Import Libraries and Load Data

```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Set plot style
sns.set(style="whitegrid")

# Load the dataset
df = pd.read_csv('supermarket_sales.csv') # Make sure the CSV file
name matches

# Display the first few rows of the dataframe
print(df.head())

    Invoice ID Branch       City Customer type  Gender  \
0  750-67-8428      A     Yangon        Member  Female
1  226-31-3081      C  Naypyitaw        Normal  Female
2  631-41-3108      A     Yangon        Normal    Male
3  123-19-1176      A     Yangon        Member    Male
4  373-73-7910      A     Yangon        Normal    Male

             Product line  Unit price  Quantity   Tax 5%     Total
Date  \
0       Health and beauty       74.69         7  26.1415  548.9715
1/5/2019
1  Electronic accessories       15.28         5   3.8200   80.2200
3/8/2019
2      Home and lifestyle       46.33         7  16.2155  340.5255
3/3/2019
3       Health and beauty       58.22         8  23.2880  489.0480
1/27/2019
4       Sports and travel       86.31         7  30.2085  634.3785
2/8/2019

    Time      Payment    cogs  gross margin percentage  gross income
Rating
0  13:08      Ewallet  522.83                 4.761905       26.1415
9.1
1  10:29         Cash   76.40                 4.761905        3.8200
9.6
2  13:23  Credit card  324.31                 4.761905       16.2155
7.4
3  20:33      Ewallet  465.76                 4.761905       23.2880
8.4
```

```
4  10:37      Ewallet  604.17                 4.761905       30.2085
5.3
```

Initial Data Inspection

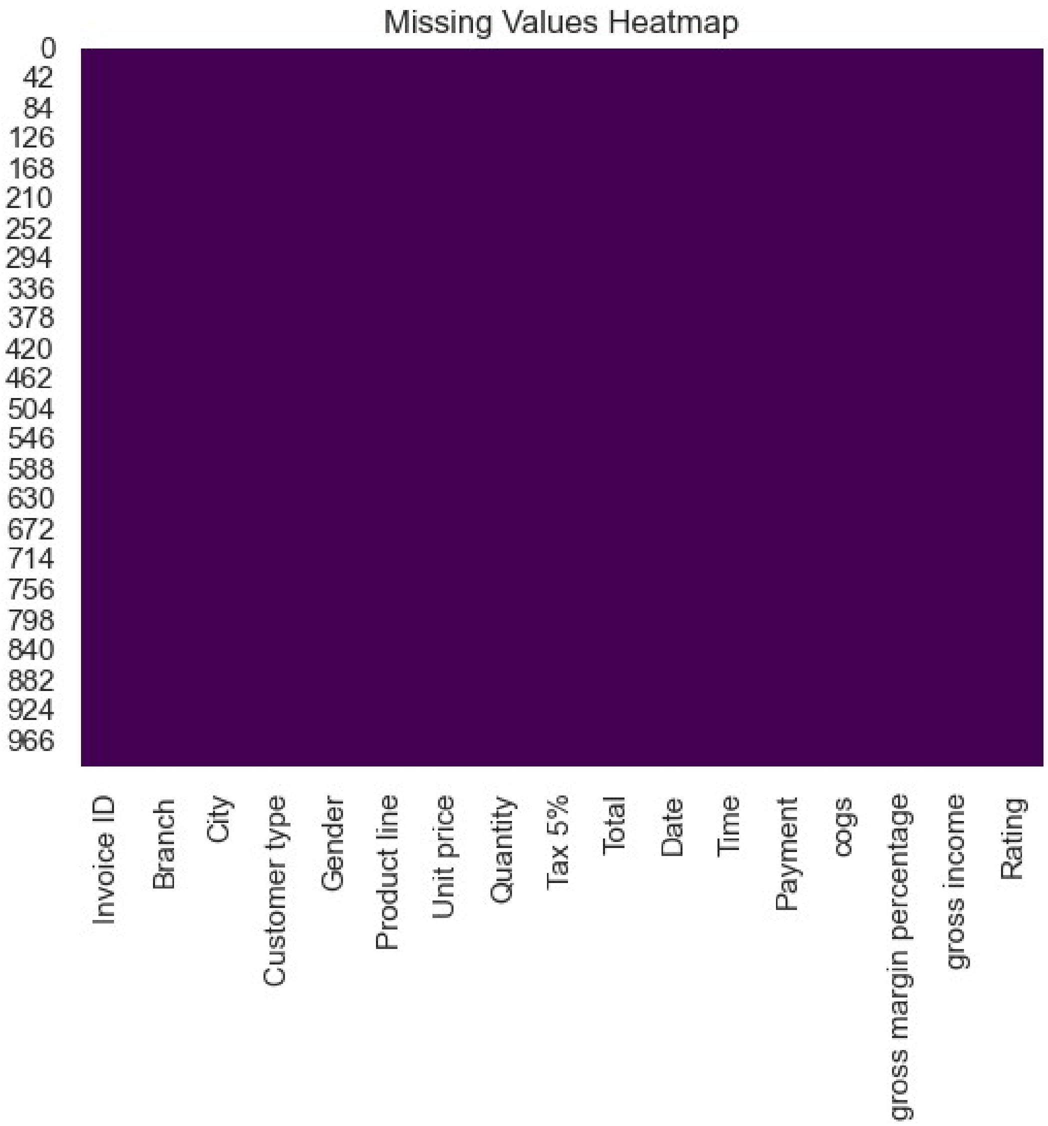
```
# Get a concise summary of the dataframe
print(df.info())

# Get descriptive statistics for numerical columns
print(df.describe())

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

