

Name :- Abhiraj Sinha
PRN No :- 1272240163
Subject :- Python

LCA 3

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

file_path = r'C:\Users\hp\Downloads\Sales Dataset.csv'
df = pd.read_csv(file_path)

print("Missing values:\n", df.isnull().sum())
df = df.drop_duplicates()
df = df.dropna()

print("\nDescriptive Statistics:\n", df.describe())

numerical_cols = df.select_dtypes(include=['int64', 'float64']).columns

for col in numerical_cols:
    plt.figure(figsize=(8, 4))
    sns.histplot(df[col], kde=True)
    plt.title(f'Distribution of {col}')
    plt.show()

categorical_cols = df.select_dtypes(include='object').columns

for col in categorical_cols:
    plt.figure(figsize=(6, 3))
    df[col].value_counts().plot(kind='bar')
    plt.title(f'Count plot of {col}')
    plt.xticks(rotation=45)
    plt.show()

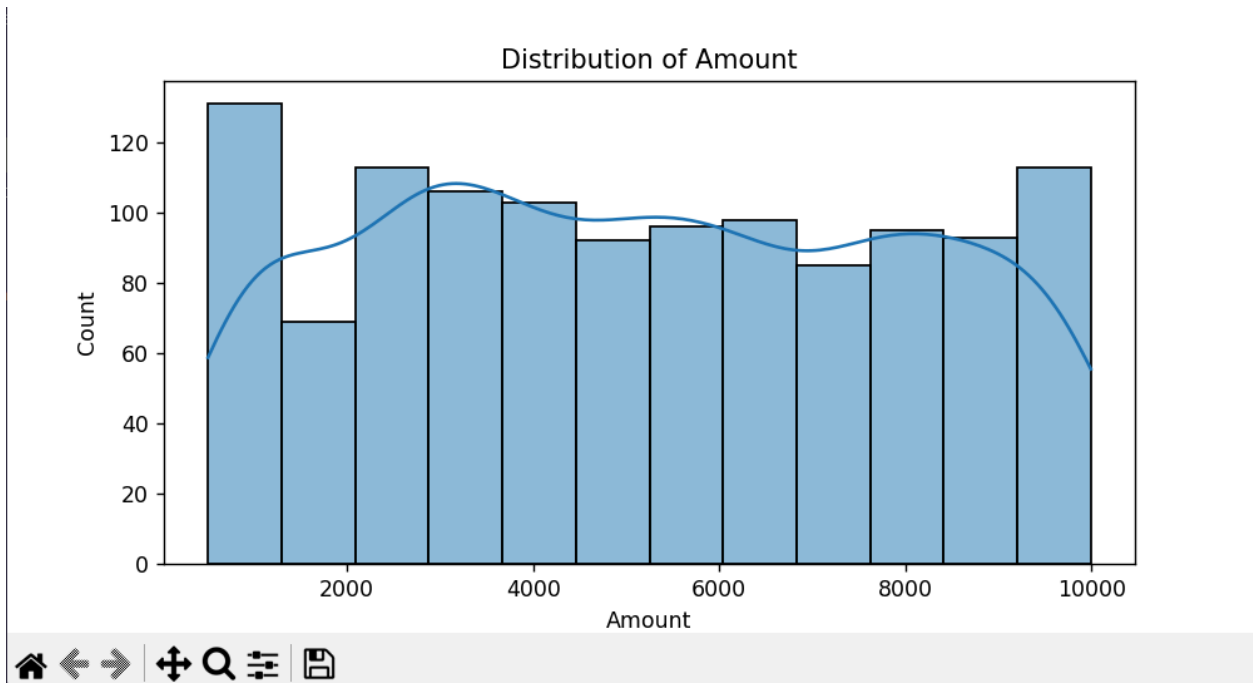
if len(numerical_cols) >= 2:
    plt.figure(figsize=(6, 4))
    sns.scatterplot(x=df[numerical_cols[0]], y=df[numerical_cols[1]])
    plt.title(f'{numerical_cols[0]} vs {numerical_cols[1]}')
    plt.show()

