Dataset Information:

- **Author:** NASA
- **Timespan:** October 1, 2023 – October 1, 2025
  (it includes very precise predictions for the future dates)
- **Link to the dataset:** NEO Earth Close Approaches Dataset
  (https://cneos.jpl.nasa.gov/ca/)
- **Rights of Usage:** Public Domain
- **Rows and Columns:** 4632 rows and 10 columns
- **Legend of table column, type and descriptions:**

	name	type	description
○	**Object**	*object*	object primary designation.
○	**Close-Approach (CA) Date**	*object*	date and time (TDB) of closest Earth approach. "Nominal Date" is given to appropriate precision. The 3-sigma uncertainty in the time is given in the +/- column in days_hours:minutes format (for example, "2_15:23" is 2 days, 15 hours, 23 minutes; "< 00:01" is less than 1 minute).
○	**View CA**	*float64*	Open the close-approach viewer and render the high-precision trajectory during the close approach.
○	**CA Distance Nominal (au)**	*float64*	the most likely (Nominal) close-approach distance (Earth center to NEO center), in astronomical units.
○	**CA Distance Minimum (au)**	*float64*	the minimum possible close-approach distance (Earth center to NEO center), in astronomical units. The minimum possible distance is based on the 3-sigma Earth target-plane error ellipse.
○	**V relative (km/s)**	*float64*	object velocity relative to Earth at close-approach.
○	**V infinity (km/s)**	*float64*	object velocity relative to a massless Earth at close-approach.
○	**H (mag)**	*float64*	asteroid absolute magnitude (in general, smaller H implies larger asteroid diameter). **Undefined for comets**.
○	**Diameter**	*object*	diameter value when known or a range (min - max) estimated using the asteroid's absolute magnitude (H) and limiting albedos of 0.25 and 0.05.
○	**Rarity**	*int64*	A measure of how infrequent 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, etc. 'n/a' means that a frequency estimate is not available.

Au : *one Astronomical Unit (au) is approximately 150 million kilometres*
LD: *one Lunar Distance (LD) is approximately 384 000 kilometres*

Visualization Protocol (data preprocessing):

1. Initial Data Familiarization:

Imported necessary libraries (pandas and re) and loaded the dataset in excel format.

2. Observation of the Dataset:

- Upon reviewing the dataset, I noticed that the **'Close-Approach (CA) Date'** column included not only the date but also additional time information. To make the data more useful, I extracted only the **date** portion and converted it into a standard YYYY-MM-DD datetime format for chronological analysis.
- Additionally, I observed that the **'Diameter'** column, which contained ranges of values (e.g. "14 m – 31 m") was categorized as string. To make this column useful for analysis, I extracted the **minimum and maximum values** of the diameter, converted them into numerical form. And I created with them a new **average diameter** column.

