Data visualization

Outline

- Data analysis of NYC Taxi Rides
  - Understanding the data
  - Exploratory data analysis
  - Temporal Analysis
  - Time-Series forecasting

- NASA Data collection & visualization
  - Data Analysis
  - Data Visualization, Part A & B

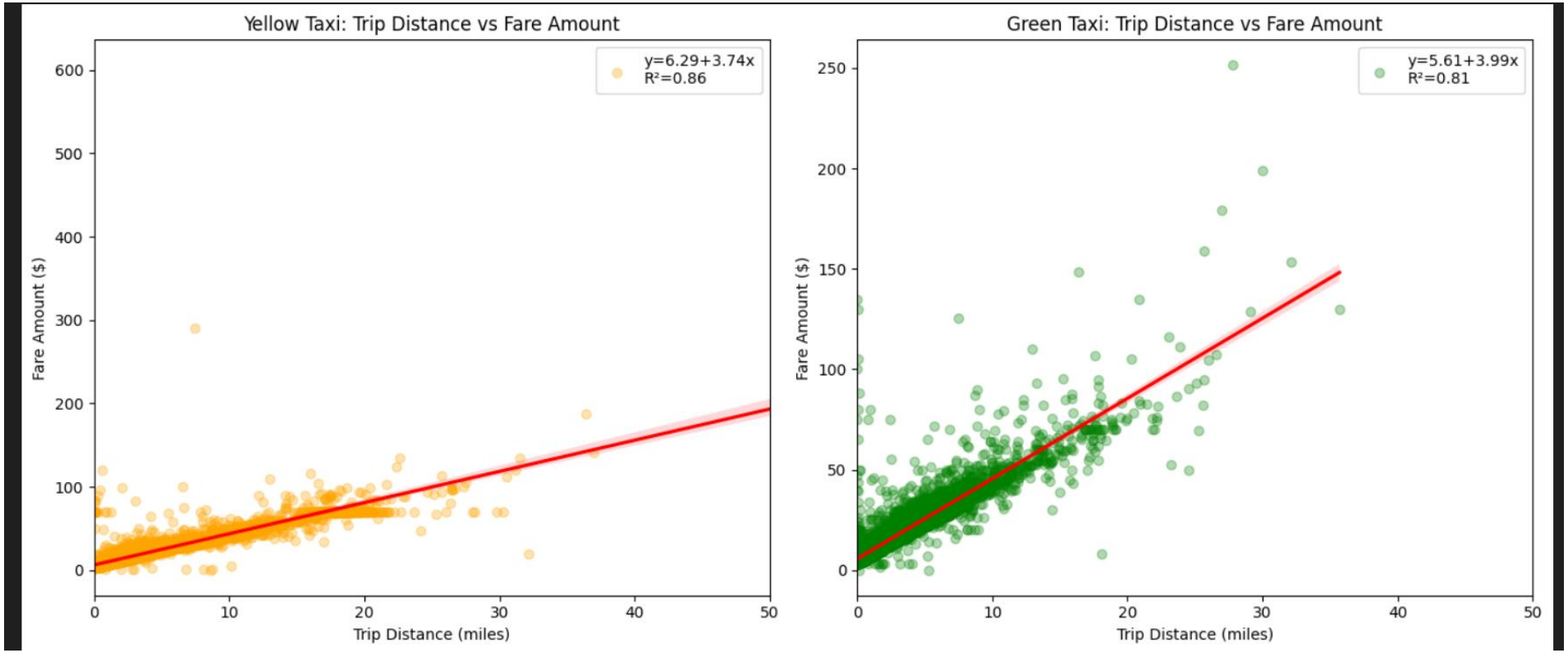
Understanding the data

NYC taxi rides

- **Using .head() reveals the data structure**
  - **Shows the top part of the data**
- **We have two seperate datasets, yellow taxi is much larger than green taxis**

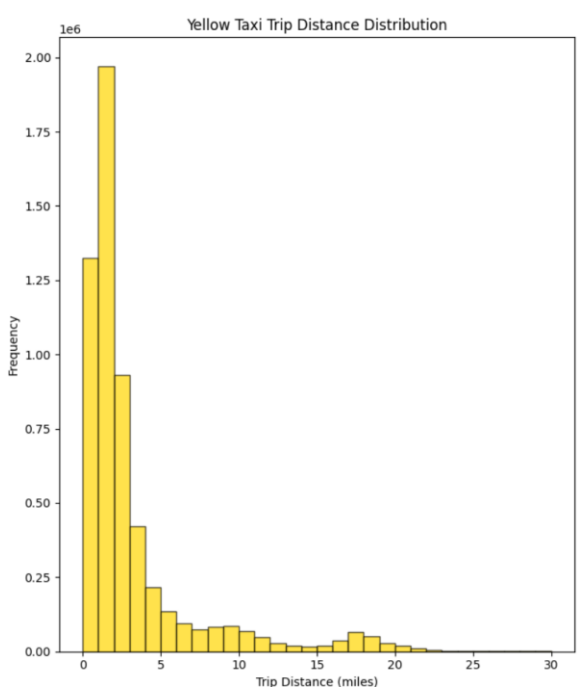
Relationship

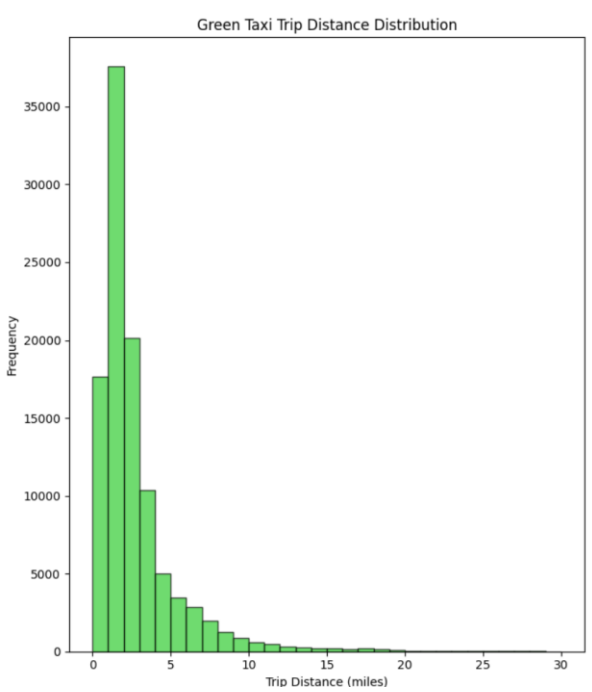
- Fare amount and trip distance is related.