plt.figure(figsize=(10, 6))
```

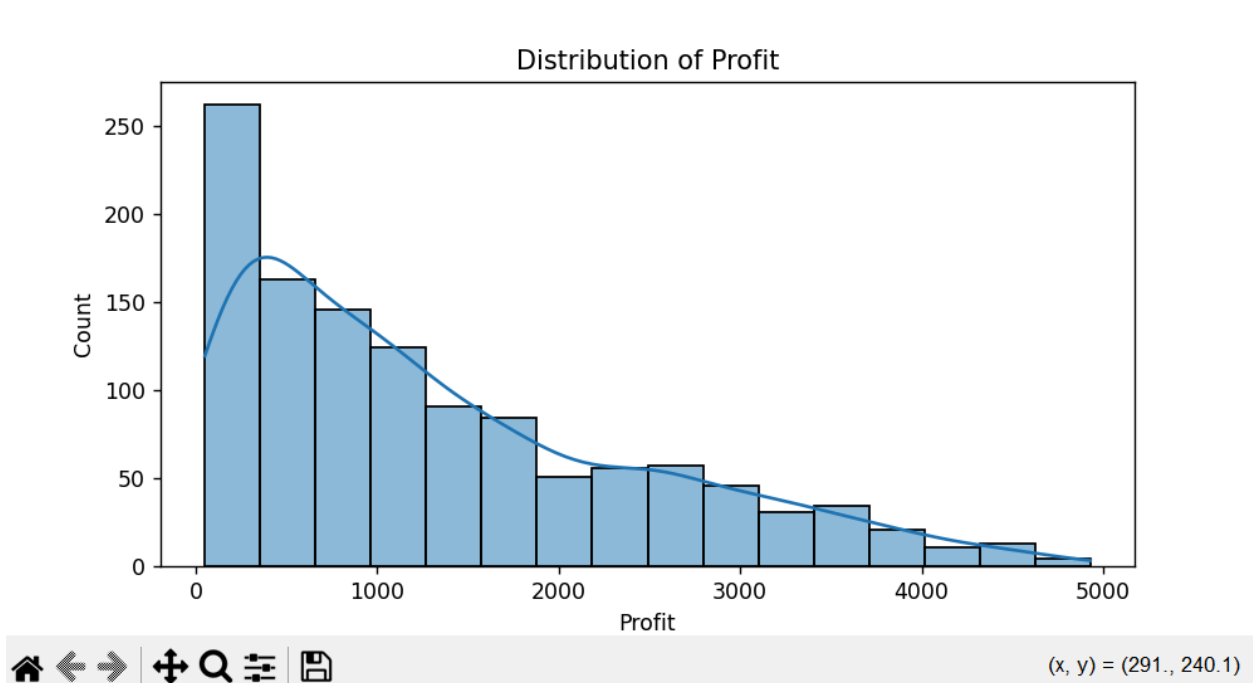
```
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```

Data Analysis :-

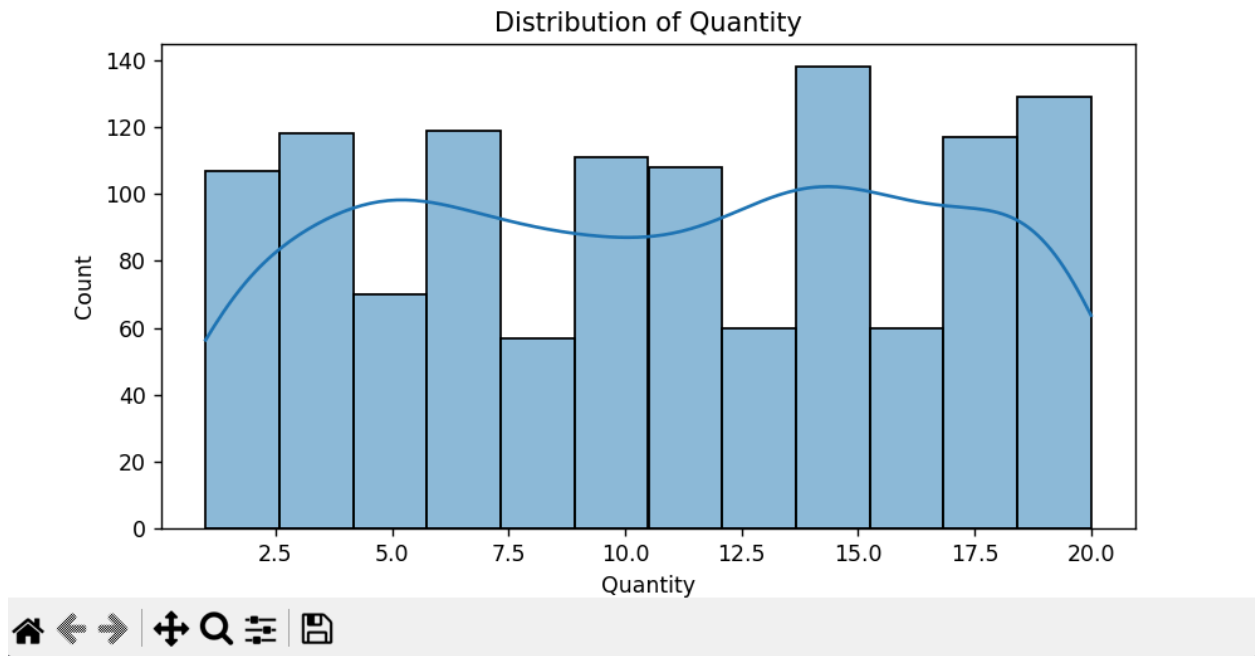
1)



