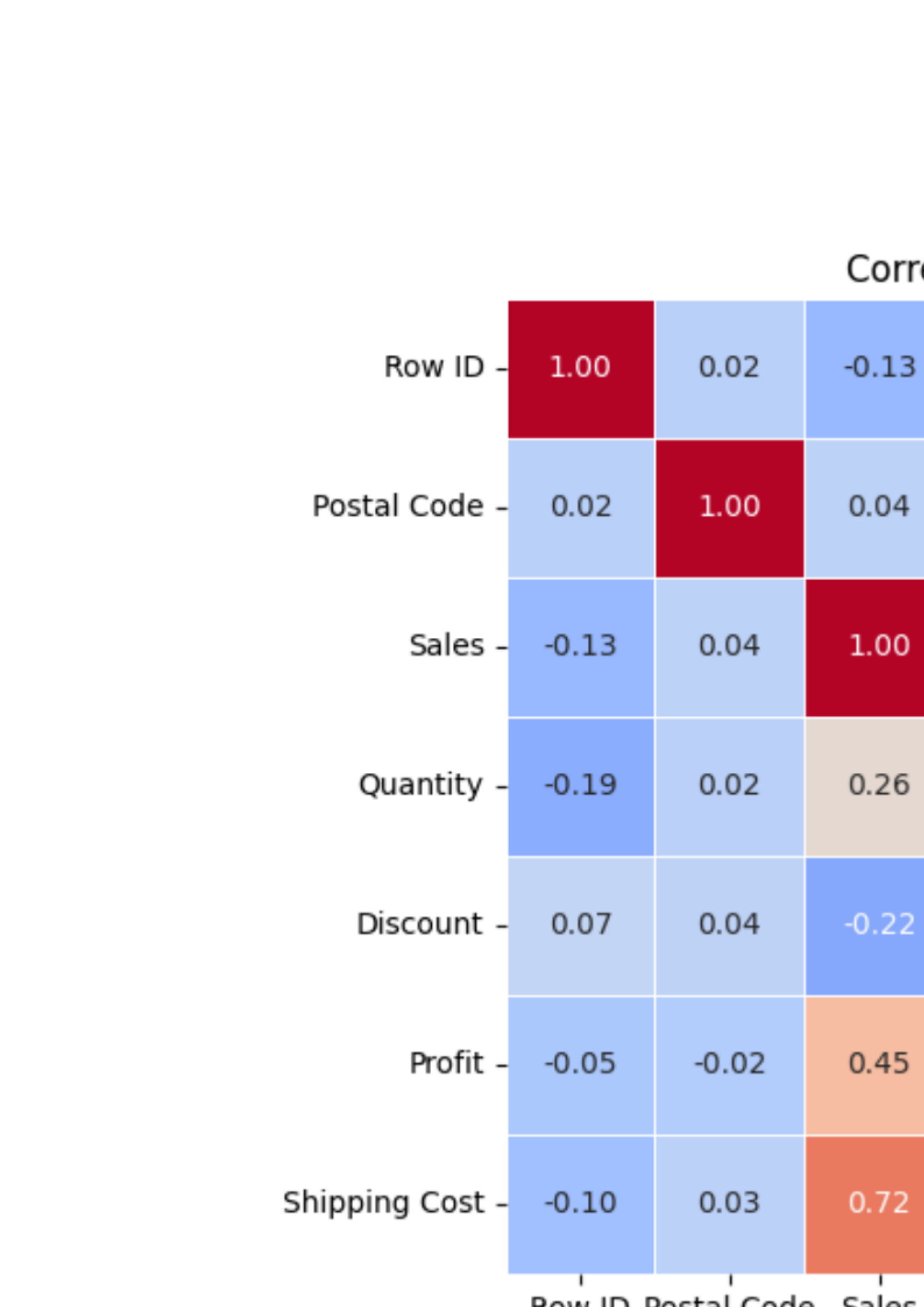
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing impor
from scipy import stats
df = pd.read_excel('Global Super
df.head()

{"type":"dataframe","variable_na

missing_values = df.isnull().sum
print(f"Missing values:\n{missin
df['Sales'] = df['Sales'].fillna
df['Profit'] = df['Profit'].fill
df['Region'] = df['Region'].fill

