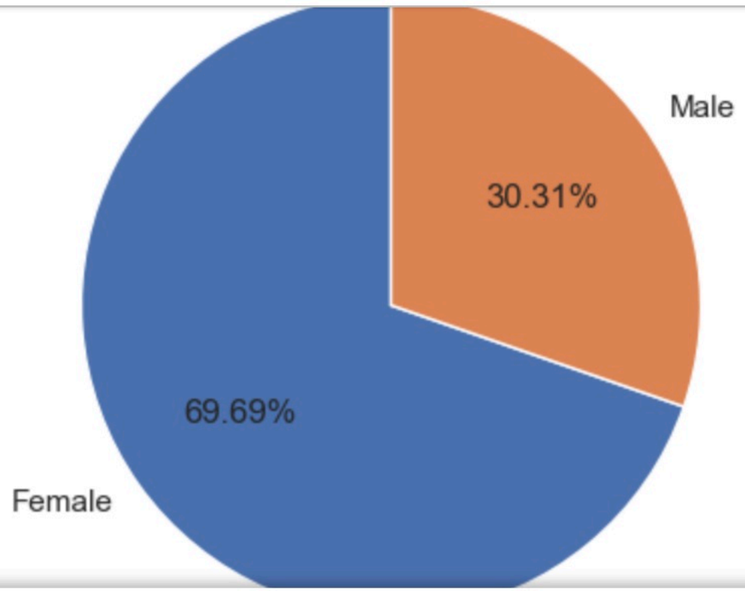
[67]:
```
#total trans count by gender
gender_count=df['Gender'].value_counts()
print(gender_count)
```

```
Gender
Female    7837
Male      3408
Name: count, dtype: int64
```

[140]:
```
#2 Gender wise distribution in pie chart
gender_counts = df['Gender'].value_counts()

plt.pie(
    gender_counts,                      # Data (counts)
    labels=gender_counts.index,         # Labels (e.g., Male, Female)
    autopct='%.2f%%',                   # Show percentages (2 decimal place)
    startangle=90,                      # Rotate for better orientation
)
# Add a title
plt.title('Gender Distribution', fontsize=14)
plt.show()
```



jupyter **Untitled** Last Checkpoint: 28 minutes ago



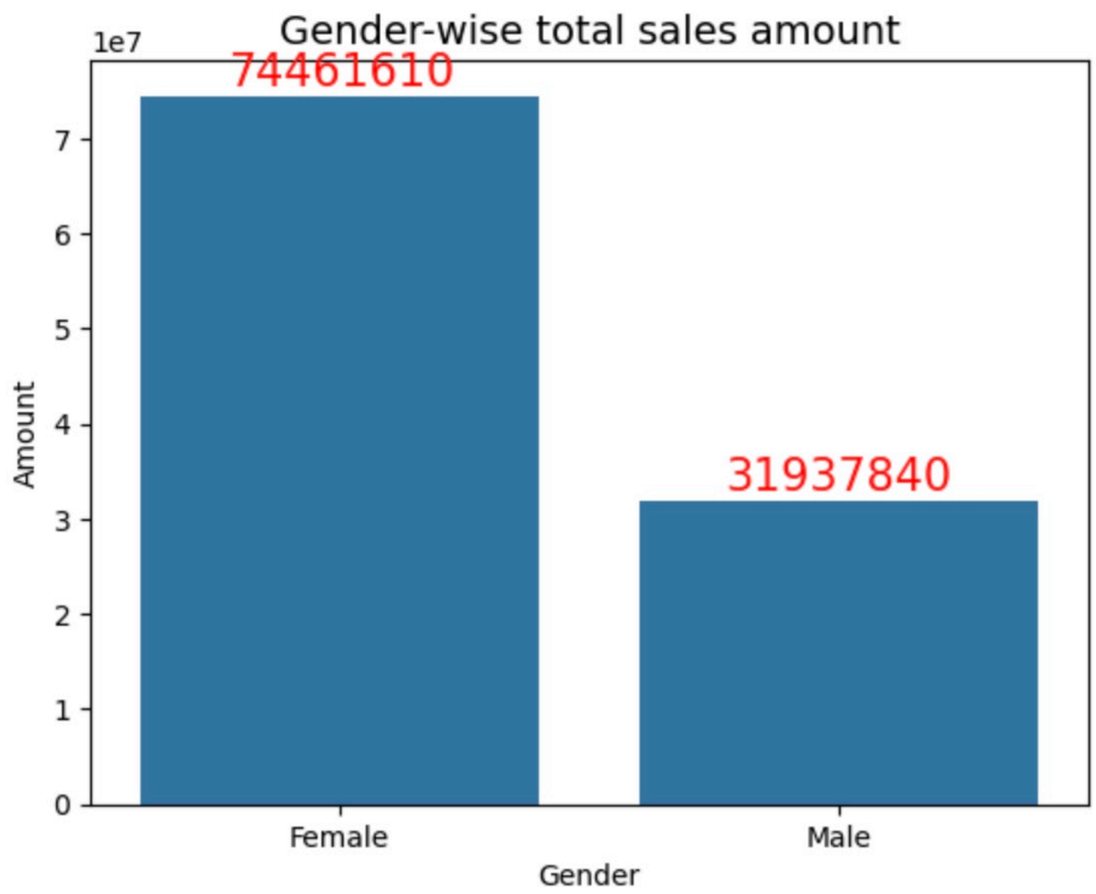
[89]:
```
df.groupby('Gender',as_index=False)['Amount'].sum()
```

