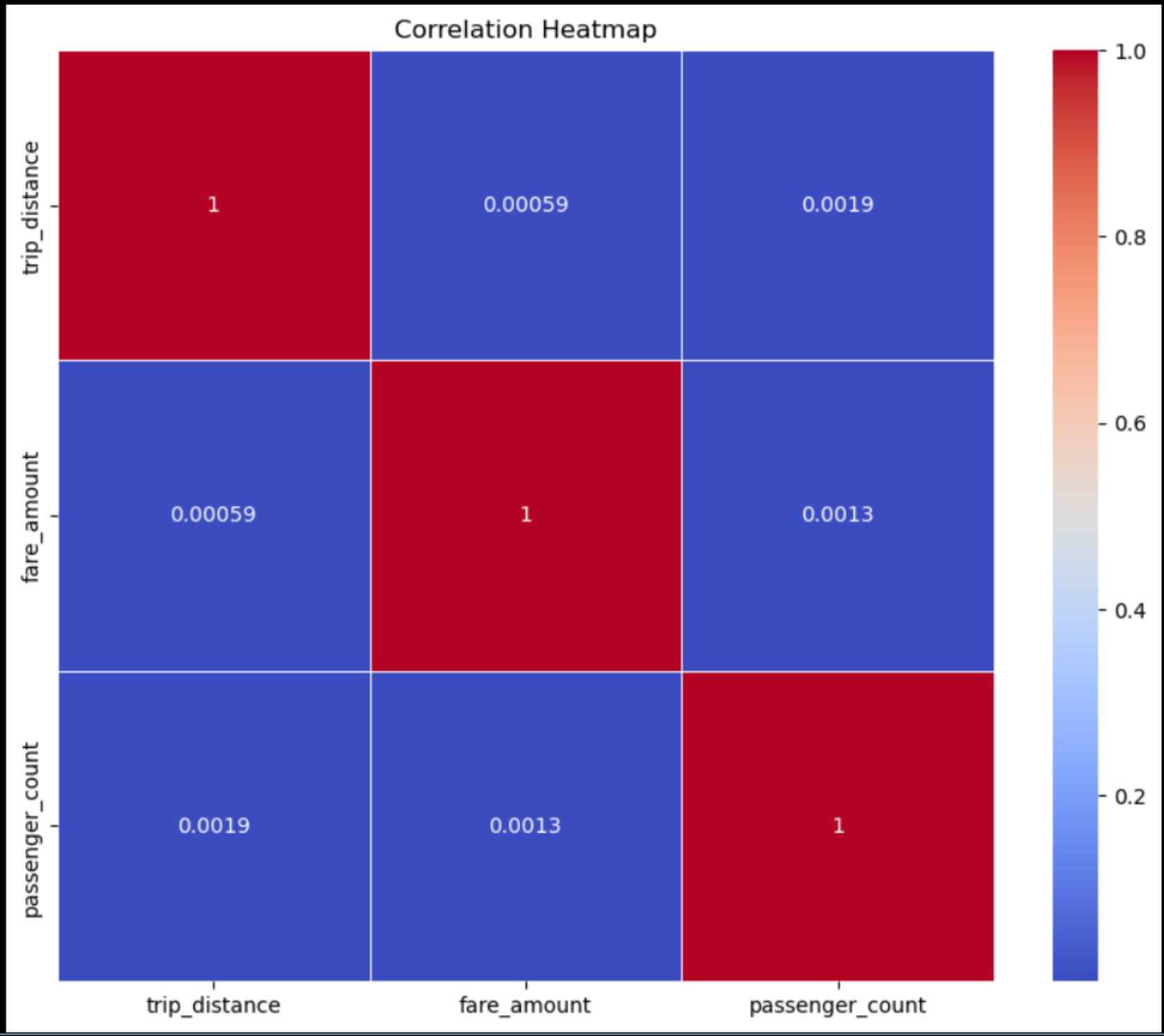
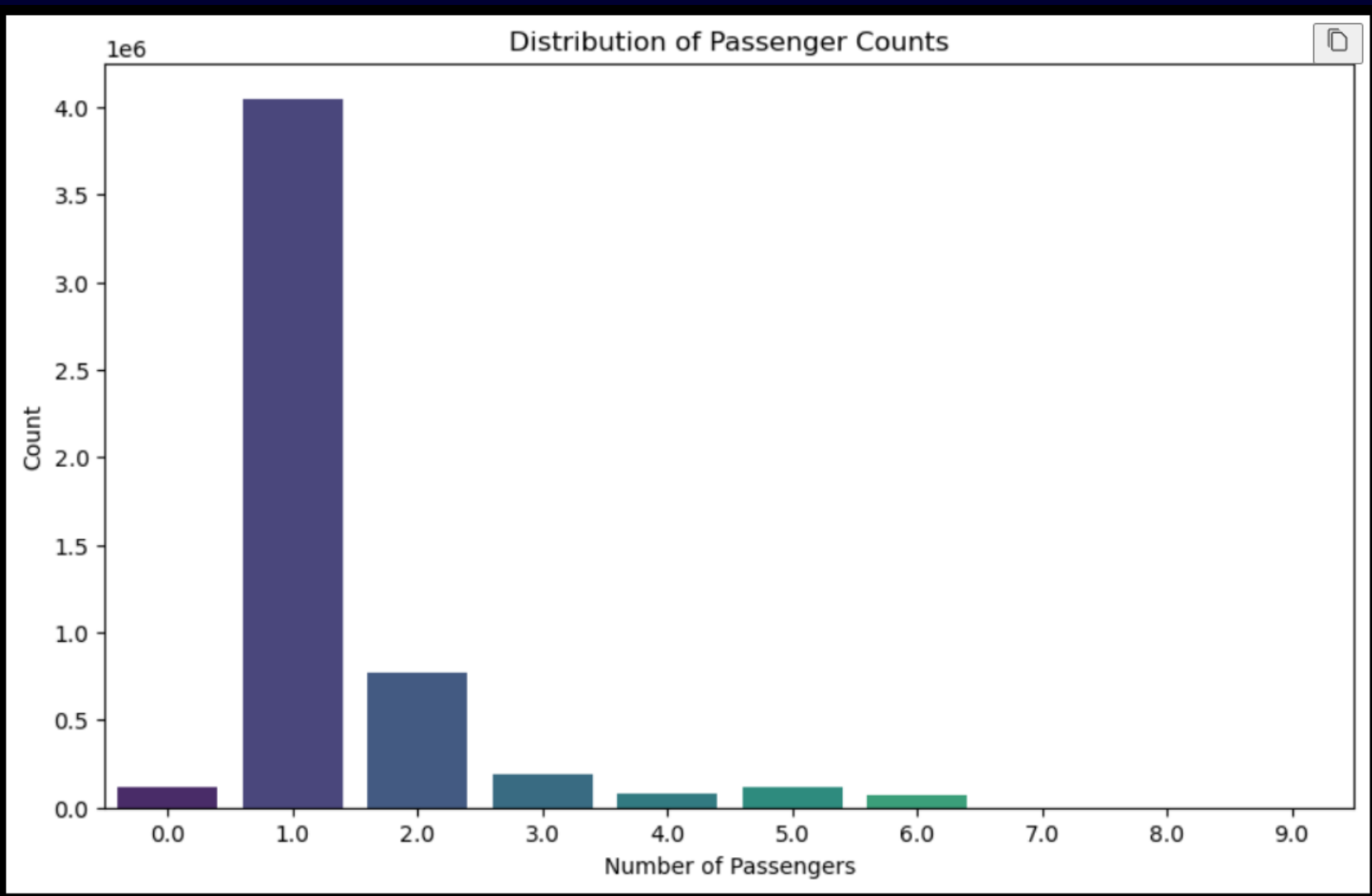
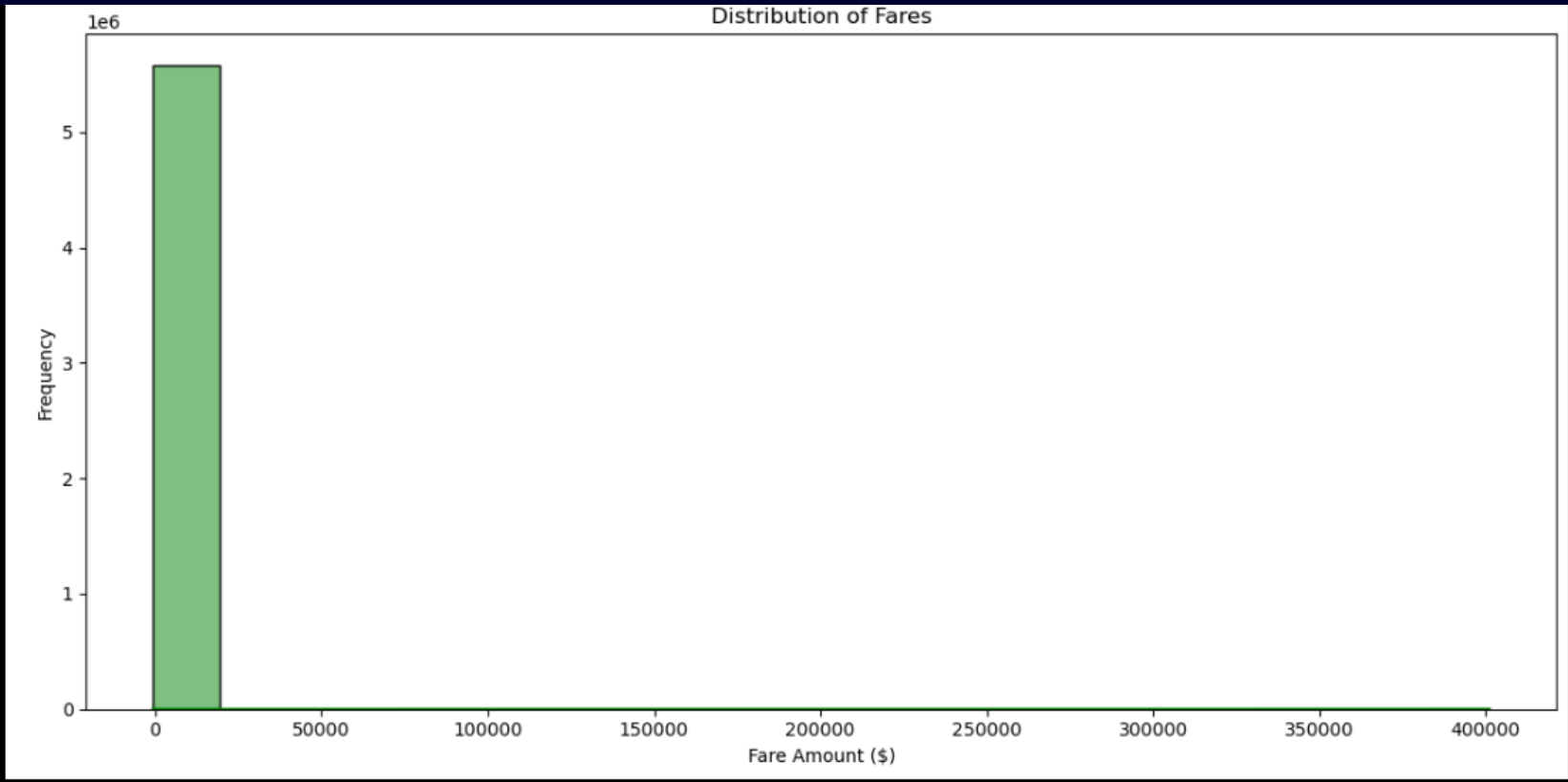
Project 1: Analysis and Forecasting of NYC Taxi Rides

