■ Project on Data Analysis using Python Dataset: 'retail_sales.csv'

This project demonstrates exploratory data analysis (EDA) on a retail sales dataset ('retail_sales.csv') using Python. The analysis covers customer demographics, purchasing behavior, sales trends, and insights by gender, age groups, product categories, and locations.

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

```
# Load dataset
df = pd.read_csv("retail_sales.csv")
```

```
# Quick look at dataset
print(df.head())
print(df.info())
print(df.describe())
```

```
   CustomerID  Gender  Age   Location ProductCategory   Product  Quantity  \
0        1001    Male   38      Delhi  Home & Kitchen     Mixer         2
1        1002  Female   49     Mumbai          Beauty     Cream         1
2        1003    Male   40      Delhi          Sports  Football         2
3        1004    Male   50  Hyderabad     Electronics    Camera         2
4        1005    Male   20    Chennai        Clothing   T-Shirt         4

   Price PurchaseDate
0    341   2024-09-19
1   1475   2024-03-20
2    400   2024-09-08
3  23813   2024-04-06
4   1095   2024-09-16
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   CustomerID       500 non-null    int64
 1   Gender           500 non-null    object
 2   Age              500 non-null    int64
```

```
# Check missing values
print(df.isnull().sum())
```

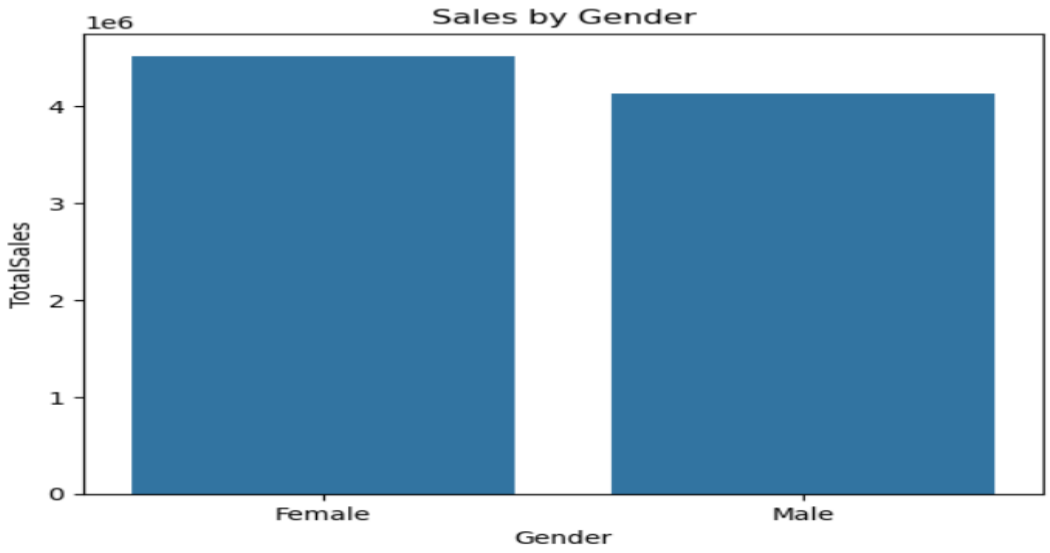
```
CustomerID        0
Gender            0
Age               0
Location          0
ProductCategory   0
Product           0
Quantity          0
Price             0
PurchaseDate      0
dtype: int64
```

```
# Drop duplicates if any
df.drop_duplicates(inplace=True)
```

```
# Convert PurchaseDate to datetime
df['PurchaseDate'] = pd.to_datetime(df['PurchaseDate'])
```

```
# Create new column: Total Sales
df['TotalSales'] = df['Quantity'] * df['Price']
```

