**Project Title:** Interactive Natural Language Processing

Dr. Mingxian Jin (Faculty Adviser)

Dr. Denny Czejdo (Faculty Adviser)

Dr. Sambit Bhattacharya (Faculty Adviser)

Songha Yoo

Charlene Brighter

Hassan Animashaun

Fall 2024

From the portion of the dataset you shared, this dataset appears to relate to **astronomical objects**, particularly **asteroids** or other celestial bodies cataloged for their orbital and physical characteristics. Here's an explanation of the dataset based on the variables mentioned:

**Overview:**

This dataset likely contains detailed information about celestial objects like asteroids in the **Main Asteroid Belt (MBA)** or other orbital regions. Each row corresponds to an object, with unique identifiers and various parameters related to its **orbital characteristics**, **classification**, and how well its trajectory has been modeled.

**Key Variables:**

1. **id**: A unique identifier for each object.
2. **spkid**: Another identifier, possibly from a specific catalog (like the JPL Small-Body Database).
3. **full\_name**: The object's name, such as 1 Ceres, 2 Pallas, which are well-known asteroids.
4. **pdes (provisional\_designation)**: Likely a temporary or provisional name given to the object before it receives a formal designation.
5. **sigma\_tp**: The uncertainty (or standard deviation) associated with the object's time of perihelion, which is the point in its orbit when it's closest to the sun.
6. **sigma\_per**: The uncertainty in the orbital period of the object, or how long it takes to complete one orbit.
7. **class**: The classification of the object, such as **MBA (Main-Belt Asteroid)**.
8. **rms**: The root mean square residuals, a measure of the error or deviation between the observed and predicted orbital positions of the object. Lower values indicate a more accurate orbital model.

**Purpose of the Dataset:**

This dataset is likely used for **tracking and analyzing the orbits of asteroids and other celestial bodies**. It may have been sourced from a database like NASA's Jet Propulsion Laboratory (JPL) or another astronomical catalog. The data can be useful for researchers in fields like **astronomy, astrophysics**, or **planetary science**, particularly for:

* Studying asteroid orbits
* Predicting future positions of these objects
* Understanding uncertainties in orbital elements
* Performing risk assessments for objects that may pass close to Earth

Would you like help analyzing specific aspects of this dataset or further insights into any particular variable?

A screenshot of a graph

Description automatically generated

**Correlation Heatmap for Key Variables:** This heatmap shows how closely related certain variables are in the dataset. Darker shades indicate stronger relationships. For example, diameter and H (magnitude) have a strong negative correlation, which means larger objects tend to be brighter (or have lower H values). Similarly, the semi-major axis (a) and orbital period (per) are positively correlated, consistent with how objects farther from the Sun take longer to complete an orbit.

A graph with orange dots

Description automatically generated

**Diameter vs. Absolute Magnitude (H):** This scatter plot highlights an inverse relationship between object size (diameter) and brightness (H magnitude). As the diameter increases, the H value decreases, meaning larger objects tend to appear brighter. This relationship makes sense because larger objects typically reflect more light, making them more visible. Objects with smaller diameters tend to have higher H values, indicating they are dimmer.

A graph of a number of points

Description automatically generated

**Semi-Major Axis (a) vs. Orbital Period (per):** This plot showcases the relationship between the semi-major axis and the orbital period, which follows Kepler's third law. As the semi-major axis (the average distance from the Sun) increases, the time it takes for the object to complete one orbit also increases. Objects farther from the Sun have longer orbital periods due to the larger paths they need to travel, which naturally results in longer years.

A graph with numbers and lines

Description automatically generated

**Trends Over Time (Average Diameter and H Magnitude):** These two line plots show trends in the average diameter and H magnitude of celestial objects over time. The average diameter plot suggests that object sizes observed fluctuate across the years, possibly due to improved technology allowing the detection of smaller objects. Meanwhile, the H magnitude trend shows similar variation, hinting that advancements in detection may have led to the observation of dimmer, smaller objects over time as well.

Classifications of celestial objects

**MBA (Main Belt Asteroids):** The most common classification, these asteroids are located in the asteroid belt between Mars and Jupiter. They typically follow relatively stable, low-eccentricity orbits.

**OMB (Outer Main Belt):** These asteroids reside in the outer regions of the main asteroid belt, closer to Jupiter, which can influence their orbits.

**IMB (Inner Main Belt):** Located in the inner region of the asteroid belt, these objects orbit closer to Mars.

**MCA (Mars-Crossing Asteroids):** These asteroids have orbits that cross Mars' orbit, posing some potential risk of close encounters with Mars.

**APO (Apollo Asteroids):** Apollo-class asteroids have orbits that cross Earth’s orbit, often posing a potential risk to Earth, though they spend most of their time outside Earth's orbit.

**AMO (Amor Asteroids):** Amor asteroids have orbits close to Earth but do not cross Earth’s orbit, as they are mostly positioned outside of it.

**ATE (Aten Asteroids):** These are near-Earth objects with orbits mostly within Earth’s orbit, making them close neighbors to our planet.

**TJN (Jupiter Trojans):** These objects share Jupiter’s orbit, located in two main groups ahead of and behind Jupiter. They are relatively stable and pose no direct risk to Earth.

**TNO (Trans-Neptunian Objects):** Located beyond Neptune, these objects reside in the outer reaches of the solar system and have little direct interaction with the inner planets.

**CEN (Centaurs):** These objects have characteristics of both asteroids and comets and typically orbit between Jupiter and Neptune.

**AST (Asteroid):** This generic label may be used for unclassified or newly identified objects that do not fit neatly into other groups.

A graph with a number of objects

Description automatically generated

Here’s a bar chart showing the percentage of each celestial object classification. This layout provides a clearer view of each category’s proportion, highlighting the dominance of **Main Belt Asteroids (MBA)**, followed by **Outer Main Belt (OMB)** and **Inner Main Belt (IMB)** asteroids. Near-Earth objects, like **Apollo (APO)** and **Amor (AMO)** asteroids, are shown in smaller but significant proportions, while classifications from the outer solar system, like **Trans-Neptunian Objects (TNO)** and **Centaurs (CEN)**, have the lowest percentages. This chart makes it easier to compare each category’s contribution. ​