Missing values:
Row ID                         0
Order ID                       0
Order Date                     0
Ship Date                      0
Ship Mode                      0
Customer ID                    0
Customer Name                  0
Segment                        0
```

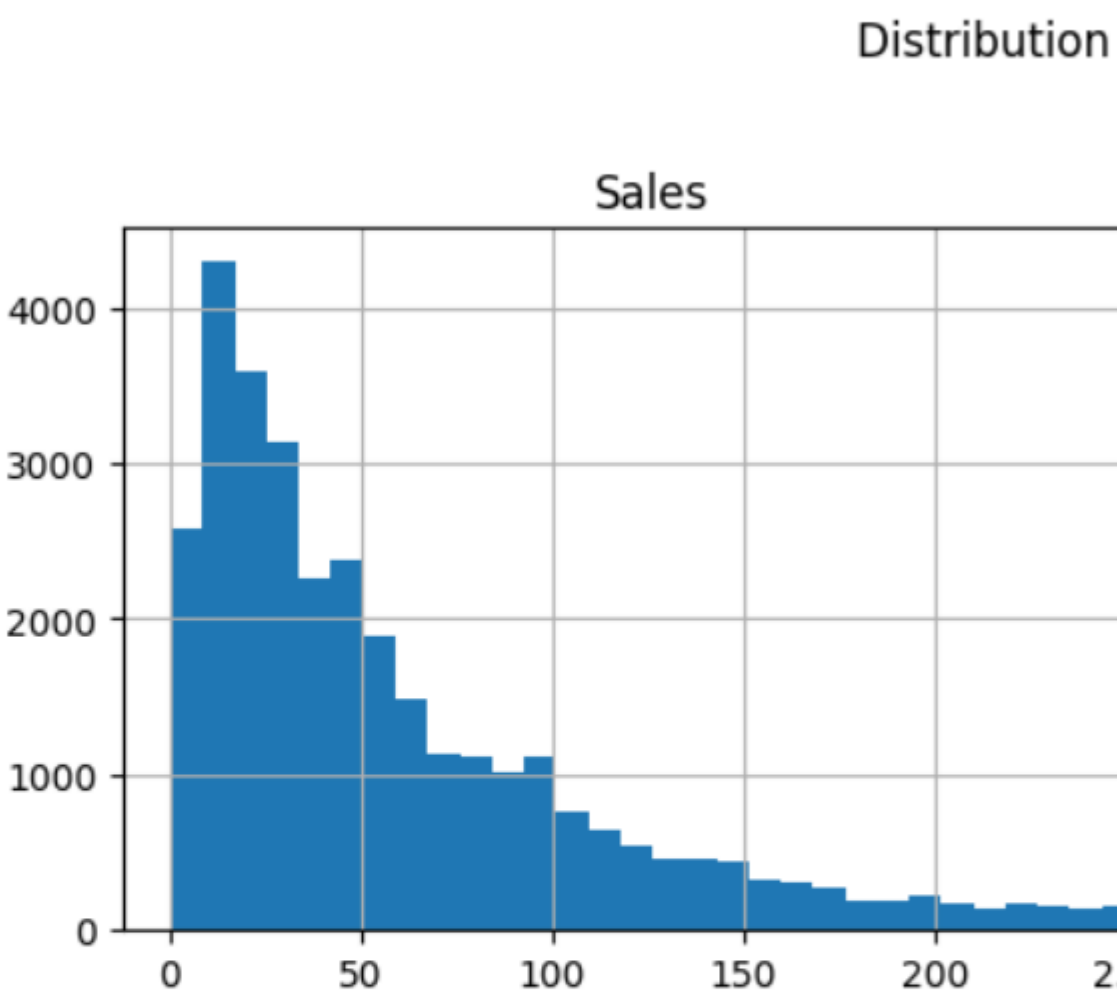
```
def remove_outliers(df, column):
