E-commerce Store Sales Analysis

Lenara Sitshayeva

Research Questions:

We will attempt to answer the following research questions:

- What is the overall sales trend for the year?
- What is the average sale size?
- Does the month of the year affect sales?
- Is there a correlation between profit and quantity sold?

Descriptive Statistics:

Next, let's look at some descriptive statistics of our data.

```
# Total Sales statistics
print('Sales Statistics')
print(f"Sales: ${df['Sales'].sum():,.2f}")
```

```
Sales Statistics
Sales: $2,297,200.86
```

```
print(f"Sale: ${df['Sales'].mean():,.2f}")
```

```
Sale: $229.86
```

```
print(f"Median Sale: ${df['Sales'].median():,.2f}")
```

```
Median Sale: $54.49
```

```
print(f"Min Sale: ${df['Sales'].min():,.2f}")
```

```
Min Sale: $0.44
```

```
print(f"Max Sale: ${df['Sales'].max():,.2f}")
```

```
Max Sale: $22,638.48
```

```
print(f"Sales Variance: ${df['Sales'].var():,.2f}")
```

```
Sales Variance: $388,434.46
```

```
print(f"Sales Std Dev: ${df['Sales'].std():,.2f}")
```

```
Sales Std Dev: $623.25
```

```
print(f"Sales Skewness: {df['Sales'].skew():,.2f}")
```

```
Sales Skewness: 12.97
```

```
print(f"Sales Kurtosis: {df['Sales'].kurt():,.2f}")
```

```
Sales Kurtosis: 305.31
```

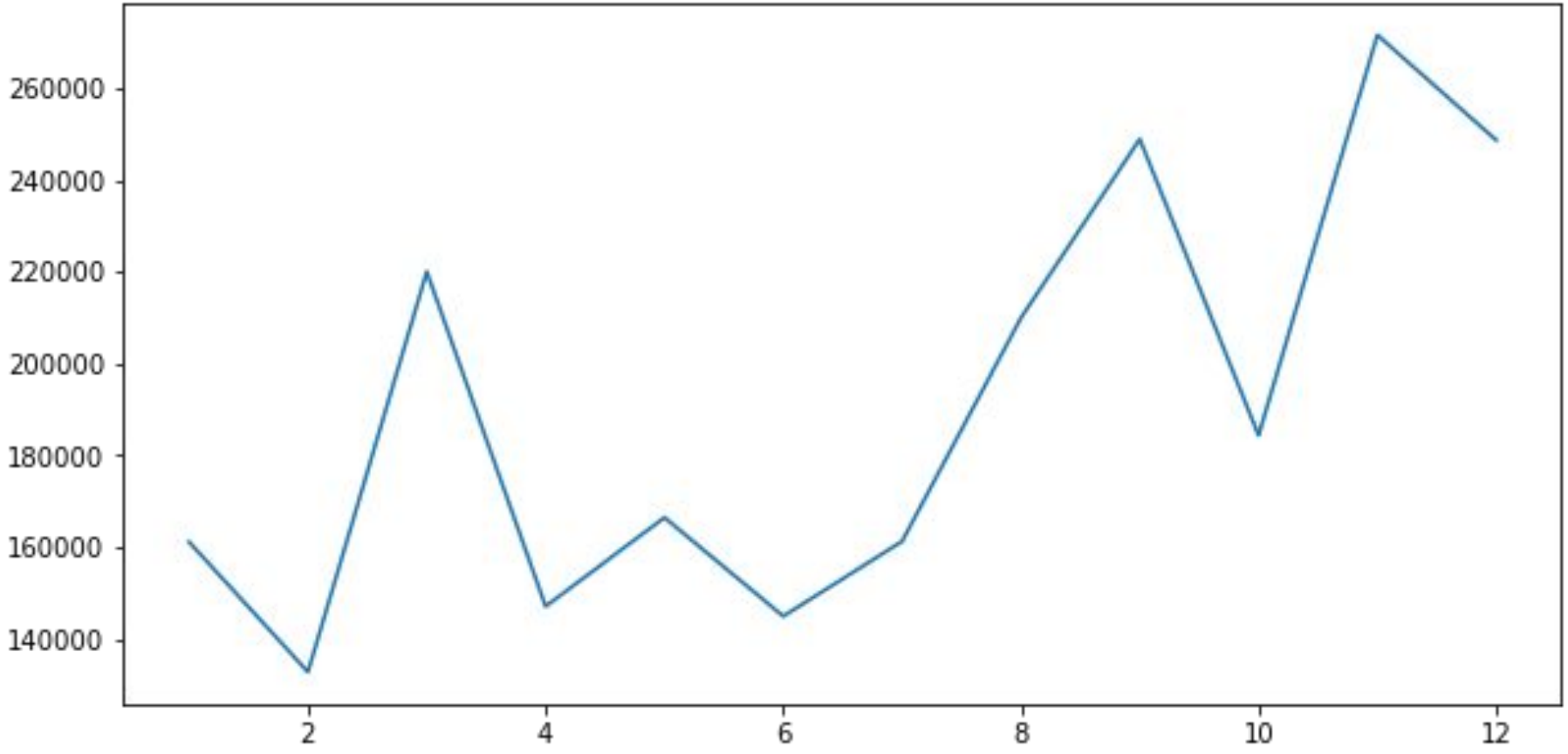
Hypotheses:

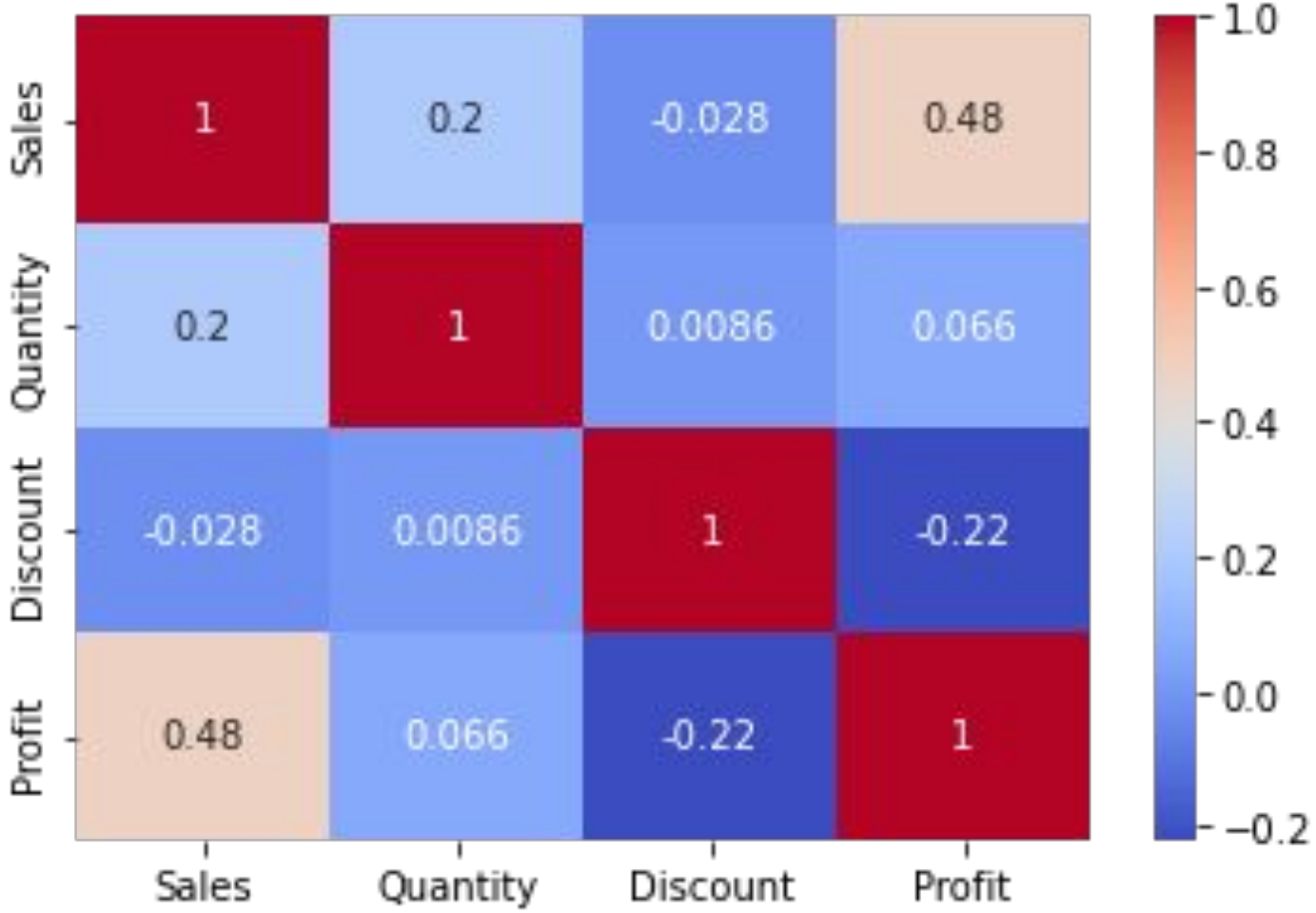
Next, we will formulate hypotheses based on our research questions and test them.