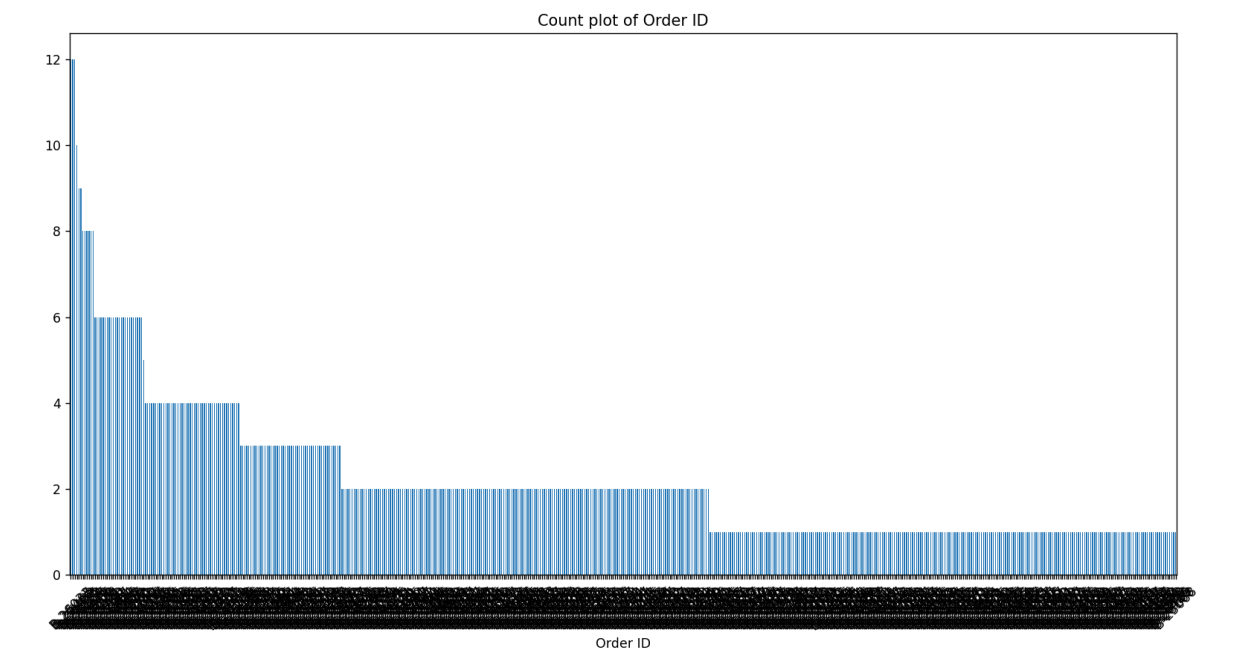
2)