    Q1 = df[column].quantile(0.25
    Q3 = df[column].quantile(0.75
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    df = df[(df[column] >= lower_
    return df
df = remove_outliers(df, 'Sales'
df = remove_outliers(df, 'Profit
print("done")
```

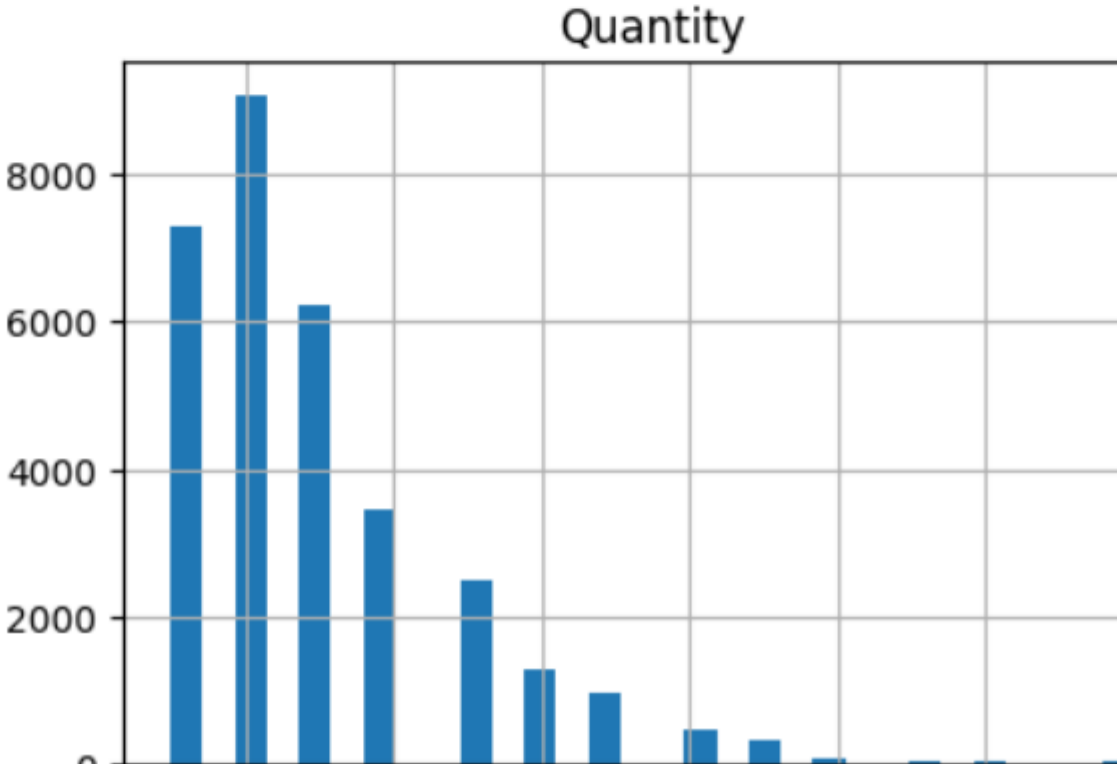
```
done
```

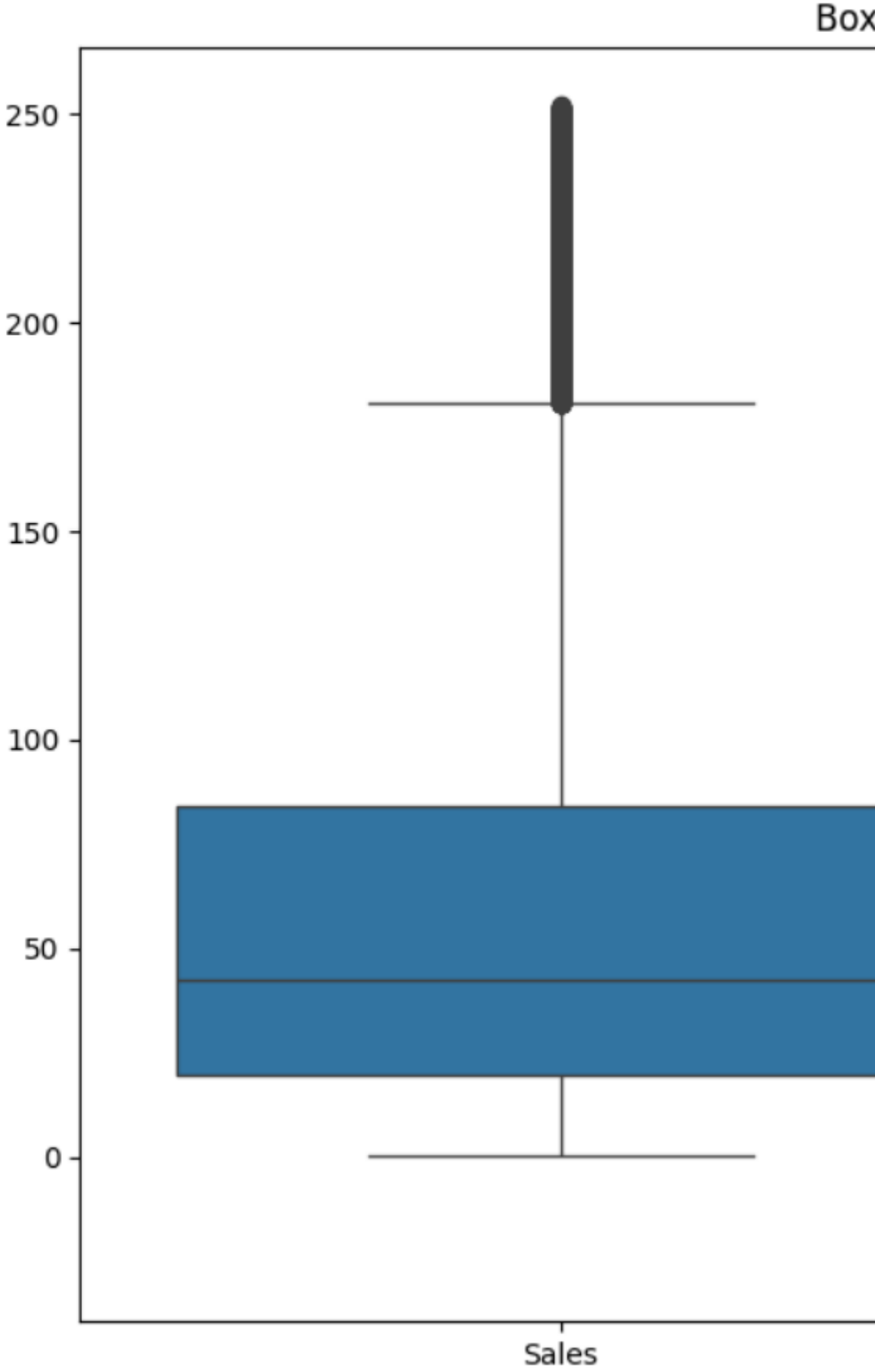
```
summary_stats = df.describe()