3. Data Cleaning:

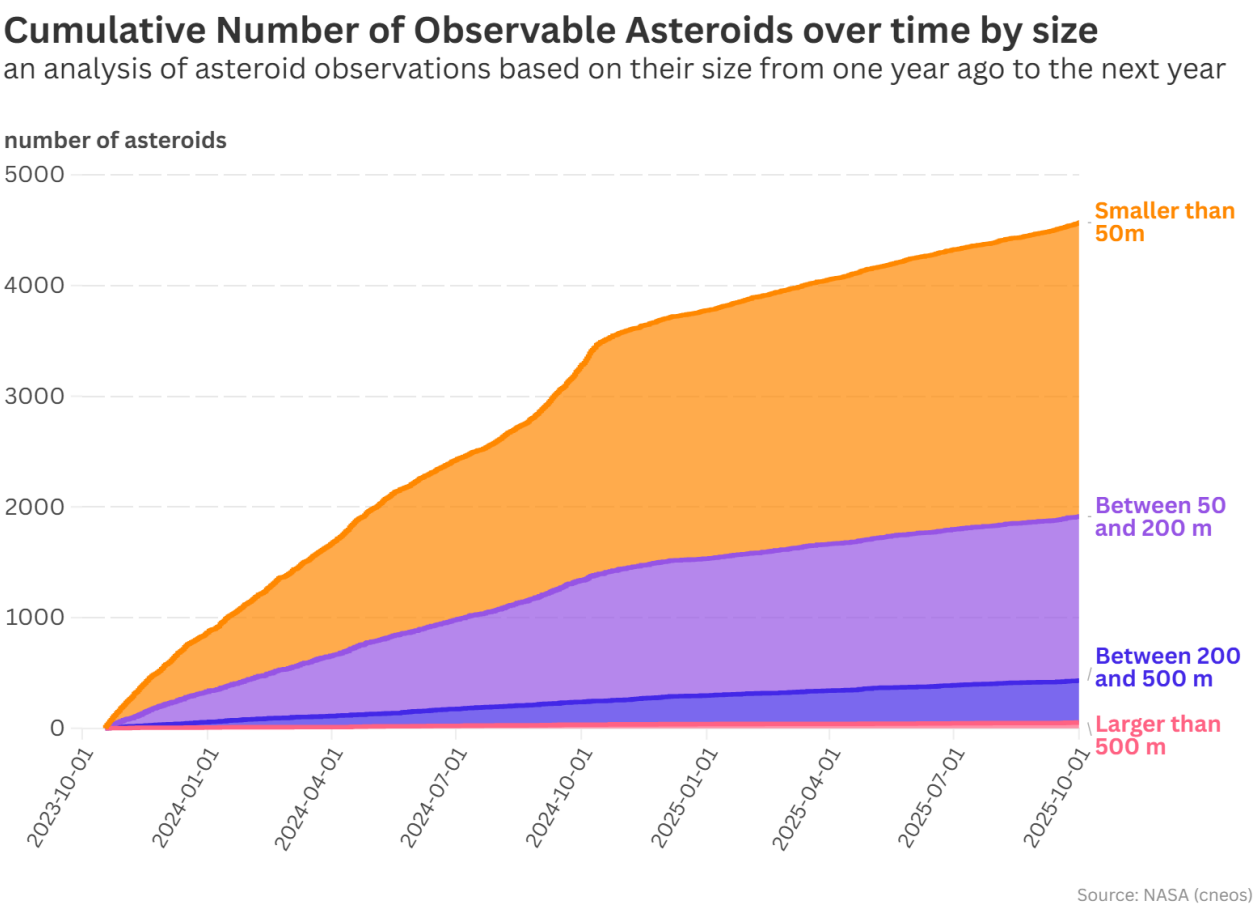
- **Removing Duplicates:** I tried to remove duplicates, but I didn't find any repeated values to drop.
- **Handling Missing Values:** missing values (probably the comets) ere dropped from the dataset to clean it for further analysis.
- **Converted Dates to Datetime Format:** the **'Close-Approach (CA) Date'** column was cleaned and formatted to YYYY-MM-DD format to ensure proper chronological sorting.
- **Extracted Min and Max Diameters:** the **diameter** column was transformed from categorical to numerical by applying a regular expression (using the re library) to extract the minimum and maximum diameter. Then the diameter column was split into two new columns, **Diameter Min (m)** and **Diameter Max (m)**.
- **Created an Average Diameter Column:** a new column named **Diameter Avg (m)** was introduced to calculate the average diameter to categorize the bodies.
- **Categorized Asteroids by Size:** asteroids were divided into categories based on diameter (***Smaller than 50m, Between 50m and 200m, Between 200m and 500m*** *and* ***Larger than 500m***). I started with an automatic categorization based on equal intervals but I noticed that the results were not ideal for visualization (they were too different from each others: 4441 elements in the first bin, 172 in the second and only 10 in the third). Then, I customized the bins to achieve a more balanced distribution (obtaining 2691 elements in the first bin, 1495 in the second, 383 in the third and 54 in the fourth).
- **Feature Engineering – One Hot Encoding:** size categories were converted into individual columns for cumulative tracking of asteroids within each category.
- **Sorting:** the dataset was sorted chronologically by the **'Close-Approach Date'**.
- **Cumulative Calculation:** cumulative counts for each size category were calculated, allowing analysis of how the number of observable asteroids in each size range grows over time.
- **Filtered Dataset:** a new dataset was created with the columns useful for the data visualization (***object, Close-Approach (CA) Date, Smaller than 50m, Between 50m and 200m, Between 200m and 500m*** *and* ***Larger than 500m****)*

4. Exportation of the dataset:

The filtered dataset was exported as csv and xlsx format with the following names:

nasa_asteroids_cleaned.csv and ***nasa_asteroids_cleaned.xlsx***

Data Visualization:



Title:
Cumulative Number of Observable Asteroids Over Time by Size

Topic:
This visualization employs an area chart to illustrate the cumulative count of near-Earth asteroids, categorized by size, over a two-year period.

Research questions:
How does the number of observable near-Earth asteroids vary over time, categorized by size?

Data Task Type:
Trend analysis and categorization.

Insights:

- ***The majority of observable asteroids are smaller than 50 metres*** *and their count increases significantly faster compared to larger asteroids.*
- ***Smaller asteroids*** *may have* ***more frequent close-Earth approaches****.*
- *The growth pattern indicates that while* ***the discovery rate of small asteroids is high, large asteroids remain relatively rare*** *over time.*