```
<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
None
        Unit price     Quantity       Tax 5%        Total        cogs
\
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.00000

mean     55.672130     5.510000    15.379369   322.966749   307.58738

std      26.494628     2.923431    11.708825   245.885335   234.17651

min      10.080000     1.000000     0.508500    10.678500    10.17000
```

```
25%      32.875000     3.000000     5.924875   124.422375   118.49750

50%      55.230000     5.000000    12.088000   253.848000   241.76000

75%      77.935000     8.000000    22.445250   471.350250   448.90500

max      99.960000    10.000000    49.650000  1042.650000   993.00000


       gross margin percentage  gross income       Rating
count             1.000000e+03   1000.000000  1000.00000
mean              4.761905e+00     15.379369     6.97270
std               6.131498e-14     11.708825     1.71858
min               4.761905e+00      0.508500     4.00000
25%               4.761905e+00      5.924875     5.50000
50%               4.761905e+00     12.088000     7.00000
75%               4.761905e+00     22.445250     8.50000
max               4.761905e+00     49.650000    10.00000
Invoice ID                 0
Branch                     0
City                       0
Customer type              0
Gender                     0
Product line               0
Unit price                 0
Quantity                   0
Tax 5%                     0
Total                      0
Date                       0
Time                       0
Payment                    0
cogs                       0
gross margin percentage    0
gross income               0
Rating                     0
dtype: int64
```

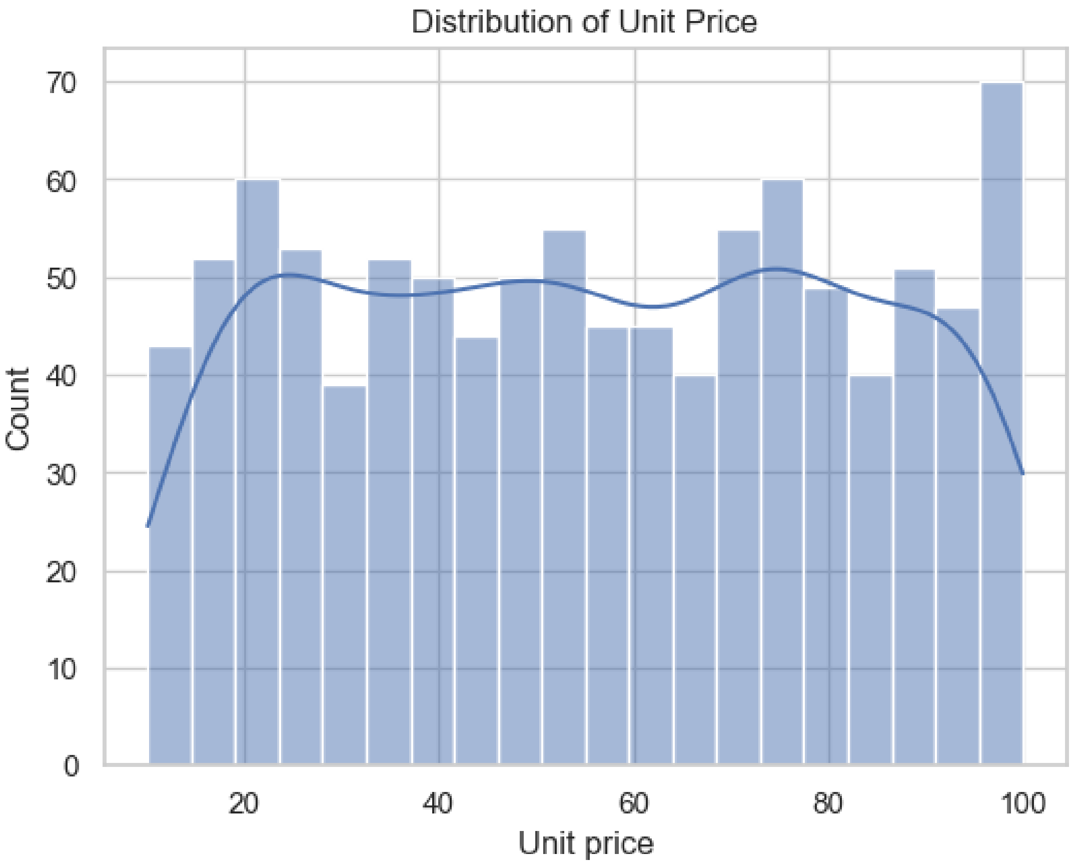
Visualize Missing Data

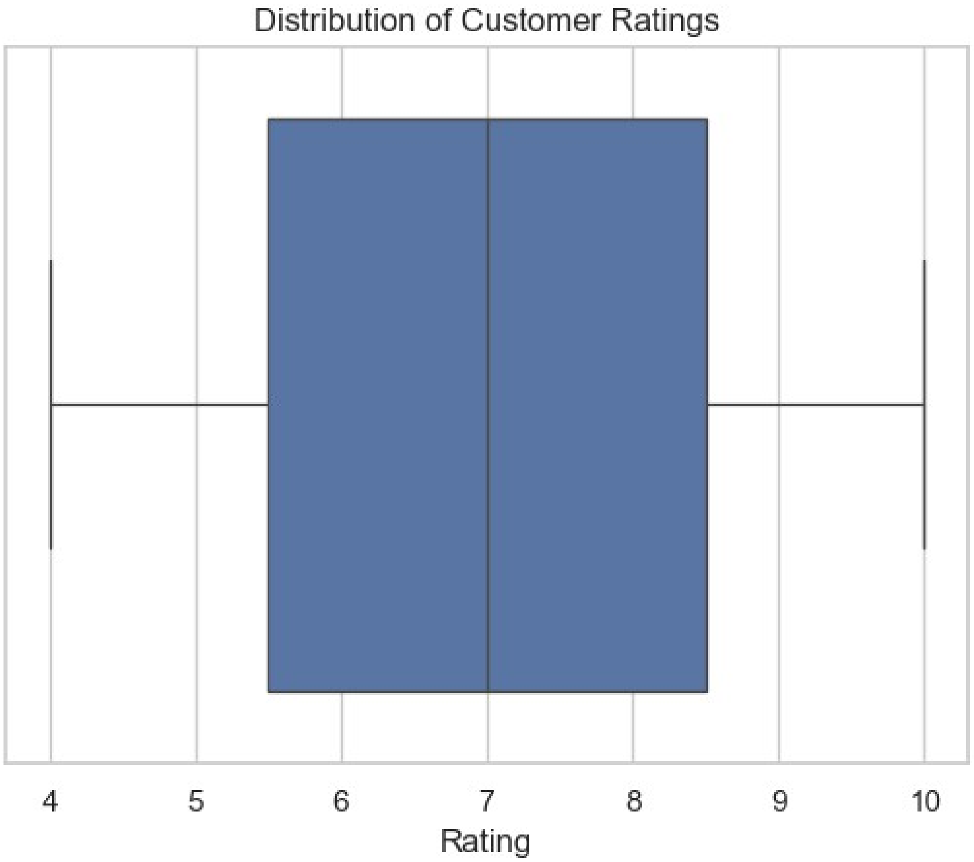
```
# Visualize missing values as a heatmap
sns.heatmap(df.isnull(), cbar=False, cmap='viridis')
plt.title('Missing Values Heatmap')
plt.show()
```

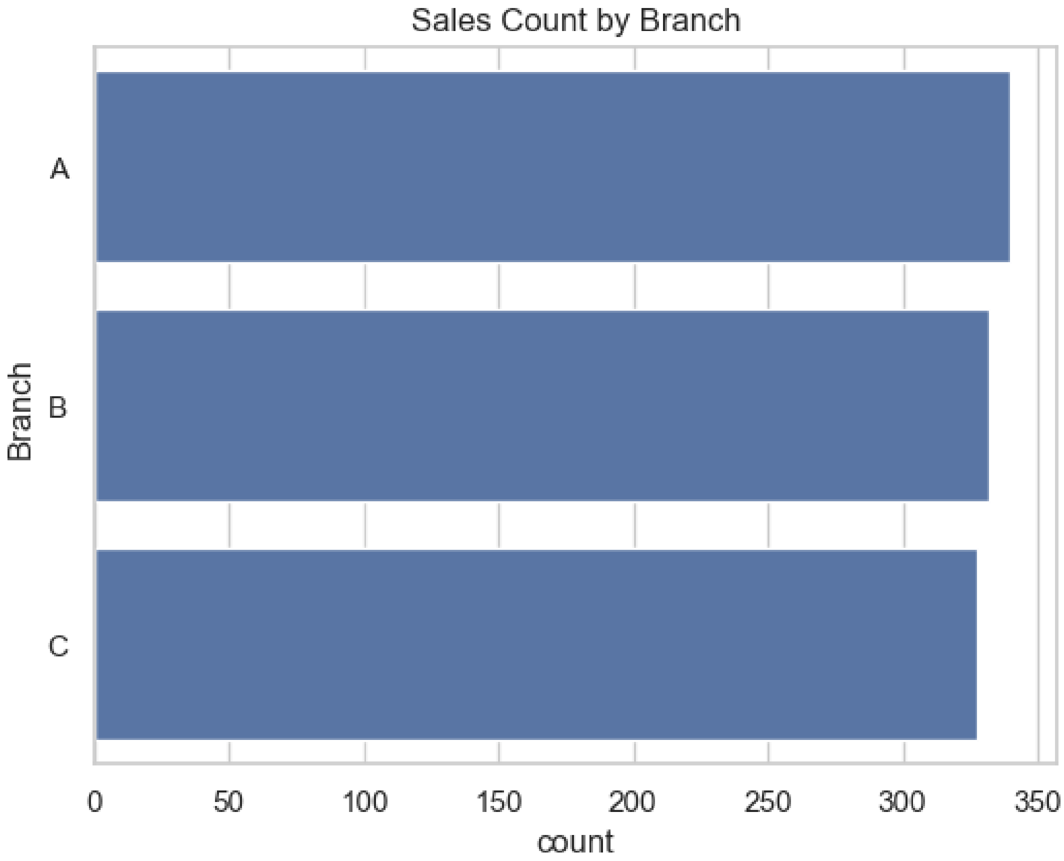


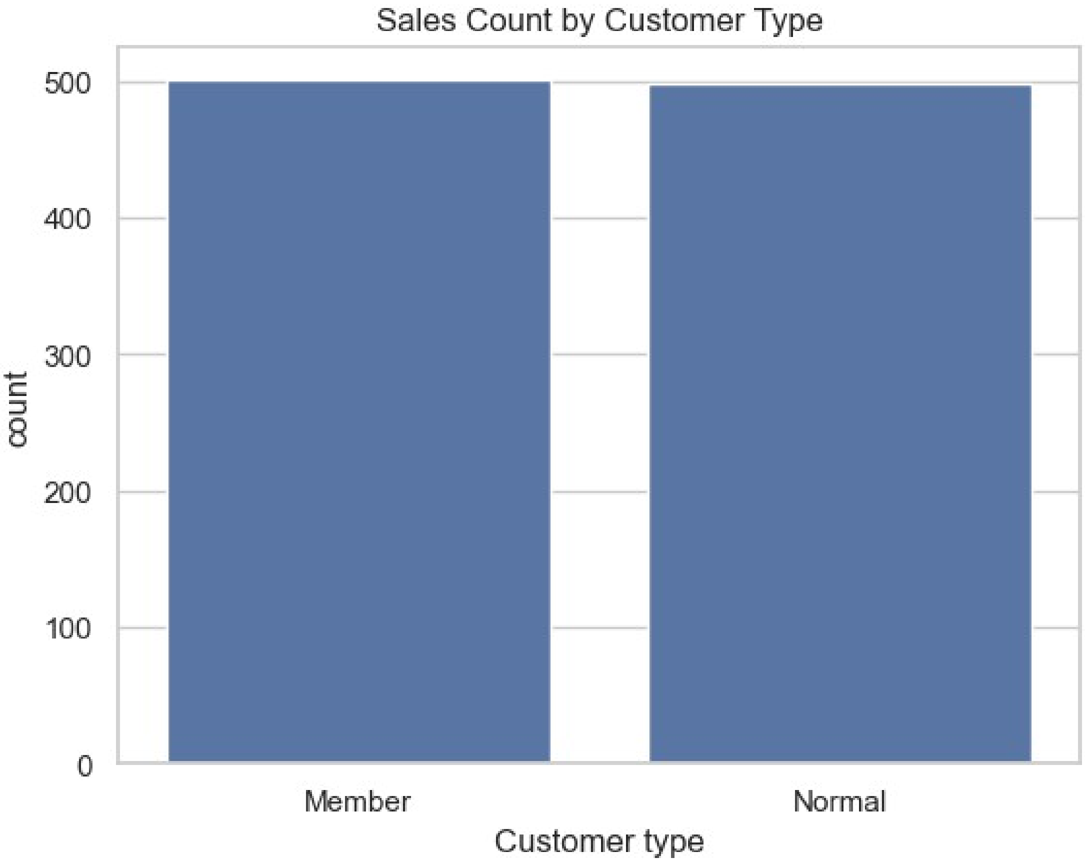
Univariate Analysis (Analyzing Single Variables)

```
# --- Numerical Variables --- #
# Distribution of Unit Price
sns.histplot(df['Unit price'], kde=True, bins=20)
plt.title('Distribution of Unit Price')
plt.show()

# Distribution of Customer Rating
sns.boxplot(x=df['Rating'])
plt.title('Distribution of Customer Ratings')
plt.show()

# --- Categorical Variables --- #
# Count of sales by Branch
sns.countplot(y='Branch', data=df, order =
df['Branch'].value_counts().index)
plt.title('Sales Count by Branch')
plt.show()

# Count of sales by Customer Type
sns.countplot(x='Customer type', data=df)
plt.title('Sales Count by Customer Type')
plt.show()
```




Distribution of Customer Ratings
4
5
6
7
8
9
10
Rating


Sales Count by Branch
A
B
C
Branch
0
50
100
150
200
250
300
350
count


Sales Count by Customer Type
500
400
300
200
100
0
count
Member
Normal
Customer type