

EXPLORATORY DATA ANALYSIS ON CUSTOMER SALES USING PYTHON

JUPYTER NOTEBOOK | PANDAS | MATPLOTLIB | SEABORN

SUBMITTED BY: QAZI HASAN AZIZ

JULY 2023

Project Brief

Our client company has been facing challenges in improving customer experiences, which has directly impacted their revenue over the past few months. To address this issue, we aim to conduct a data analytics project using Python, to analyze the vast amount of sales data accumulated from the last few months. The goal is to identify patterns, extract valuable insights, and provide actionable recommendations to enhance customer experiences and subsequently boost revenue.

Objectives

The primary objectives of this data analytics project are as follows:

- Analyze the sales data to identify trends and patterns related to customer behaviors.
- Identify potential pain points and areas of improvement in the customer journey.
- Correlate customer experiences with revenue fluctuations to understand the impact.
- Recommend data-driven strategies to enhance customer experiences and increase revenue.

Tools used in this project

1. Python Jupyter Notebook
2. Pandas
3. Matplotlib
4. Seaborn

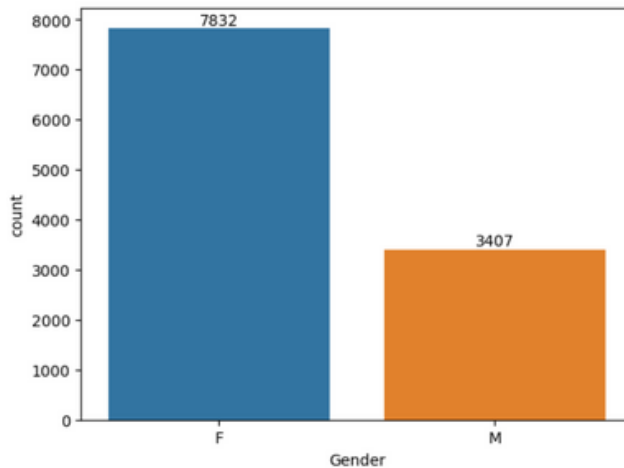
Results and Insights

No of purchases by Gender

```
[15]: # plotting a bar chart for Gender and it's count
```

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

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



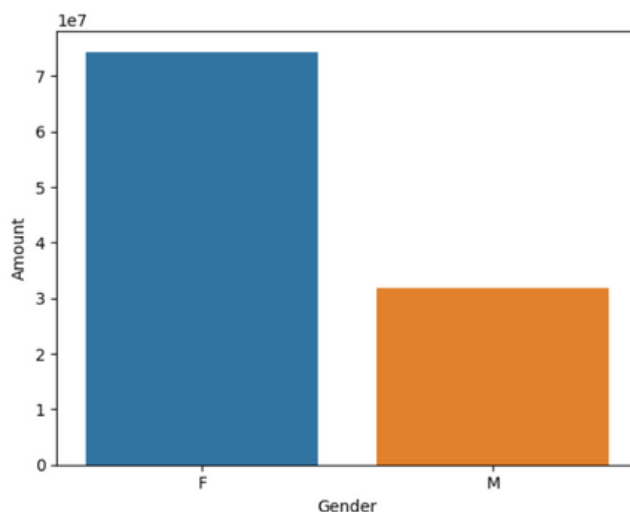
Purchase Power by Gender

```
# plotting a bar chart for gender vs total amount
```

```
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)
```

