```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns


df = pd.read_csv('supermarket_sales.csv')


df.head()
```

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%
0	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415
1	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.8200
2	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155
3	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.2880
4	373-73-	A	Yangon	Normal	Male	Sports and	86.31	7	30.2085

```
df.shape
```

```
(1000, 17)
```

```
print("Number of rows : ",df.shape[0])
print("Number of columns : ",df.shape[1])
print("Columns : ",df.columns)
```

```
Number of rows :  1000
Number of columns :  17
Columns :  Index(['Invoice ID', 'Branch', 'City', 'Customer type', 'Gender',
       'Product line', 'Unit price', 'Quantity', 'Tax 5%', 'Total', 'Date',
       'Time', 'Payment', 'cogs', 'gross margin percentage', 'gross income',
       'Rating'],
      dtype='object')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Invoice ID              1000 non-null   object
 1   Branch                  1000 non-null   object
 2   City                    1000 non-null   object
```

```
 3   Customer type            1000 non-null   object
 4   Gender                   1000 non-null   object
 5   Product line             1000 non-null   object
 6   Unit price               1000 non-null   float64
 7   Quantity                 1000 non-null   int64
 8   Tax 5%                   1000 non-null   float64
 9   Total                    1000 non-null   float64
 10  Date                     1000 non-null   object
 11  Time                     1000 non-null   object
 12  Payment                  1000 non-null   object
 13  cogs                     1000 non-null   float64
 14  gross margin percentage  1000 non-null   float64
 15  gross income             1000 non-null   float64
 16  Rating                   1000 non-null   float64
dtypes: float64(7), int64(1), object(9)
memory usage: 132.9+ KB
```

```
df.describe()
```

	Unit price	Quantity	Tax 5%	Total	cogs	gross margin percentage	
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.00000	1.000000e+03	10
mean	55.672130	5.510000	15.379369	322.966749	307.58738	4.761905e+00	
std	26.494628	2.923431	11.708825	245.885335	234.17651	6.131498e-14	
min	10.080000	1.000000	0.508500	10.678500	10.17000	4.761905e+00	
25%	32.875000	3.000000	5.924875	124.422375	118.49750	4.761905e+00	
50%	55.230000	5.000000	12.088000	253.848000	241.76000	4.761905e+00	
75%	77.935000	8.000000	22.445250	471.350250	448.90500	4.761905e+00	

```
df.describe(include='object')
```

	Invoice ID	Branch	City	Customer type	Gender	Product line	Date	Time	Payment
count	1000	1000	1000	1000	1000	1000	1000	1000	1000
unique	1000	3	3	2	2	6	89	506	3
top	849-09-3807	A	Yangon	Member	Female	Fashion accessories	2/7/2019	19:48	Ewallet

