

DTU



Group 16

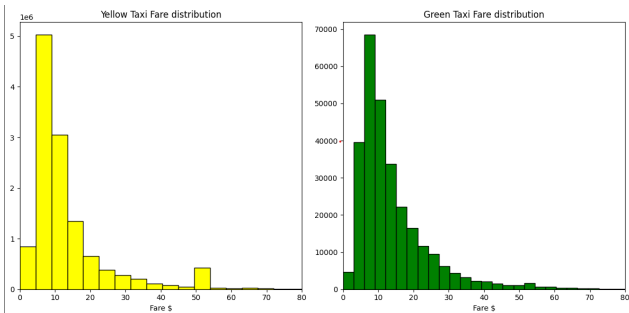
Data Visualization and Analysis

Project 1 - Taxis in New York City

Understanding the data

- Two taxi companies operating in New York City
- January 1, 2022, to April 30, 2022.
- Only trip under 1000 miles
- fare amounts less than \$1000

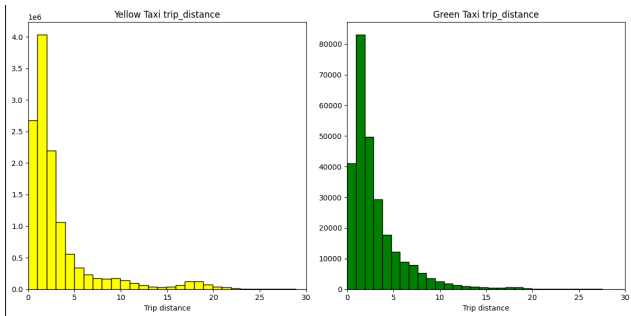
Exploratory Data Analysis 1



Fare distribution

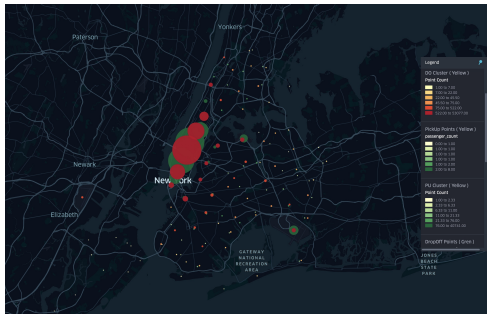
- This plot shows the distribution of taxi fares for two different types of taxis: **yellow taxis** and **green taxis**.
- The data used for the plot has been filtered to include fares between 0 and 1000.

Exploratory Data Analysis 2



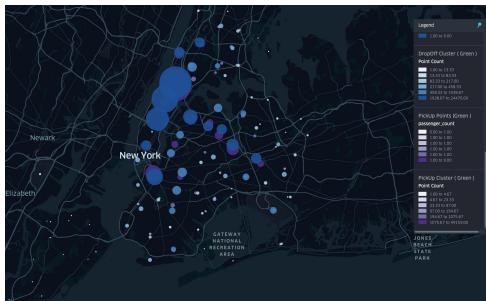
Trip Distance

In each subplot, a histogram is used to visualize the distribution of trip distances. The x-axis represents the trip distances in miles, while the y-axis represents the frequency or count of trips falling within each distance range.



Yellow Taxi PickUps and DropOffs

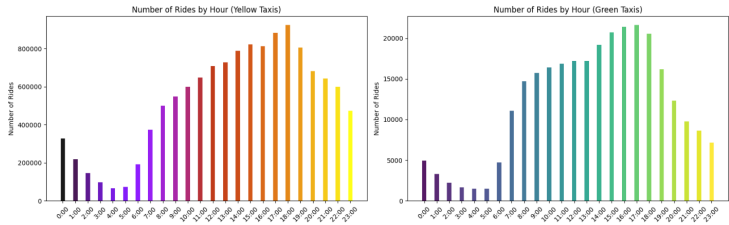
- Our data shows, that the majority of pickups and dropoffs, are around city center.
- The data shows at least two points outside city center, where more people are being dropped off then picked up.



Green Taxi Pick Up and Drop Off

- Like the data from the yellow taxi, we can also see here that, the majority of the pickups and dropoffs, happens inside NYC center.
- But where the data shows a difference is in the amount of pickups and dropoffs outside of NYC Center.

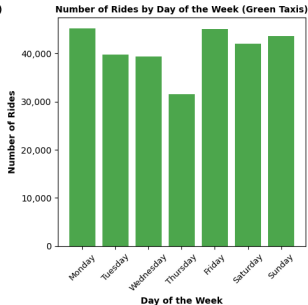
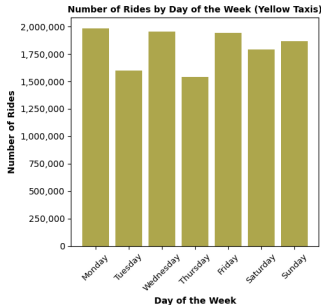
Data analysis 1



Rides by hour

The first plot shows the number of rides by hour for yellow taxis, and the second plot shows the number of rides by hour for green taxis. The x-axis represents the hours of the day, and the y-axis represents the count of rides.

