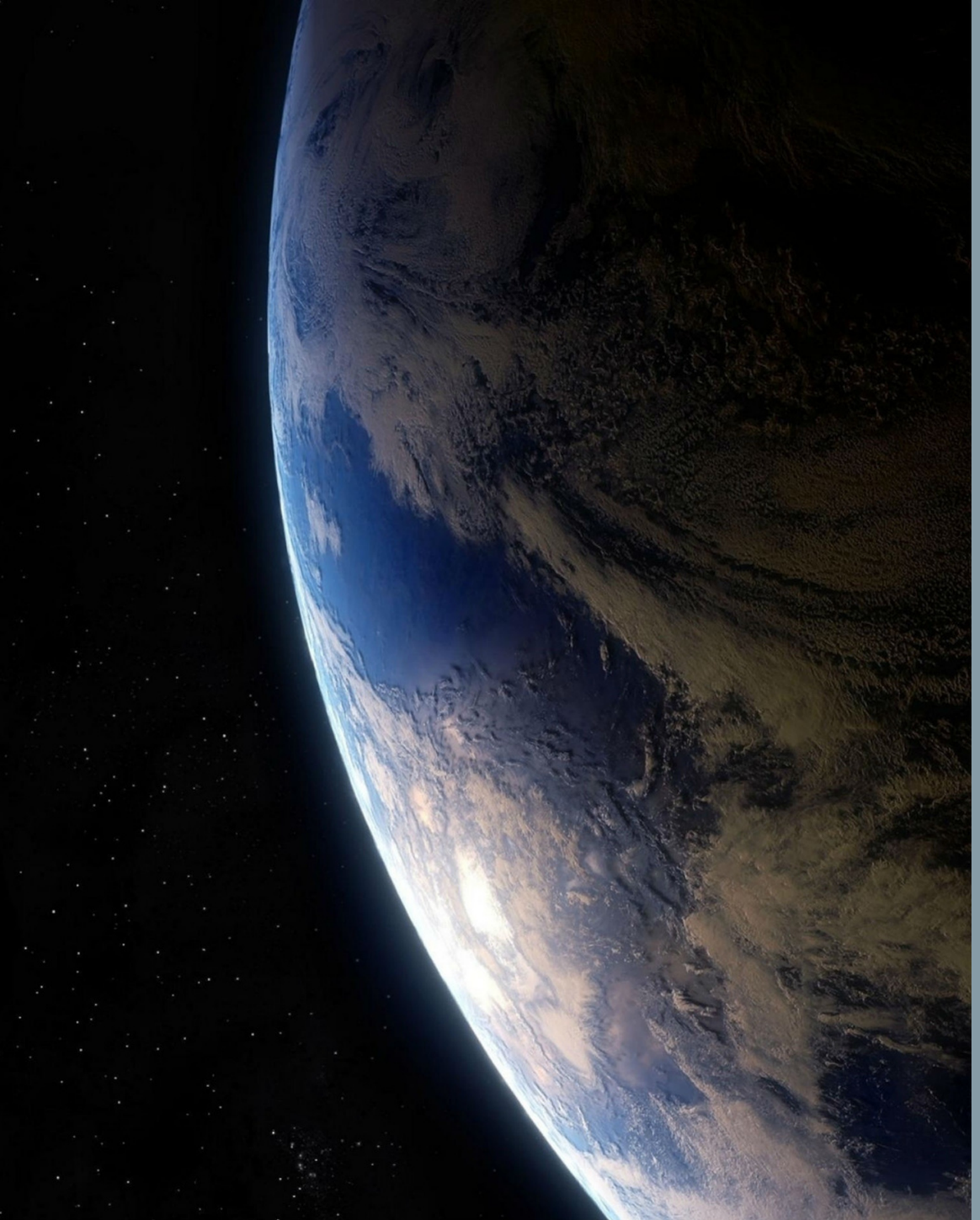
FINAL DATA SCIENCE PROJECT PRESENTATION

Exploring Potentially Hazardous Near-Earth Objects

Presented by Kiara A. Richardson

Spring 2023

RCEL 506 Applied Statistics and Data Science for Engineering Leaders



Project Overview

DISCUSSION POINTS

- Project Goals
- Literature Review
- The Dataset
- Data Cleaning
- Exploratory Data Analysis
- Machine Learning Model
- Results and Discussion
- Hypothesis Conclusion
- Future Work