Hypothesis 1:

- Null hypothesis: There is no significant trend in sales throughout the year.
- Alternative hypothesis: There is a significant trend in sales throughout the year.

Monthly sales




Sales
Quantity
Discount
Profit
1
0.2
-0.028
0.48
0.2
1
0.0086
0.066
-0.028
0.0086
1
-0.22
0.48
0.066
-0.22
1
Sales
Quantity
Discount
Profit
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Based on the correlation matrix provided:

There is a strong positive correlation (0.48) between Sales and Profit. This indicates that as Sales increase, so does Profit. This is a desirable relationship for businesses as it means that they can increase their profits by increasing their sales.

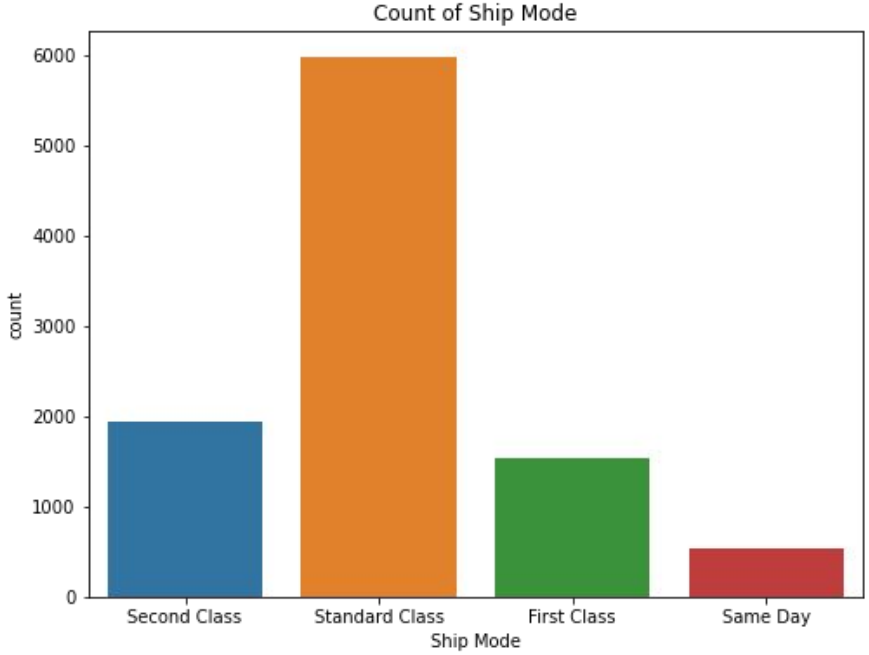
There is a weak positive correlation (0.20) between Sales and Quantity. This indicates that there is some relationship between the two variables, but it is not very strong. This could be due to various factors such as pricing, seasonality, etc.