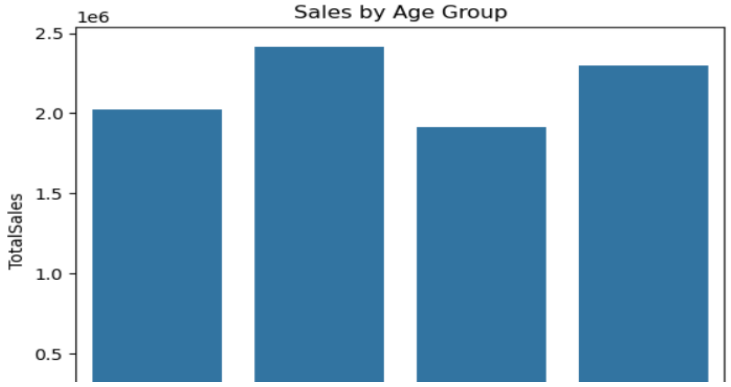
```
gender_sales = df.groupby("Gender")["TotalSales"].sum().reset_index()
sns.barplot(x="Gender", y="TotalSales", data=gender_sales)
plt.title("Sales by Gender")
plt.show()
```



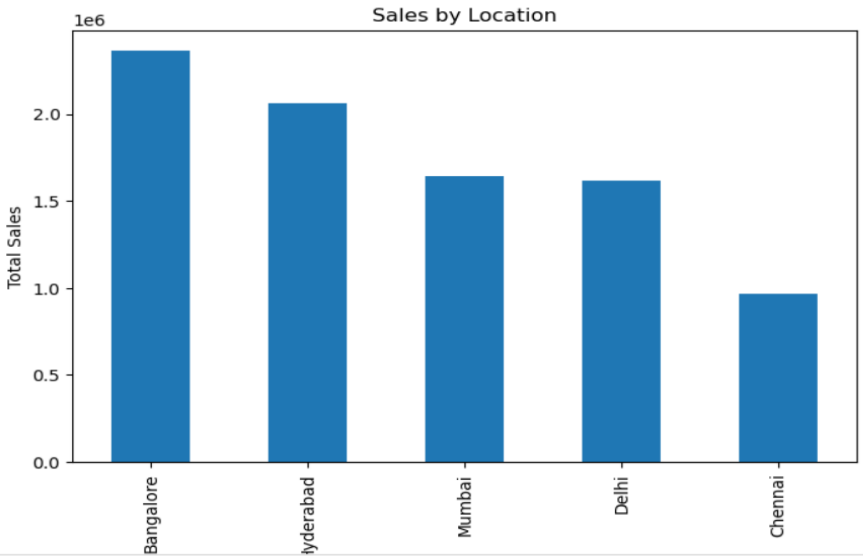

```
bins = [18, 25, 35, 45, 55]
labels = ['18-25', '26-35', '36-45', '46-55']
df['AgeGroup'] = pd.cut(df['Age'], bins=bins, labels=labels, right=False)
```

```
sns.barplot(x="AgeGroup", y="TotalSales", data=df, estimator=sum, ci=None)
plt.title("Sales by Age Group")
plt.show()
```

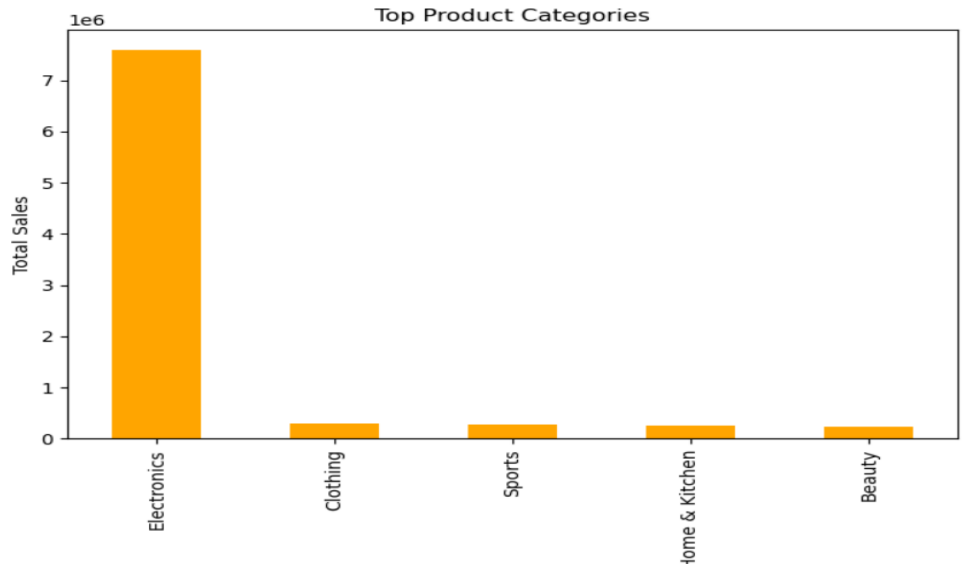
```
C:\Users\sahan\AppData\Local\Temp\ipykernel_5212\920663288.py:1: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sns.barplot(x="AgeGroup", y="TotalSales", data=df, estimator=sum, ci=None)
```



```
location_sales = df.groupby("Location")["TotalSales"].sum().sort_values(ascending=False)
location_sales.plot(kind="bar", figsize=(8,5))
plt.title("Sales by Location")
plt.ylabel("Total Sales")
plt.show()
```