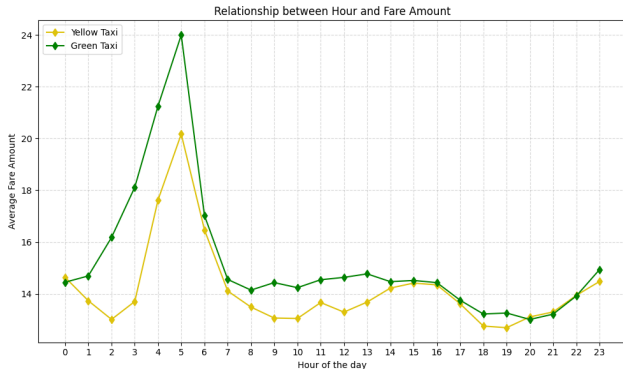
Data analysis 2



Rides by day

These plots show the number of rides by day of the week for yellow taxis and green taxis, respectively. The x-axis represents the days of the week (Monday to Sunday), and the y-axis represents the count of rides. Each bar in the plot represents the number of rides for a specific day of the week.

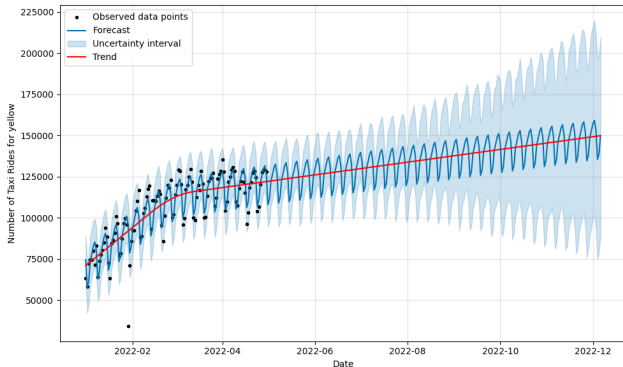
Data analysis 4



Relationship between Hours and Fare Amount

These plots show the average trip distance for different time periods for yellow taxis and green taxis, respectively. The x-axis represents the time periods, and the y-axis represents the average trip distance.

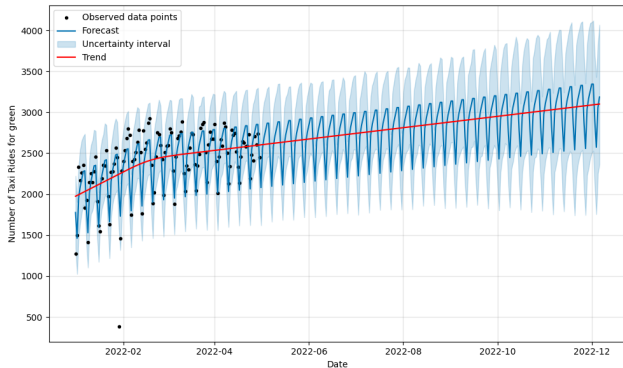
Time-Series Forecasting 1



Time-Series Forecasting yellow

This plot shows the forecasted number of **yellow** taxi rides over time. The plot includes a trend line (in red) that represents the expected overall pattern or direction of the data.

Time-Series Forecasting 2



Time-Series Forecasting green

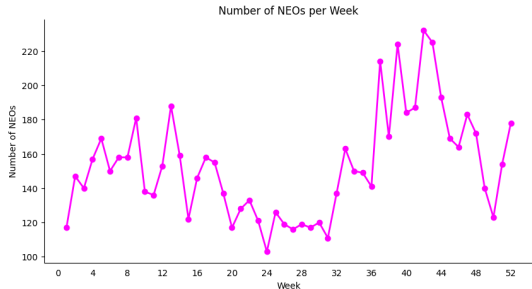
This plot shows the forecasted number of **green** taxi rides over time. The plot includes a trend line (in red) that represents the expected overall pattern or direction of the data.

Project 2 - NASA and Near-Earth Object.

Understanding the data

- Using an API
- A year from January 1st 2022.
- Using Pandas data frame
- Behavioral patterns of NEOs

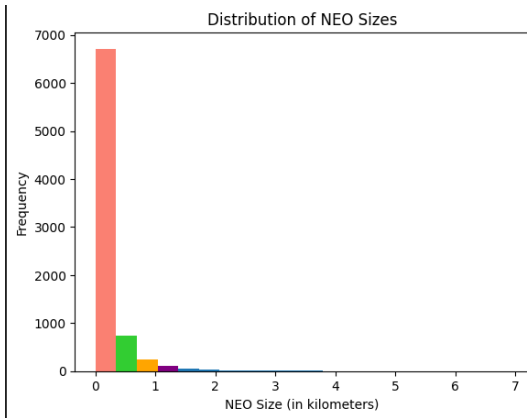
Data Visualization Part A



Line plot of the number of NEOs

- It's a line plot, which shows how many NEO there are per week.
- The Number of NEOs is highest between week 36 to week 45, indicating a peak in NEO activity from early September to mid-November. This could suggest a seasonal trend or specific astronomical events during that period.