print(f"Summary Statistics:\n{sum
mean_sales = df['Sales'].mean()
median_sales = df['Sales'].media
std_sales = df['Sales'].std()
mean_profit = df['Profit'].mean(
median_profit = df['Profit'].med
std_profit = df['Profit'].std()
print(f"Sales Mean: {mean_sales}
{std_sales}")
print(f"Profit Mean: {mean_profi
{std_profit}")
```

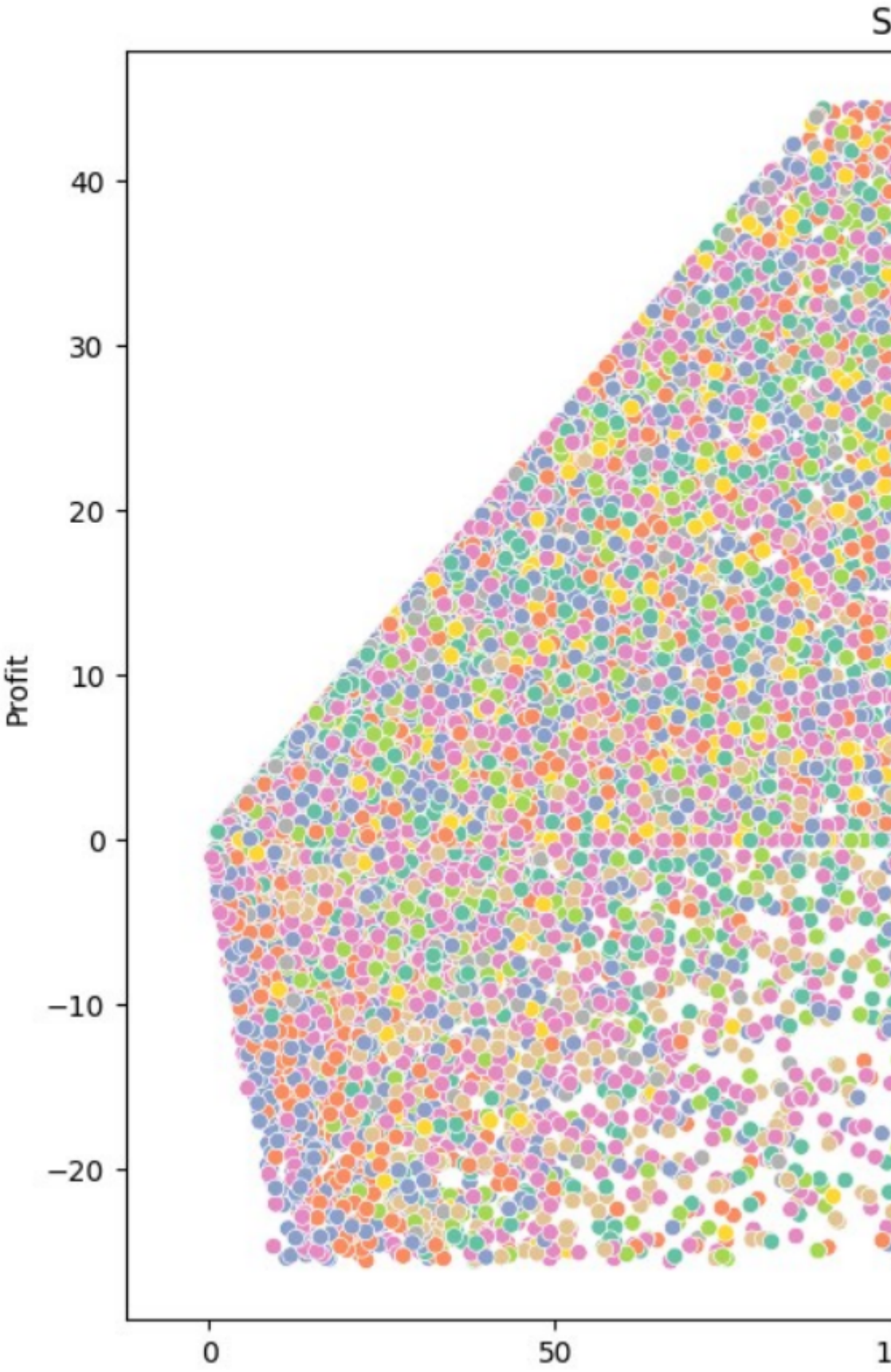
```
mean        2.959353       0.126369
min         1.000000       0.000000
25%         2.000000       0.000000
50%         2.000000       0.000000
75%         4.000000       0.200000
max        14.000000       0.800000
std         1.931421       0.20143
Sales Mean: 59.329159581560944, M
53.226193628132 23
Profit Mean: 8.366079337681 57, Me
13.626970520837 58