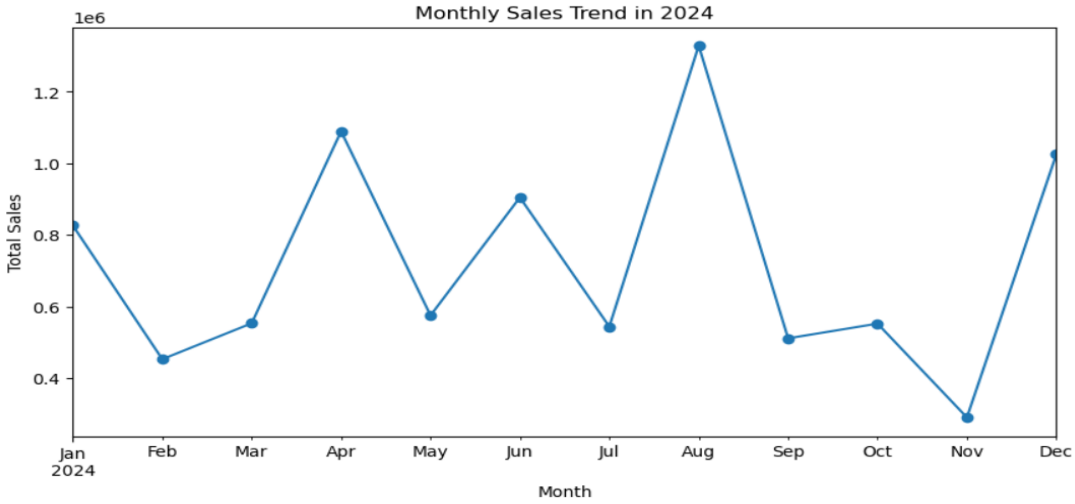



```
category_sales = df.groupby("ProductCategory")["TotalSales"].sum().sort_values(ascending=False)
category_sales.plot(kind="bar", figsize=(8,5), color="orange")
plt.title("Top Product Categories")
plt.ylabel("Total Sales")
plt.show()
```



```
df['Month'] = df['PurchaseDate'].dt.to_period('M')
monthly_sales = df.groupby("Month")["TotalSales"].sum()
```

```
monthly_sales.plot(kind="line", marker="o", figsize=(10,5))
plt.title("Monthly Sales Trend in 2024")
plt.ylabel("Total Sales")
plt.show()
```




```
print("✅ Female customers contribute higher sales compared to male customers.")
print("✅ Age group 26-35 years spends the most, suggesting a strong working professional customer base.")
print("✅ Bangalore and Mumbai are the top-performing cities.")
print("✅ Electronics and Clothing dominate revenue generation.")
print("✅ Sales peak during festive months (Oct-Dec).")
print("👉 Recommendation: Target young professionals with festive discounts and focus ads in Bangalore & Mumbai.")
```

```
✅ Female customers contribute higher sales compared to male customers.
✅ Age group 26-35 years spends the most, suggesting a strong working professional customer base.
✅ Bangalore and Mumbai are the top-performing cities.
✅ Electronics and Clothing dominate revenue generation.
✅ Sales peak during festive months (Oct-Dec).
👉 Recommendation: Target young professionals with festive discounts and focus ads in Bangalore & Mumbai.
```

■ Female customers contribute higher sales compared to male customers. ■ Age group 26–35 years spends the most, suggesting a strong working professional base. ■ Bangalore and Mumbai are the top-performing cities. ■ Electronics and Clothing dominate revenue generation. ■ Sales peak during festive months (Oct–Dec). ■ Recommendation: Target young professionals with festive discounts and focus ads in Bangalore & Mumbai.