Project Goals

Tisserand Parameter

$$T_J = \frac{a_j}{a} + 2\cos(i)\sqrt{\frac{a}{a_J}(1-e^2)}$$

$T_J > 3$ are asteroids
$T_J < 3$ are comets

a = small body semi-major axis
i = small body orbital inclination
e = small body eccentricity
a_J = perturbing large body semi-major axis (Jupiter)

Know the difference

- A body is a NEO if its closest approach to the Sun is less than 1.5 astronomical units (AU)
- If a body crosses the Earth's orbit, and is larger than 140 meters across, it is considered a PHO

01 TOPIC

To study

- What Near Earth Objects (NEOs) are asteroids and which are comets
- What comet and asteroid Near Earth Objects (NEOs) are Potentially Hazardous Objects (PHOs)
- How PHOs behave graphically
- What seasons does Earth experience the most PHO interactions

02 HYPOTHESIS

- There are more asteroid Near-Earth Objects (NEOs)
- There are more asteroid Near-Earth Objects (NEO) that are also Potentially Hazardous Objects (PHOs)
- The Semi-Major axis of a Potentially Hazardous Object (PHO) should decrease as it approaches the Sun and it's Orbital Speed (or Mean Motion) should increase as it approaches the Sun
- Earth will experience the most Potentially Hazardous Object (PHO) interactions in the Fall months

Literature Review

Origin and Evolution of Near Earth Asteroids by A. Morbidelli - Cambridge University

Comets in the near-Earth object population by Francesca DeMeo and Richard P. Binzel - Massachusetts Institute of Technology

Potential impact detection for Near-Earth asteroids: the case of 99942 Apophis (2004 MN4) - Cambridge University

The Dataset

KAGGLE

NASA Asteroid Classification

https://www.kaggle.com/datasets/shrutimehta/nasa-asteroids-classification?resource=download&select=nasa.csv

NASA JPL NEO

Jet Propulsion Laboratory NEO Library

http://neo.jpl.nasa.gov/

DATAFRAME SIZE

4687 rows × 40 columns

THE DATAFRAME

The data types for the columns are: bool(1), float64(30), int64(5), object(4).

The memory usage is 1.4+ MegaBytes (MB).

Original Dataset Snippet

	Neo Reference ID	Name	Absolute Magnitude	Est Dia in KM(min)	Est Dia in KM(max)	Est Dia in M(min)	Est Dia in M(max)	Est Dia in Miles(min)	Est Dia in Miles(max)	Est Dia in Feet(min)	Est Dia in Feet(max)	Close Approach Date	Epoch Date Close Approach
0	3703080	3703080	21.600	0.127220	0.284472	127.219879	284.472297	0.079051	0.176763	417.388066	933.308089	1995-01-01	788947200000
1	3723955	3723955	21.300	0.146068	0.326618	146.067964	326.617897	0.090762	0.202951	479.225620	1071.581063	1995-01-01	788947200000
2	2446862	2446862	20.300	0.231502	0.517654	231.502122	517.654482	0.143849	0.321655	759.521423	1698.341531	1995-01-08	789552000000
3	3092506	3092506	27.400	0.008801	0.019681	8.801465	19.680675	0.005469	0.012229	28.876199	64.569144	1995-01-15	790156800000
4	3514799	3514799	21.600	0.127220	0.284472	127.219879	284.472297	0.079051	0.176763	417.388066	933.308089	1995-01-15	790156800000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4682	3759007	3759007	23.900	0.044112	0.098637	44.111820	98.637028	0.027410	0.061290	144.723824	323.612307	2016-09-08	1473318000000
4683	3759295	3759295	28.200	0.006089	0.013616	6.089126	13.615700	0.003784	0.008460	19.977449	44.670934	2016-09-08	1473318000000
4684	3759714	3759714	22.700	0.076658	0.171412	76.657557	171.411509	0.047633	0.106510	251.501180	562.373736	2016-09-08	1473318000000
4685	3759720	3759720	21.800	0.116026	0.259442	116.025908	259.441818	0.072095	0.161210	380.662441	851.187094	2016-09-08	1473318000000
4686	3772978	3772978	19.109	0.400641	0.895860	400.640618	895.859655	0.248946	0.556661	1314.437764	2939.172192	2016-09-08	1473318000000