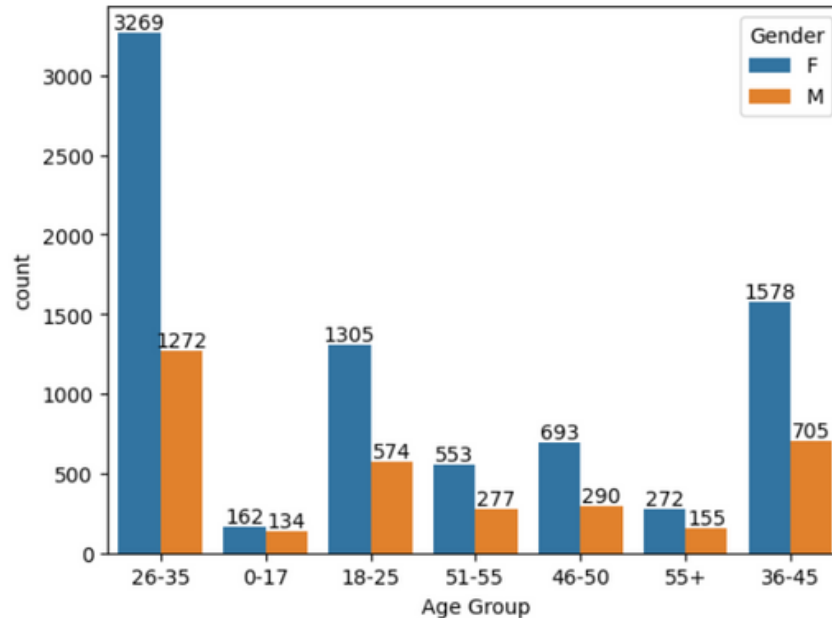
```
<Axes: xlabel='Gender', ylabel='Amount'>
```



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

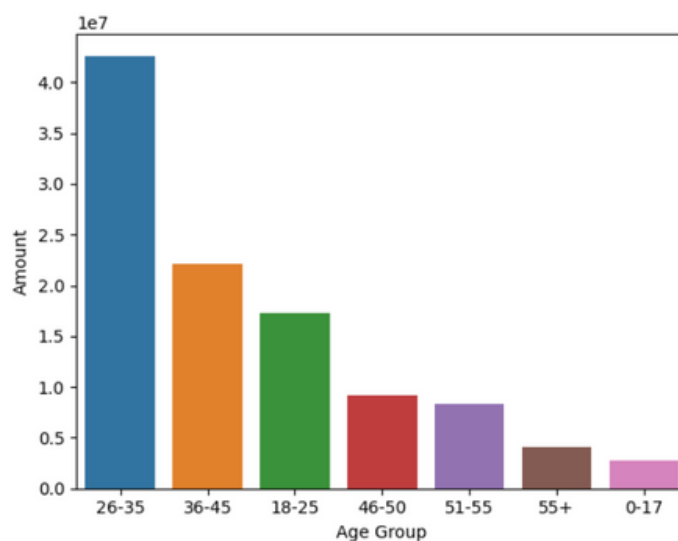
No of purchases by Age-group

```
ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender')  
  
for bars in ax.containers:  
    ax.bar_label(bars)
```



Purchase power by Age-Group

```
# Total Amount vs 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)  
  
<Axes: xlabel='Age Group', ylabel='Amount'>
```



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

No of purchases by State

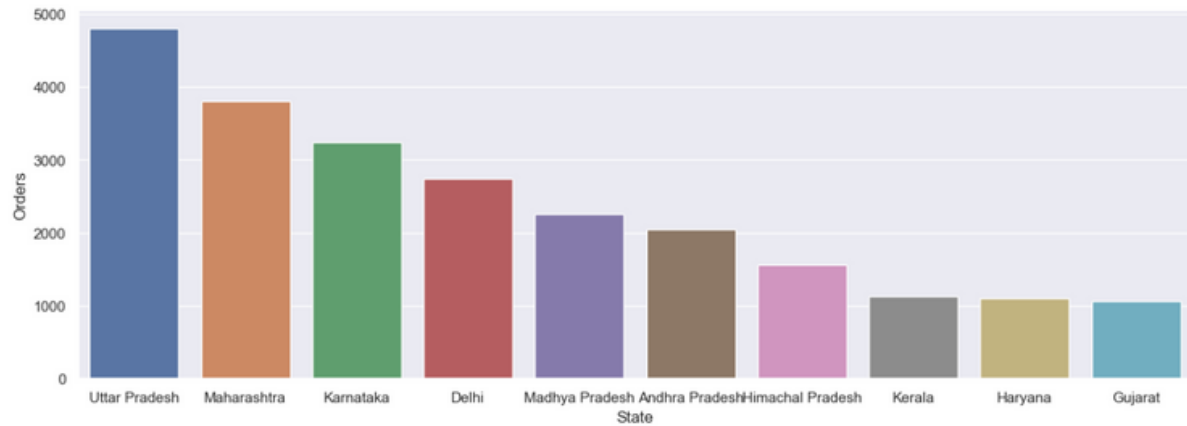
```
# total number of orders from top 10 states
```

