DSCI 510: Final Report

1. What is the name of your project and who is in the team? Please describe it as a research question and provide a short description.

Name: Whew… That Was Close
Team: Raquel Haddad
Research Question: What do we know about Near-Earth Objects (NEOs) and does the media update the world on new discoveries?

NEOs are comets and asteroids with a perihelion distance q, their closest distance from the Sun, less than 1.3 au (astronomical units). 1 au is roughly equal to 150 million kilometers. The vast majority of NEOs are asteroids, as comets are restricted to having an orbital period P less than 200 years. These NEOs have been nudged by the gravitational attraction of nearby planets into orbits that allow them to enter the Earth's neighborhood. As they orbit the Sun, NEOs occasionally approach close to Earth. CNEOS (Center for Near Earth Object Studies) at NASA calculates the motion of all NEOs forwards to 2200 A.D. and backwards to 1900 A.D., and determines the times and distances of the Earth close approaches.

In my project, I aim to answer my research questions by using data from the Jet Propulsion Laboratory (JPL) at NASA, specifically from the Small-Body Database API, and Google News articles. I will analyze asteroid data, including physical characteristics and orbital elements, and cross-reference new articles with NASA asteroid data to understand the diversity of asteroids in our solar system and if we are kept up to date with new discoveries.

The data in this API for all asteroids and comets in JPL's SBDB are dynamically generated, as its default query parameters will return NEO close-approaches less than 0.05 au in the next 60 days sorted by date. The user can update the parameters depending on what records they would like to see returned.

2. What data did you collect? How did you collect it? How many data samples did you collect?
    a. Specify exact data sources and your approach
    b. Describe what has been changed from your original plan, what challenges you encountered or resolved.

Asteroid Data
Source: https://ssd-api.jpl.nasa.gov/doc/cad.html
Other References: https://cneos.jpl.nasa.gov/ca/ &
https://ssd.jpl.nasa.gov/tools/sbdb_lookup.html#/
Approach: I used the `requests` library to query data. I included information on diameter, absolute magnitude, distance, relative velocity, and orbital elements (eccentricity, semi-major axis, inclination, longitude of ascending node, argument of perihelion, and mean anomaly) for asteroids in the database.

Data Samples Collected: 30+
For simplicity reasons, I kept the default parameters that will return data into the next 60 days. In my code, the user has the opportunity to change these parameters and supplement however many of their own small body IDs to display data on specific asteroids.

News Data
Source:
https://news.google.com/search?q=nasa%20asteroid&hl=en-US&gl=US&ceid=US%3Aen
Approach: I accessed HTML content using `requests` and parsed the data with `BeautifulSoup`.
Data Samples Collected: 10

From my initial project proposal, my purpose was to use the data I collect to warn scientists of asteroid impact risks to Earth. I slightly deviated from this because I thought it would be interesting to delve into how the media currently portrays these risks in news articles. Additionally, I did not end up creating any of the visualizations I stated in my project proposal. I initially said that I would make orbit class distributions, heat maps/risk matrices, and time series plots, but I decided to not produce these.

My data retrieval steps when using the API were fairly straightforward. I had to make sure to set the parameters that I wanted (diameter, small-body ID, etc.), but other than that it was easy. In order to analyze this data though, I had to convert it to a dataframe and ensure that the columns were numeric and not strings, since I am dealing with mostly numeric characteristics of the asteroids. When dealing with the news articles, I had trouble including the absolute URL link so that the user could click on it from the dictionary and read the article. However, I mitigated the issues so that the user has access to the API data and recent news articles on asteroids.

3. What kind of analysis and visualizations did you do?
    a. What analysis techniques did you use, and what are your findings?
    b. Describe the figures you made. Explain its setup, meaning of each element.
    c. Describe your observations and conclusion.
    d. Describe the impact of your findings.

Analysis Techniques Used:
1. Descriptive statistics for asteroid characteristics (mean, median, mode, range, standard deviation)
2. Diameter display if available
3. NumPy for correlation analysis between semi-major axes and eccentricities
4. Dictionary with title, link, and date of “NASA asteroid” news articles after web scraping

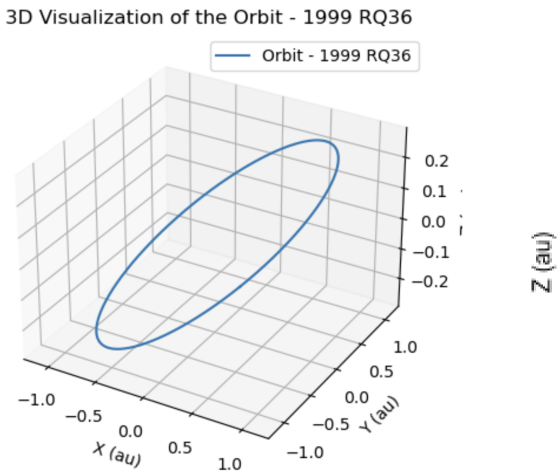
Findings: My key findings from my analyses were that the mean absolute magnitude H value for the given query search was about 24 au. An asteroid’s absolute magnitude is the visual magnitude an observer would record if the asteroid were placed 1 au away, and 1 au from the Sun and at a zero phase angle. I found this interesting because asteroid’s can be considered potentially hazardous if they have an H value of 22.0 or less. This makes sense because in general, a smaller H implies a larger diameter.