[89]:

	Gender	Amount
0	Female	74461610.49
1	Male	31937840.00

[95]:
```
#3 gender wise total sales amount bar chart

#step 1: Group data by gender and sum the amount
Gen_Wise_Sales = df.groupby('Gender',as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False)
#step 2: create the bar plot
ax=sns.barplot(x='Gender',y='Amount',data=Gen_Wise_Sales)
#step 3: add data labels on top of bars
for bar in ax.containers:
    ax.bar_label(bar,fmt='%.0f',fontsize=16,color='red')
```

```
#step 3: add data labels on top of bars
for bar in ax.containers:
    ax.bar_label(bar,fmt='%.0f',fontsize=16,color='red')
#step 4: add titles
plt.title('Gender-wise total sales amount',fontsize=14)
plt.show()
```

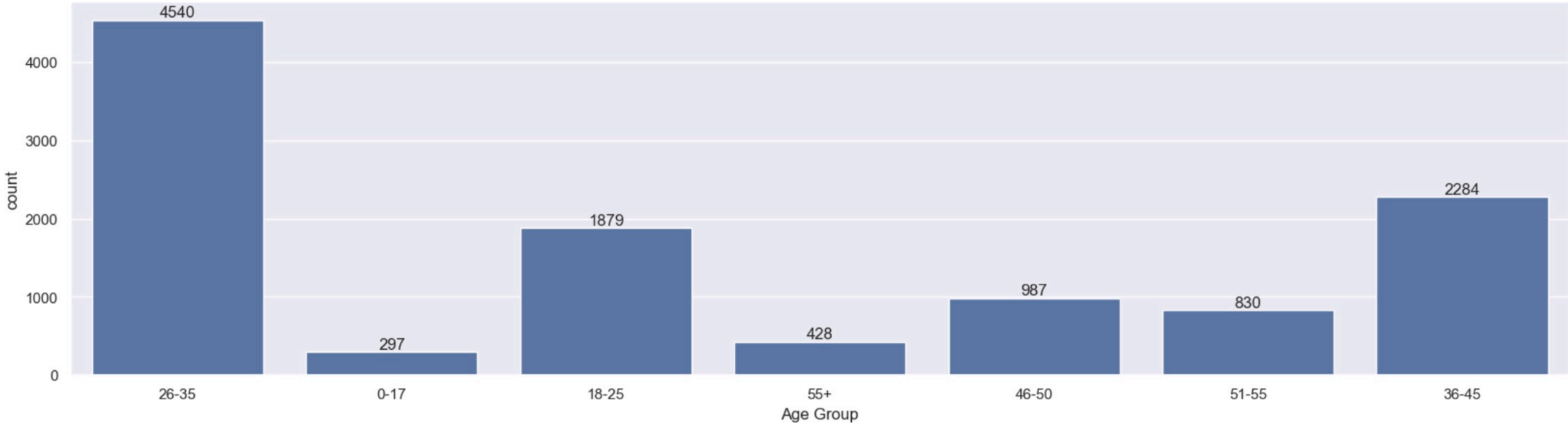


--> Age group wise total transactions

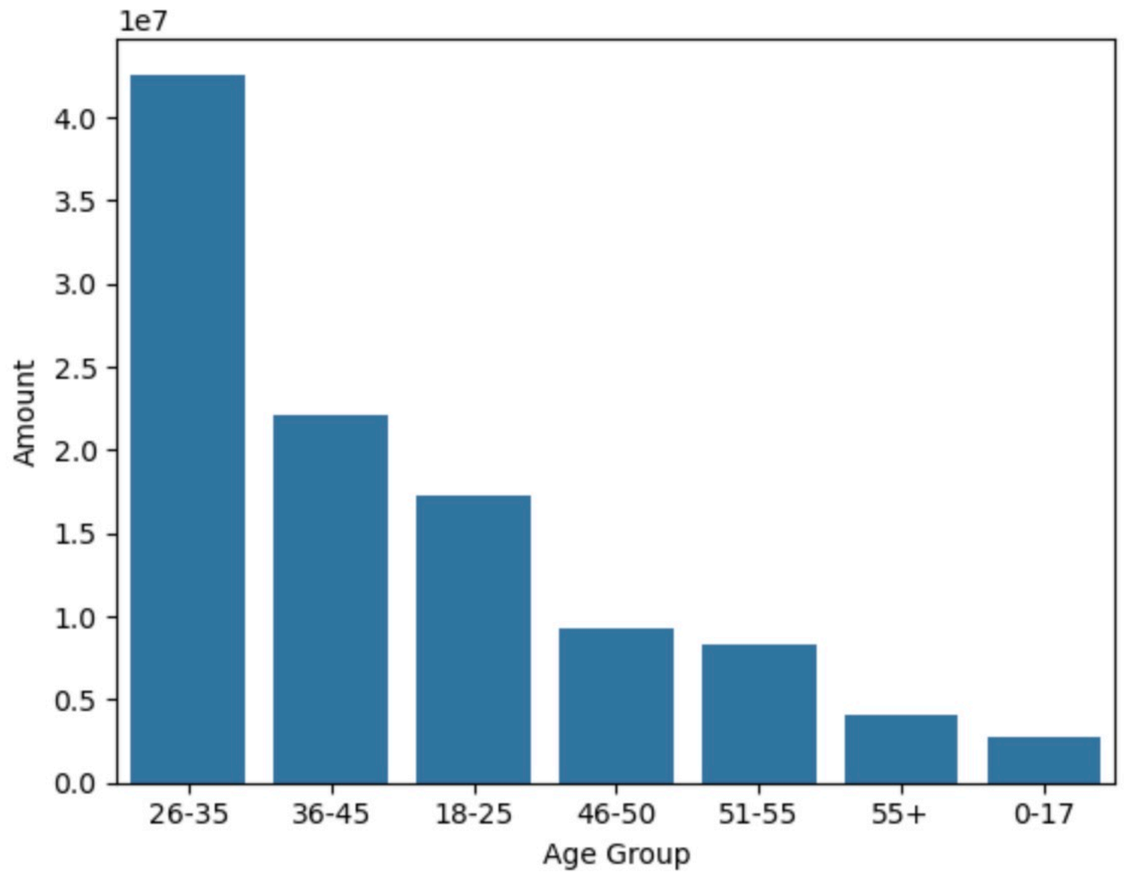
jupyter **Untitled** Last Checkpoint: 28 minutes ago

--> Age group wise total transactions

[139]:
```
#Age group wise trans. count in bar chart
ax = sns.countplot(data =df, x='Age Group')
for bars in ax.containers:  #show data labels
    ax.bar_label(bars)
plt.show()
```



[106]:
```
#5 age group wise total amount
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()
```