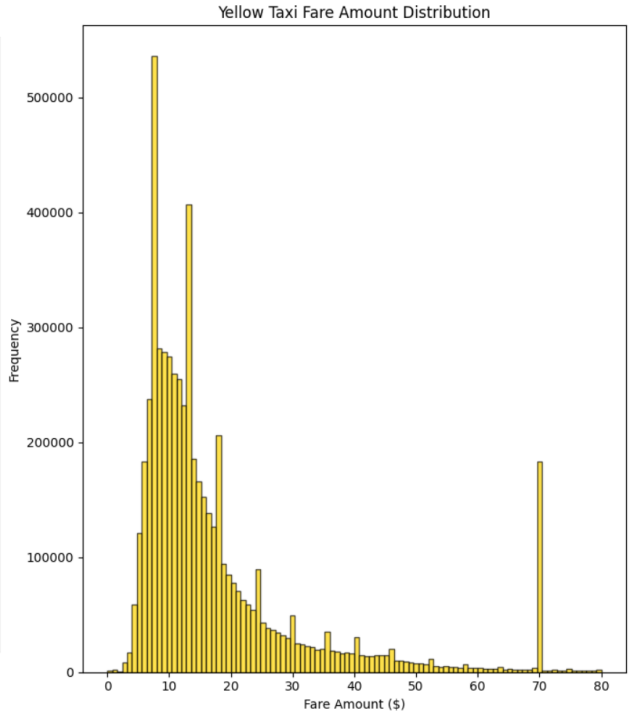


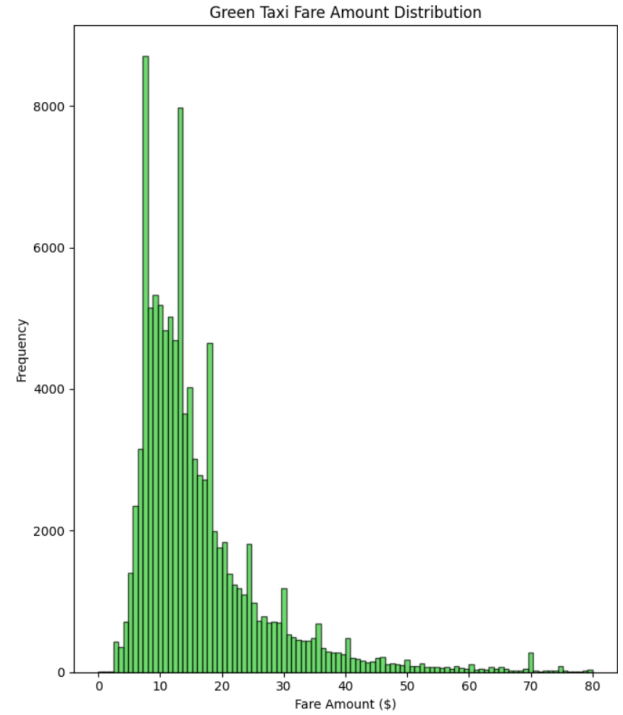
Trips

- Fare amount is fairly low $10 <
- Most trips are also fairly short.