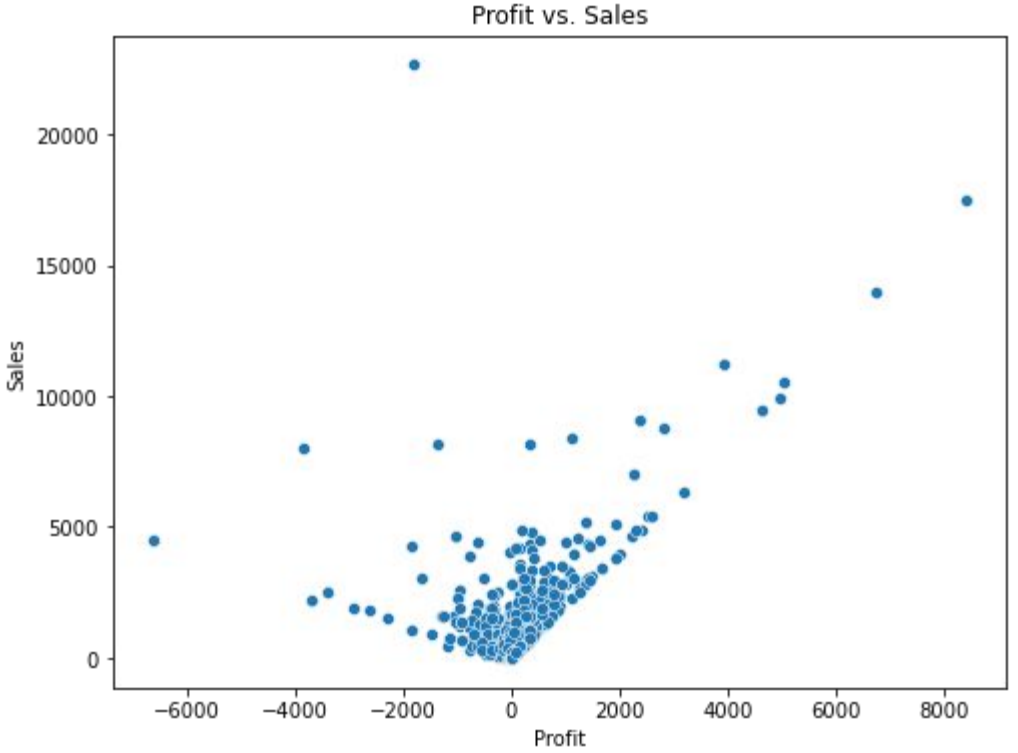
There is a weak positive correlation (0.07) between Quantity and Profit. This indicates that there is some relationship between the two variables, but it is not very strong. This could be due to various factors such as the cost of goods sold, pricing, etc.