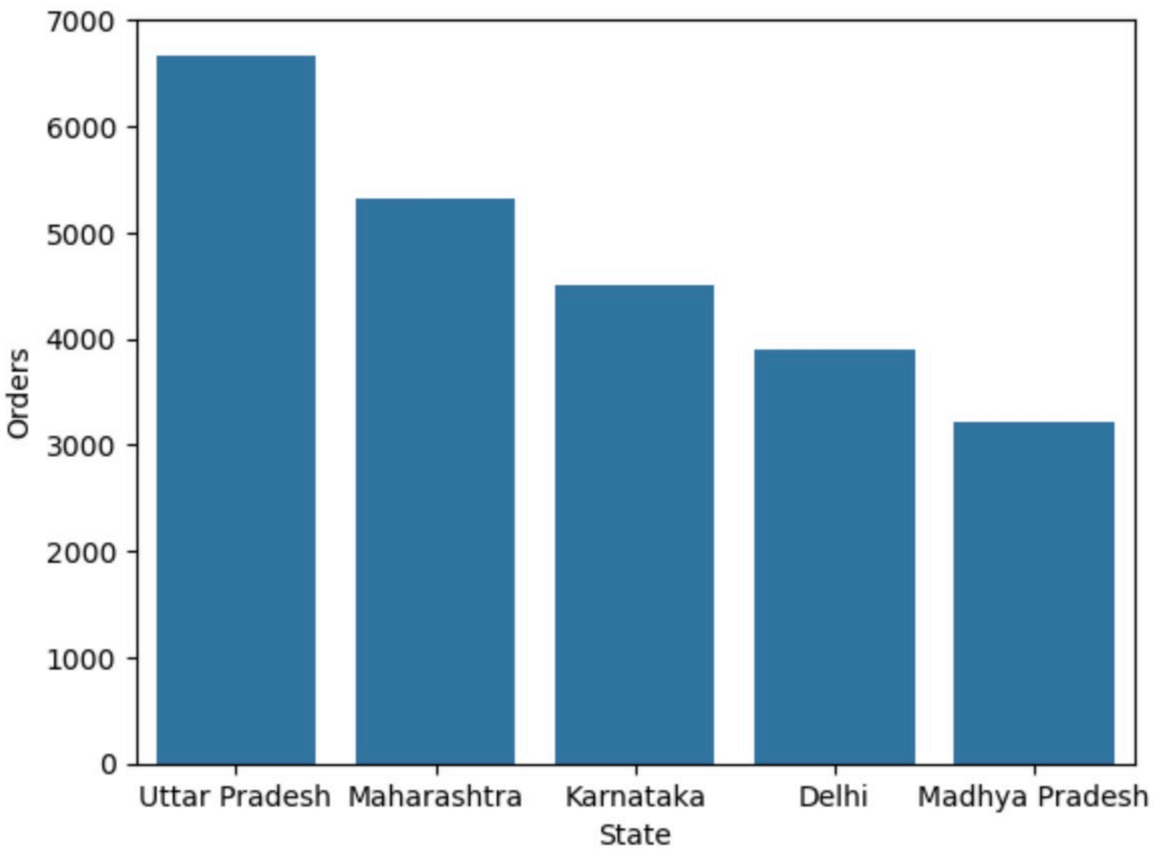
--> State wise analysis

[112]:
```
#6 top 5 stateby total quantity  ordered
sales_state=df.groupby(['State'],as_index=False)['Orders'].sum().sort_values(by='Orders',ascending=False).head()
sns.barplot(data=sales_state,x='State',y='Orders')
plt.show()
```

jupyter **Untitled** Last Checkpoint: 28 minutes ago

--> State wise analysis