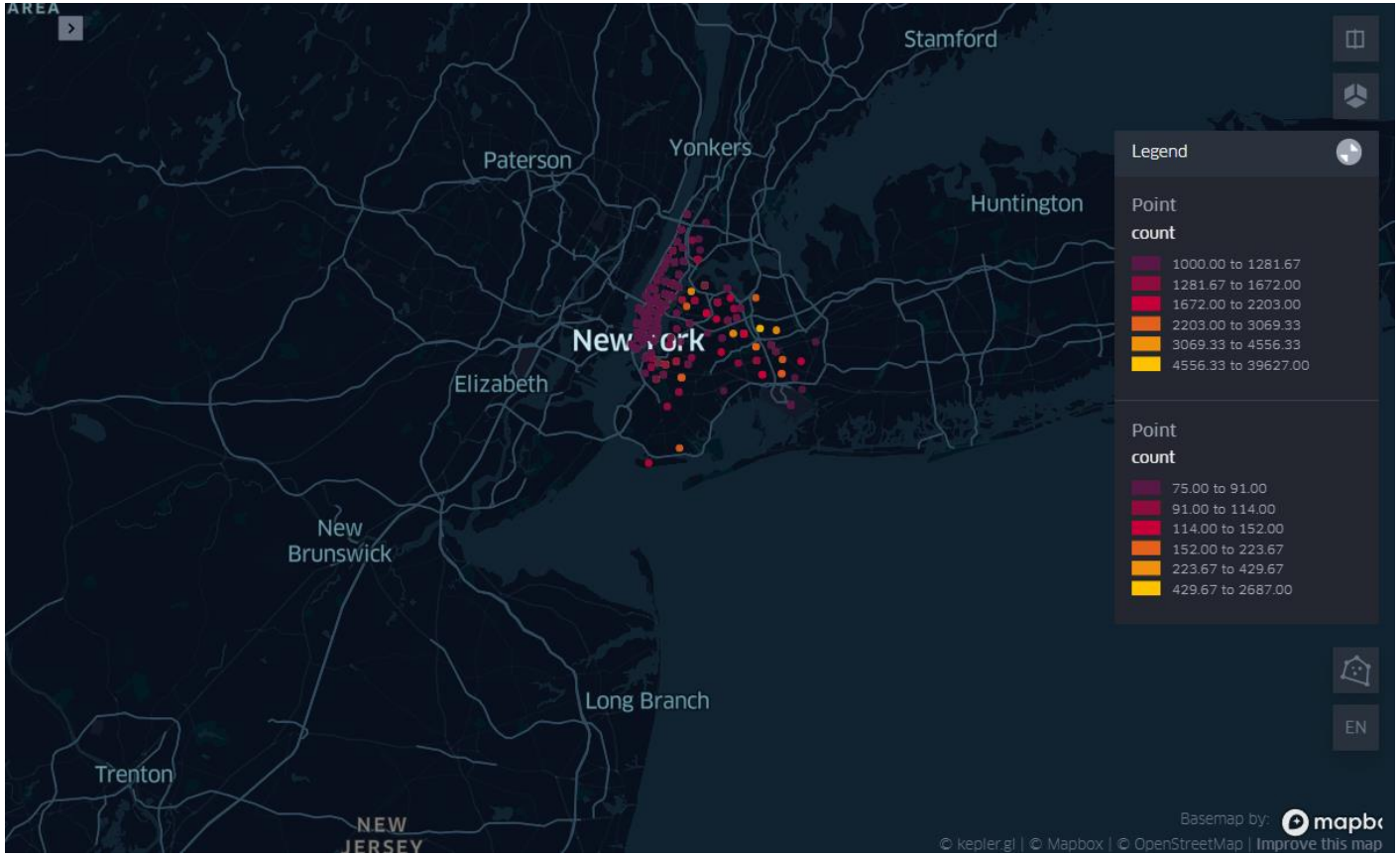






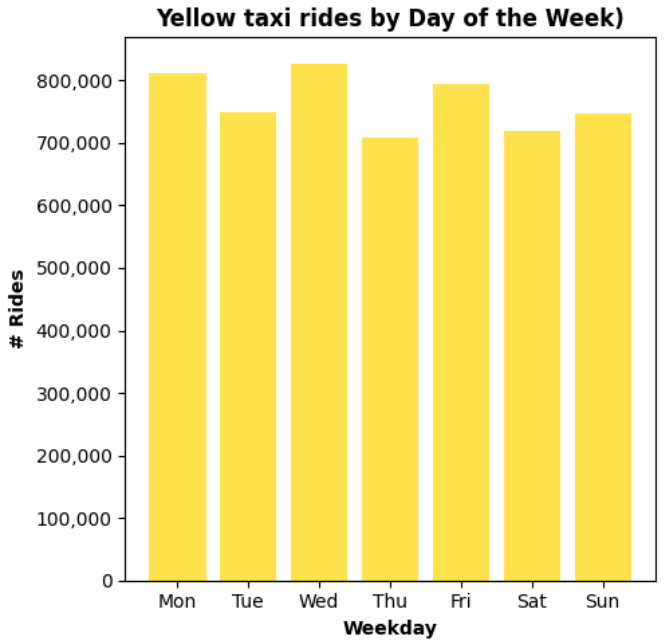
3D Data visualization

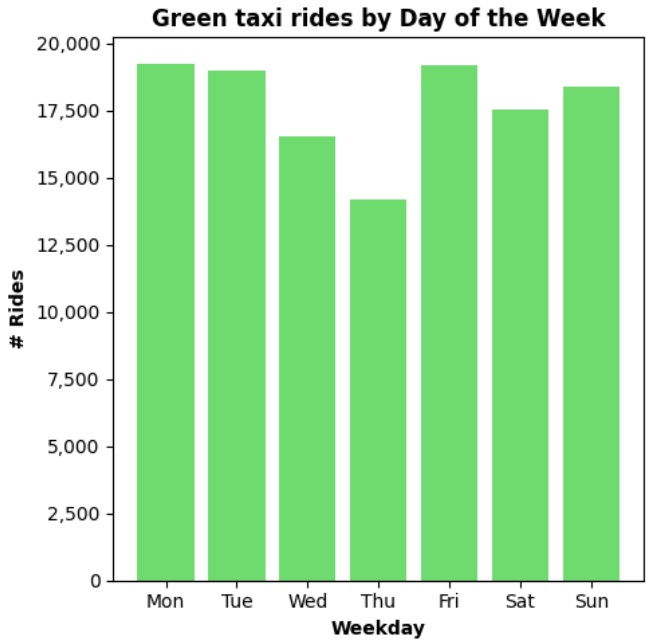
- Interactive Map



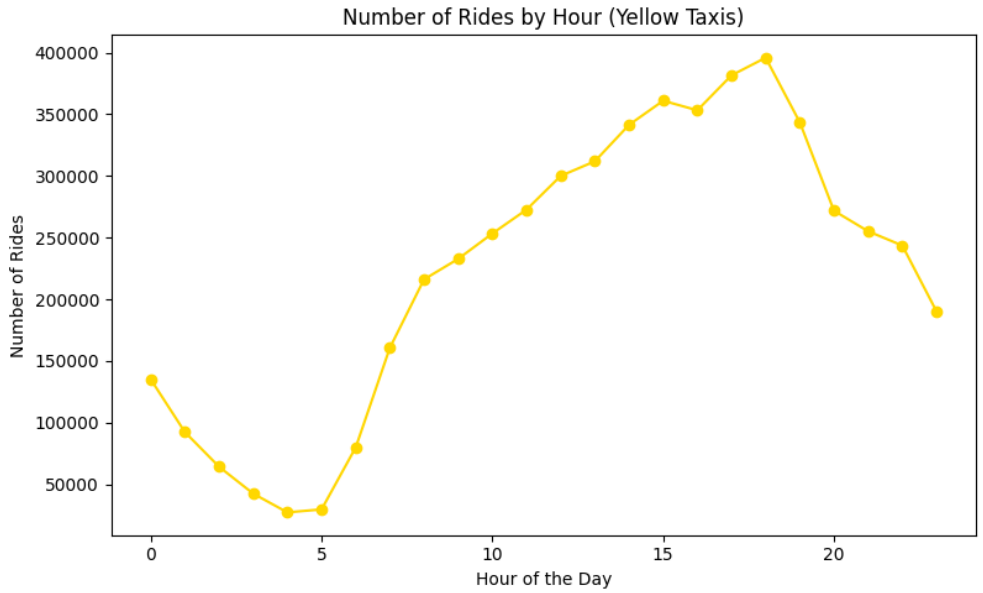
Taxi rides by day of the week yellow and green