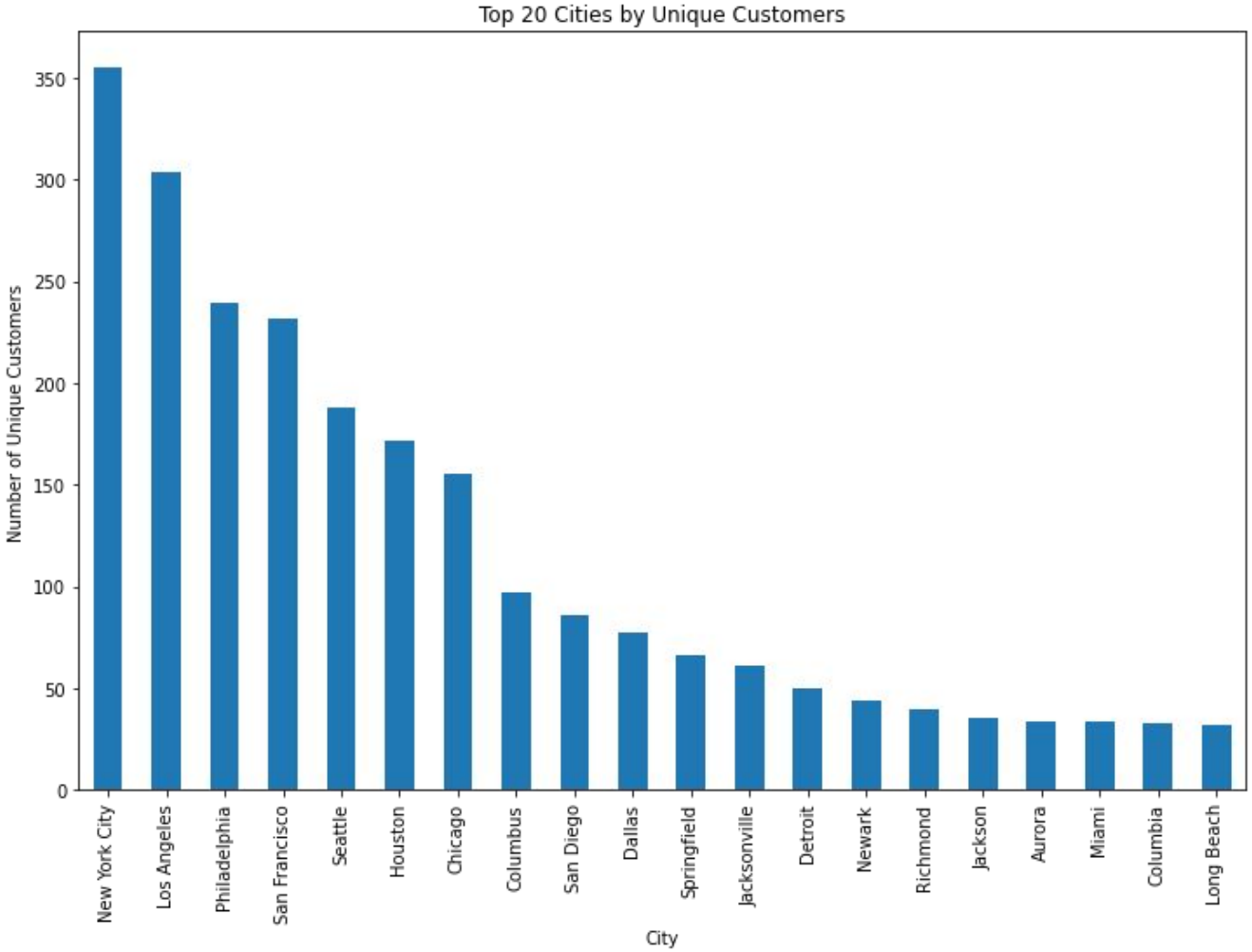
There is a weak negative correlation (-0.02) between Sales and Discount. This indicates that there is some relationship between the two variables, but it is not very strong. This could be due to various factors such as pricing strategy, promotions, etc.

