DIWALI SALES ANALYSIS

By Chirag Modha

CONTENTS

INTRODUCTION

- Hello! I'm Chirag Modha and I've performed the following Data Analytics Project on Diwali Sales with Python using numpy, pandas, matplotlib and seaborn.
- The dataset was provided by Rishabh Mishra
- Dataset Link –
- The data was loaded into jupyter notebook, it required some minor cleaning and transformation and further EDA and Visualizations were performed.
- The objective of this analysis is to identify customer demographics, including geographic distribution by area, determine the states that generate the highest revenue, and analyze which product categories drive the most sales and revenue.

ORIGINAL DATASET

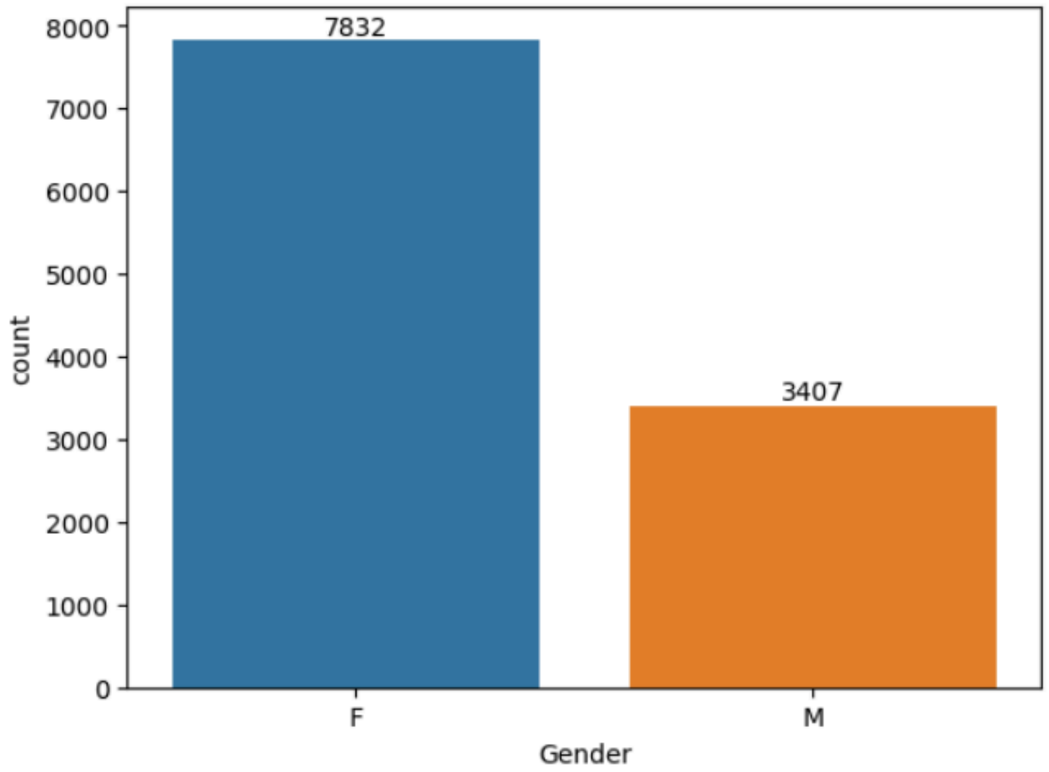
	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	unnamed
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.00	NaN	Nal
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.00	NaN	Nal
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.00	NaN	Nal
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.00	NaN	Nal
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.00	NaN	Nal
5	1000588	Joni	P00057942	M	26-35	28	1	Himachal Pradesh	Northern	Food Processing	Auto	1	23877.00	NaN	Nal
6	1001132	Balk	P00018042	F	18-25	25	1	Uttar Pradesh	Central	Lawyer	Auto	4	23841.00	NaN	Nal
7	1002092	Shivangi	P00273442	F	55+	61	0	Maharashtra	Western	IT Sector	Auto	1	NaN	NaN	Nal
8	1003224	Kushal	P00205642	M	26-35	35	0	Uttar Pradesh	Central	Govt	Auto	2	23809.00	NaN	Nal
9	1003650	Ginny	P00031142	F	26-35	26	1	Andhra Pradesh	Southern	Media	Auto	4	23799.99	NaN	Nal

CLEANED DATA

	User_ID	Cust_name	Product_ID	Gender	Age_Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877
...	...	...	...	...	...	...	...	...	...	...	...	...	...
11246	1000695	Manning	P00296942	M	18-25	19	1	Maharashtra	Western	Chemical	Office	4	370
11247	1004089	Reichenbach	P00171342	M	26-35	33	0	Haryana	Northern	Healthcare	Veterinary	3	367
11248	1001209	Oshin	P00201342	F	36-45	40	0	Madhya Pradesh	Central	Textile	Office	4	213
11249	1004023	Noonan	P00059442	M	36-45	37	0	Karnataka	Southern	Agriculture	Office	3	206
11250	1002744	Brumley	P00281742	F	18-25	19	0	Maharashtra	Western	Healthcare	Office	3	188