[112]:
```
#6 top 5 stateby total quantity  ordered
sales_state=df.groupby(['State'],as_index=False)['Orders'].sum().sort_values(by='Orders',ascending=False).head()
sns.barplot(data=sales_state,x='State',y='Orders')
plt.show()
```



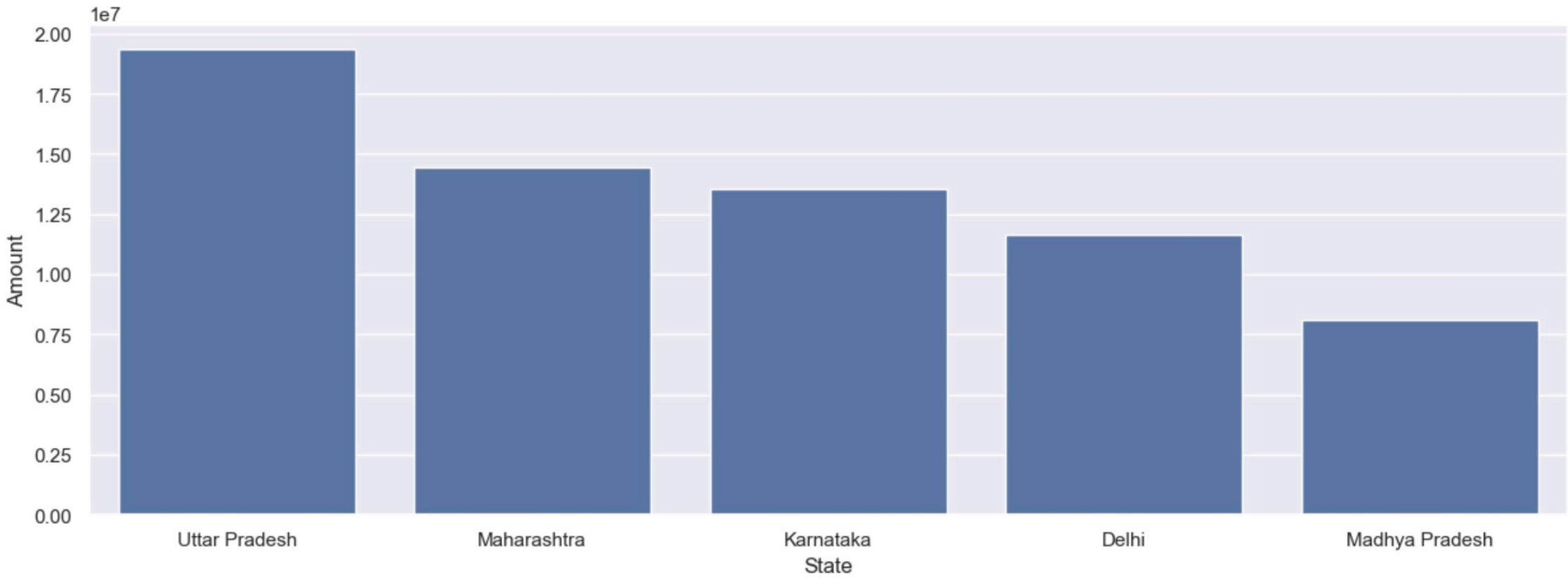
Home | Untitled | localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 28 minutes ago