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

```
sns.set(rc={'figure.figsize':(15,5)})
```

```
sns.barplot(data = sales_state, x = 'State', y = 'Orders')
```

<Axes: xlabel='State', ylabel='Orders'>



Purchase power by State

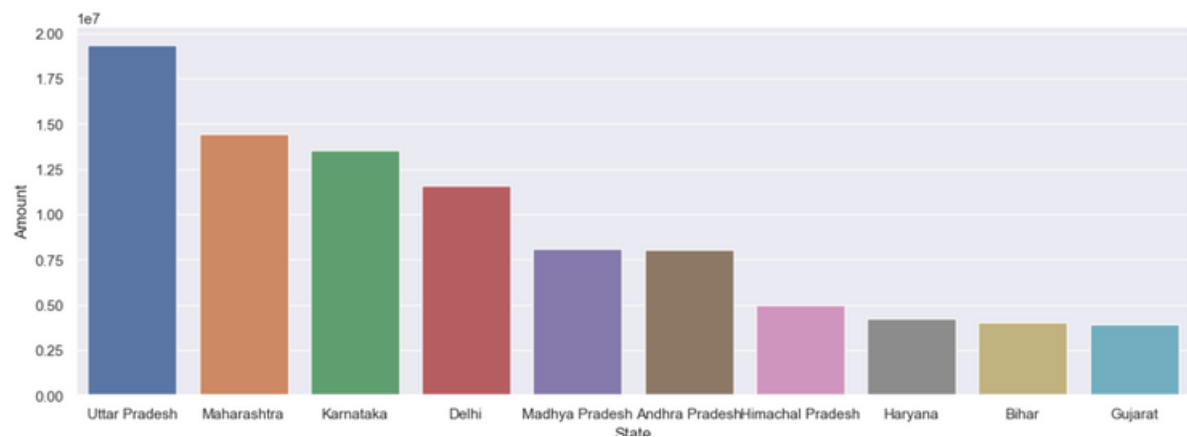
```
# total amount/sales from top 10 states
```

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

```
sns.set(rc={'figure.figsize':(15,5)})
```