correlation_matrix = df.drop(colu
'Order Date', 'Ship Date', 'Custo
'State', 'Country', 'Market', 'Re
'Sub-Category', 'Product Name',
print(f"Correlation Matrix:\n{co
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, a
fmt='.2f', linewidths=0.5)
plt.title('Correlation Heatmap')
plt.show()

Correlation Matrix:
                Row ID  Postal C
Profit  \
```


Corr
Row ID
Postal Code
Sales
Quantity
Discount
Profit
Shipping Cost
1.00
0.02
-0.13
0.02
1.00
0.04
-0.13
0.04
1.00
-0.19
0.02
0.26
0.07
0.04
-0.22
-0.05
-0.02
0.45
-0.10
0.03
0.72


Distribution
Sales
4000
3000
2000
1000
0
0
50
100
150
200


Quantity
8000
6000
4000
2000
0


Box
250
200
150
100
50
0
Sales


S
40
30
20
10
0
−10
−20
Profit
0
50

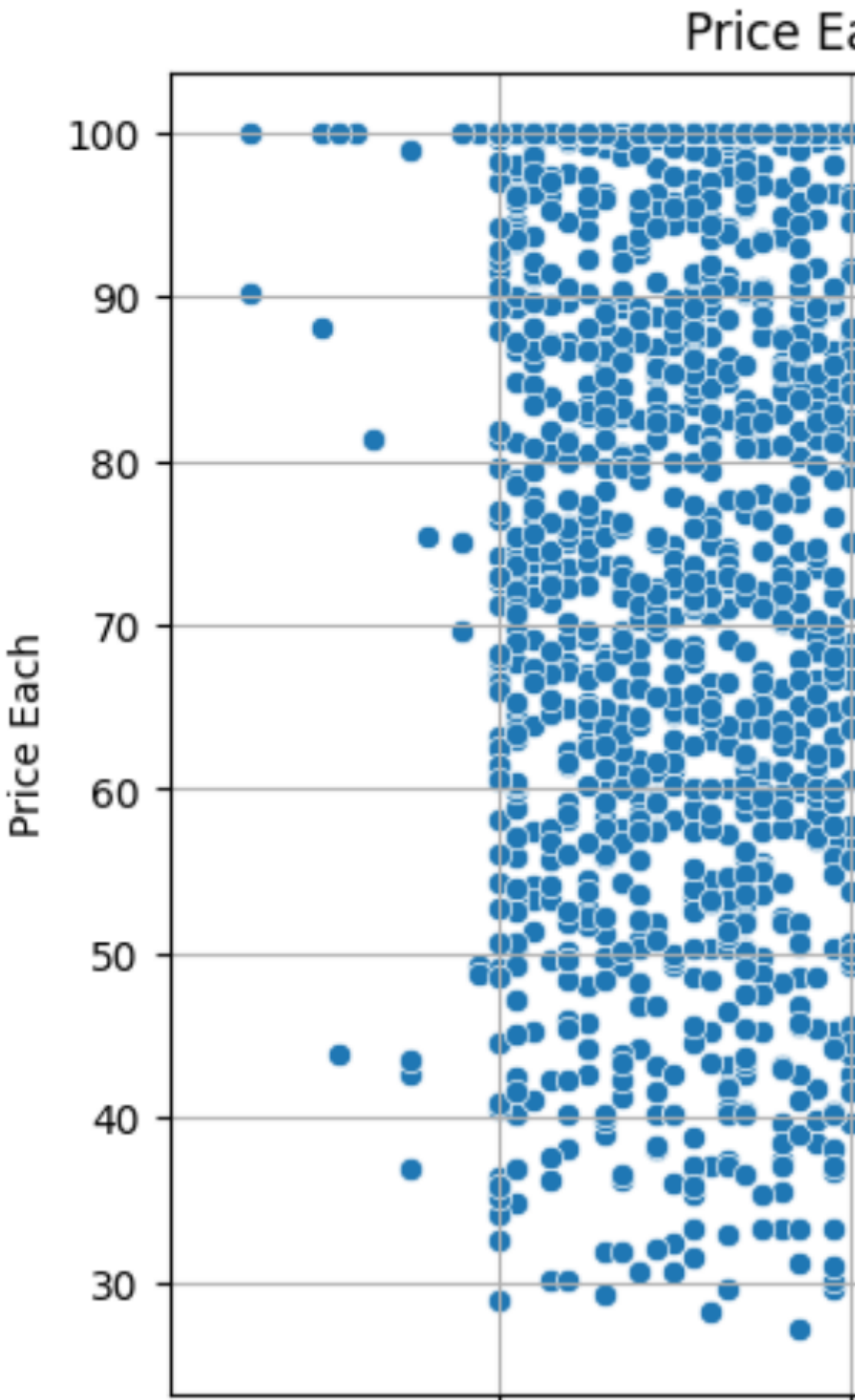
```
plt.figure(figsize=(10, 6))
df.groupby('ORDERDATE')['SALES']
plt.title("Sales Trend Over Time
plt.xlabel("Date")
plt.ylabel("Total Sales")
plt.grid(True)
plt.show()
plt.figure(figsize=(8, 6))
sns.scatterplot(x='QUANTITYORDER
plt.title("Price Each vs Quantity
plt.xlabel("Quantity Ordered")
plt.ylabel("Price Each")
plt.grid(True)
plt.show()
plt.figure(figsize=(8, 6))
df.groupby('COUNTRY')['SALES'].su
plt.title("Sales by Country")
plt.xlabel("Country")
plt.ylabel("Total Sales")
plt.xticks(rotation=45)
plt.show()