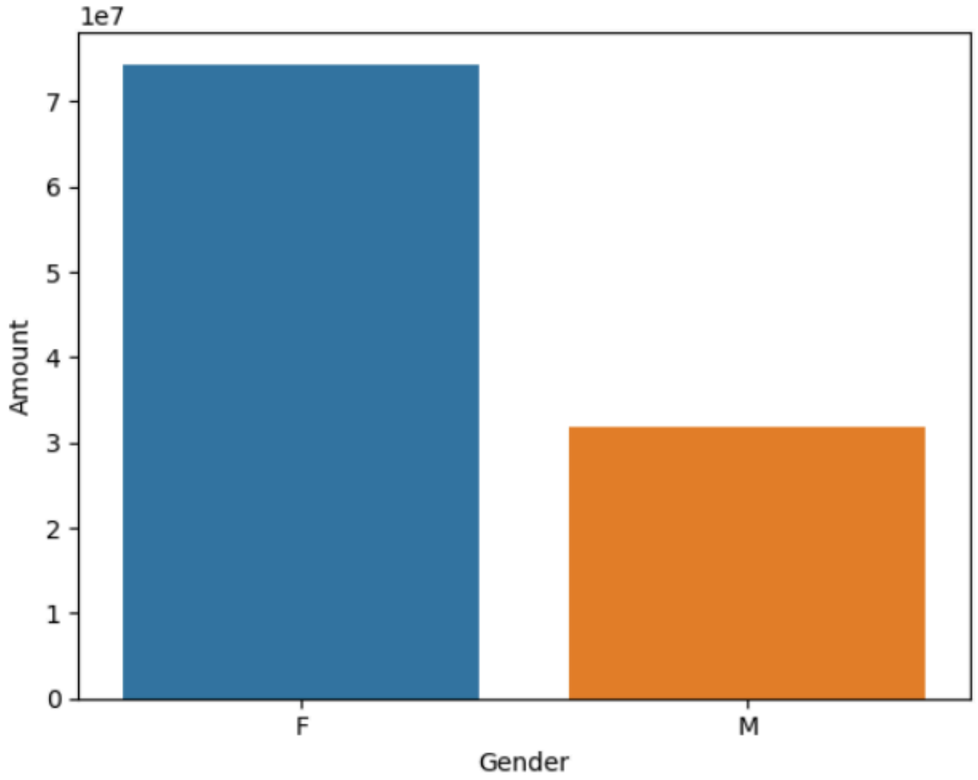
CUSTOMER COUNT BASED ON GENDER

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

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



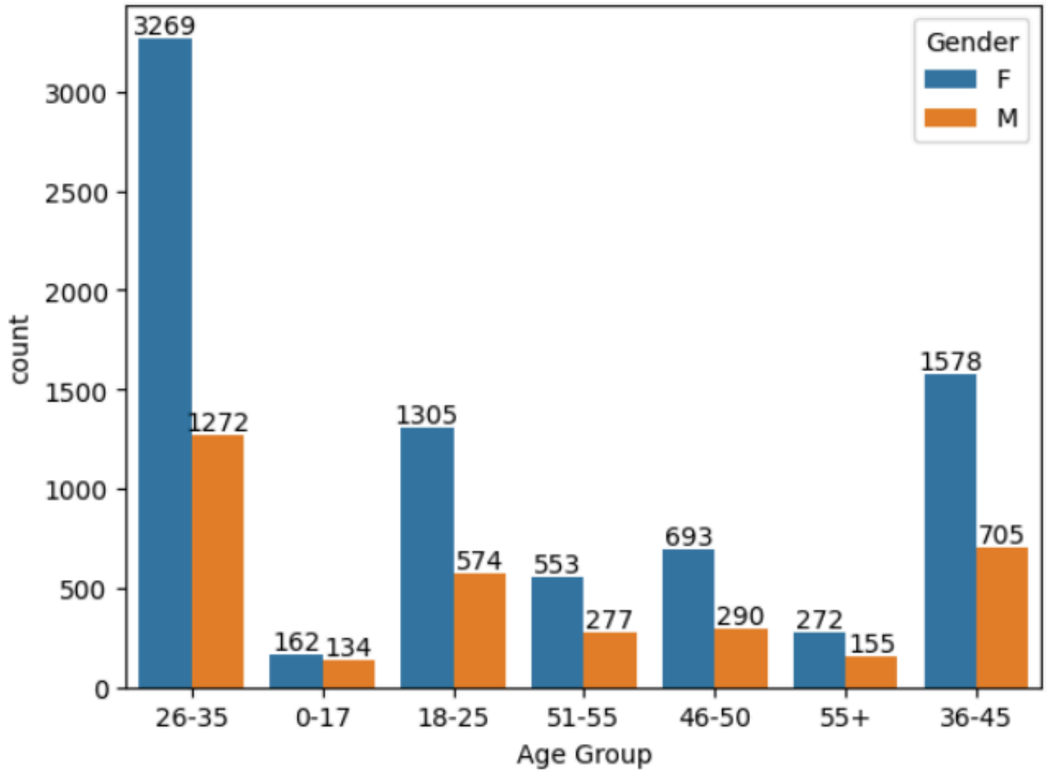
AMOUNT SPENT GENDER WISE

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

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



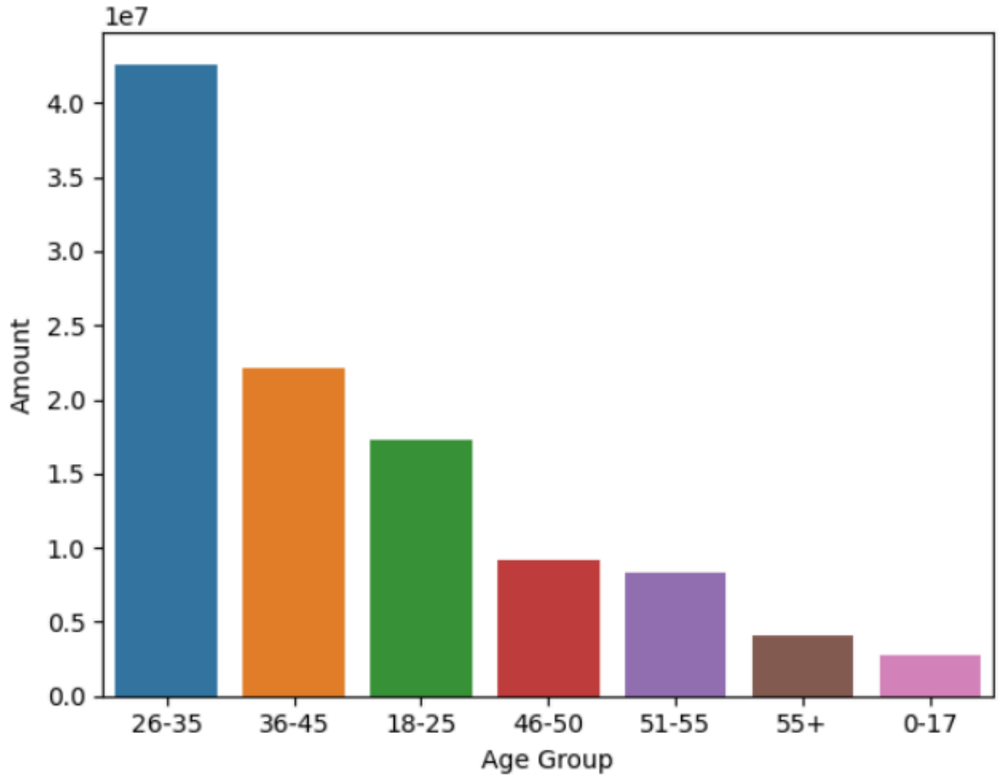
CONSUMERS COUNT ON AGE GROUP FURTHER DIVIDED BY GENDER