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

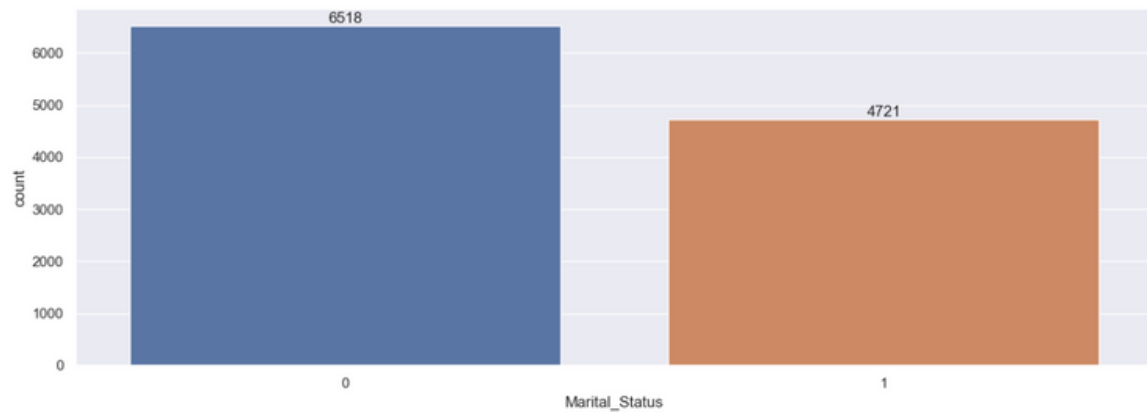
<Axes: xlabel='State', ylabel='Amount'>



We can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively

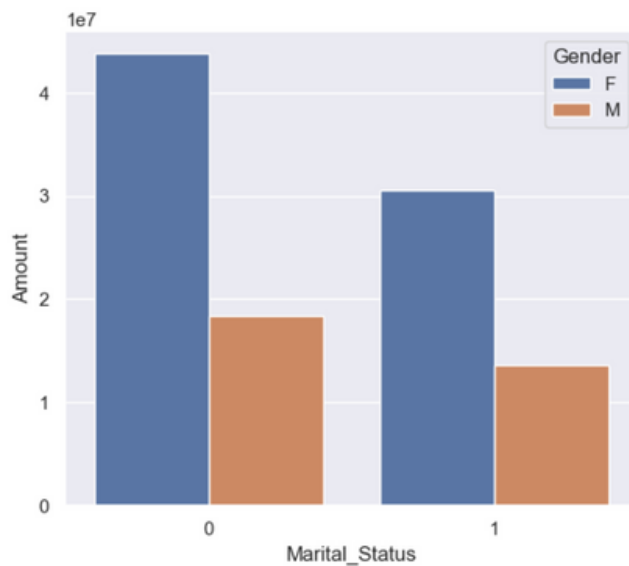
No of purchases by Marital Status

```
ax = sns.countplot(data = df, x = 'Marital_Status')  
  
sns.set(rc={'figure.figsize':(7,5)})  
for bars in ax.containers:  
    ax.bar_label(bars)
```



Purchase power by Marital Status

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

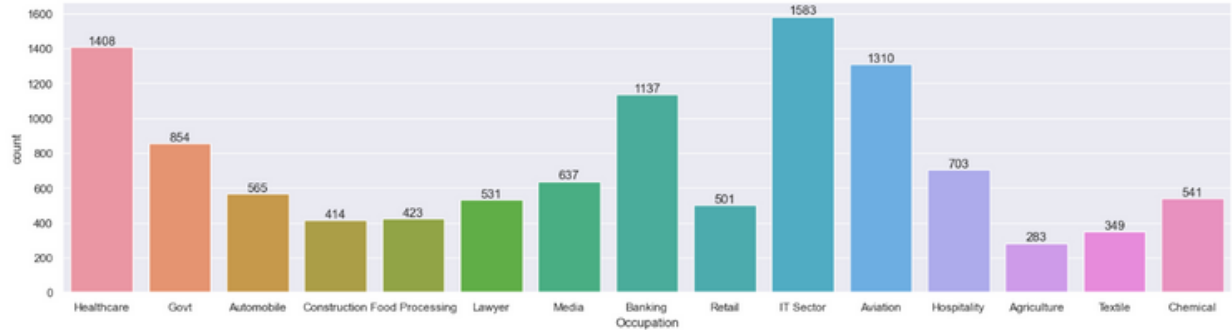


We can see that most of the buyers are married (women) and they have high purchasing power

No of purchases by Occupation

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

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

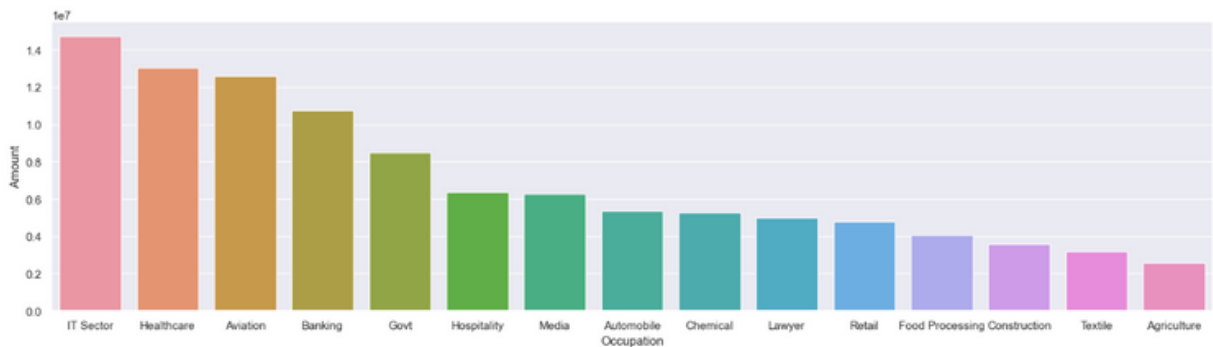


Purchase power by Occupation

```
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')
```