Cleaning the Data

1. Import the data with the help of pandas
2. Use .info() and .isna() to check for NaN values
3. Use .drop() to remove unnecessary columns
4. Rename the cleaned dataset into a new variable
5. Cleaned dataset is left with
    a. The data types for the columns are: bool(1), float64(16), object(1)
    b. The memory usage is 627.2+ KiloBytes

Note: NASA JPL keeps very clean datasets so there were no NaN values

4687 ROWS × 40 COLUMNS

ORIGINAL DATASET

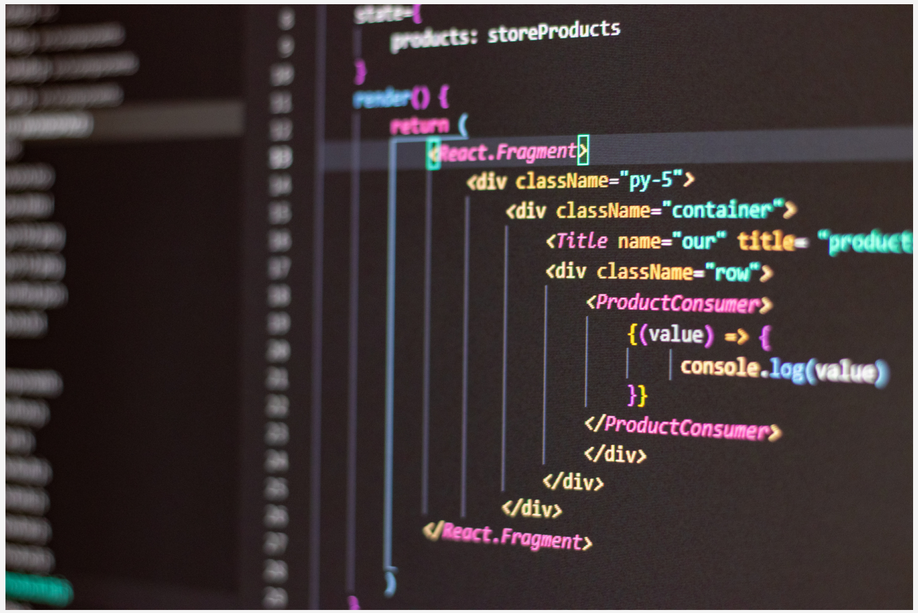
4687 ROWS × 18 COLUMNS

CLEANED DATASET

Cleaned Dataset Snippet

	Absolute Magnitude	Est Dia in KM(max)	Close Approach Date	Relative Velocity km per hr	Miss Dist. (Astronomical)	Jupiter Tisserand Invariant	Epoch Osculation	Eccentricity	Semi Major Axis	Inclination	Asc Node Longitude	Orbital Period
0	21.600	0.284472	1995-01-01	22017.003799	0.419483	4.634	2458000.5	0.425549	1.407011	6.025981	314.373913	609.599786
1	21.300	0.326618	1995-01-01	65210.346095	0.383014	5.457	2458000.5	0.351674	1.107776	28.412996	136.717242	425.869294
2	20.300	0.517654	1995-01-08	27326.560182	0.050956	4.557	2458000.5	0.348248	1.458824	4.237961	259.475979	643.580228
3	27.400	0.019681	1995-01-15	40225.948191	0.285322	5.093	2458000.5	0.216578	1.255903	7.905894	57.173266	514.082140
4	21.600	0.284472	1995-01-15	35426.991794	0.407832	5.154	2458000.5	0.210448	1.225615	16.793382	84.629307	495.597821
...	...	...	...	...	...	...	...	...	...	...	...	...
4682	23.900	0.098637	2016-09-08	79755.354273	0.041361	5.156	2457637.5	0.361512	1.161429	39.880491	164.183305	457.179984
4683	28.200	0.013616	2016-09-08	11610.539577	0.006469	5.742	2458000.5	0.073200	1.075134	5.360249	345.225230	407.185767