Task 2: Exploratory Data Analysis

- Conduct exploratory data analysis to understand the patterns and relationships in the data.
- Analyze the distribution of trip distances, fares, and passenger counts.
- Explore the relationship between trip distances, fares, and passenger counts.

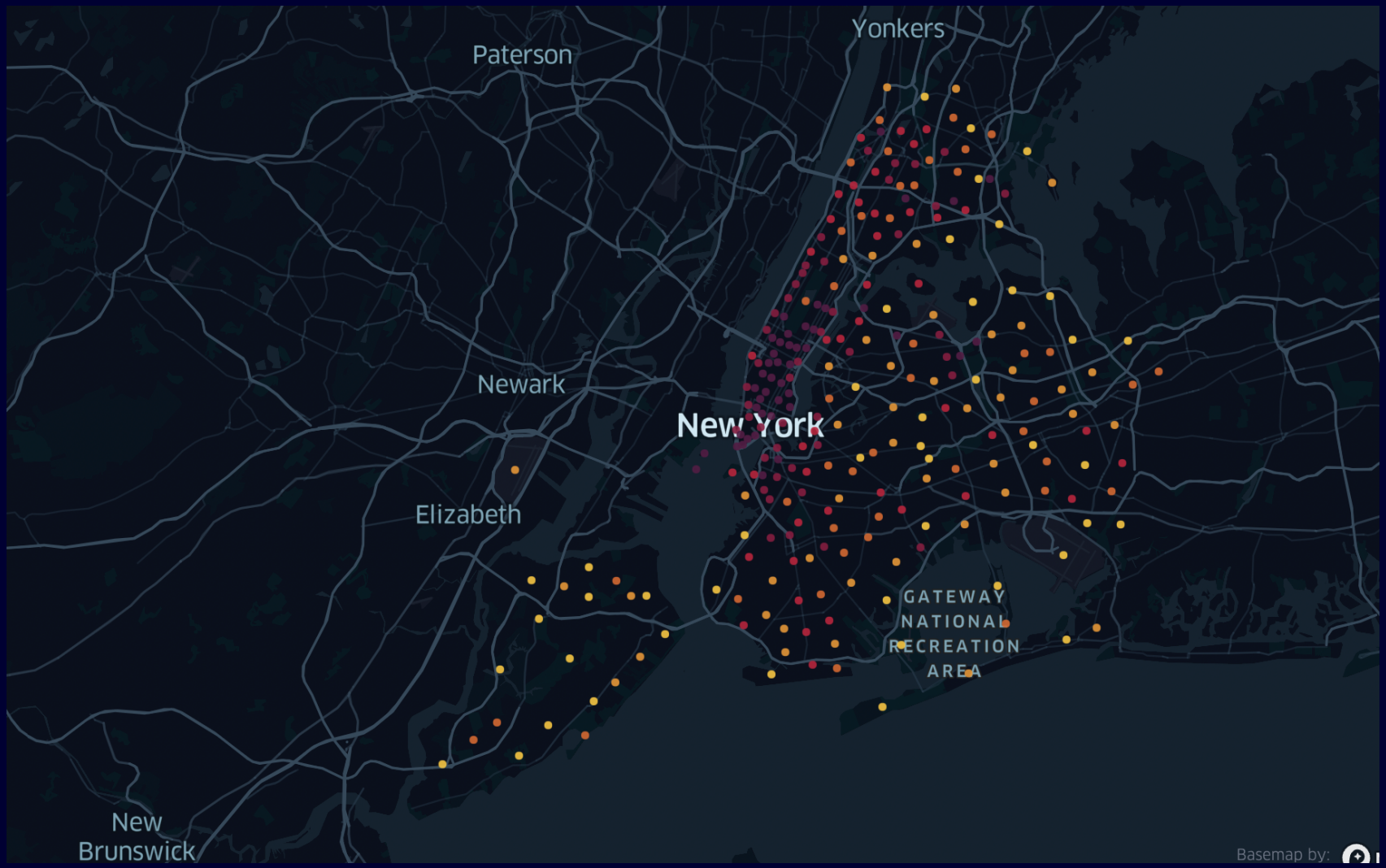

Correlation Heatmap
trip_distance
fare_amount
passenger_count
1
0.00059
0.0019
0.00059
1
0.0013
0.0019
0.0013
1
trip_distance
fare_amount
passenger_count
1.0
0.8
0.6
0.4
0.2