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

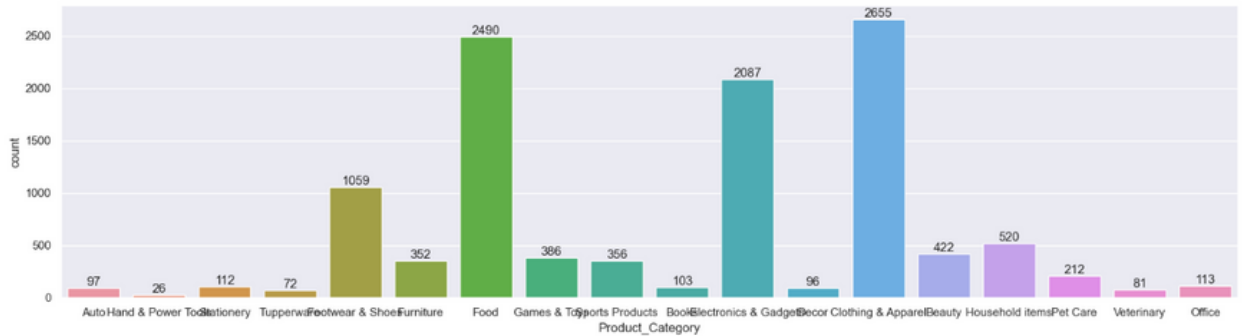


From above graphs we can see that most of the buyers are working in IT, Healthcare and Aviation sector

No of purchases by Product Catagory

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

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

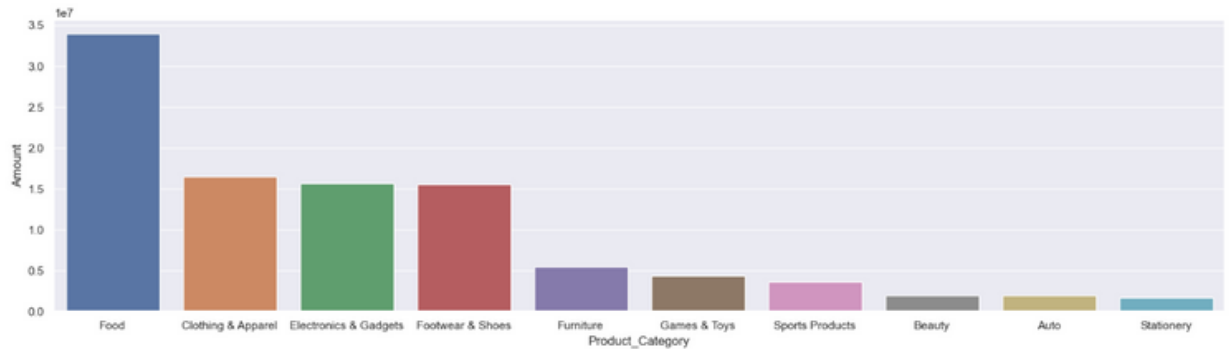


Purchase power by Product Catagory

```
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')
```

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



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

Final Results & Recommendations

- Married women age group 26-35 yrs from UP, Maharastra and Karnataka are more likely to buy our products.
- Most of our top customers are working in IT, Healthcare and Aviation
- Our most sold products are from Food, Clothing and Electronics category.

1. Based on our findings, we can recommend that, the company should focus on increasing product items on the categories like Food and Clothing mainly as most of our customers are female within 35 age.
2. Marketing Activities needs to be increased in the states Kerala, Bihar, Gujrat to increase visibly and revenue.