4684	22.700	0.171412	2016-09-08	25889.910626	0.061009	4.410	2458000.5	0.368055	1.528234	4.405467	37.026468	690.054279
4685	21.800	0.259442	2016-09-08	40867.522309	0.260760	4.477	2458000.5	0.202565	1.486600	21.080244	163.802910	662.048343
4686	19.109	0.895860	2016-09-08	129408.666253	0.462372	4.108	2458000.5	0.405642	1.474045	53.574923	187.642183	653.679098

Exploratory Data Analysis



- Description
- Pandas Profiling
- Correlation

Description

	Absolute Magnitude	Est Dia in KM(max)	Relative Velocity km per hr	Miss Dist. (Astronomical)	Jupiter Tisserand Invariant	Epoch Osculation	Eccentricity	Semi Major Axis	Inclination	Asc Node Longitude	Orbital Period
count	4687.000000	4687.000000	4687.000000	4687.000000	4687.000000	4.687000e+03	4687.000000	4687.000000	4687.000000	4687.000000	4687.000000
mean	22.267865	0.457509	50294.919829	0.256778	5.056111	2.457724e+06	0.382569	1.400264	13.373844	172.157275	635.582076
std	2.890972	0.826391	26255.601377	0.145798	1.237818	9.202975e+02	0.180444	0.524154	10.936227	103.276777	370.954727
min	11.160000	0.002260	1207.814804	0.000178	2.196000	2.450164e+06	0.007522	0.615920	0.014513	0.001941	176.557161
25%	20.100000	0.074824	30358.313370	0.133420	4.049500	2.458000e+06	0.240858	1.000635	4.962341	83.081208	365.605031
50%	21.900000	0.247765	46504.401181	0.265029	5.071000	2.458000e+06	0.372450	1.240981	10.311836	172.625393	504.947292
75%	24.500000	0.567597	65079.535405	0.384154	6.019000	2.458000e+06	0.512411	1.678364	19.511681	255.026909	794.195972
max	32.100000	34.836938	160681.487851	0.499884	9.025000	2.458020e+06	0.960261	5.072008	75.406667	359.905890	4172.231343