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

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



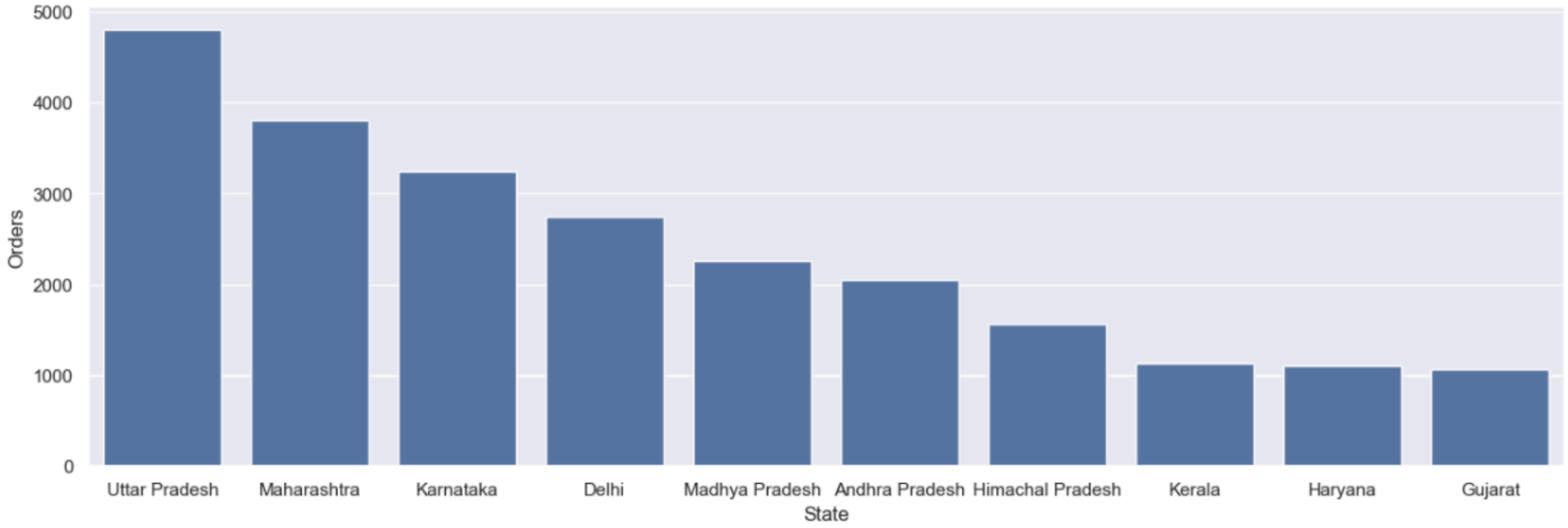
AMOUNT SPENT BY EACH AGE GROUP

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

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



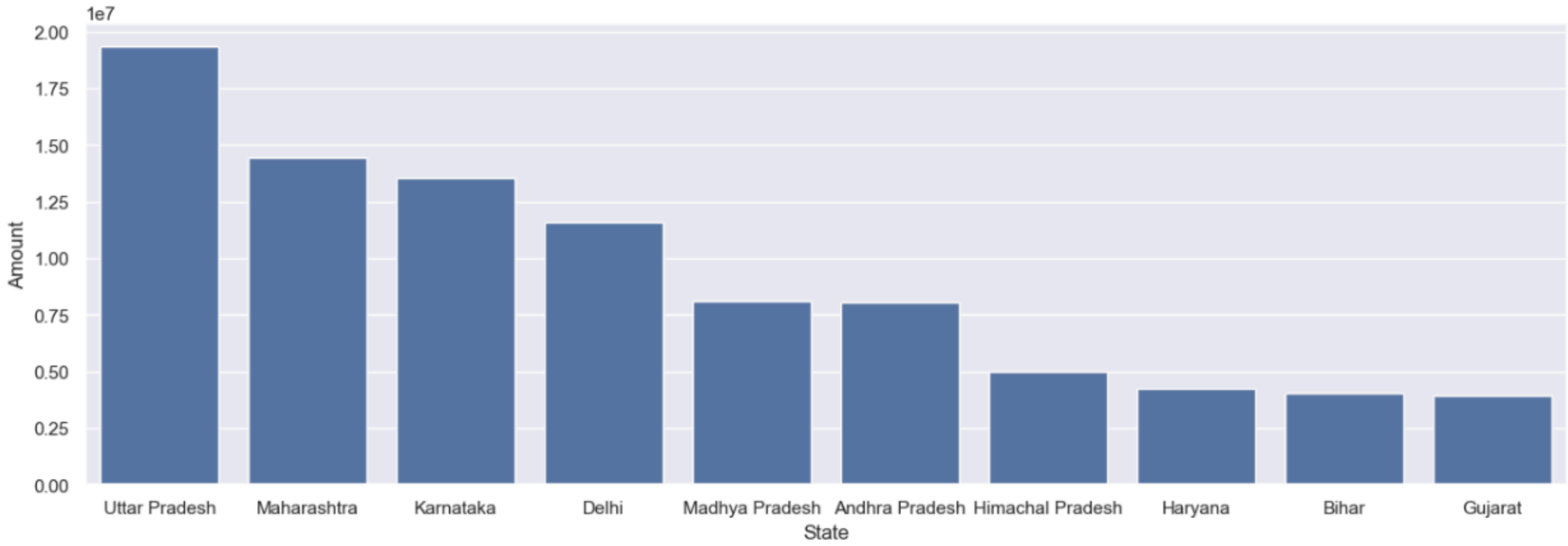
NUMBER OF ORDER FROM TOP 10 STATES