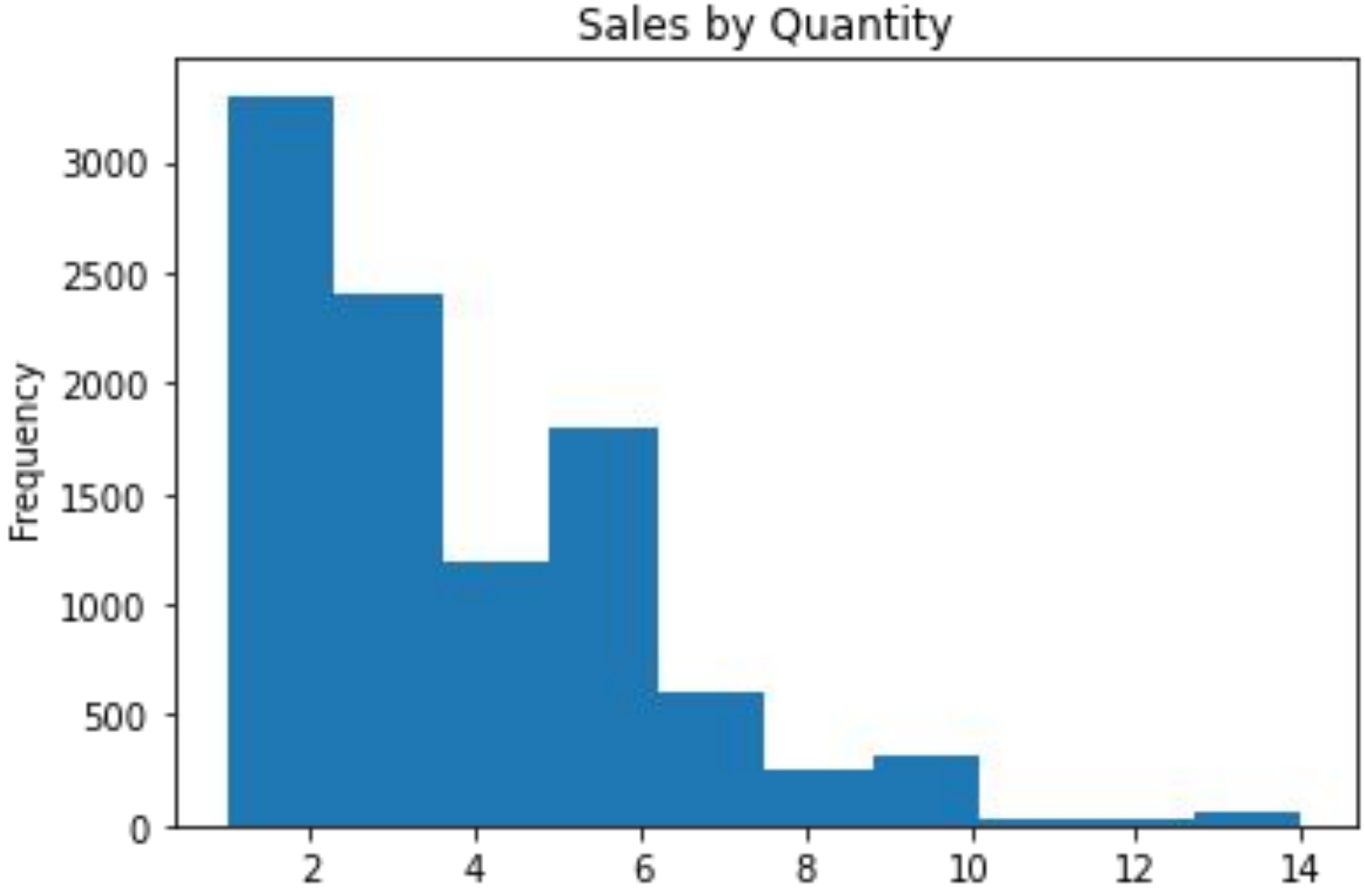
There is a weak negative correlation (-0.22) between Discount and Profit. This indicates that as Discount increases, Profit decreases. This is an undesirable relationship for businesses as it means that they are sacrificing their profits in order to make sales.

Overall, the correlation matrix suggests that Sales is the most important variable in determining Profit. However, the relationships between the variables are not very strong, which suggests that there are other factors that could be influencing Profit as well. Further analysis and modeling would be needed to identify these factors and to develop a more accurate model for predicting Profit based on Sales, Quantity, and Discount.


Count of Ship Mode
count
6000
5000
4000
3000
2000
1000
0
Second Class
Standard Class
First Class
Same Day
Ship Mode


Profit vs. Sales
Sales
20000
15000
10000
5000
0
−6000
−4000
−2000
0
2000
4000
6000
8000
Profit


Top 20 Cities by Unique Customers
Number of Unique Customers
0
50
100
150
200
250
300
350
New York City
Los Angeles
Philadelphia
San Francisco
Seattle
Houston
Chicago
Columbus
San Diego
Dallas
Springfield
Jacksonville
Detroit
Newark
Richmond
Jackson
Aurora
Miami
Columbia
Long Beach
City


Sales by Quantity
Frequency
3000
2500
2000
1500
1000
500
0
2
4
6
8
10
12
14