Distribution of Passenger Counts
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Count
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
Number of Passengers


Distribution of Fares
1e6
5
4
3
2
1
0
Frequency
0
50000
100000
150000
200000
250000
300000
350000
400000
Fare Amount ($)

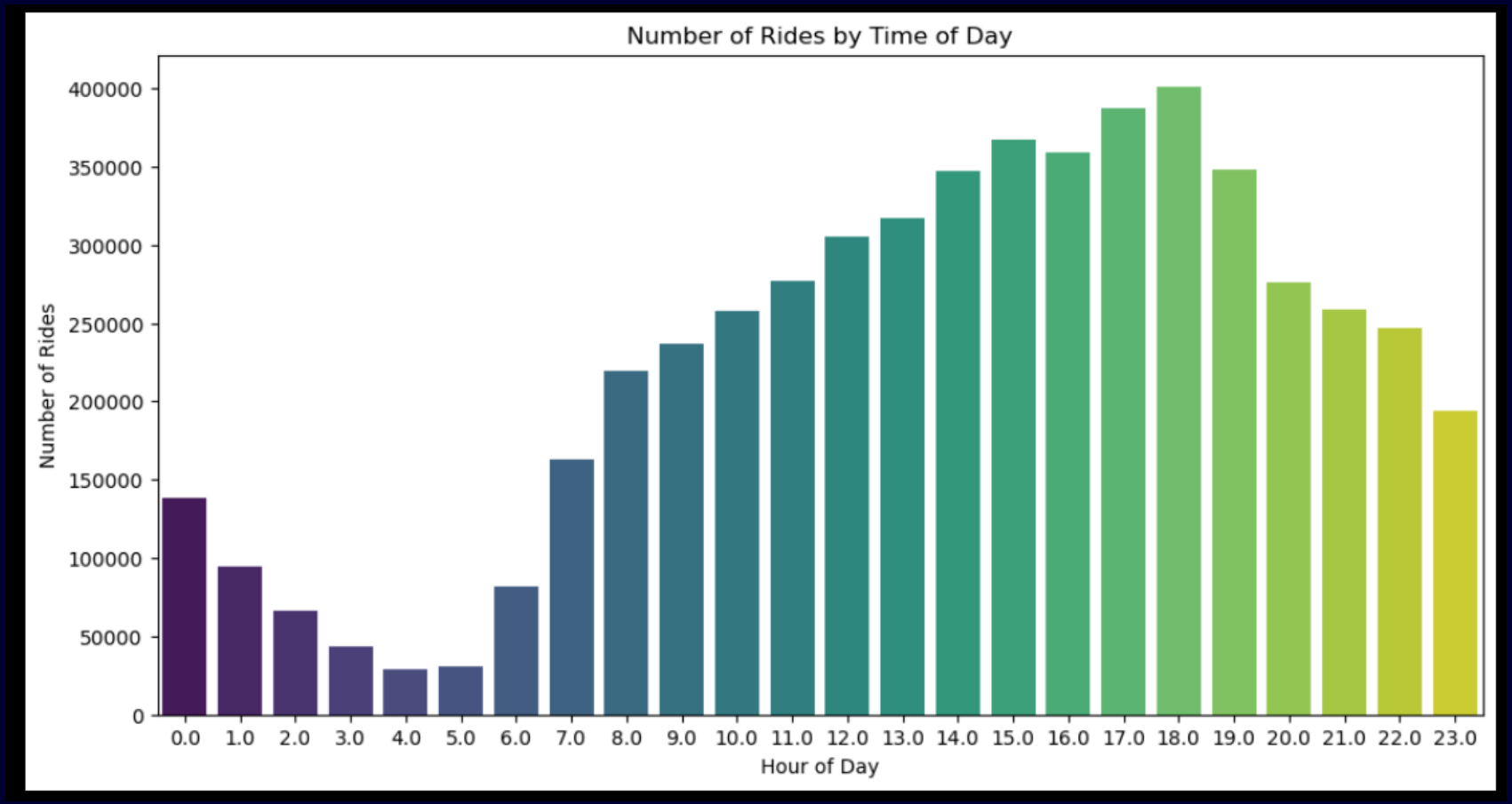
Task 3: Spatial Analysis

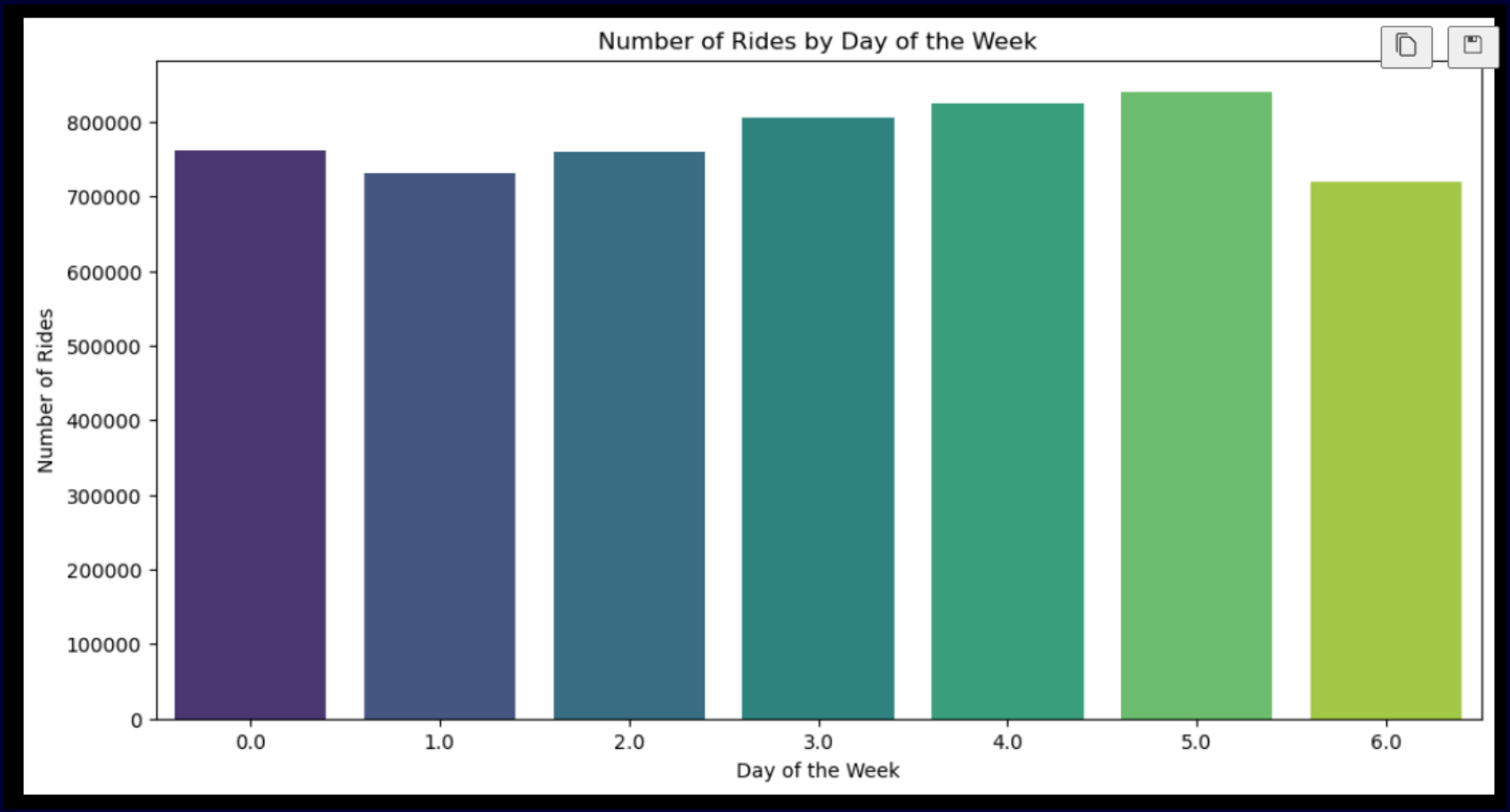
- Use Kepler.gl (https://kepler.gl/) or similar tools/packages to visualize the spatial patterns of taxi rides.
- Explore pickup and dropoff locations, as well as routes taken during taxi rides.
- Analyze spatial patterns to identify hotspots of taxi demand.