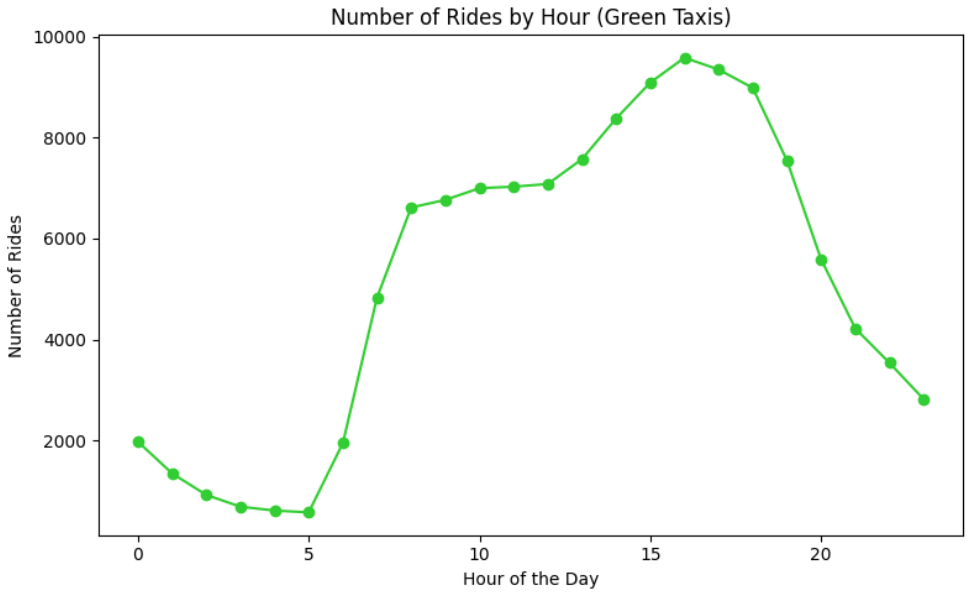
- It’s evident that Yellow Taxi are much more common than the green taxi.
- Thursday is the most busiest day for yellow taxis, while it’s the least busiest day for green taxi