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

sns.set(rc={'figure.figsize':(16,5)})
sns.barplot(data = sales_state, x = 'State',y = 'Orders')
```



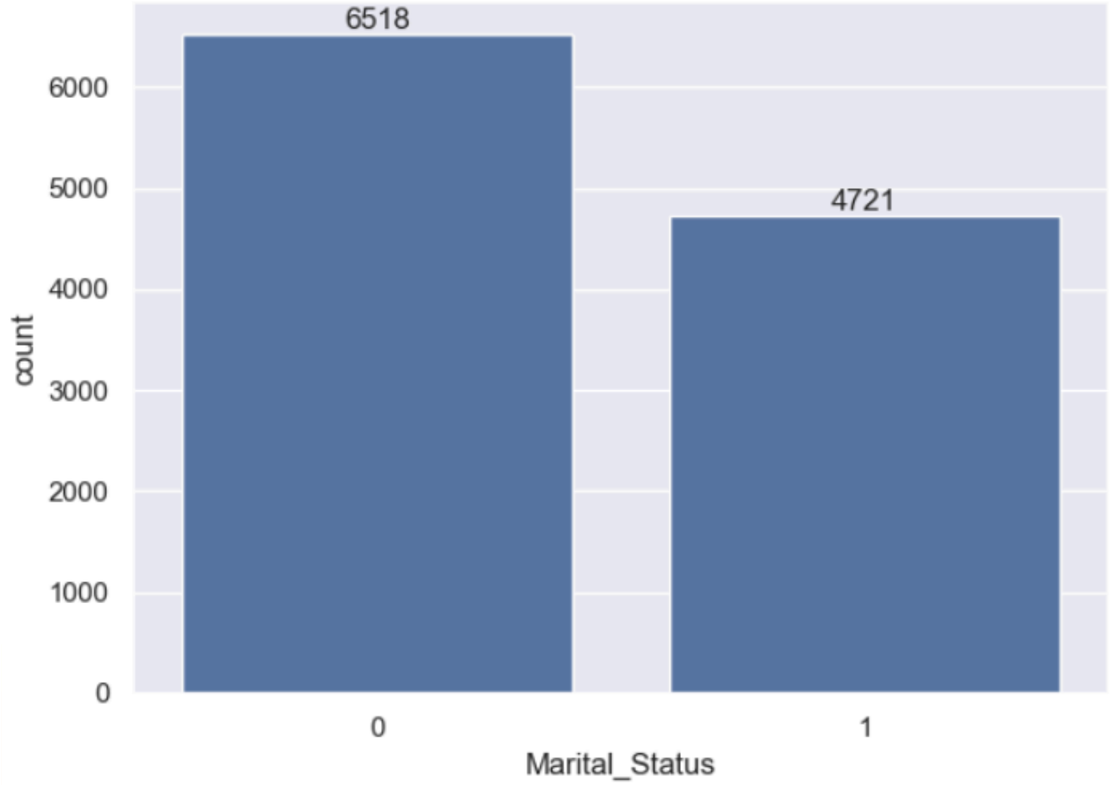
TOTAL SALES FROM TOP 10 STATES

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

sns.set(rc={'figure.figsize':(16,5)})
sns.barplot(data = sales_state, x = 'State',y = 'Amount')
```