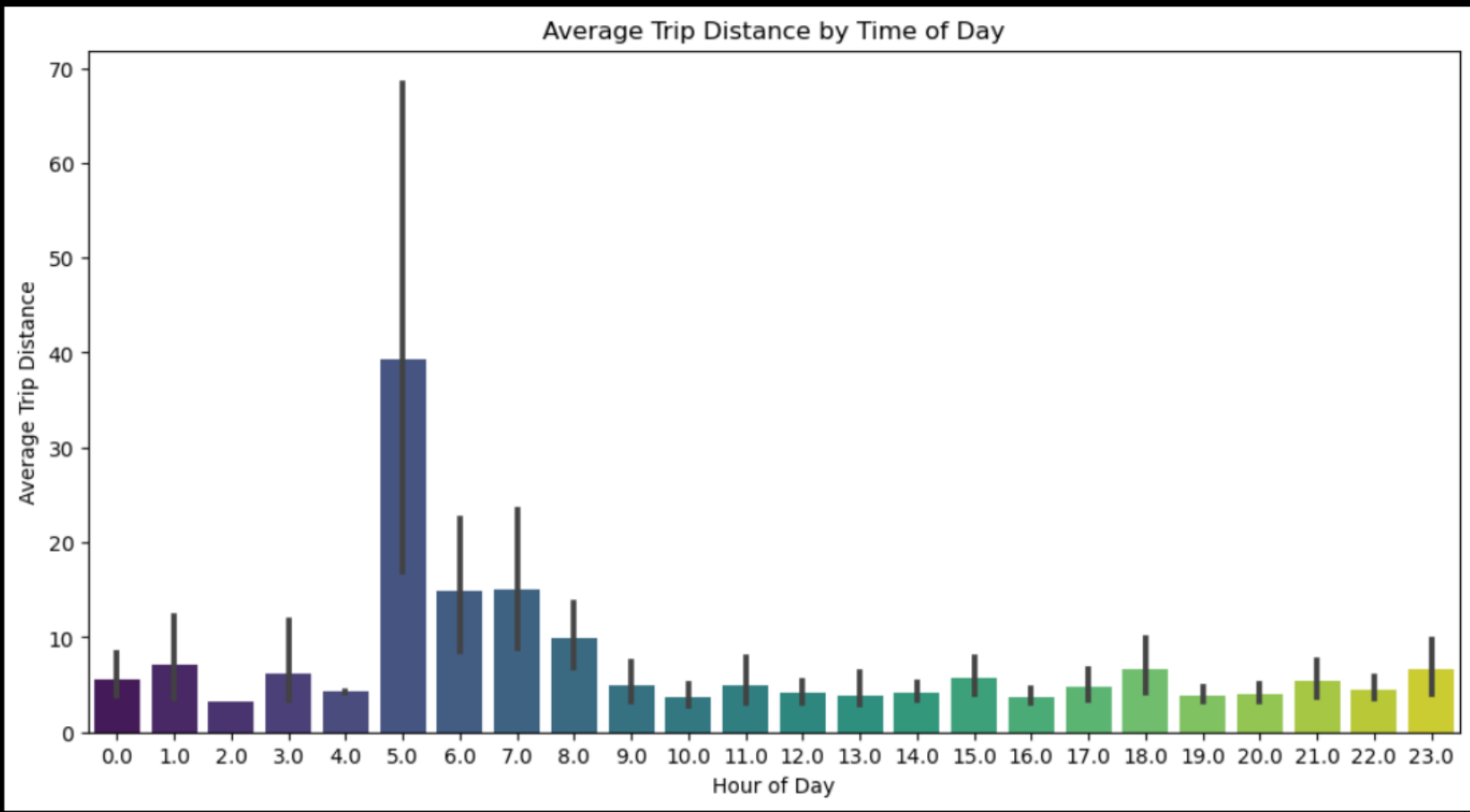

Yonkers
Paterson
Newark
New York
Elizabeth
GATEWAY
NATIONAL
RECREATION
AREA
New
Brunswick
Basemap by:

Task 4: Temporal Analysis

- Analyze the temporal patterns of taxi rides.
- Examine the number of rides by time of day, day of the week, and month of the year.
- Investigate the relationship between temporal patterns and other variables, such as trip distance and fare.

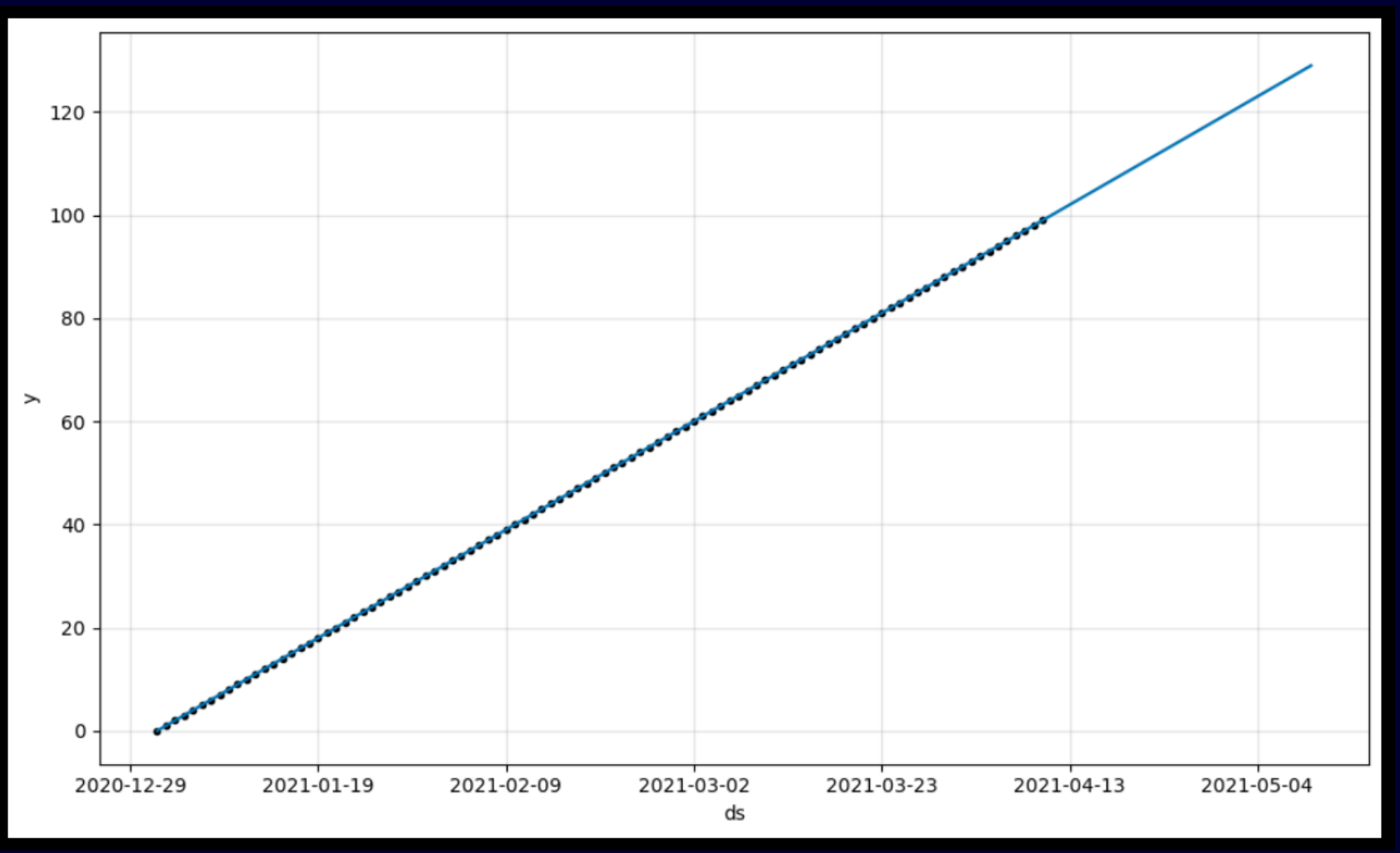

Number of Rides by Time of Day
Number of Rides
400000
350000
300000
250000
200000
150000
100000
50000
0
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0
Hour of Day