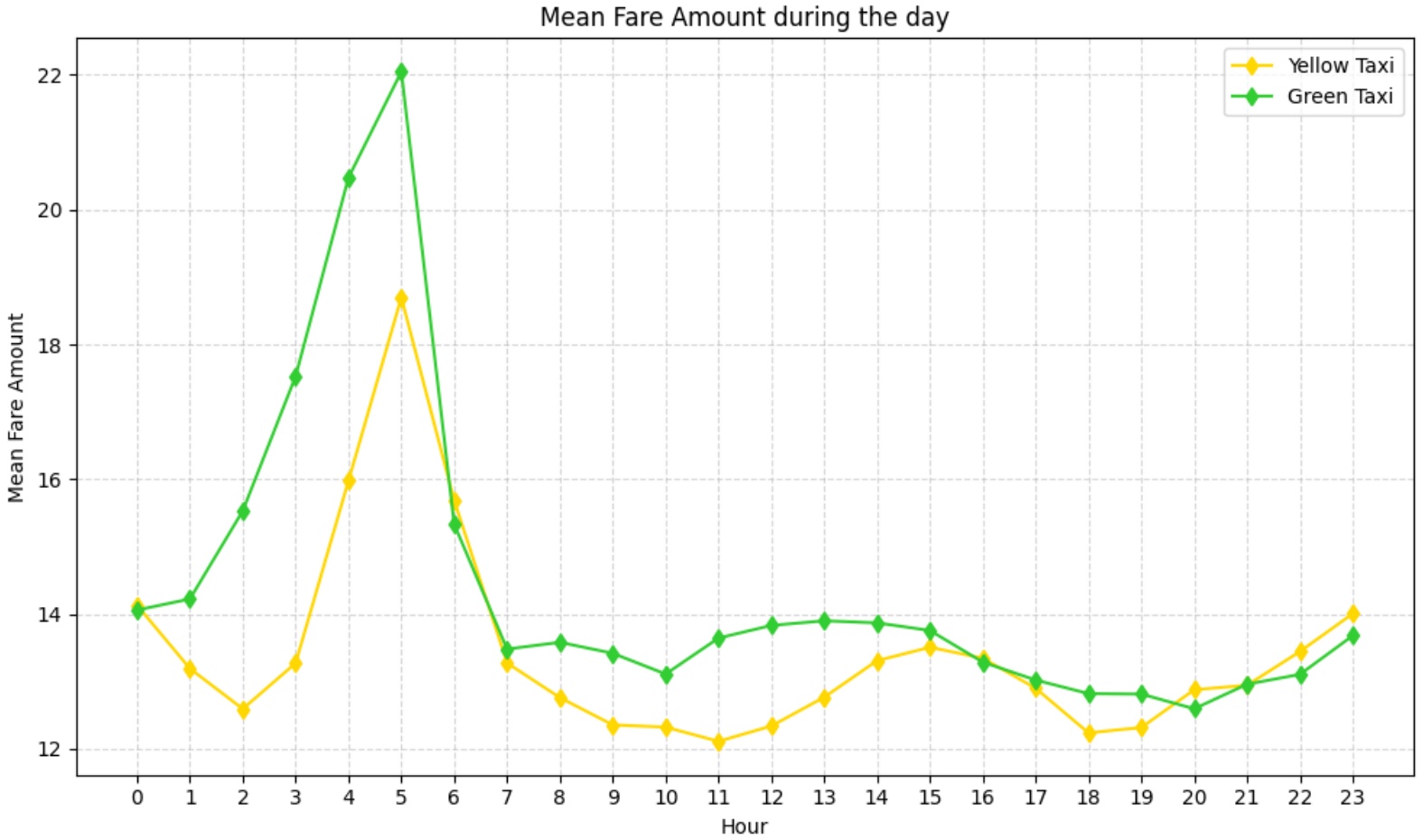
Rides during the day





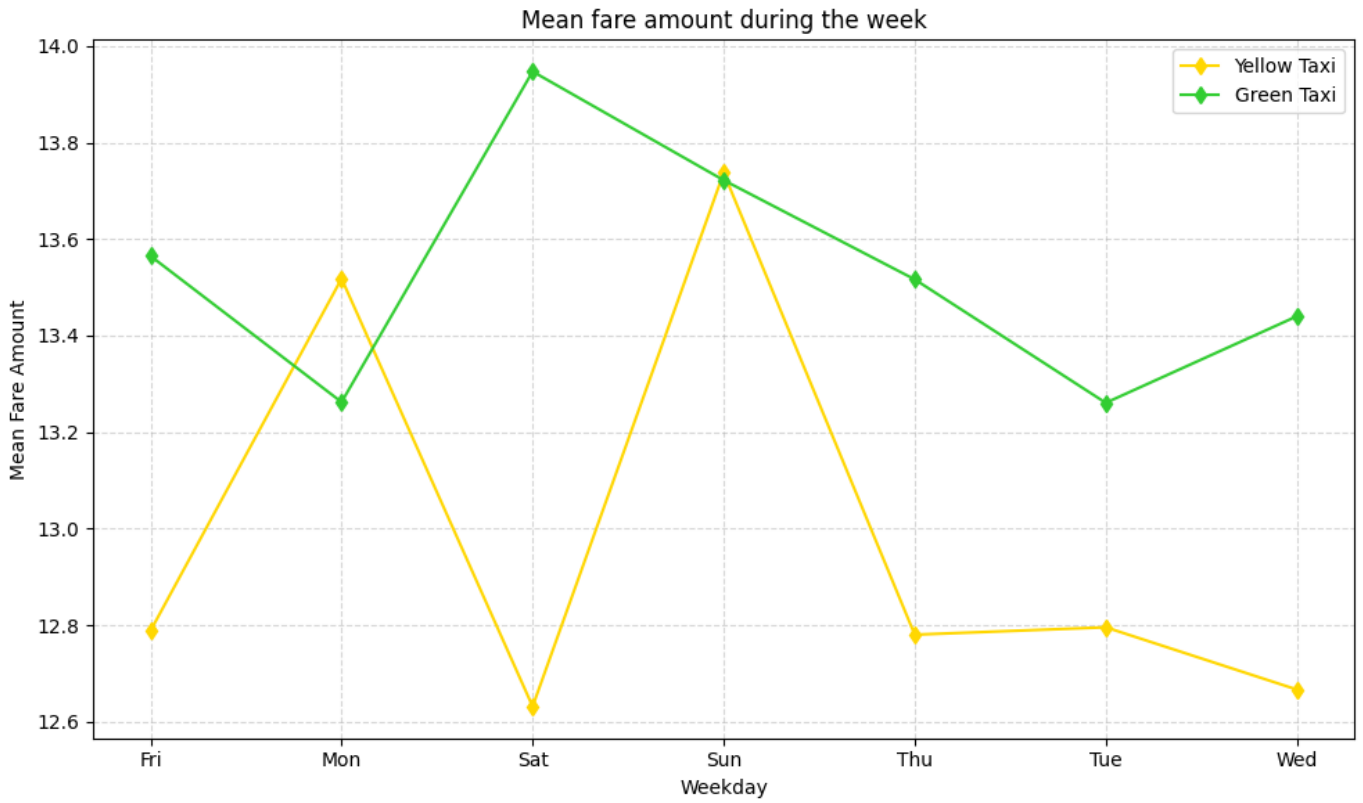
Mean Fare during the day

- While there is a lower number of trips during midnight and 6AM, the mean fare is much higher.

