ASTEROID OBRIT ANALYSIS

Why choose this dataset?

Because it enables us to understand more about asteroids in space, we selectedthe "Asteroid Orbit Parameters" dataset. Imagine that you are interested in learning how these space rocks orbit the Sun. This dataset provides us with a wealth of knowledge regarding that.

Our selection criteria for this dataset are rather simple. First of all, we have a strong curiosity in space and wish to comprehend how things function there. Second, we are thinking about specific queries. If any asteroids could be hazardous to Earth, if there are any patterns in how they move, and if there is arelationship between how brilliant an asteroid seems and its motion.

The collection contains a wealth of knowledge. It provides information on the shape of asteroids' orbits as well as their distance from and time spent around the Sun.

Asteroid research has applications beyond pure scientific curiosity. For instance,it aids in our preparation for the potential impact of an asteroid on Earth, which is quite significant.

The data is also easily accessible and well-organized, which helps us save a ton of time. Consequently, it is a wonderful match for our investigation.

The dataset was retrieved from Kaggle.

EDA analysis:

Statistical analysis:

Epoch (TDB): This functions as a sort of timestamp for the moment we discovered these asteroids. The dataset includes observations from various eras.

Consider the Orbit Axis (AU) as the average distance between these asteroids and the Sun. They are, on average, 1.78 times further away from the Sun than Earth.

Consider the asteroid's route around the Sun to get a sense of its eccentric orbit.The majority of pathways have an average value around 0.45, which makes them rather elliptical in shape.

Imagine an asteroid orbiting like a roller coaster. Orbit Inclination (deg). This reveals how much they incline in relation to a flat path. They lean 12.94 degreeson average.

Finding the location along an asteroid's course where it is closest to the Sun isanalogous to the perihelion argument (deg). This point typically has a temperature of 181.51 degrees.

Consider the Node Longitude (deg) as the compass direction of an asteroid'scourse in space. The asteroid appears to be moving at a direction of about
172.67 degrees on average.

Mean Anomaly (deg): This identifies the location of an asteroid in its journey ata specific moment. It's averaging around 172.83 degrees.

The term "perihelion distance" (AU) refers to how close an asteroid approachesthe Sun throughout its orbit. They are typically 0.91 times closer to Earth than the Sun on average.

Imagine an asteroid's orbital period as a yearlong excursion. This informs usthat the journey takes, on average, 2.47 years.

Checking the Minimum Orbit Intersection Distance (AU) is equivalent to seeinghow closely an asteroid crosses Earth's path. They go, on average, 0.10 times farther than the distance between Earth and the Sun.

Space Reference: The labelling or classification of these asteroid routes is likedeciphering a hidden code. There are various hidden codes.

Object brightness is gauged by the asteroids' magnitude. With an average brightness of roughly 22.29, they range in brightness from brilliant to very faint.

Count plot of a categorical feature:

We next categorized the asteroids based on the type and count of asteroids which are Amor Asteroid which are 6000 in number, Apollo Asteroid which are8000 in number , Aten Asteroid which are nearly 1000 in number, Apohele Asteroid which are very close to zero in number.

Define a mapping of original class labels to new labels

We defined the category of asteroids with new labels so that we can get the correlation matrix and check if there are any relation of category with other features.

Data visualization with heatmap.

From the heatmap we get to know Strong Positive Correlation Between Orbit Axis and Orbital Period (Yr): You might notice that the average distance from the Sun (Orbit Axis) and the amount of time it takes for an orbit to complete (Orbit Period) are strongly positively correlated. Given that objects farther fromthe Sun often have longer orbital periods.

In general, asteroids with bigger Orbit Axis values (indicating they are farther from the Sun) tend to have larger Aphelion Distances (indicating they reach farther from the Sun during their orbits). This is because Aphelion Distance andOrbit Axis have a positive association. Simply put, an asteroid's maximum distance from the Sun increases in direct proportion to its average distance fromthe Sun.

Now since we performed the data visualization and got to know that Orbit Axisand Orbital Period (Yr), Aphelion Distance and Orbit Axis are higly co related we will focus on them.

Data visualization with histogram

Most asteroids have temporary aphelion distances, with a peak between 1.0 and 3.5 AU (distance from the Sun).

Orbit Axis: Most asteroids are located at orbit axis between 1.0 and 2.5 AU,indicating orbits that are quite close to the Sun.

Orbital Period: Most asteroids have brief orbital durations (between one and five years), but a few have very lengthy periods, such as one outlier with almosta hundred years.

By highlighting common tendencies and spotting outliers, these distributionsprovide information on the features of asteroid orbits.

Data visualization with boxplot

The middle 50% of asteroids have aphelion distances between around 1.0 and 3.4 AU, according to the box plot for aphelion distance. Some outliers haveaphelion points that are extremely far away.

Orbit Axis (AU): According to the box plot for orbit axis, 50% of asteroids haveaxes that are between around 1.3 and 2.2 AU. There are outliers with axes that are noticeably larger, similar to aphelion distance.

Orbital Period (Years): According to the box plot, the majority of asteroids have periods between 1.5 and 3.3 years. There are a few anomalies with significantly longer orbital periods, though.

Box plot before data preprocessing.

Asteroid distance from the Sun is measured in terms of aphelion distance (AU).It is typically 2.5 times further from the Sun than Earth. The box's middle line depicts the average separation, while the box itself displays the region where most asteroids gather, which is between 1.75 AU and 3.4 AU. However, we notice some dots outside of the boxes, which resemble insane travellers travelling extremely far or extremely close.

Data preprocessing

now we performed data preprocessing to remove the outliers and also removefew anomalies from the data, so that we perform better Analysis and make a good model.

Box plot after data preprocessing.

The form of the path asteroids follow as they orbit the Sun is revealed to us by Orbit Axis (AU). The majority of asteroids travel between 1.4 AU and 2.2 AU, with the average distance being around 1.75 AU. But once more, there are thosedots outside of the boxes, which are sort of the outliers.

Finally, Orbital Period (yr) is the time it takes an asteroid to orbit the Sun once.It takes roughly 2.2 years on average. The majority of asteroids travel in between 1.8 and 3.2 years. There are some dots outside of those boxes, as you probably have suspected.

These plots enable us to determine the locations of the majority of asteroids andidentify any odd ones with distinctive patterns.

Pairplot analysis

The clustering of points in the "Aphelion Distance vs. Orbit Axis" plot, whichsuggests a possible link between these two variables, is the most noticeable pattern seen in the pair plot. The other two figures, on the other hand, do not clearly show any linear relationships, indicating weaker or non-linear relationships between aphelion distance, orbit axis, and orbital period.

Conclusion:

In summary, our exploratory data analysis (EDA) of asteroid properties has shedimportant light on the connections between important variables. We discovered a clear clustering pattern in the aphelion distance and orbit axis, pointing to a possible relationship between asteroids' distance from the Sun and the contours of their orbital routes. Both the aphelion distance and the orbital period, as well as the relationship between the orbit axis and the orbital period, are not clearly linear. This implies that there may be more complicated or nonlinear processes affecting an asteroid's orbital period.

Additional EDA and statistical analysis may reveal more linkages and insights in the data. These discoveries are crucial for developing forecasting models andimproving our comprehension of how asteroids behave inside our solar system.