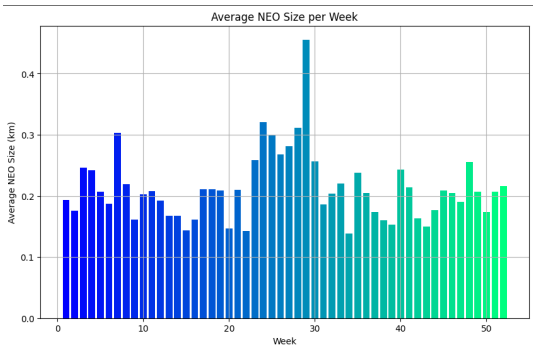
Data Visualization Part A



Histogram of the distribution of NEO sizes

- The majority 4/5 parts of NEOs fall into the category of sizes less than 500 m diameter.
- The largest NEO observed measuring approximately 7 kilometers for diameter.

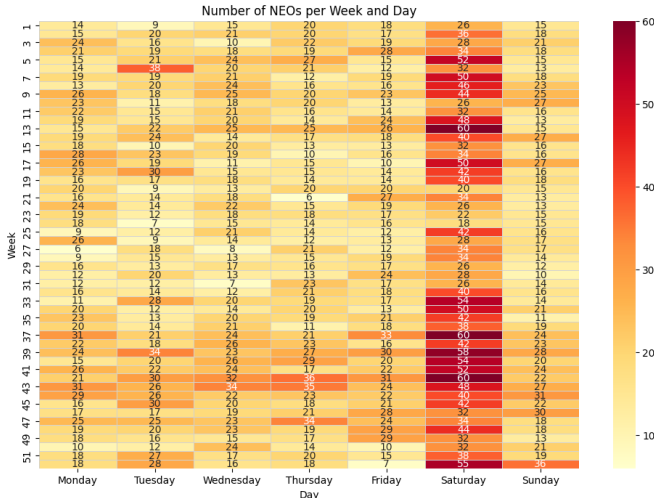
Data Visualization Part A



Bar plot of the average NEO size

- Overall, there is a more even distribution of data in the different weeks
- The average NEO diameter size peaked at around 0.48 kilometers during week 29. This suggests that there was a relatively larger NEO present during that time, potentially indicating an anomalous event or a specific NEO of significant size.

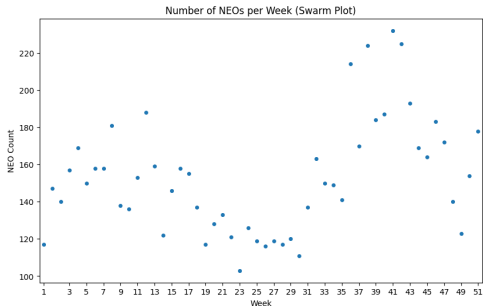
Data Visualization Part A



Heat map of NEO

- A heat map of the number of NEOs per day per week.
- is a graphical representation of data in which values are displayed as colors
- Higher values are usually represented by "warmer" colors such as red

Data Visualization Part A

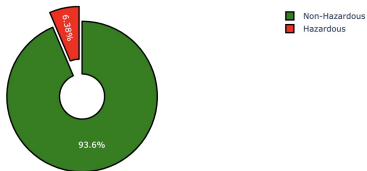


Swarm Plot

- The individual dots represent the number of NEOs in the different weeks
- It makes it possible to see where the trends are in the data and get an overview of where the data is lumped together or has diversion
- Quickly identify patterns and outliers in the data

Data Visualization Part B

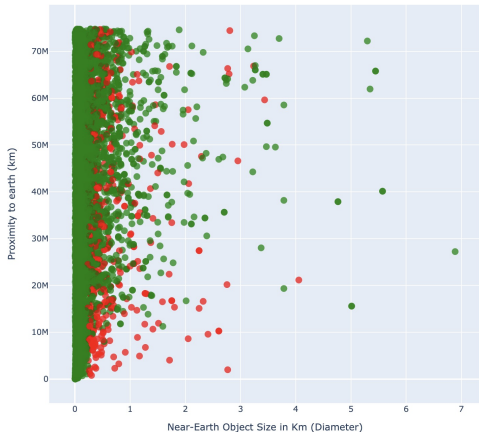
Proportions of Hazardous & Non-Hazardous Near-Earth Objects.



Proportions of hazardous Neo's

- The pie chart depicts the composition of Neo objects into two categories: non-hazardous and hazardous.
- Non-hazardous Neo's account for a significant majority, comprising 93.6% of the total.
- The remaining portion represents the hazardous Neo objects.

Data Visualization Part B



Size-Distance Relationship and Hazard Classification of Neo Objects

- The scatterplot illustrates the relationship between Neo size and proximity to Earth.
- The color-coded points allow identification of hazardous and non-hazardous Neo objects.