File Edit View Run Kernel Settings Help | Trusted

Code | JupyterLab | Python [conda env:base] *

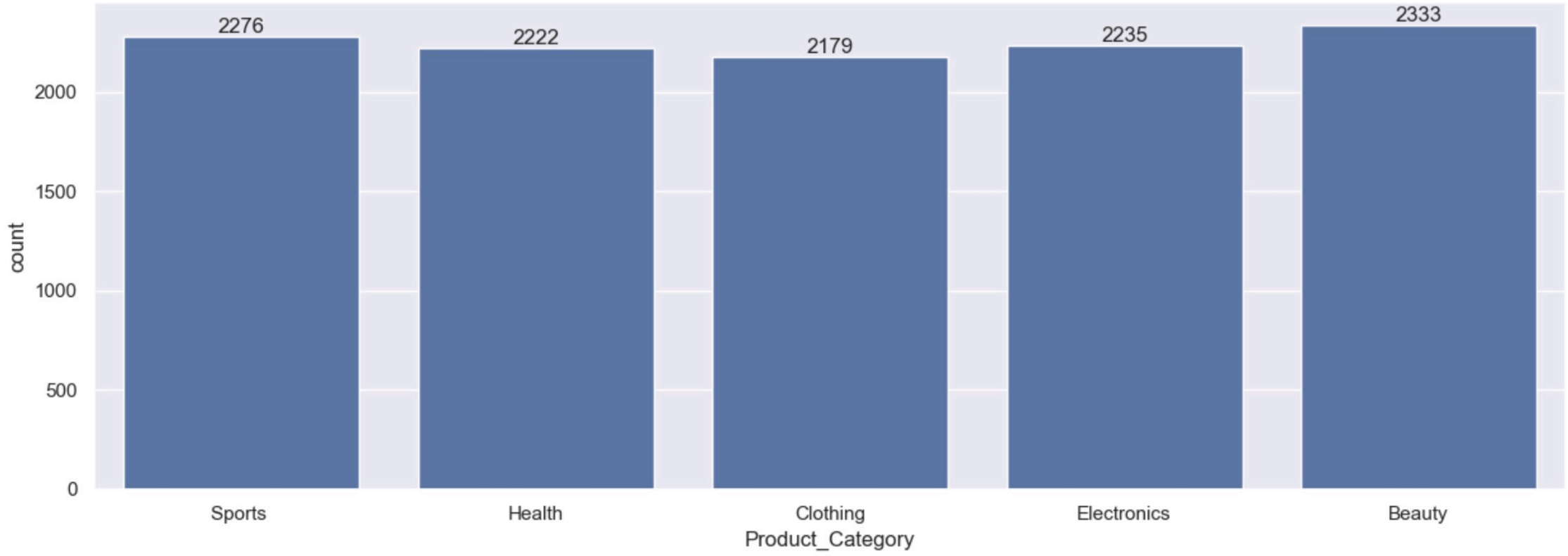
[117]:
```
#7 top 5 state as per total sales amount
sales_state=df.groupby(['State'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False).head()
#print(sales_state)
sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data=sales_state,x='State',y='Amount')
plt.show()
```



--> Product_Category Analysis

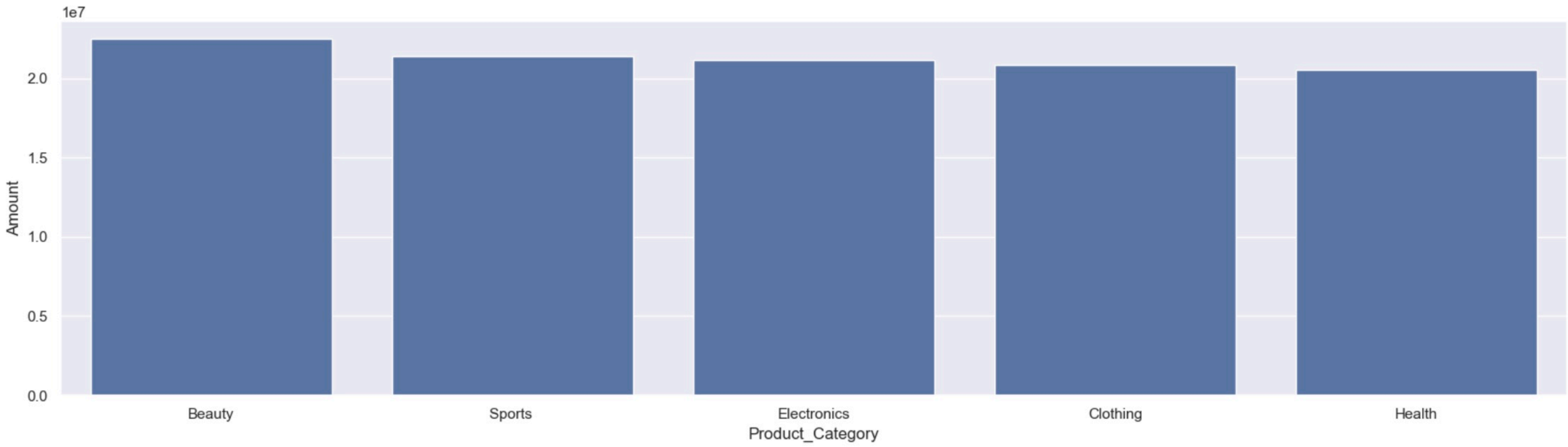
--> Product_Category Analysis

[119]:
```
#8 product category wise transcation count
ax=sns.countplot(data=df,x='Product_Category')
sns.set (rc={'figure.figsize':(2,6)})
for bars in ax.containers:
    ax.bar_label(bars)
    plt.show()
```