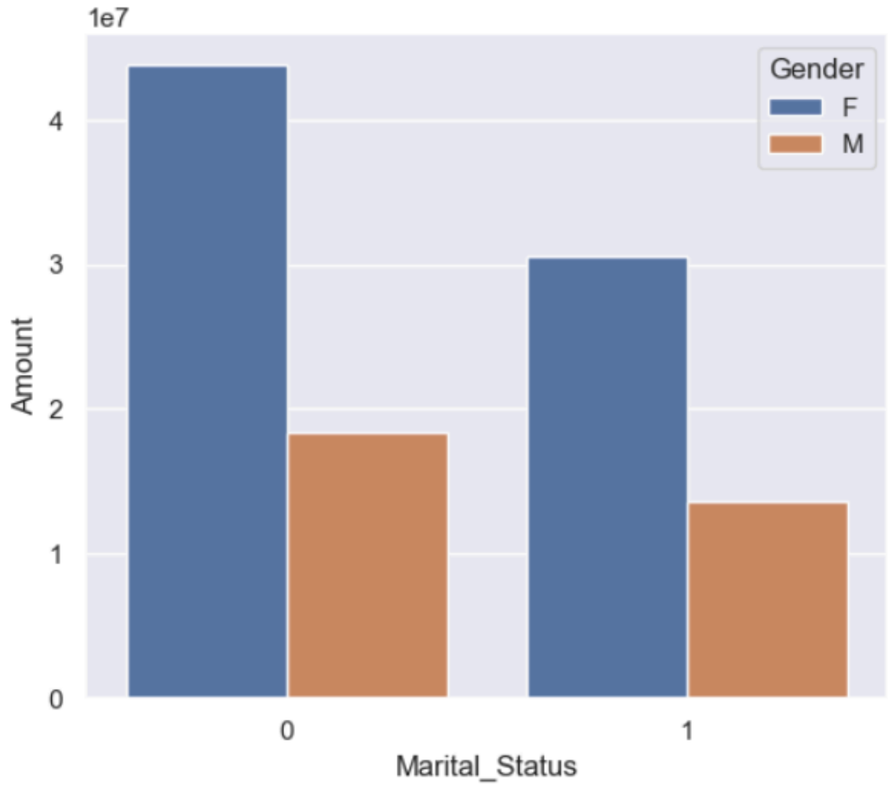
MARITAL STATUS O F CONSUMERS

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

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



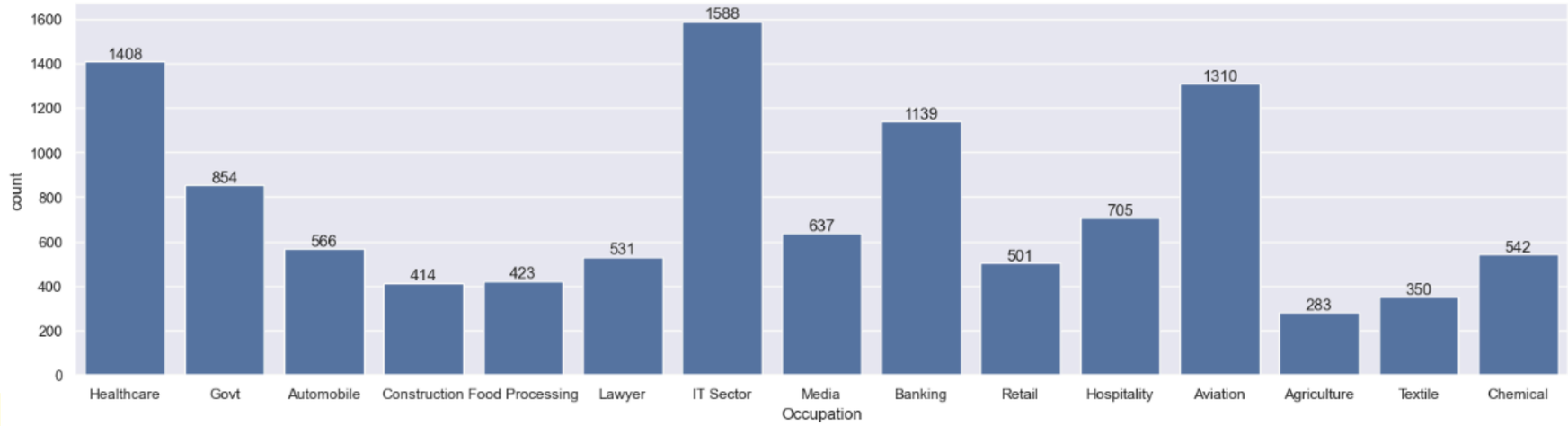
AMOUNT SPENT BASED ON MARITAL STATUS

```
sales_mstatus = 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_mstatus, x = 'Marital_Status',y = 'Amount', hue='Gender')