Observation

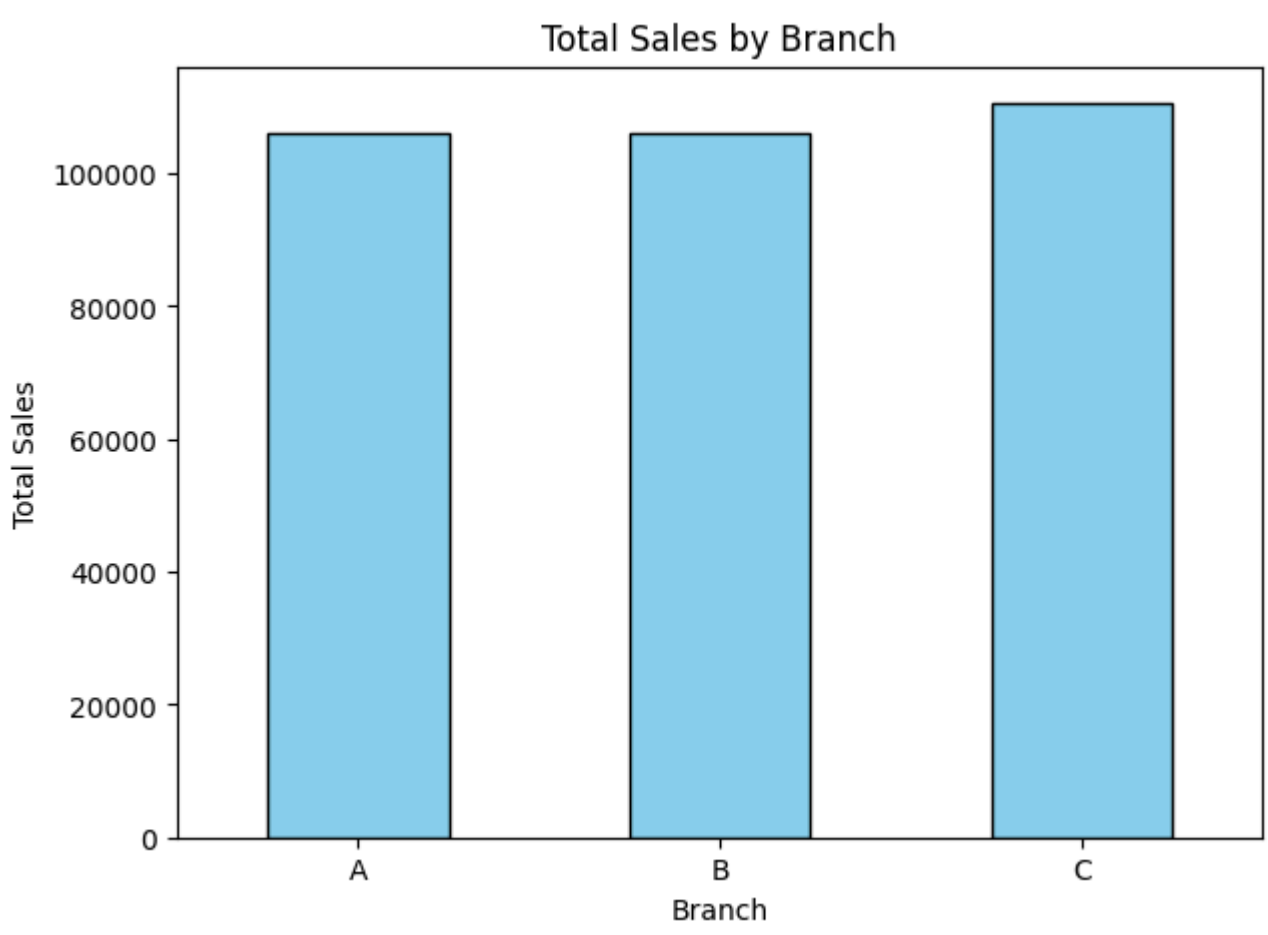
- The dataset contains 1000 rows and 17 columns.
- Dataset has no missing values.
- Data Types:

- Numeric Columns: Unit price, Quantity, Tax 5%, Total, COGS, Gross margin %, Gross income, Rating.
- Categorical Columns: Invoice ID, Branch, City, Customer Type, Gender, Product line, Date, Time, Payment.
- The average unit price is approximately $55.67.
- The average quantity purchased per transaction is about 5.51 items.
- The average customer rating is around 6.97 out of 10.

```
branch = df.groupby('Branch')['Total'].sum()
print(branch)
```

```
Branch
A    106200.3705
B    106197.6720
C    110568.7065
Name: Total, dtype: float64
```

```
plot = branch.plot(kind='bar', color='skyblue',edgecolor="black", figsize=(7,5))
plt.title('Total Sales by Branch')
plt.xlabel('Branch')
plt.ylabel('Total Sales')
plt.xticks(rotation=0)
plt.show()
```



- Branch C is performing Great in terms of revenue, with the highest total sales.

- Branch A is behind the Branch after which is followed by Branch B

```
purchase = df.groupby('Customer type')['Total'].mean()
purchase
```