[129]:
```
#9 amount wise product category in bar chart
```

[129]:
```
#9 amount wise product category in bar chart
sales_state = df.groupby(['Product_Category'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False).head()

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



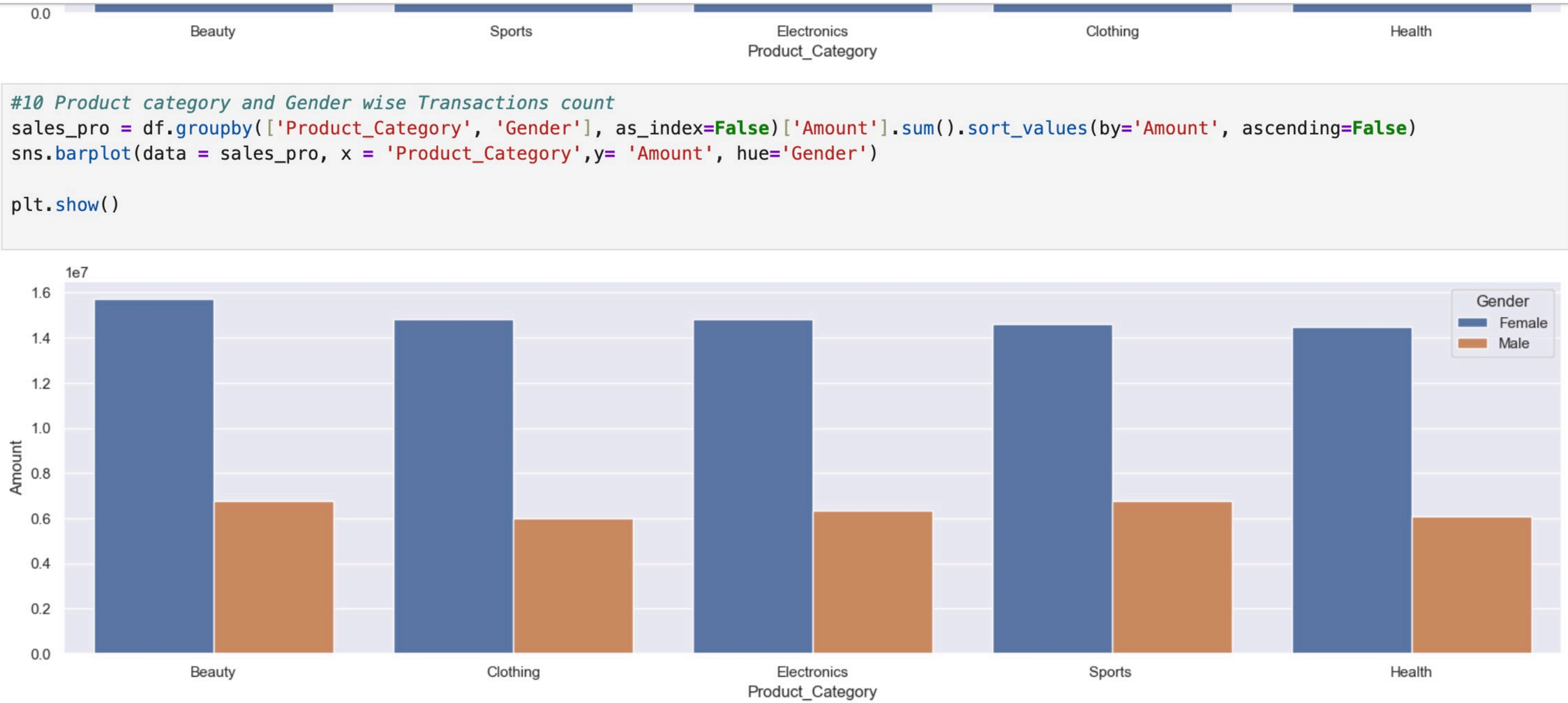
[133]:
```
#10 Product category and Gender wise Transactions count
sales_pro = df.groupby(['Product_Category', 'Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(data = sales_pro, x = 'Product_Category',y= 'Amount', hue='Gender')