Pandas Profiling

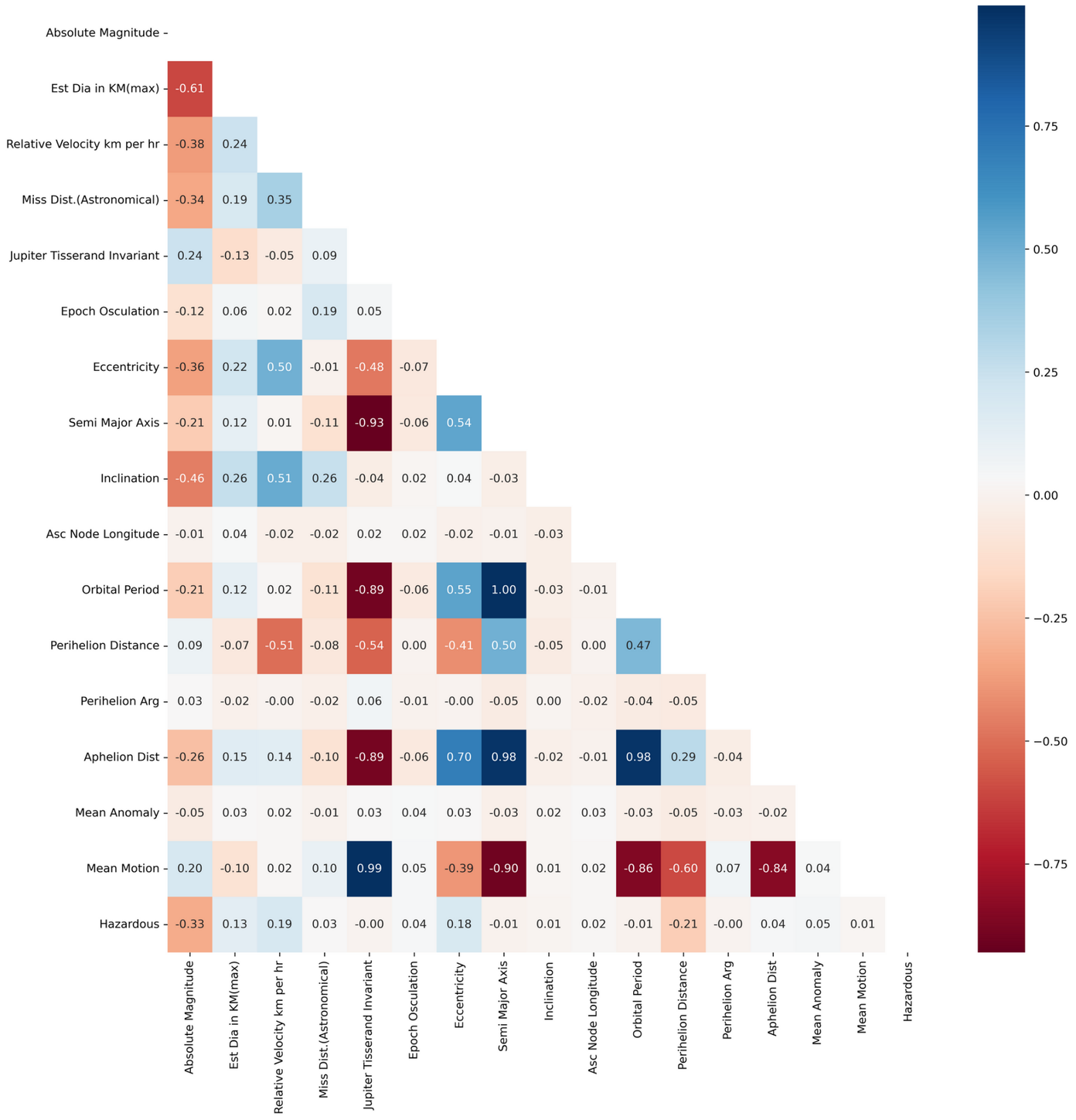
Dataset statistics

Number of variables	18
Number of observations	4687
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	627.2 KiB
Average record size in memory	137.0 B