Number of Rides by Day of the Week
Number of Rides
800000
700000
600000
500000
400000
300000
200000
100000
0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
Day of the Week


Average Trip Distance by Time of Day
Average Trip Distance
0
10
20
30
40
50
60
70
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
21.0
22.0
23.0
Hour of Day

Task 5: Time-Series Forecasting

- Use Prophet or a similar tool to forecast the number of taxi rides in the future for green and yellow taxis.
- Create a time-series model, tune its parameters, and validate its performance.
- Interpret the model's predictions and identify the factors driving the forecasted trends.


120
100
80
60
40
20
0
y
2020-12-29
2021-01-19
2021-02-09
2021-03-02
2021-03-23
2021-04-13
2021-05-04
ds

Project 2: NASA Data Acquisition, Visualization, and Analysis

Task 2: Data Analysis (NASA)

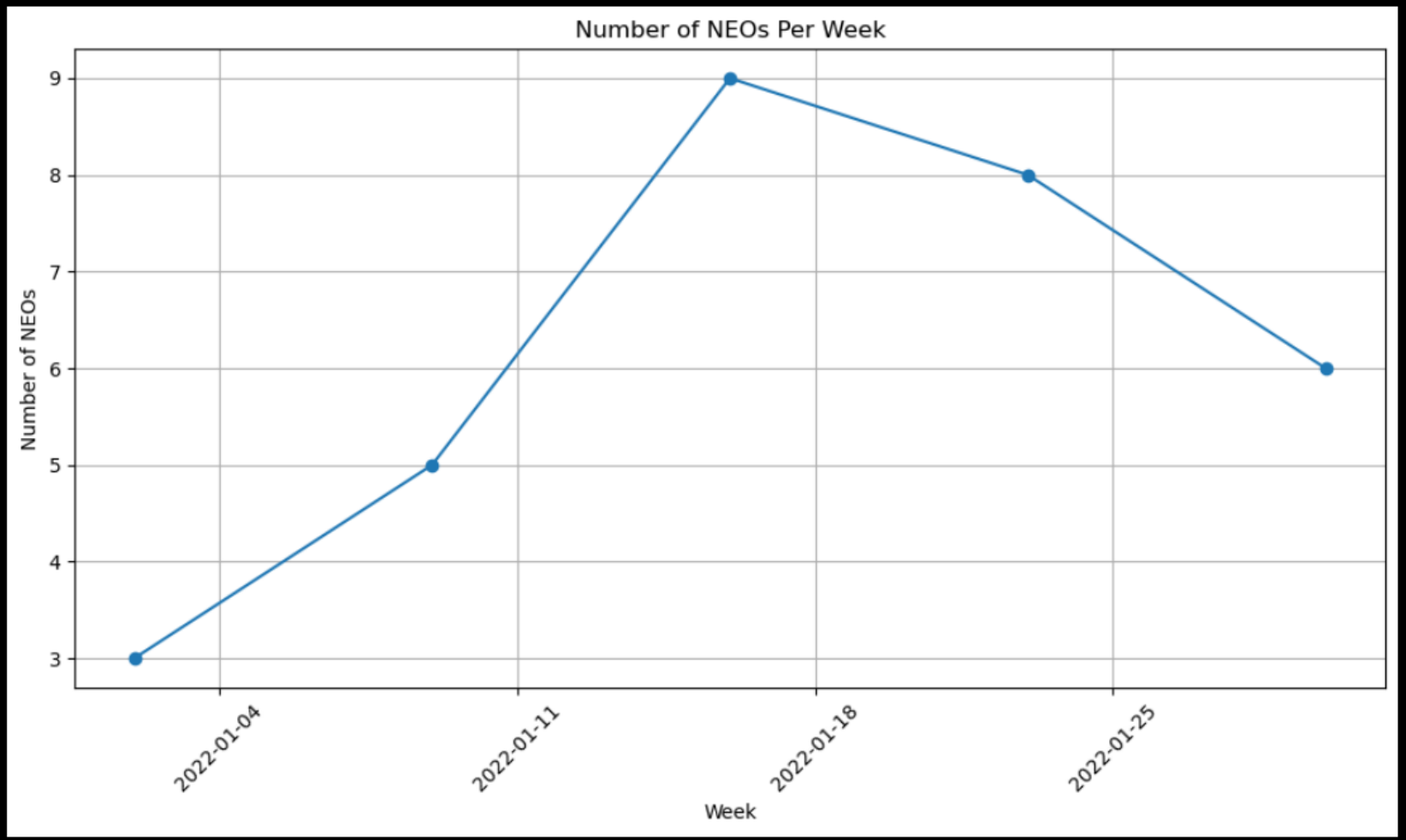
- Calculate the average size of the NEOs for each day.
- Determine the proportion of NEOs that are potentially hazardous.
- Find the NEO with the closest approach distance for each day.
- Use statistical methods to analyze the data. Calculate the mean, median, mode, and standard deviation of the NEO sizes.
- Define at least two other statistics and compute them for analyzing the data further.
- Use statistical methods to determine if the size of a NEO is correlated with whether it is potentially hazardous.