plt.figure(figsize=(8, 6))
df.groupby('PRODUCTLINE')
['SALES'].sum().sort_values().plo
plt.title("Sales by Product Line
```

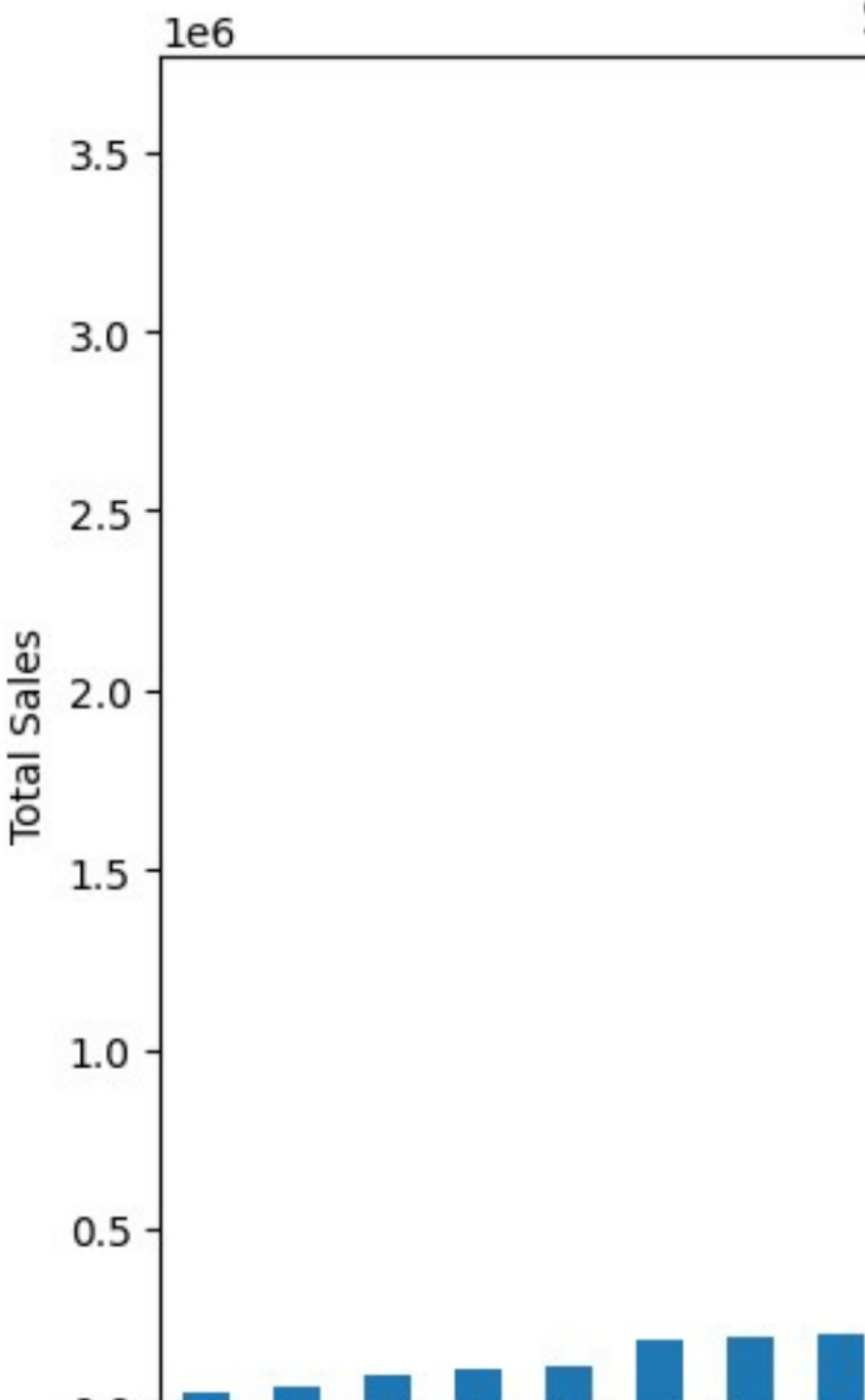
```
print(f"Quantity Ordered Coeffic