plt.show()
```

Home | Untitled | localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 28 minutes ago

File Edit View Run Kernel Settings Help | Trusted

Code | JupyterLab | Python [conda env:base] *



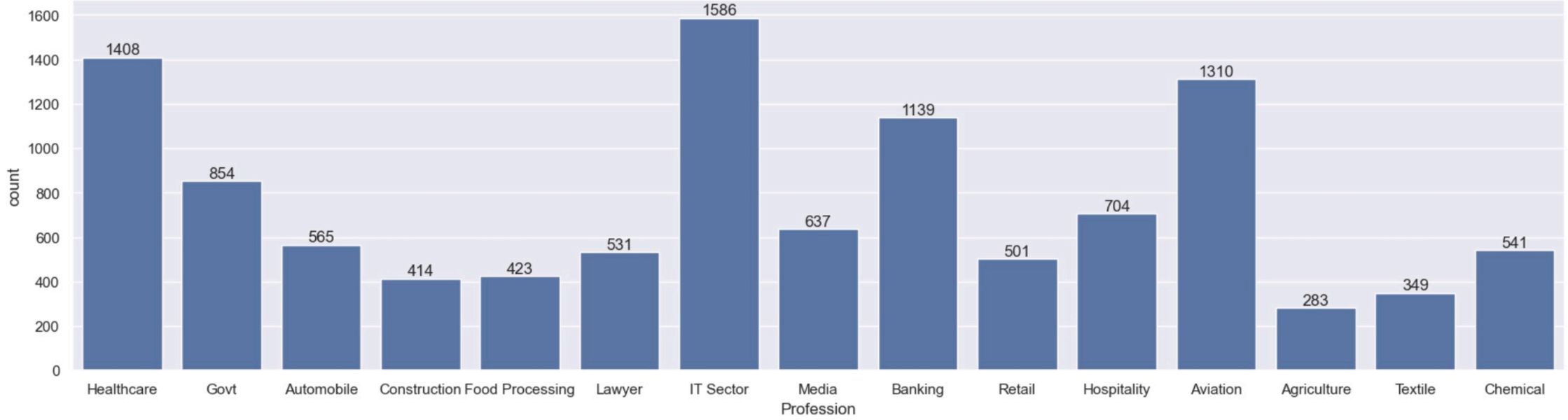
[133]:
```
#10 Product category and Gender wise Transactions count
sales_pro = df.groupby(['Product_Category', 'Gender'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(data = sales_pro, x = 'Product_Category',y= 'Amount', hue='Gender')

plt.show()
```


--> Profession wise Analysis

[134]:
```
#11 Profession wise Transactions count
```

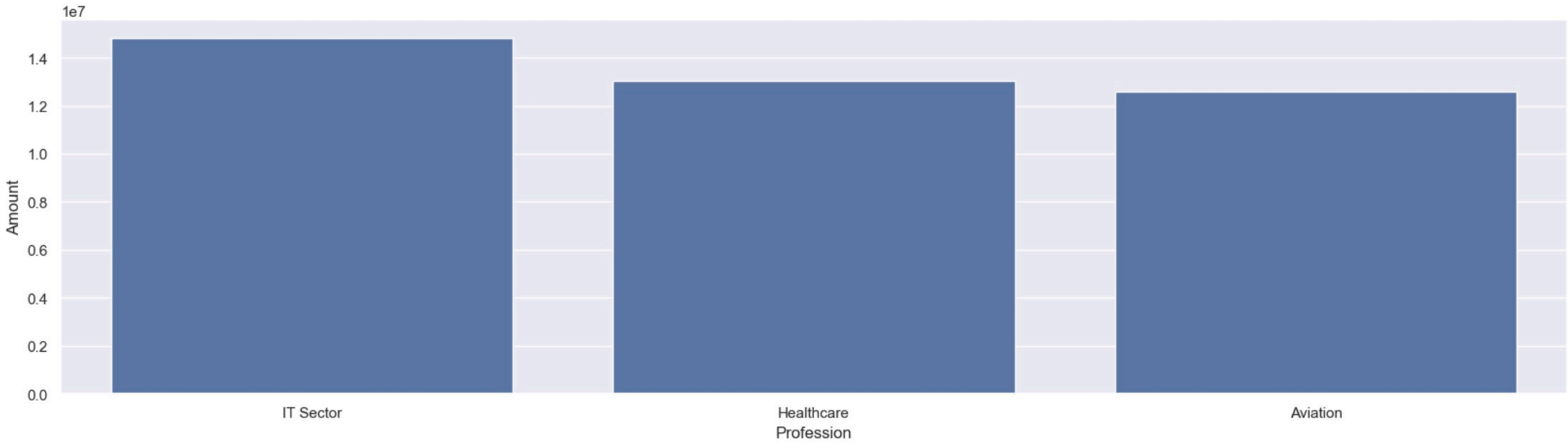
--> Profession wise Analysis

[134]:
```
#11 Profession wise Transactions count