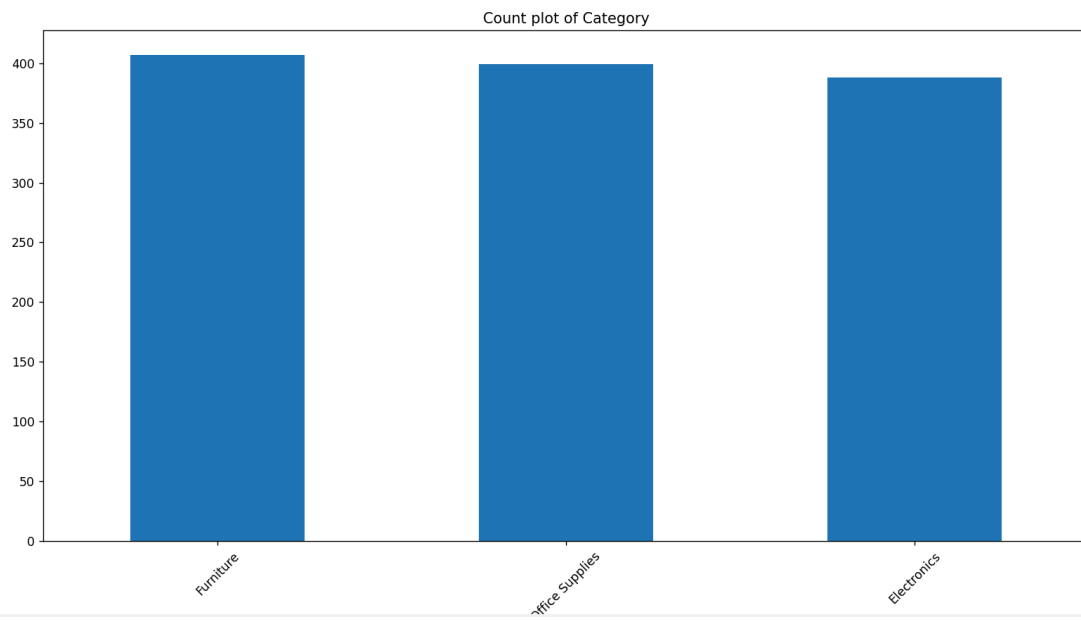
3)



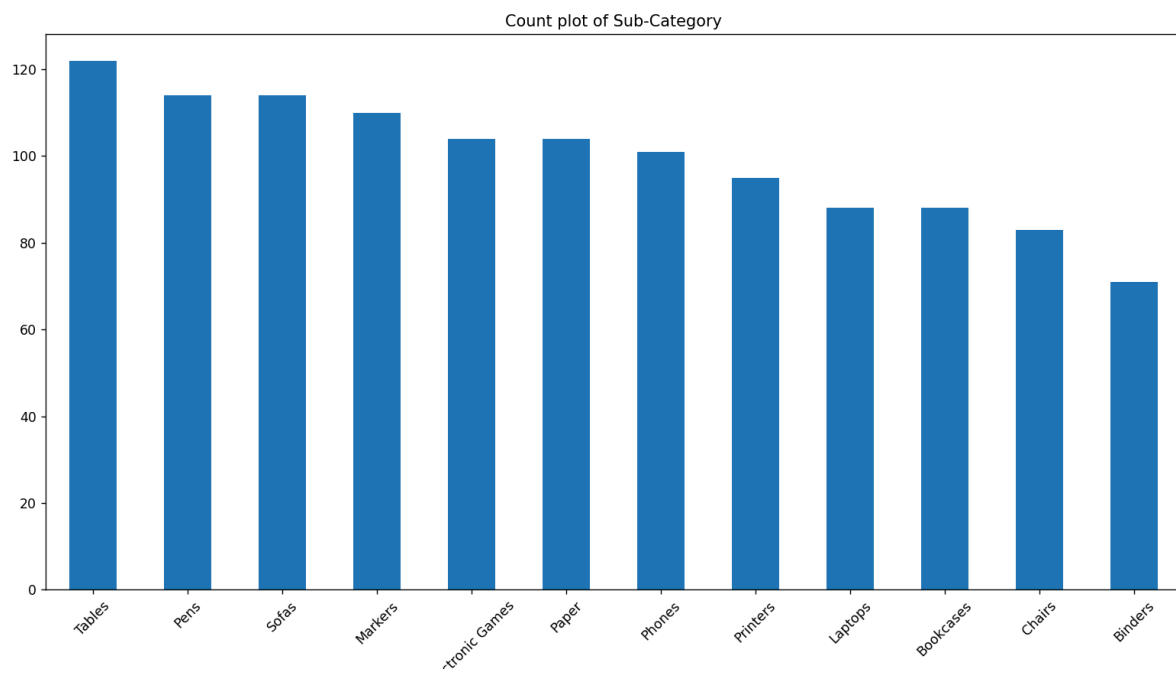
4)