print("\nRecommendations:")
if model.coef_[0] > 0:
    print("Increasing Price Each
high-margin products.")
else:
    print("Increasing Price Each
Consider other strategies.")

if model.coef_[1] > 0:
    print("Increasing Quantity O
However, be mindful of profit ma
else:
    print("Increasing Quantity O
increase Sales. Re-evaluate disc

best_country = df.groupby('COUNT
best_productline = df.groupby('P

print(f"\nTop performing country
print(f"Top performing product l
```

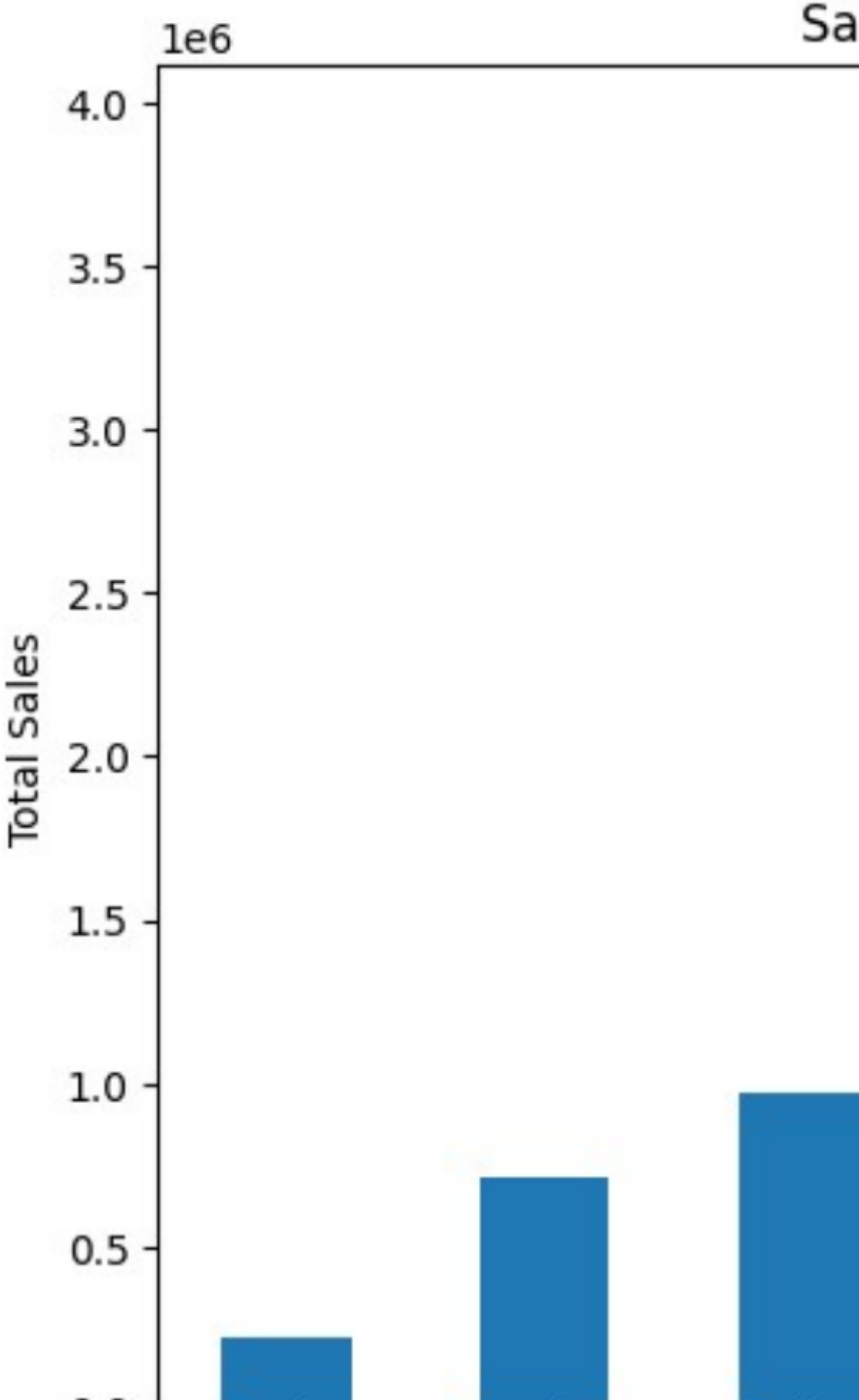
```
Shape of the dataset: (2823, 24)
Data types:
 ORDERNUMBER              int64
```

```
Missing values:
 ORDERNUMBER              0
QUANTITYORDERED           0
PRICEEACH                 0
ORDERLINENUMBER           0
SALES                     0
ORDERDATE                 0
STATUS                    0
QTR_ID                    0
MONTH_ID                  0
YEAR_ID                   0
PRODUCTLINE               0
MSRP                      0
PRODUCTCODE               0
CUSTOMERNAME              0
PHONE                     0
ADDRESSLINE1              0
ADDRESSLINE2           2521
CITY                      0
STATE                  1486
POSTALCODE               76
COUNTRY                   0
TERRITORY              1074
CONTACTLASTNAME           0
CONTACTFIRSTNAME          0
dtype: int64
```

```
4  10/10/2003 0:00  Shipped
6505551386

                      ADDRESSLINE1
POSTALCODE  \
0         897 Long Airport Avenue
10022
1              59 rue de l'Abbaye
51100
2  27 rue du Colonel Pierre Avia
75508
3              78934 Hillside Dr.
90003
4                 7734 Strong St.
NaN

          COUNTRY TERRITORY CONTAC
0  United States       NaN
1         France      EMEA
2         France      EMEA
3            USA       NaN
4  United States       NaN

[5 rows x 24 columns]
```


Price Ea
Price Each
100
90
80
70
60
50
40
30


1e6
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Total Sales


Sa
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Total Sales