ax = sns.countplot(data = df, x = 'Profession')
for bars in ax.containers:
    ax.bar_label(bars)
plt.show()
```



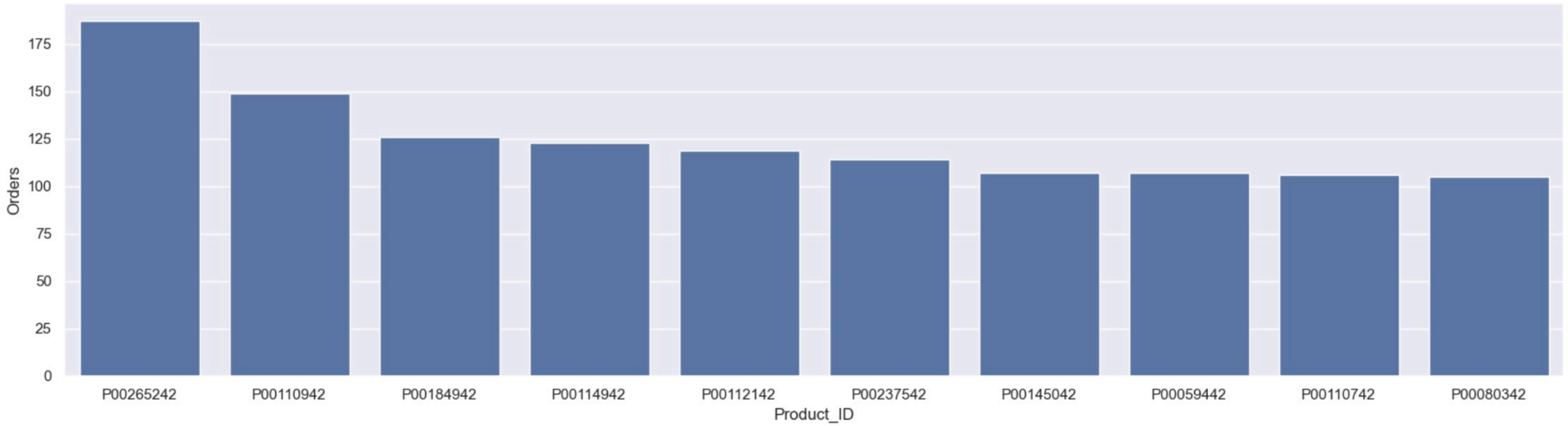
[137]:
```
#12 Amount Wise Top Professions
sales_state = df.groupby(['Profession'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(3)

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

[137]:
```
#12 Amount Wise Top Professions
sales_state = df.groupby(['Profession'], as_index=False)['Amount'].sum().sort_values(by='Amount', ascending=False).head(3)

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



[138]:
```
#13 Order wise Top 10 Product
sales_state = df.groupby(['Product_ID'], as_index=False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)

sns.set(rc={'figure.figsize':(20,5)})
```

[138]:
```
#13 Order wise Top 10 Product
sales_state = df.groupby(['Product_ID'], as_index=False)['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)

sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Product_ID',y= 'Orders')
plt.show()
```



Conclusion:

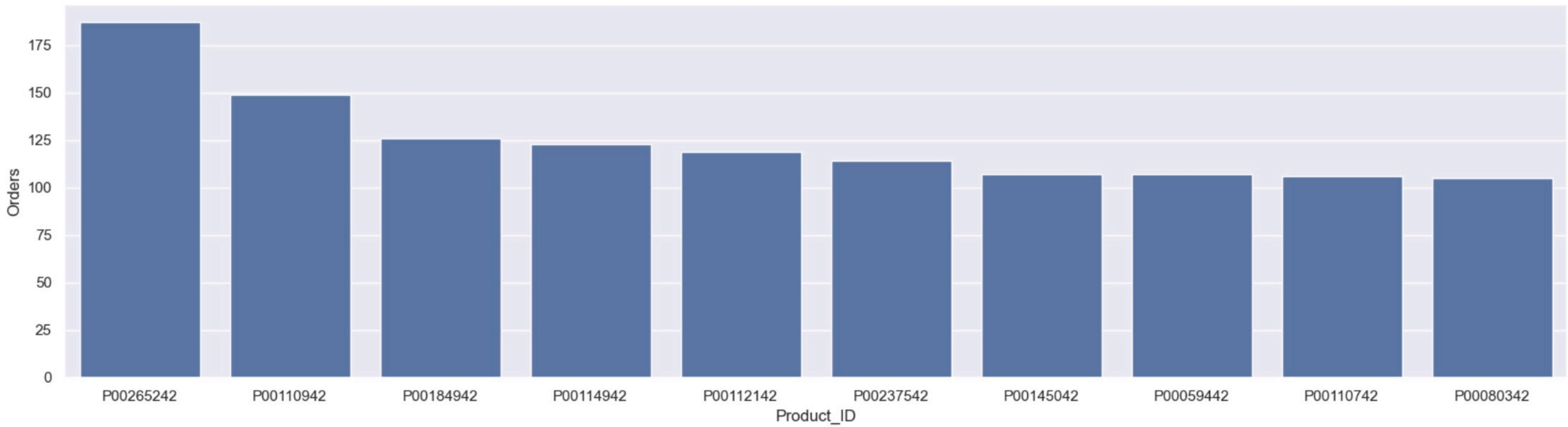
Home | Untitled | localhost:8888/notebooks/Desktop/anaylsis/Untitled.ipynb?

jupyter **Untitled** Last Checkpoint: 28 minutes ago

File Edit View Run Kernel Settings Help | Trusted

Code | JupyterLab | Python [conda env:base] *

```
sns.barplot(data = sales_state, x = 'Product_ID',y= 'Orders')
plt.show()
```



Conclusion:

Female age group 26-35 yrs from UP, Maharastra and Karnataka working in IT, Healthcare and Aviation are more likely to buy products from Beauty, Sports and Electronics