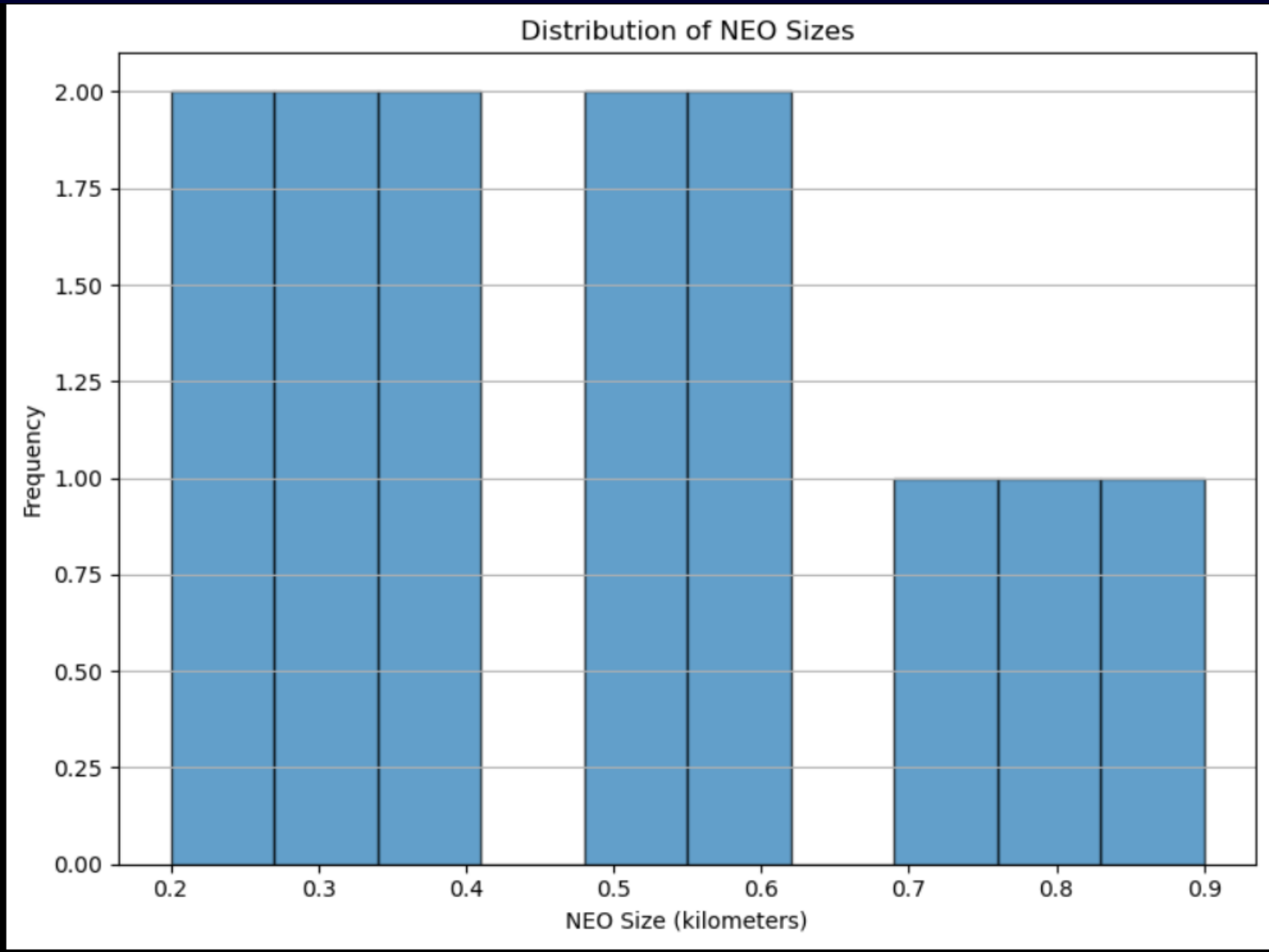
```python
data = {
    'element_count': 125,
    'near_earth_objects': {
        '2022-01-07': [{
            # ... (other data for each near-earth object)
            'estimated_diameter': {
                'kilometers': {
                    'estimated_diameter_min': 0.1925550782,
                    'estimated_diameter_max': 0.4305662442
                }
            }
        }]
    }
}

neo_list = data['near_earth_objects']['2022-01-07']
total_diameter = 0
count = 0

for neo in neo_list:
    diameter_min = neo['estimated_diameter']['kilometers']['estimated_diameter_min']
    diameter_max = neo['estimated_diameter']['kilometers']['estimated_diameter_max']
    total_diameter += diameter_min
    count += 1

average_diameter = total_diameter / count

print("Average Diameter (kilometers):", average_diameter)
```

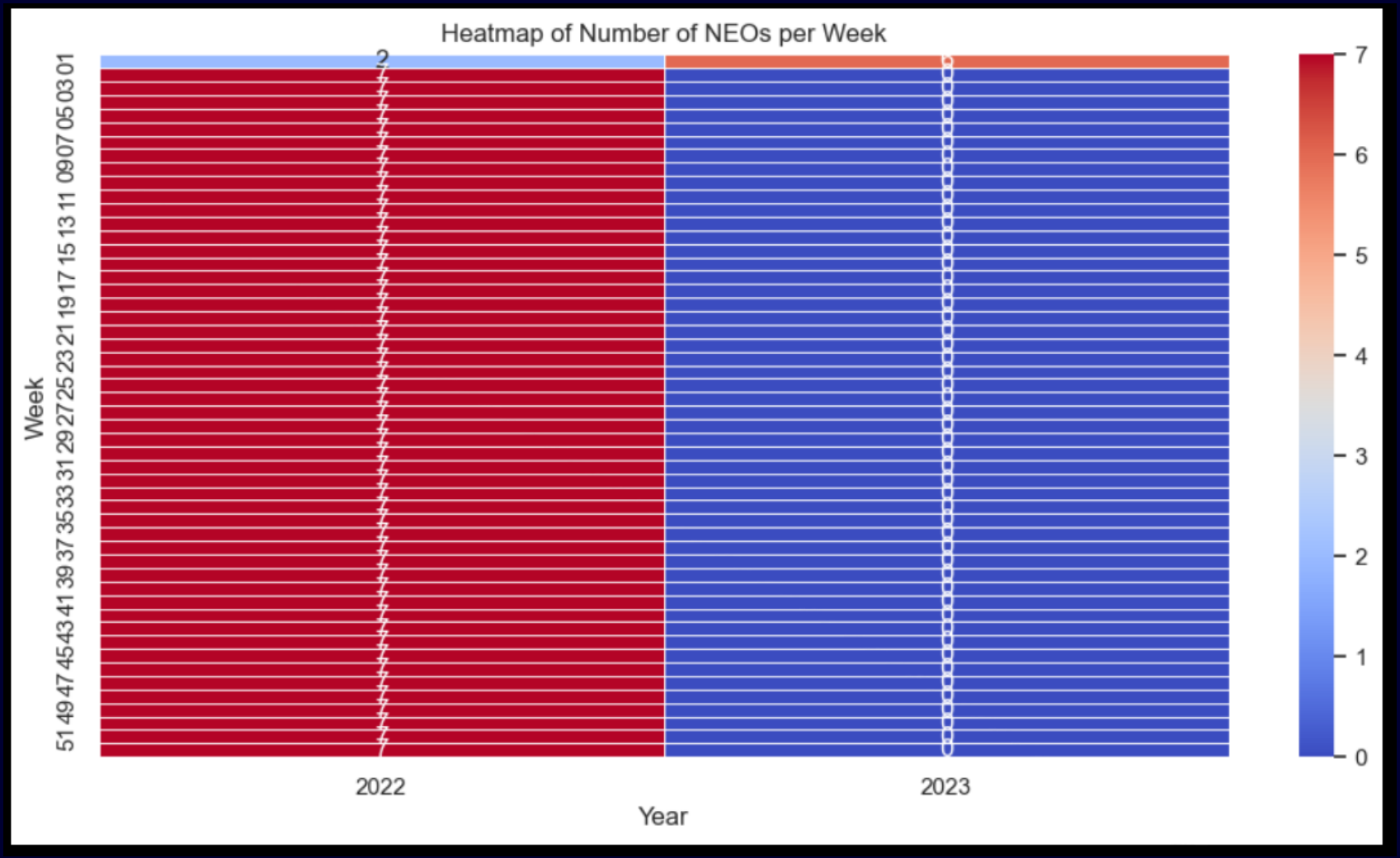
```
Average Diameter (kilometers): 0.1925550782
```

Task 3: Data Visualization Part A (NASA)

- Create a line plot of the number of NEOs per week.
- Create a histogram of the distribution of NEO sizes.
- Create a bar plot of the average NEO size per week.
- Use a library like Seaborn to create more complex visualizations, such as a box plot of the NEO sizes or a heat map of the number of NEOs per week.