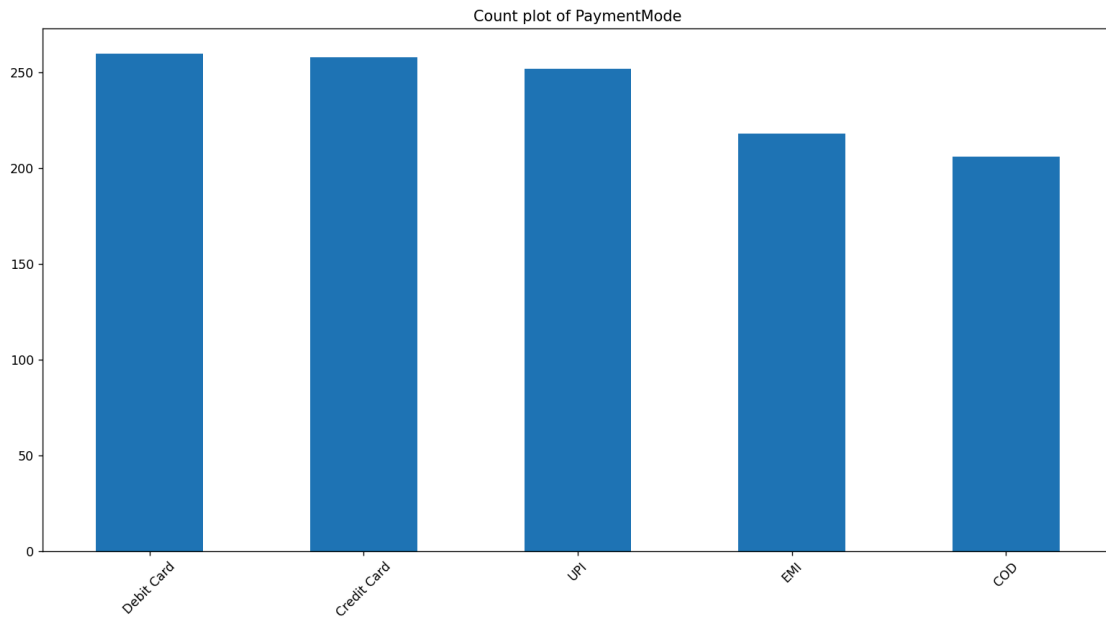
5)



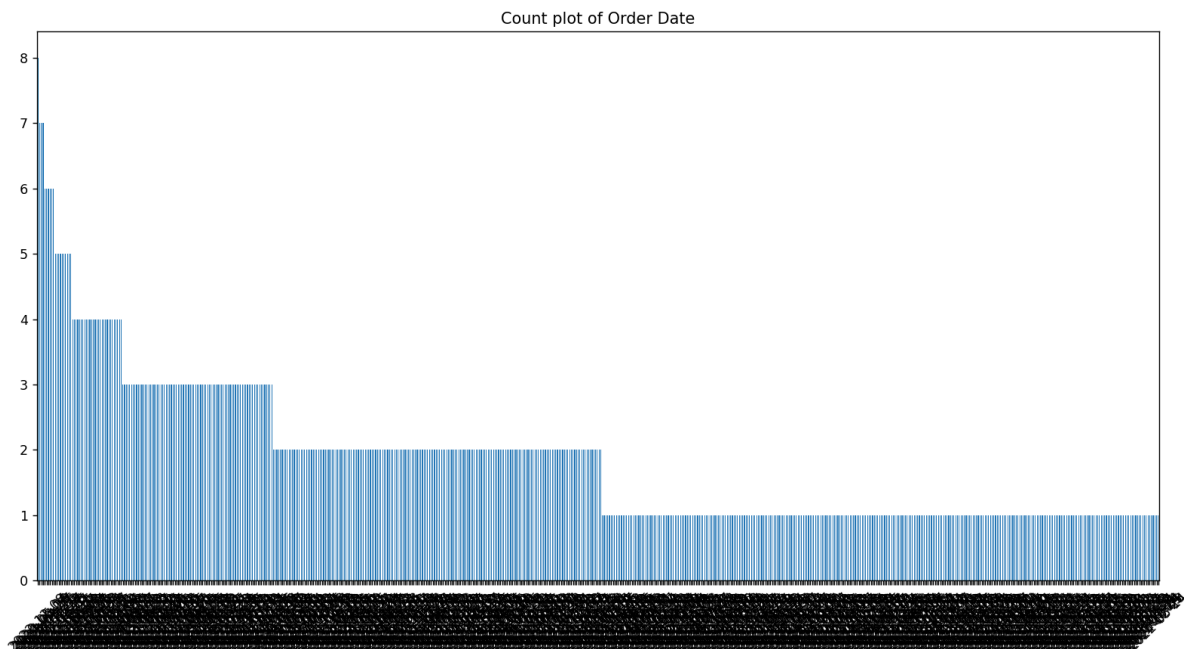
6)