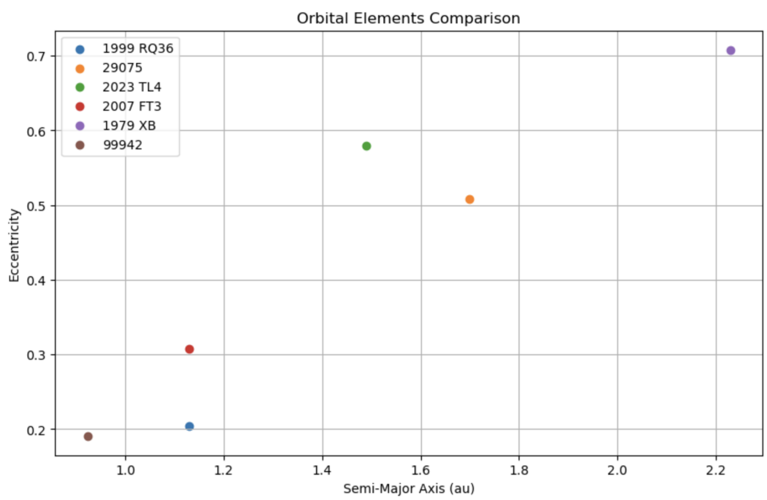
Eccentricity describes the amount by which an orbit deviates from a perfect circle. A value of 0 indicates a perfectly circular orbit, and between 0 and 1 indicate an elliptical orbit. The semimajor axis refers to an asteroid's average orbital distance from the Sun (measured in au). I found it practical that in my data, the correlation coefficient between semi-major axis and eccentricity of an asteroid was 0.80, showing a strong positive correlation.

Figures:
Distance vs. Velocity Scatter Plot: For the given query search, I generated a scatter plot of distance vs. relative velocity. Each point on the graph represents a different small body ID.
Orbital Elements Comparison: The user inputs several small body IDs and each asteroid's semi-major axis and eccentricity value will be plotted accordingly. This easily visualizes similarities and differences between a given cluster of asteroids.
3D visualization of asteroid orbits: Using several orbital element characteristics, I visualized a 3D asteroid orbit for any asteroid present in the database.

Observations and Conclusions: There does not seem to be statistical significance to the distance vs. relative velocity plot in the given query search. This is because the data is portraying small bodies from up to 60 days in the future, so it is seemingly random and not connected. For the orbital elements scatterplot, the results will vary depending on which small body IDs the user inputs. Some asteroids could have very similar eccentricity and semi-major axis values to each other, but most of the time they will be different. I love the 3D visualization graph of the asteroid orbits because each asteroid has its own unique path.

Impact of Findings: The significance of my project is that the user can take what they read in asteroid news articles, and can then use my code to find statistics and the orbital visualization of the asteroid of their choice. For example, in the dictionary that I provide, the user can find an article called "NASA's most wanted: The 5 most dangerous asteroids in the solar system". The #1 asteroid in this article is Bennu (ID: 1999 RQ36). The user can take Bennu's ID and put it in my 3D visualization code to see its orbit, and also put it in my orbital elements comparison code and compare it to the other asteroids mentioned in that article, for example. The use of the code in my project has many opportunities for the exploration of the solar system.





4. Future work
    a. Given more time, what direction would you take to improve your project?

Given more time, I would have worked more with diameter data. From the CNEOS website that is linked to the JPL API, I probably could have web scraped more data about the asteroids, specifically the diameter column. In the API, if the diameter was not known, it would be considered “null”. However, the CNEOS website has data that estimates the diameter as a range using H and limiting albedos (fraction of light reflected by the asteroid) 0.25 and 0.05. I could have taken the midpoint of this range for each asteroid and plotted it on a scatter plot.

To improve upon my descriptive statistics and correlation analysis, I would have liked to also use the “rarity” column from the CNEOS website. For each asteroid in the database, it is assigned a rarity score from 0-3 that indicates how unusual the Earth close approach is for asteroids of the same size and larger. 0 means an average frequency of 100 per year, i.e., roughly every few days or less, 1 corresponds to roughly once a month, 2 to roughly once a year, 3 to roughly once a decade. I would have wanted to perform a regression analysis to see which factors best predict an asteroid’s rarity, however I did not go through with this because I was having trouble obtaining the diameter data.

Lastly, I would expand my project to look at the other planets besides Earth and see what data I can find on their close approaches. In the API, I can query “body=Juptr” for example, and see what is returned for Jupiter’s close approaches. For a given time period, do these asteroids have a larger diameter and a faster relative velocity than Earth close approaches, on average?