```



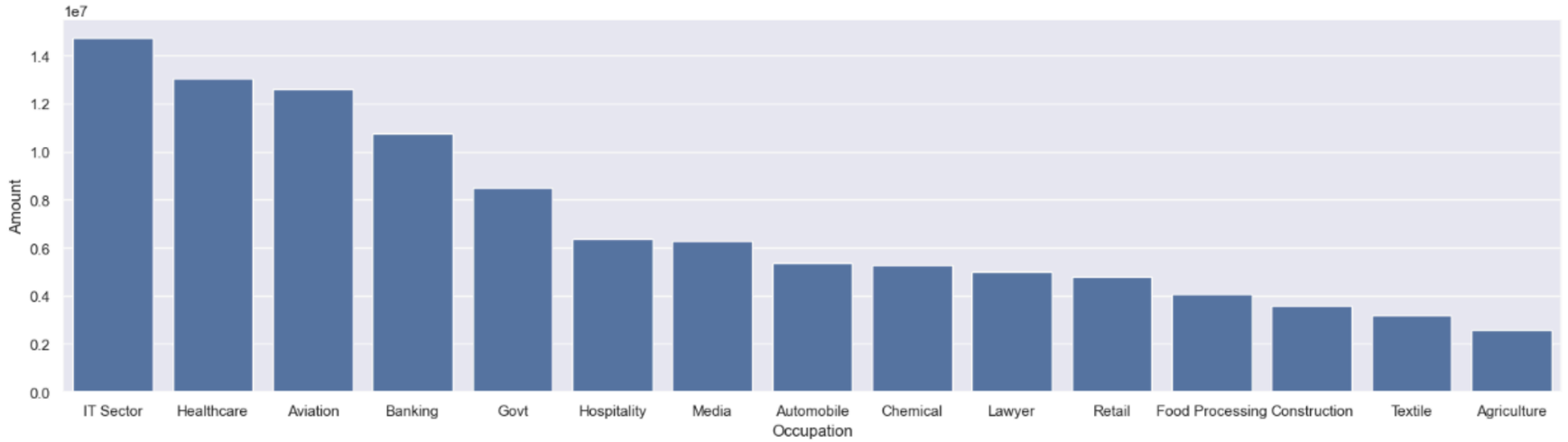
TOTAL COUNT OF ORDERS BASED ON OCCUPATION OF CONSUMERS

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

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



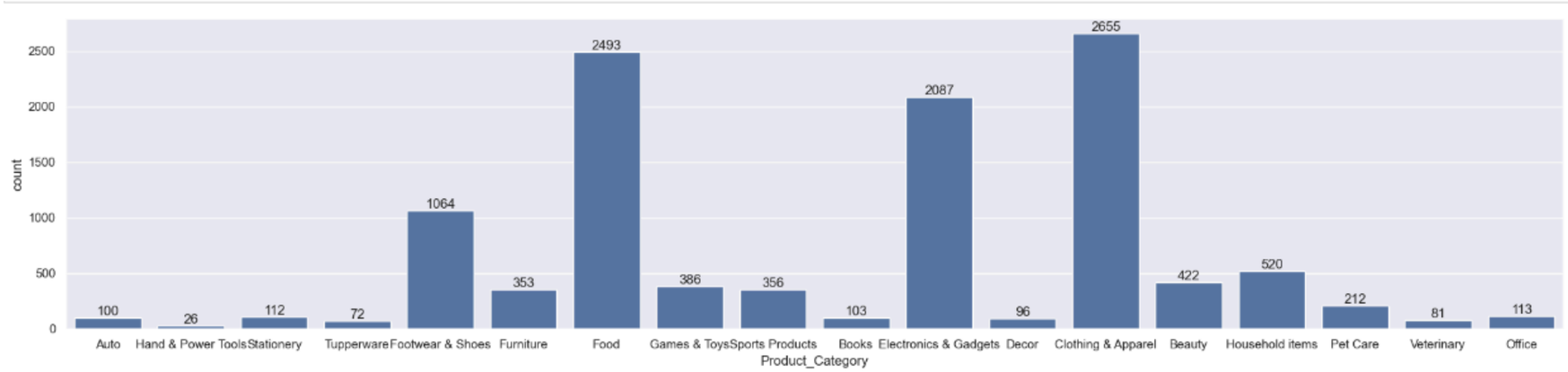
TOTAL AMOUNT OF SALES BASED ON OCCUPATION

```
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')
```