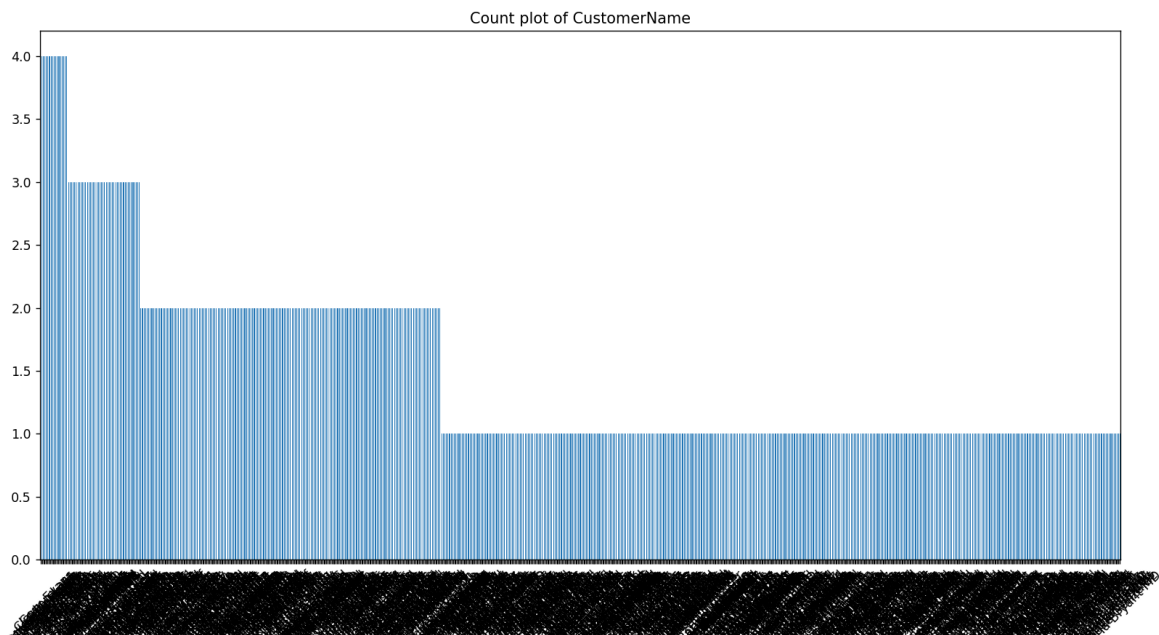
7)



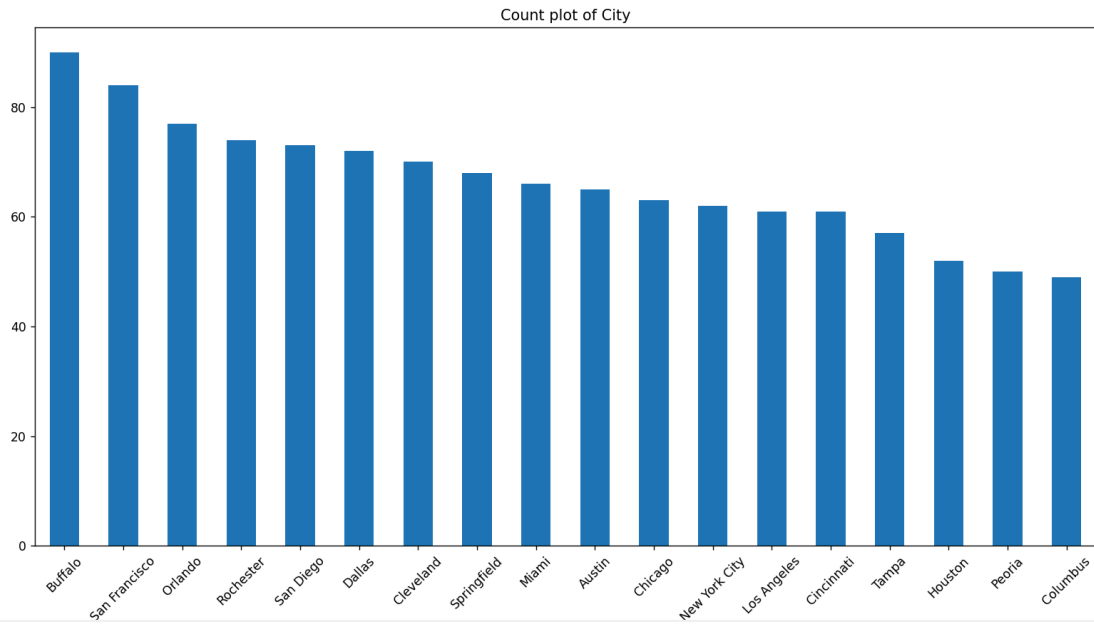
8)