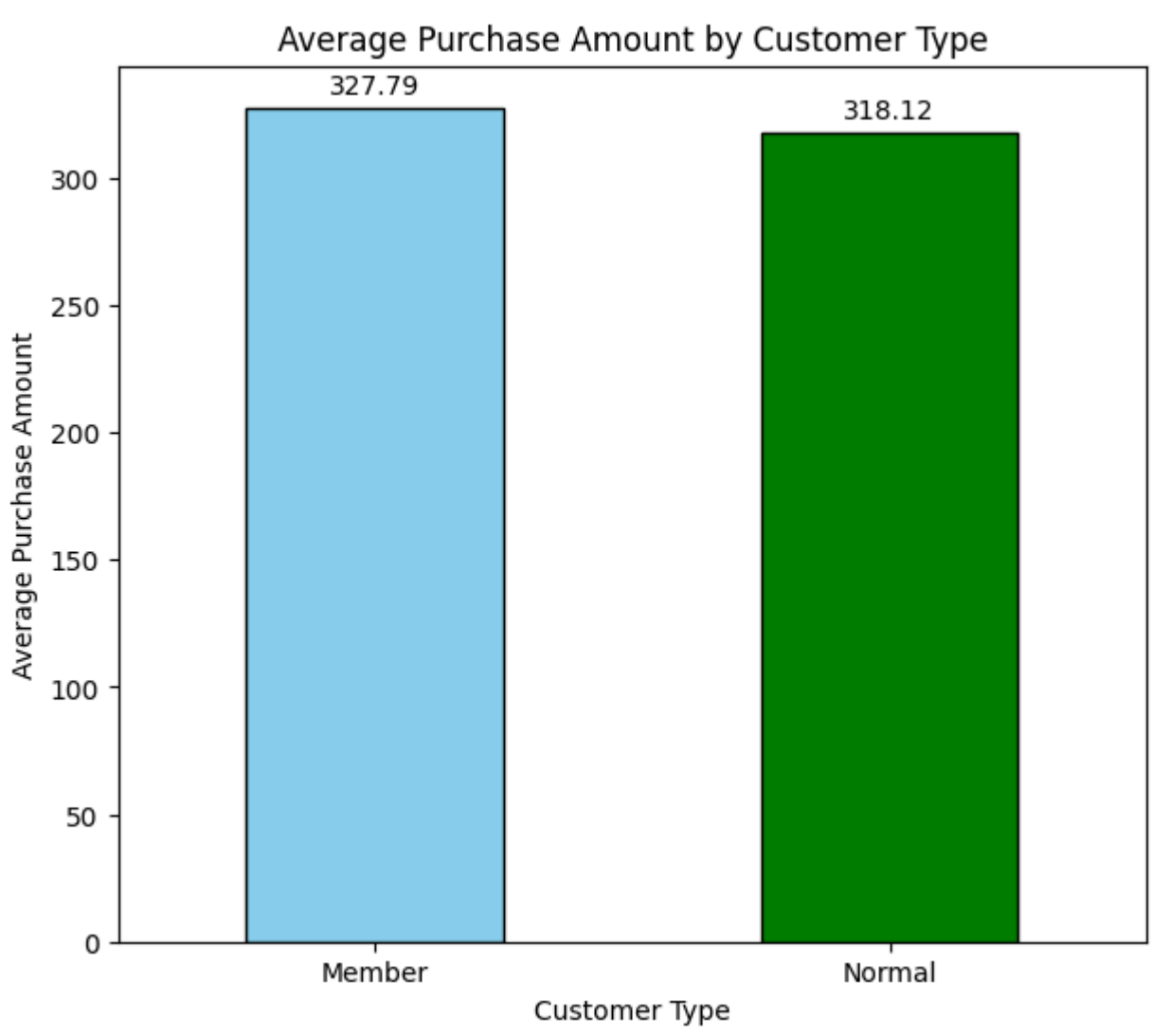
Customer type	Total
Member	327.791305
Normal	318.122856

dtype: float64

```
plot = purchase.plot(kind='bar', color=['Skyblue', 'green'],edgecolor="black", figsize=(7
plot.bar_label(plot.containers[0], fmt='%.2f', label_type='edge', padding=3)

plt.title('Average Purchase Amount by Customer Type')
plt.xlabel('Customer Type')
plt.ylabel('Average Purchase Amount')
plt.xticks(rotation=0)

plt.show()
```



- Members spend more per transaction than Normal customers.

```
product = df.groupby('Product line')['Total'].sum()
product
```