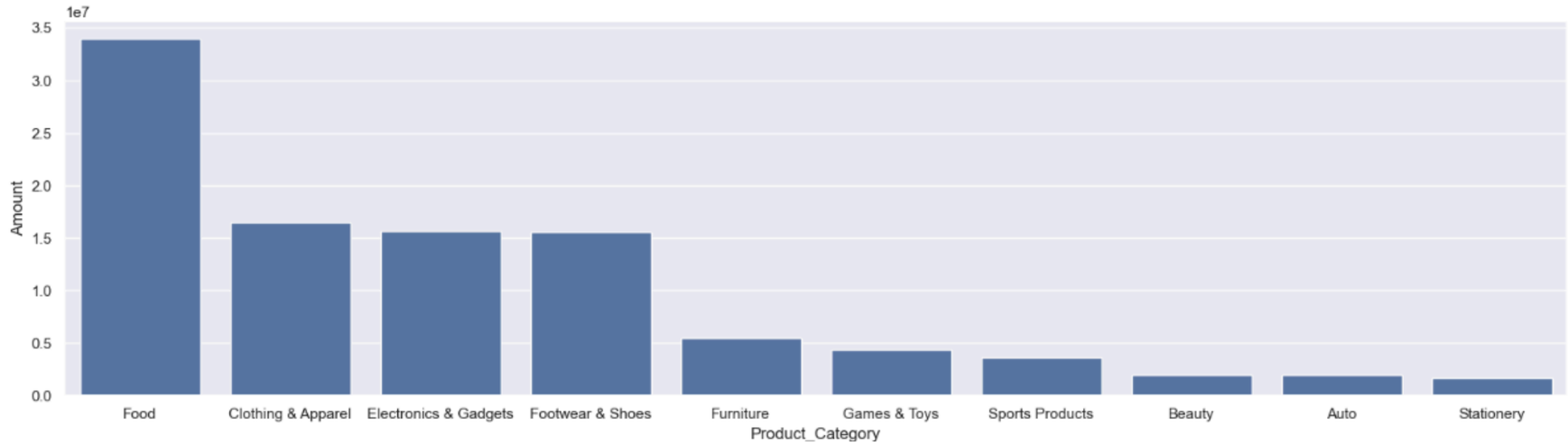
COUNT OF PRODUCT BASED ON PRODUCT CATEGORY

```
sns.set(rc={'figure.figsize':(24,5)})
ax = sns.countplot(data = df, x = 'Product_Category')

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



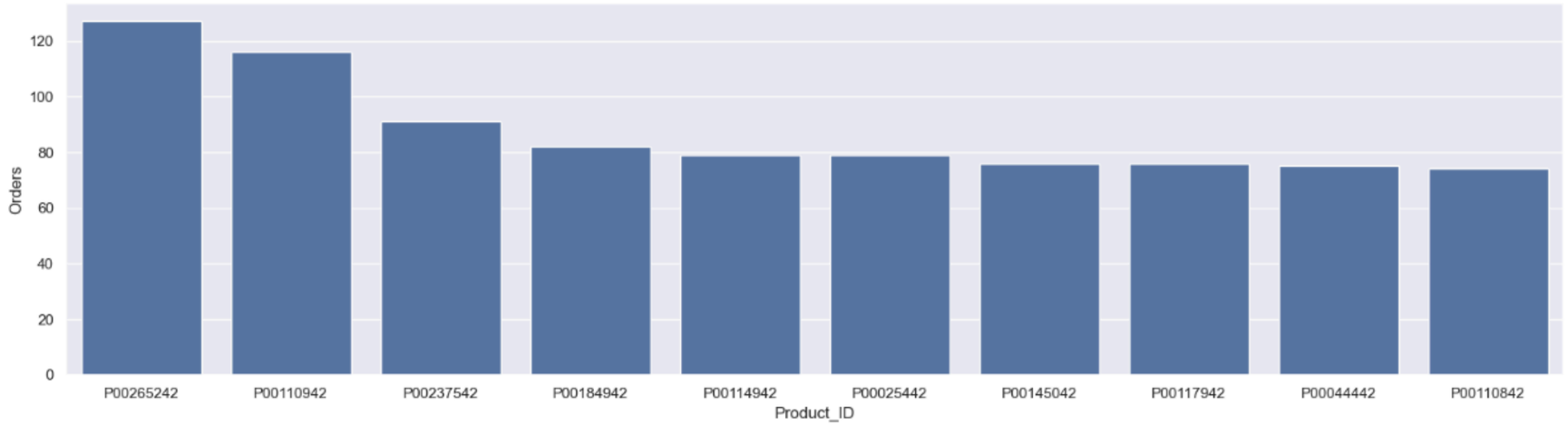
TOTAL REVENUE BASED ON PRODUCT_CATEGORY

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

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



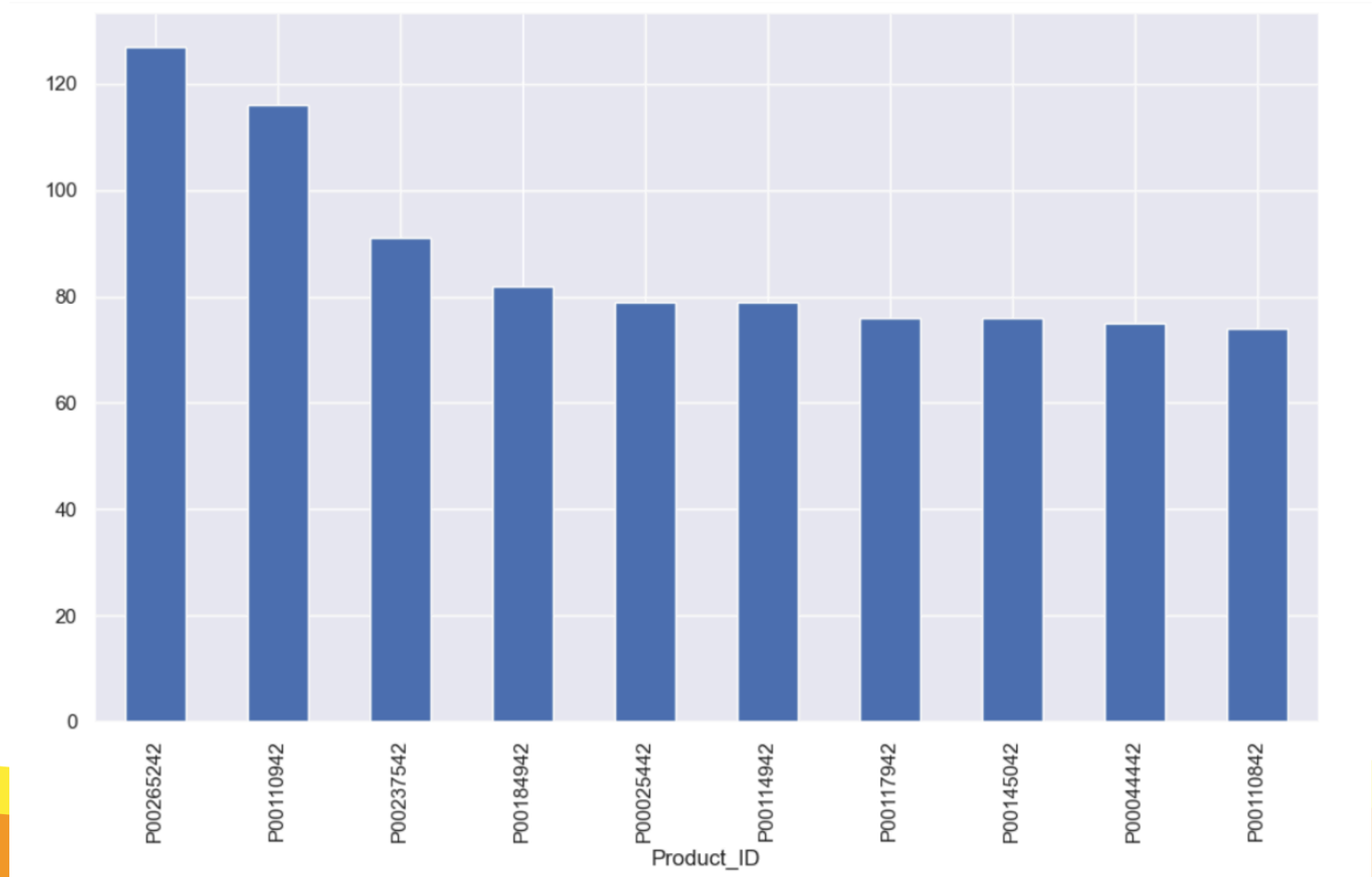
ORDER COUNT BASED ON PRODUCT ID

```
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')
```



TOP 10 MOST SOLD PRODUCT

```
fig1, ax1 = plt.subplots(figsize=(12,7))
df.groupby('Product_ID')['Orders'].sum().nlargest(10).sort_values(ascending=False).plot(kind='bar')
```



CONCLUSION

- Married women aged 26–35 from Uttar Pradesh, Maharashtra, and Karnataka, employed in the IT, Healthcare, and Aviation sectors, are more likely to purchase products in the Food, Clothing, and Electronics categories.

THANK YOU