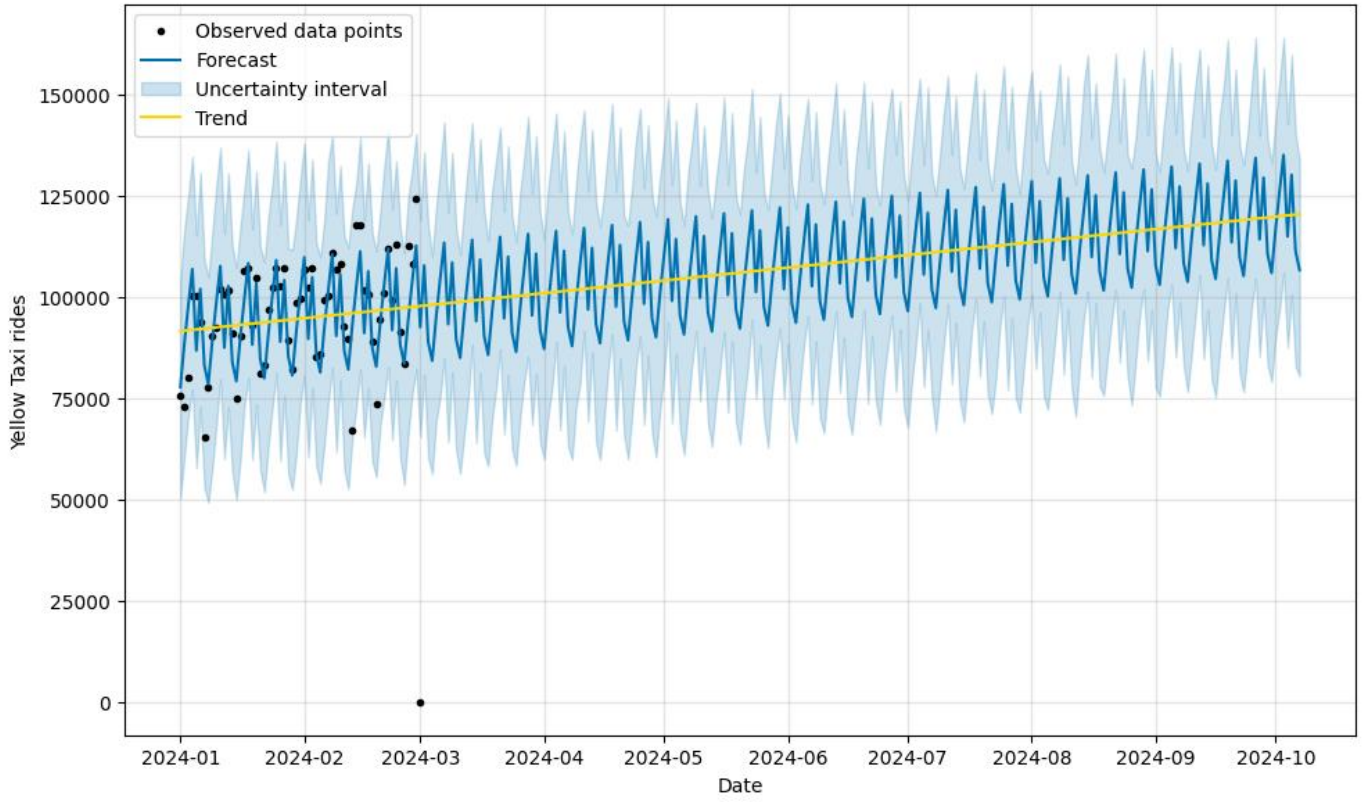


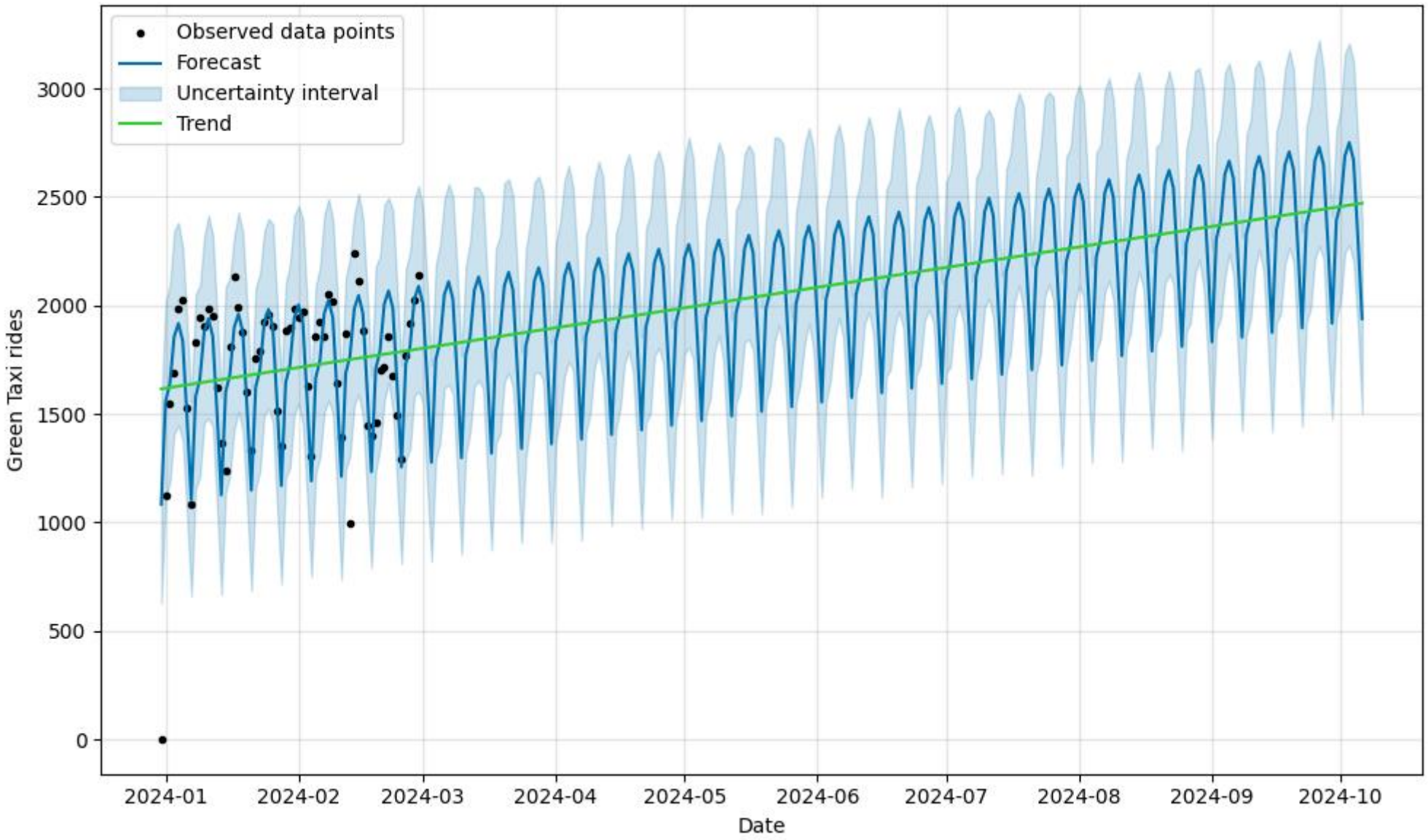
Mean fare during the week

- Saturday and Sunday brings most revenue.
- From previous graph, we also understood fares are much higher between midnight and 6 AM



Forecasting




Observed data points
Forecast
Uncertainty interval
Trend
Green Taxi rides
3000
2500
2000
1500
1000
500
0
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06
2024-07
2024-08
2024-09
2024-10
Date