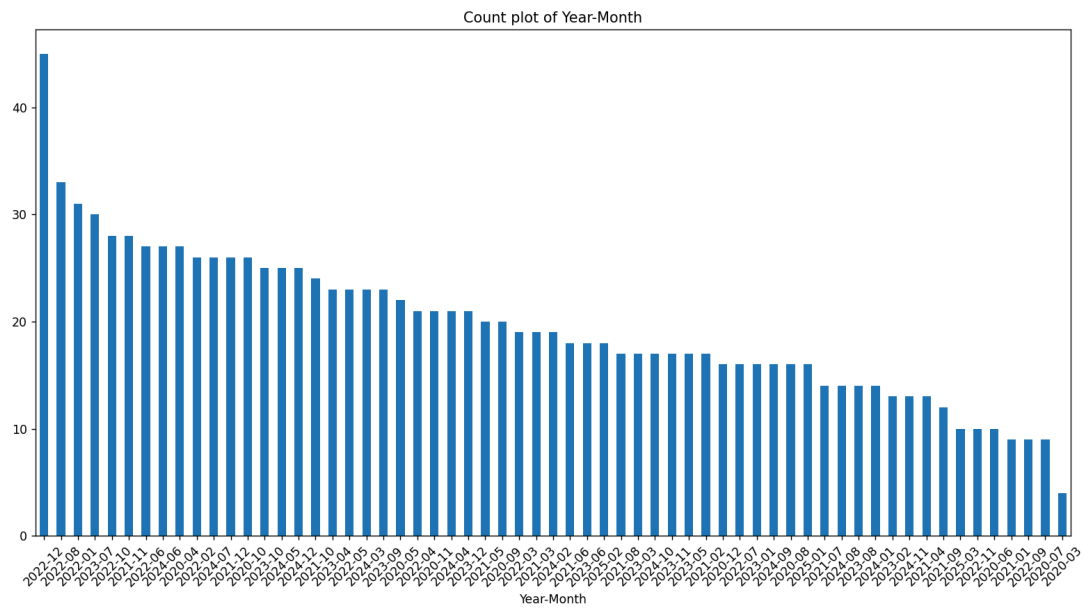
9)



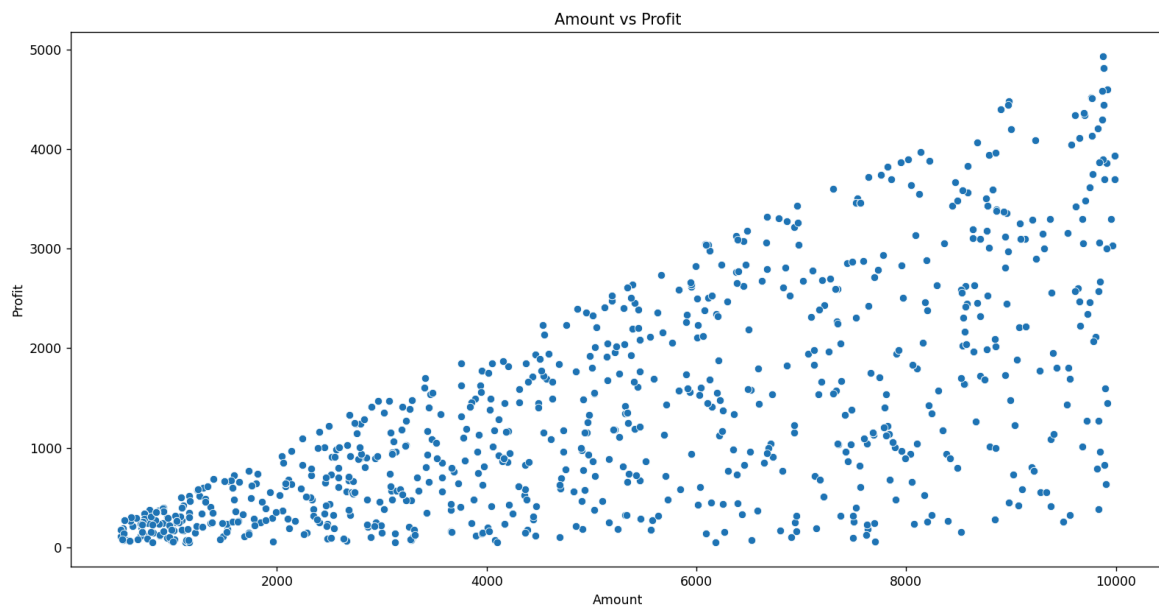
11)