Variable types

Numeric	16
Categorical	1
Boolean	1

Correlation



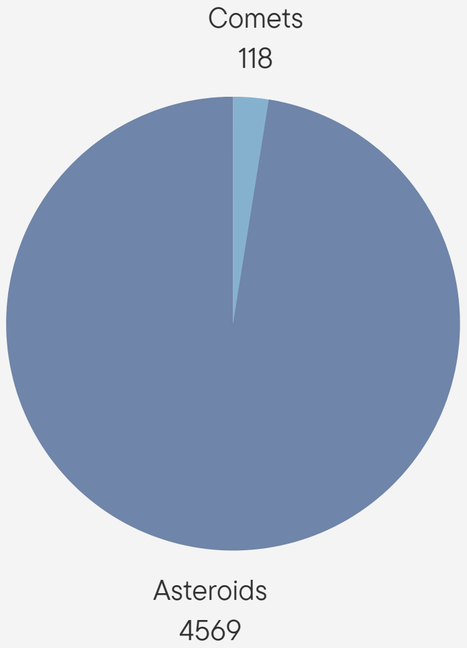
Machine Learning Model

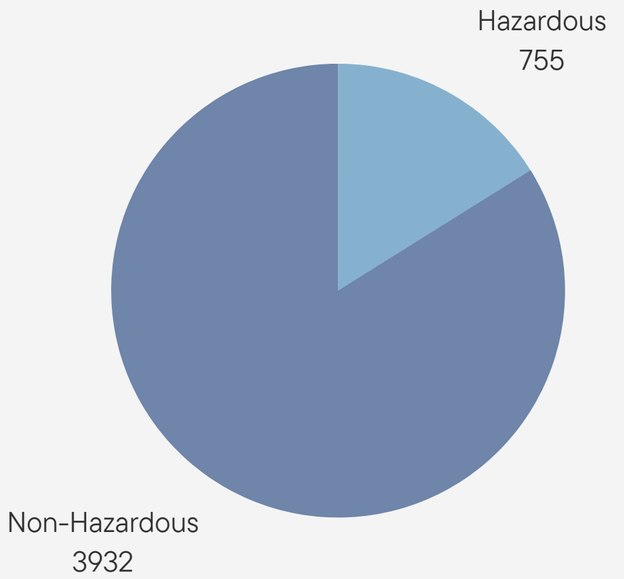
There is none :)

- used Exploratory Data Analysis methods

Results

OVERALL DATASET





COMETS VS. ASTEROIDS

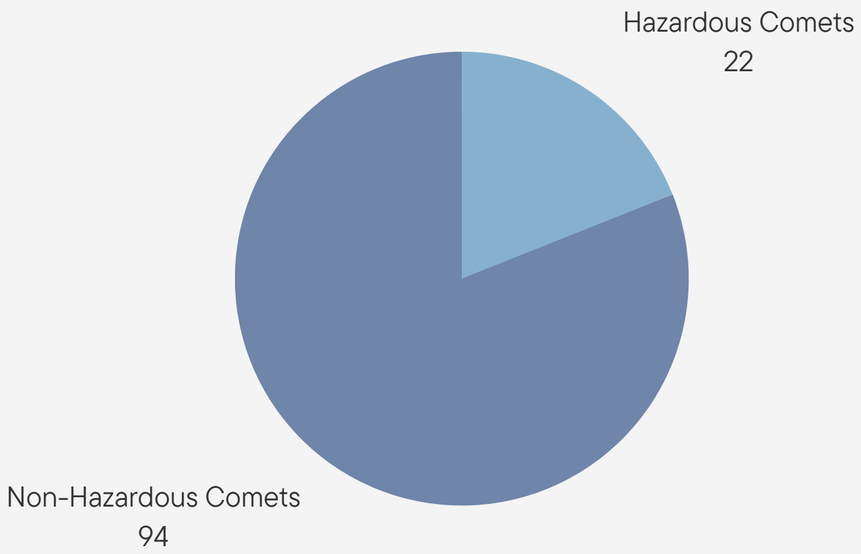
Clearly there are more asteroid-like objects in the overall dataset