Project NASA

NEO data and potential hazards

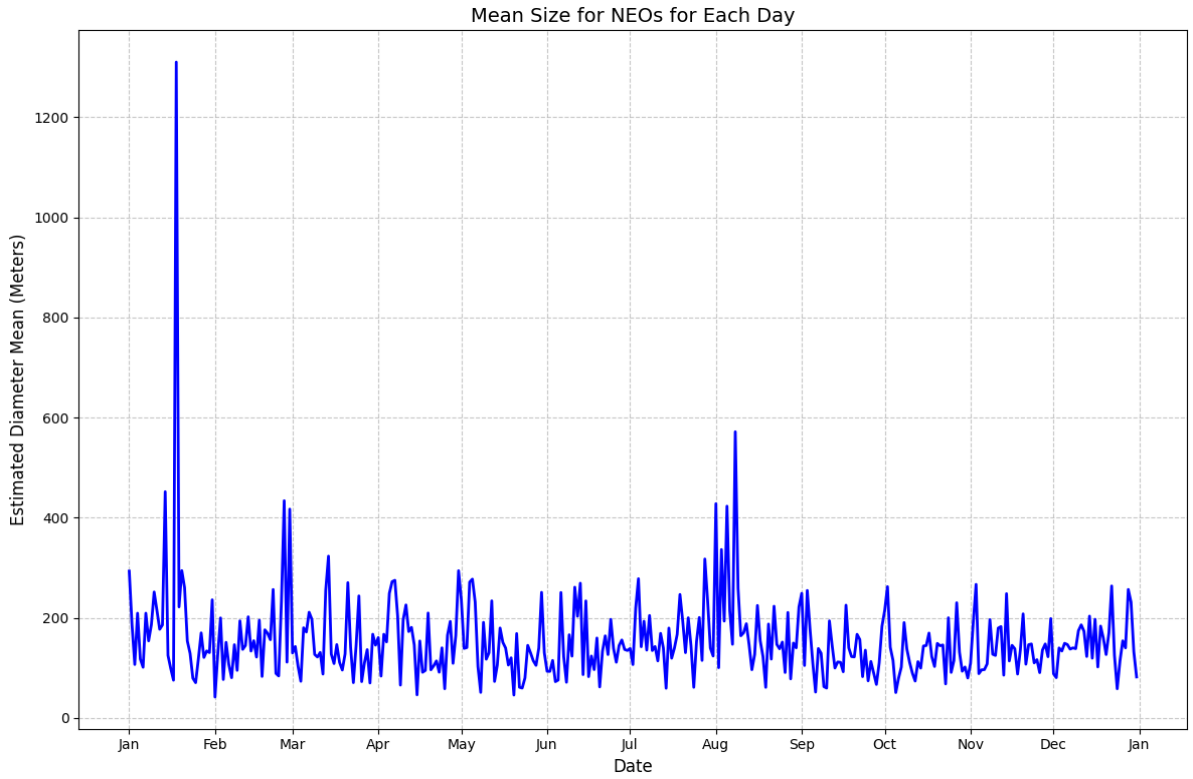
- In order to analyze the data we wish to examine, the average size of NEOs per day is calculated. Data was plotted using the principle of 'choosing an effective visual

Calculating mean size from data

```
df.groupby("date")["estimated_diameter.meters.estimated_diameter_mean"].mean()
```

[11] ✓ 0.0s

```
date
2023-01-01    293.252253
2023-01-02    185.803280
2023-01-03    106.628617
2023-01-04    209.134181
2023-01-05    117.550816
                 ...
2023-12-27    139.824445
2023-12-28    256.643697
2023-12-29    231.144942
2023-12-30    132.109357
2023-12-31     81.721031
Name: estimated_diameter.meters.estimated_diameter_mean, Length: 365, dtype: float64
```

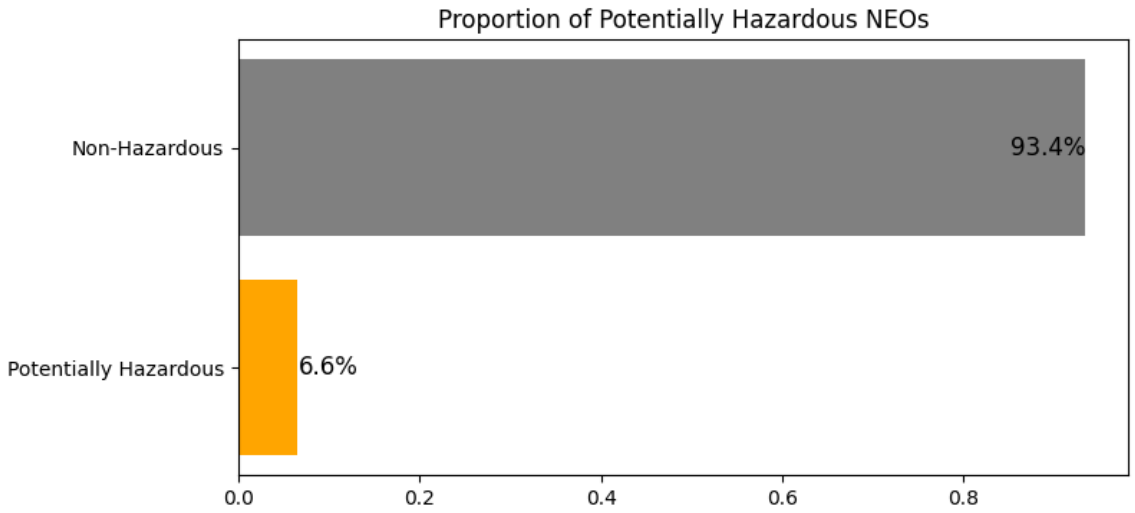


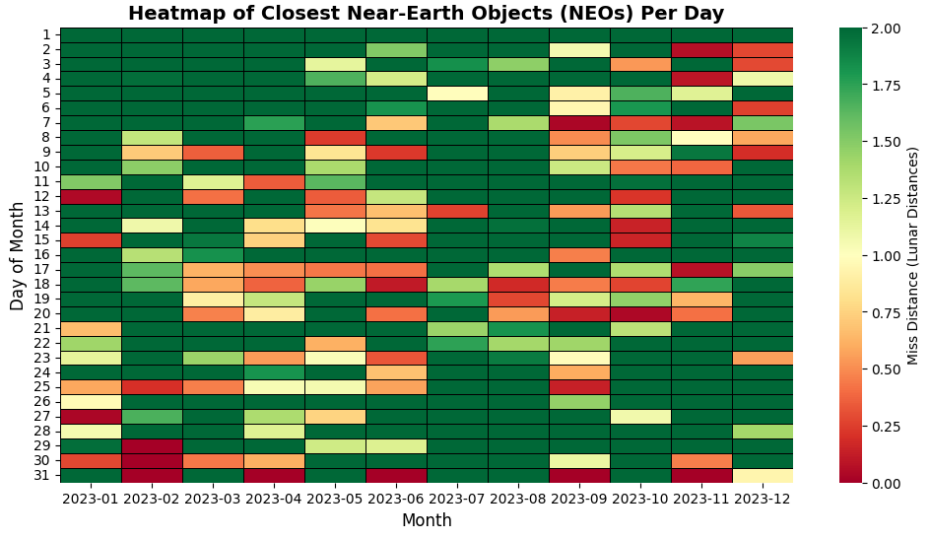
Understanding the hazards

Barchart over the proption of potentionally harzadous NEOs and heatmap of closest near- earth objects.