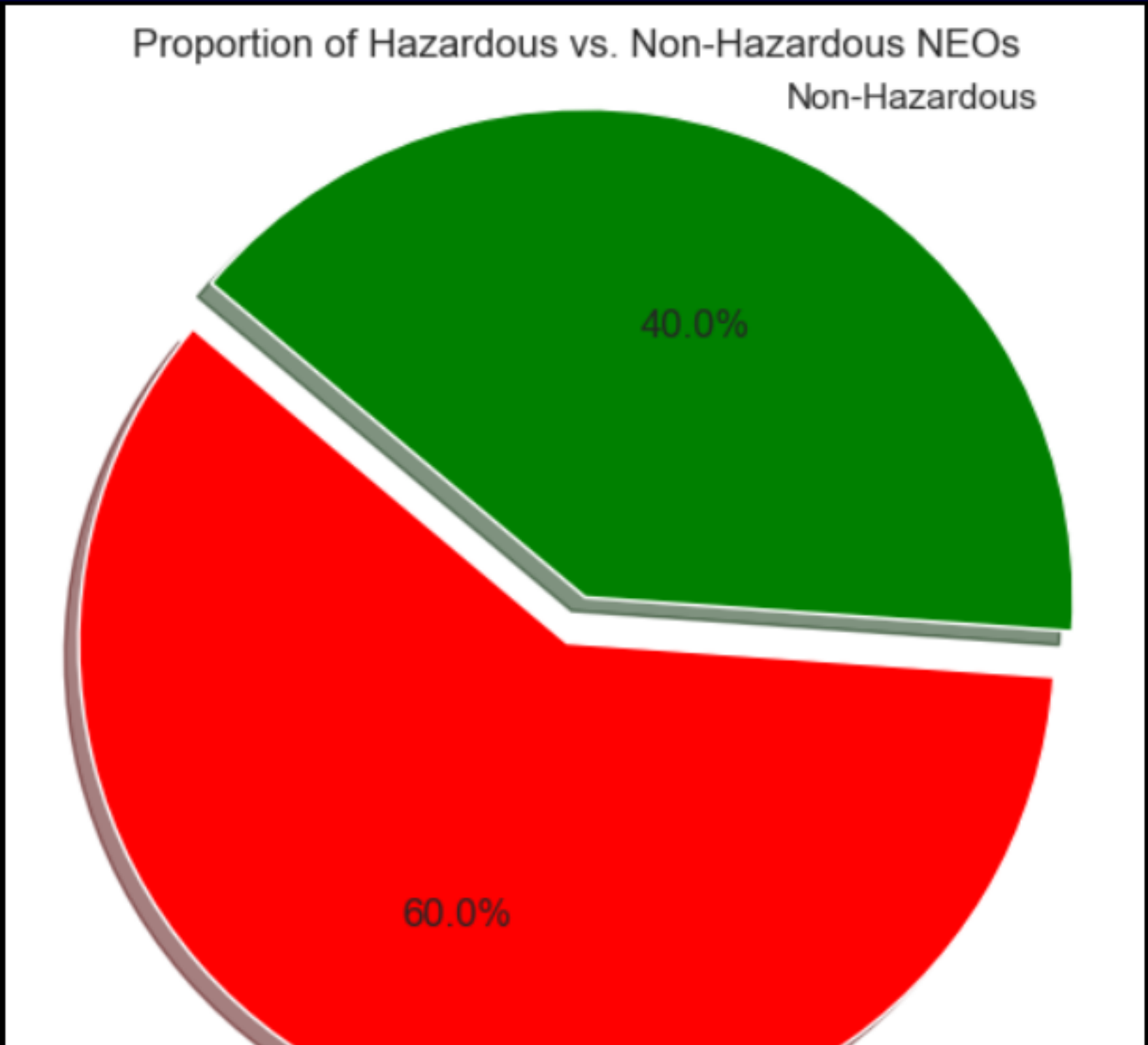

Number of NEOs Per Week
Number of NEOs
9
8
7
6
5
4
3
2022-01-04
2022-01-11
2022-01-18
2022-01-25
Week

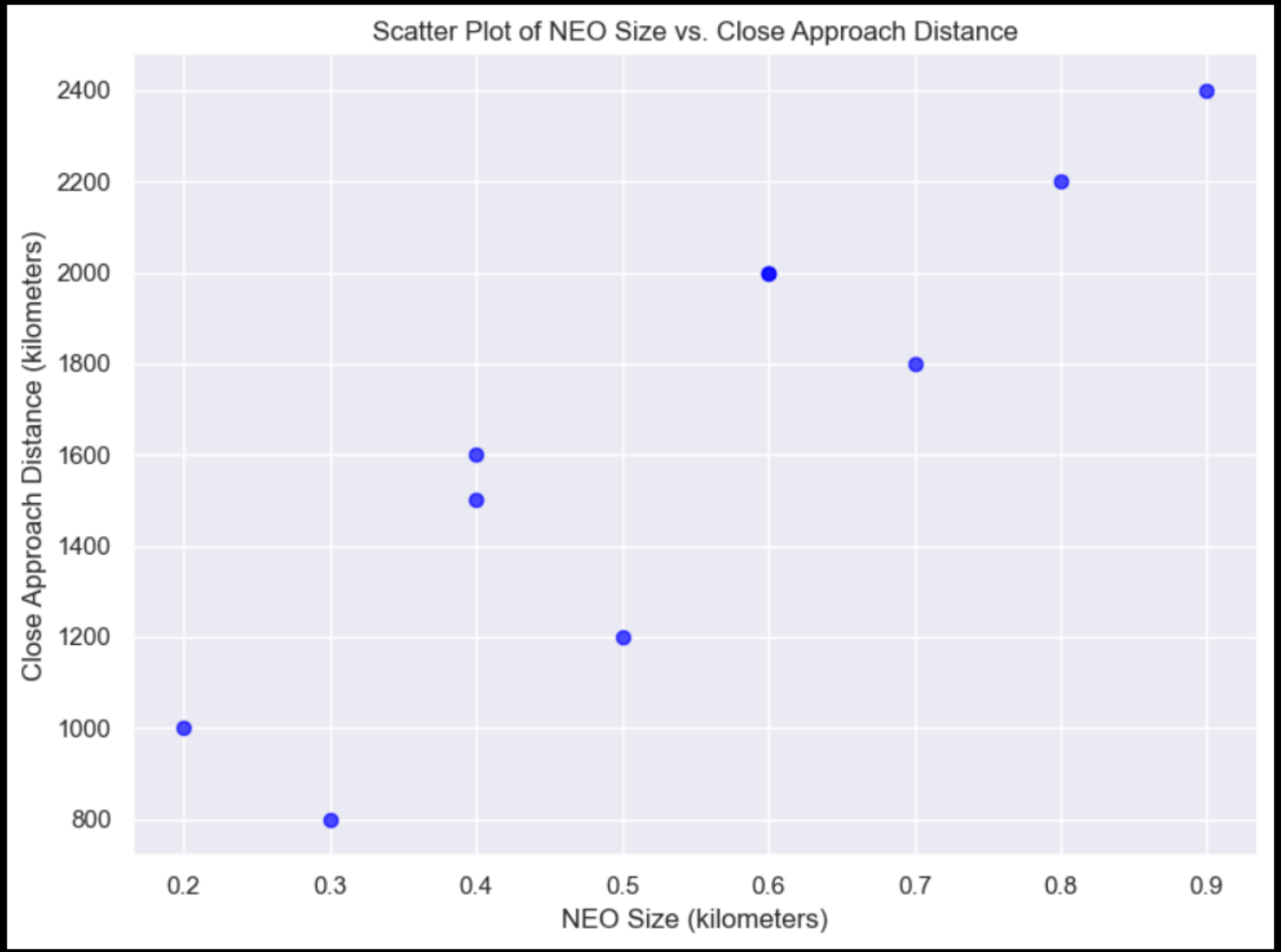

Distribution of NEO Sizes
Frequency
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
NEO Size (kilometers)