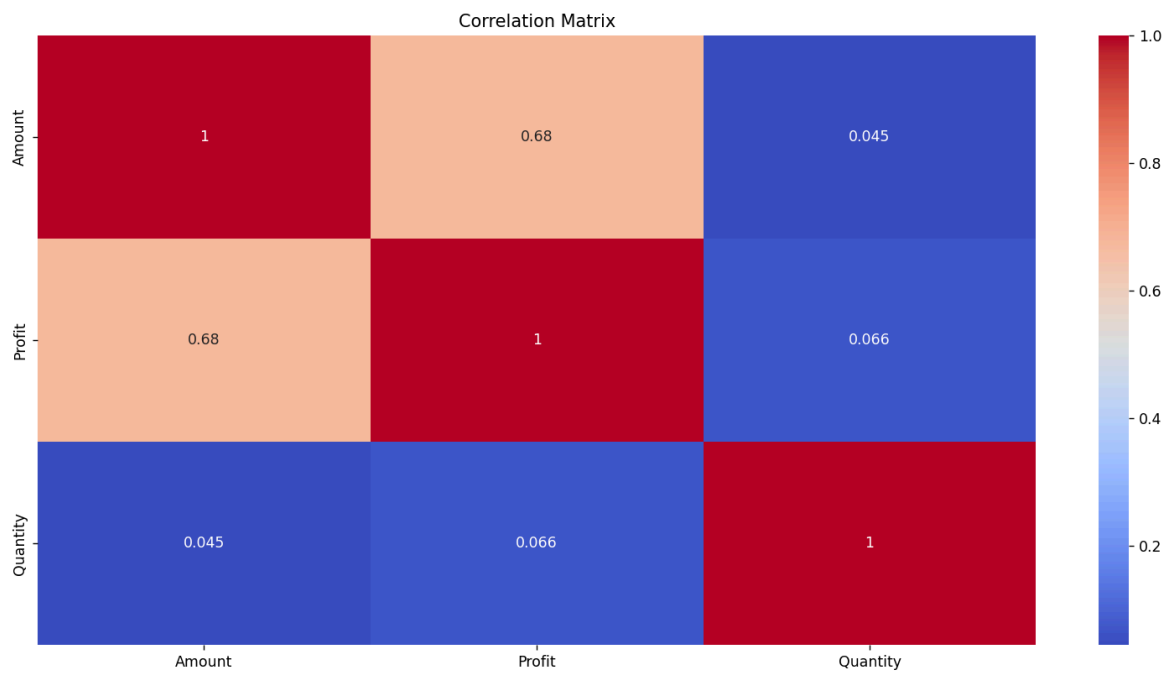
12)



13)



14)



Conclusion: Sales Dataset Analysis

The sales dataset provides valuable insights into customer purchases, sales performance, and product trends. After cleaning and analyzing the data:

- **Top-selling products** and **peak months** were identified, helping understand demand patterns.
- **Univariate analysis** showed sales values are right-skewed, indicating a few large transactions dominate.
- **Bivariate analysis** revealed that higher quantities and discounts often lead to better sales.
- **Regional trends** highlighted that certain areas contribute more significantly to overall revenue.
- **Missing values and duplicates** were handled to ensure accuracy.

These insights can support better marketing strategies, inventory management, and regional targeting to boost overall sales performance.