Green = Far away, Safe

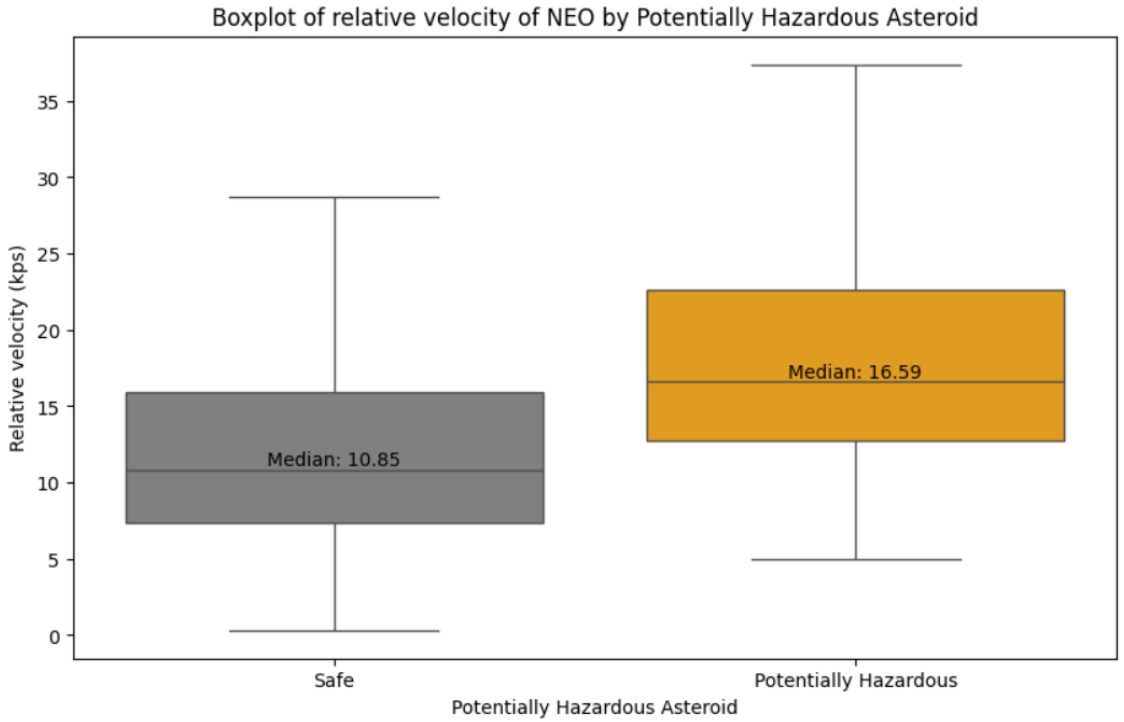
Red = Close, hazard

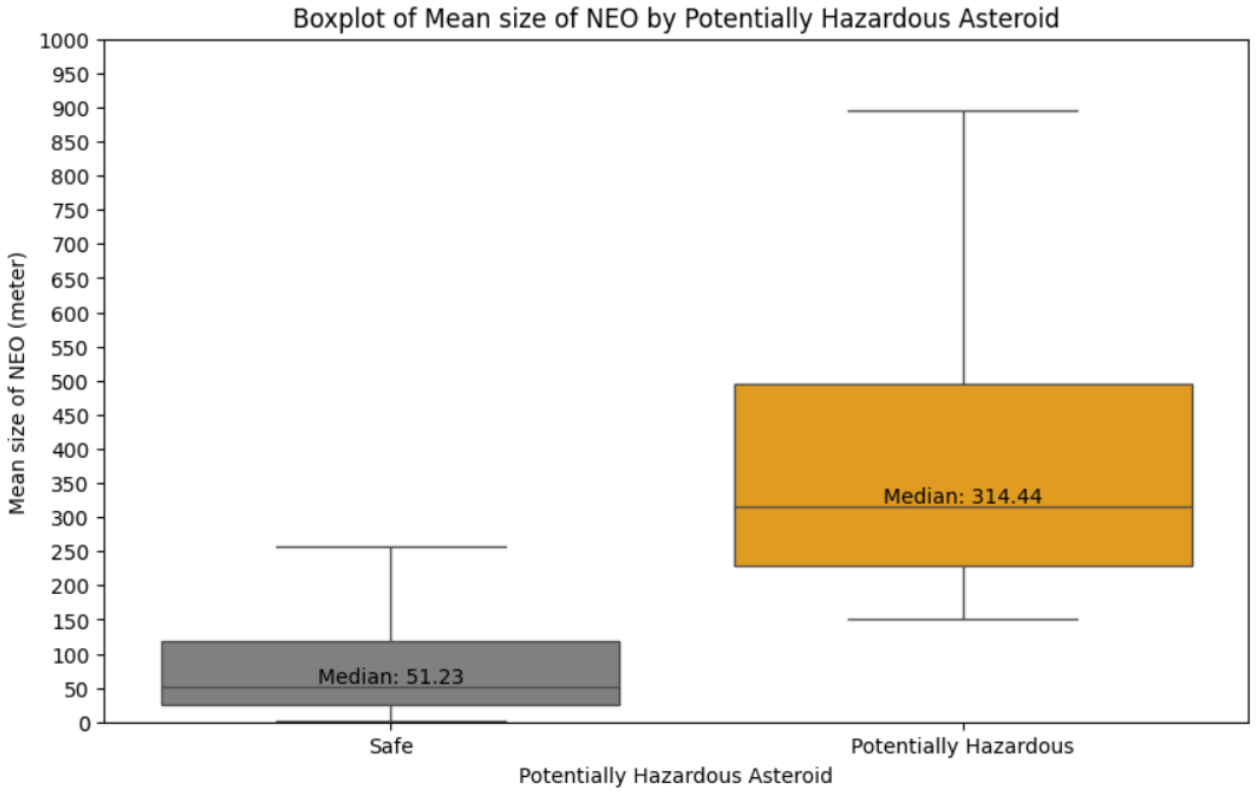




Size of hazardous NEO’s

- Median is 6x times higher
- Velocity is 0.6x higher
- Hazardous NEO’s are both bigger and faster than non-hazardous NEO’s





NEO’s frequency per week

- Fairly low between week 21 and 33