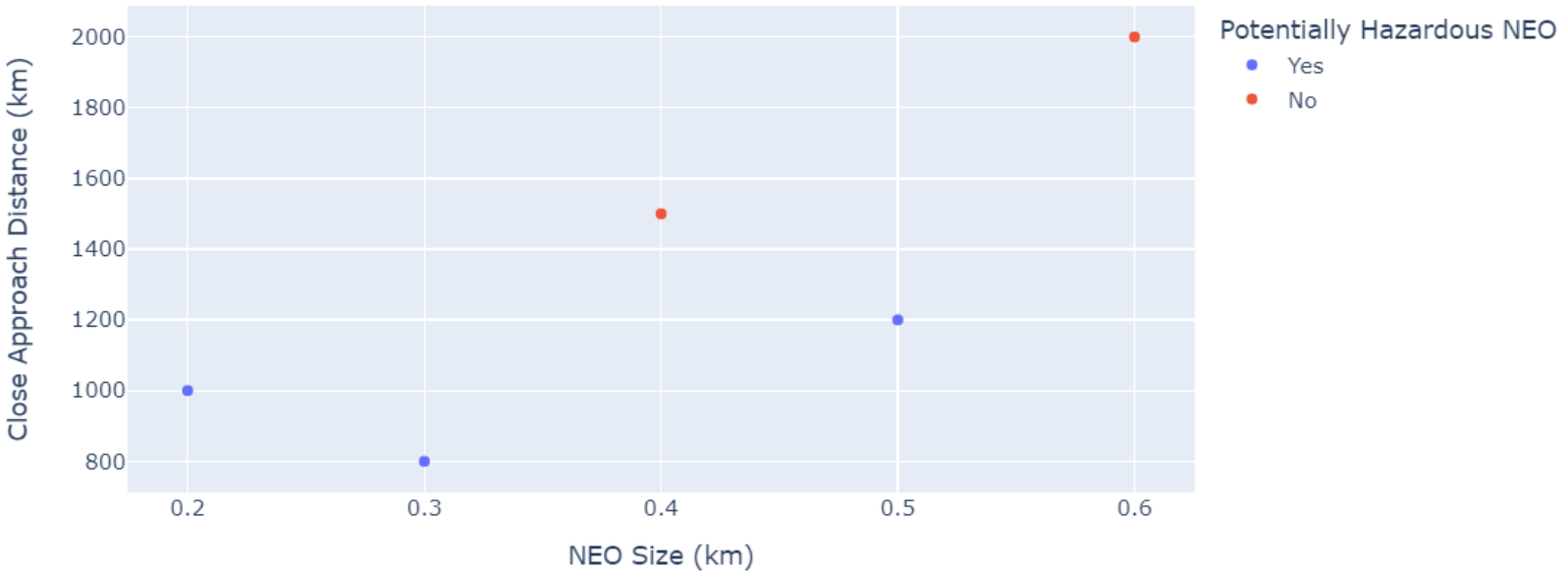

Heatmap of Number of NEOs per Week
Week
01 03 05 07 09 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51
2022
2023
Year
7
6
5
4
3
2
1
0

Task 4: Data Visualization Part B (NASA)

- Create a pie chart of the proportion of hazardous vs non-hazardous NEOs.
- Create a scatter plot of the correlation between NEO size and close approach distance.
- Customize the appearance of your plots (e.g., colors, labels, titles).
- Create interactive visualizations using a library like Plotly. For example, create an interactive scatter plot where you can hover over each point to see more information about the NEO.

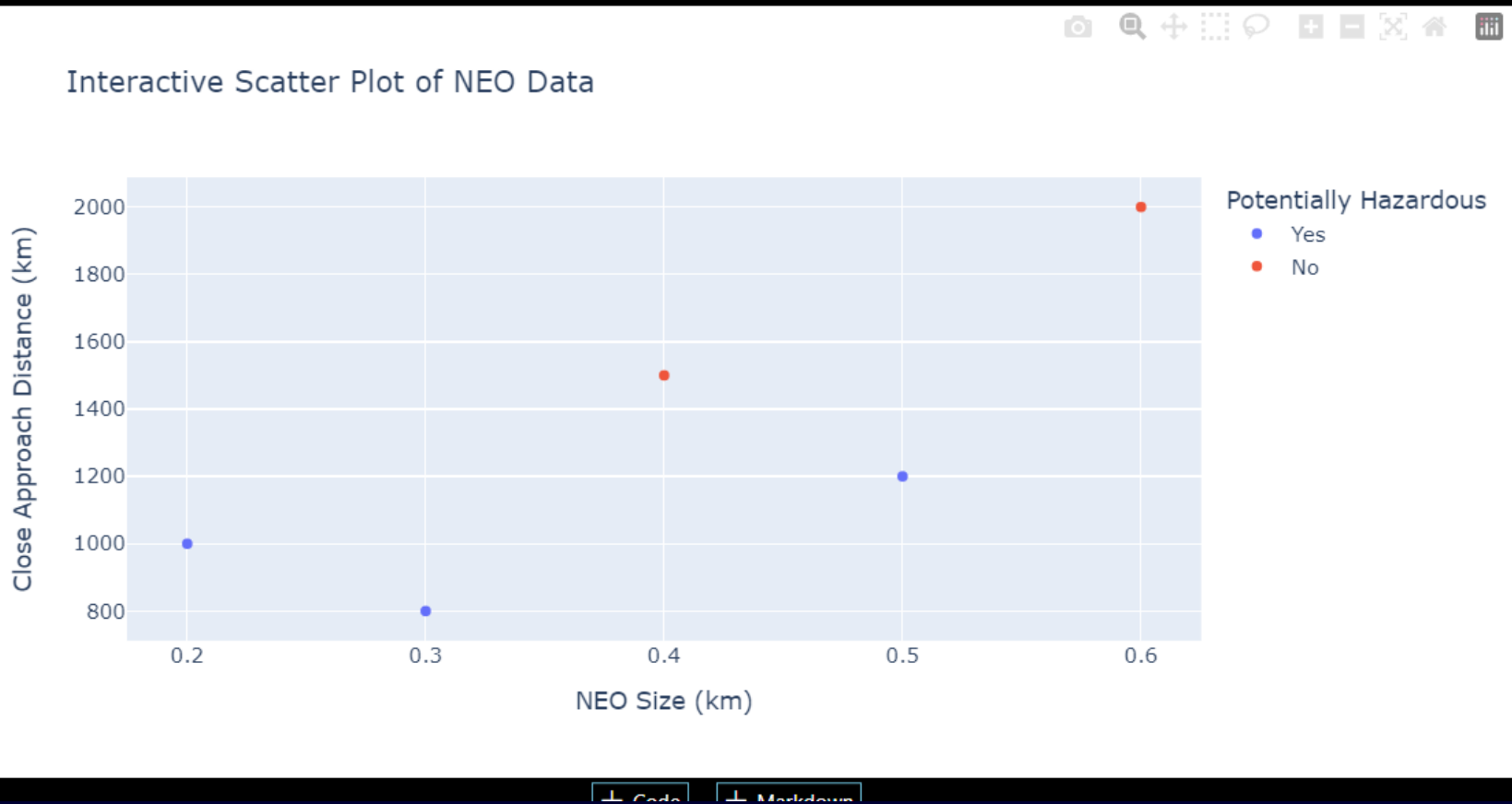

Proportion of Hazardous vs. Non-Hazardous NEOs
Non-Hazardous
40.0%
60.0%


Scatter Plot of NEO Size vs. Close Approach Distance
Close Approach Distance (kilometers)
2400
2200
2000
1800
1600
1400
1200
1000
800
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
NEO Size (kilometers)


Interactive Scatter Plot of NEO Data
Close Approach Distance (km)
2000
1800
1600
1400
1200
1000
800
0.2
0.3
0.4
0.5
0.6
NEO Size (km)
Potentially Hazardous NEO
Yes
No

Task 5: Interpretation of Results (NASA)

- Interpret the results of your data visualization in part A and B.
- What insights can you gain about NEOs from your results? Summarizing your main findings with a single visualization plot.


Interactive Scatter Plot of NEO Data
Close Approach Distance (km)
2000
1800
1600
1400
1200
1000
800
0.2
0.3
0.4
0.5
0.6
NEO Size (km)
Potentially Hazardous
Yes
No