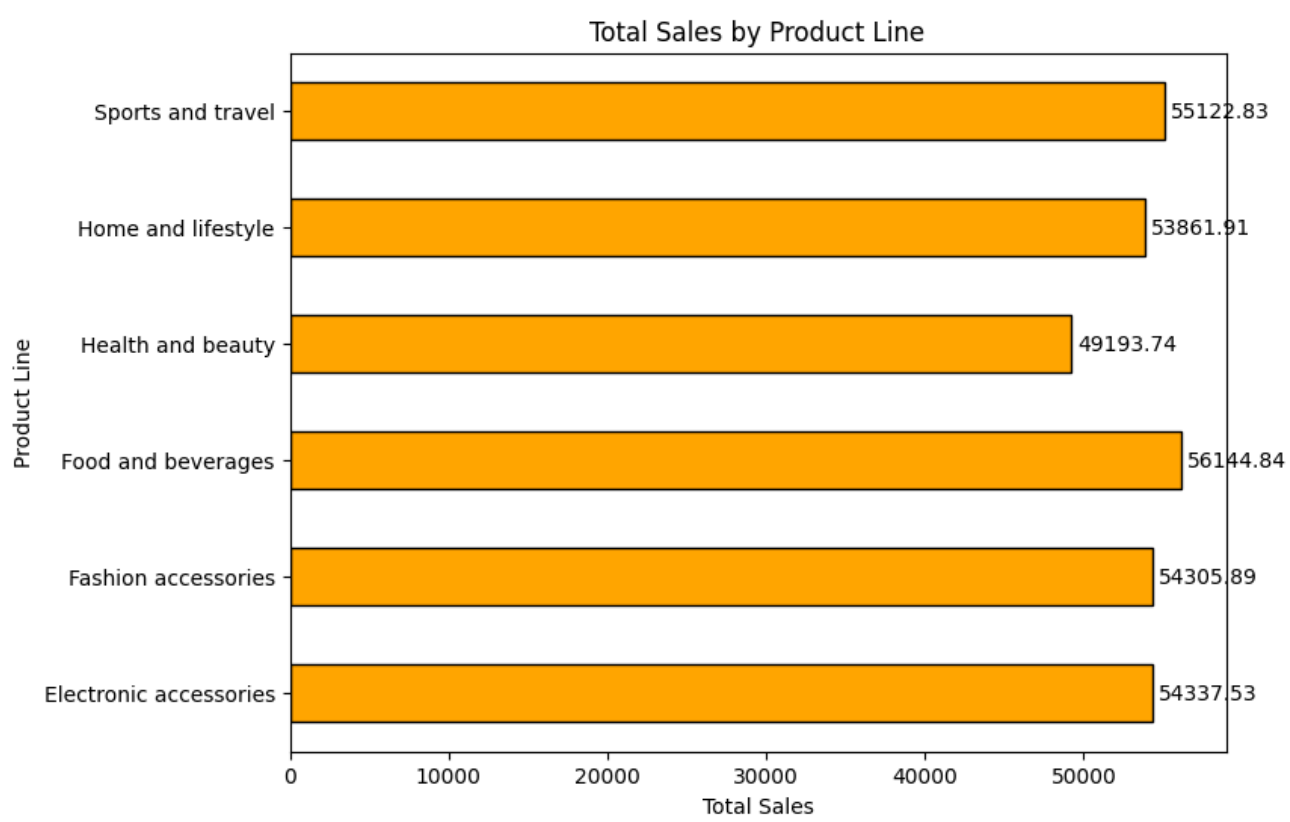
Product line	Total
Electronic accessories	54337.5315
Fashion accessories	54305.8950
Food and beverages	56144.8440
Health and beauty	49193.7390
Home and lifestyle	53861.9130
Sports and travel	55122.8265

dtype: float64

```
plot = product.plot(kind='barh', color='orange',edgecolor="black",orientation="horizontal
plot.bar_label(plot.containers[0], fmt='%.2f', padding=3)

plt.title('Total Sales by Product Line')
plt.xlabel('Total Sales')
plt.ylabel('Product Line')

plt.show()
```



- Food and beverages are the top product category in terms of revenue.
- This insight suggests that the company's current strategies in this product category are effective. By analysing and understanding what's driving success here—whether it's pricing, promotions, product quality, or customer preferences—the company can replicate similar approaches across underperforming product lines to enhance overall sales and profitability

```
plt.figure(figsize=(8, 6))

sns.histplot(data=df, x='Rating', bins=10, color='r', edgecolor="black")

plt.title('Distribution of Customer Ratings')
plt.xlabel('Rating Score')
plt.ylabel('Frequency')

plt.show()
```

Ratings are evenly distributed, it indicates varied customer experiences

Summary

Key Business Insights

- Branch Performance
  - **Branch B** outperformed the others in total sales, establishing itself as the **top revenue-generating location**.
  - This reflects strong **operational efficiency**, effective local marketing, or a strategically advantageous location.
- Customer Spending Behavior
  - **Loyalty program members** consistently made **higher-value purchases** than non-members.
  - This highlights the **positive impact of the loyalty program** on customer spending and engagement.

- Best-Selling Product Line
    - The **Food and Beverages** category led all product lines in revenue contribution.
    - **Fashion Accessories** and **Electronic Accessories** followed, indicating a strong customer preference for daily-use and lifestyle products.
- Customer Satisfaction Ratings
    - Customer ratings were **evenly distributed**:
        - **50%** fell in the **moderate range (4–7)**