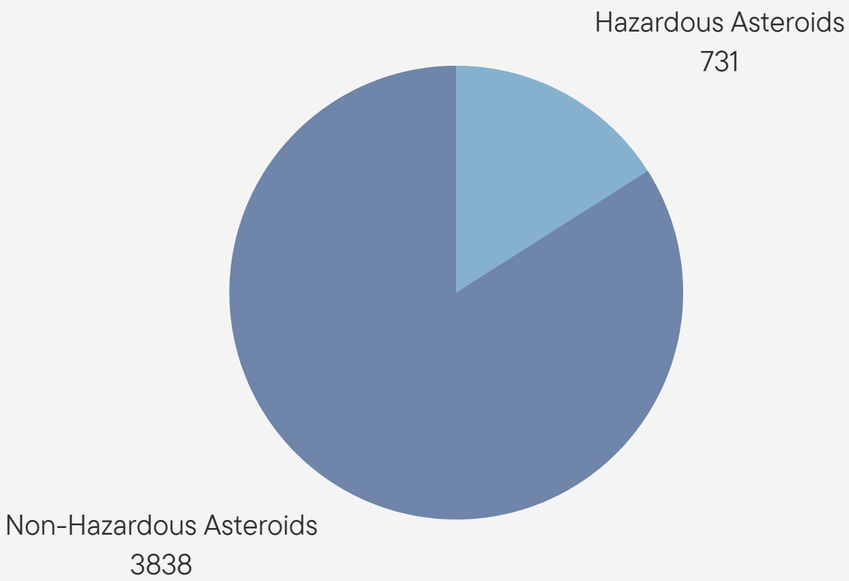
HAZARDOUS VS. NONHAZARDOUS

Clearly there are more Non-Hazardous objects in the overall dataset

Results

SUBDIVIDED DATASET





HAZARDOUS COMETS VS. NONHAZARDOUS COMETS

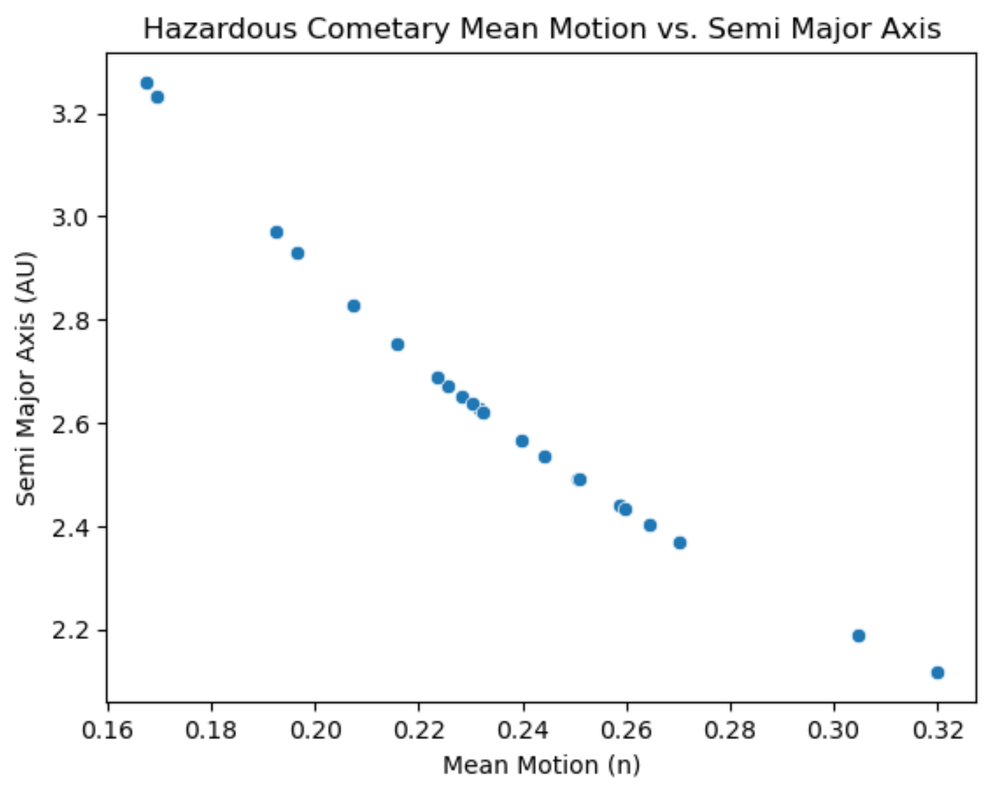
Clearly there are more Non-Hazardous comets in the subdivided dataset

HAZARDOUS ASTEROIDS VS. NONHAZARDOUS ASTEROIDS

Clearly there are more Non-Hazardous asteroids in the subdivided dataset

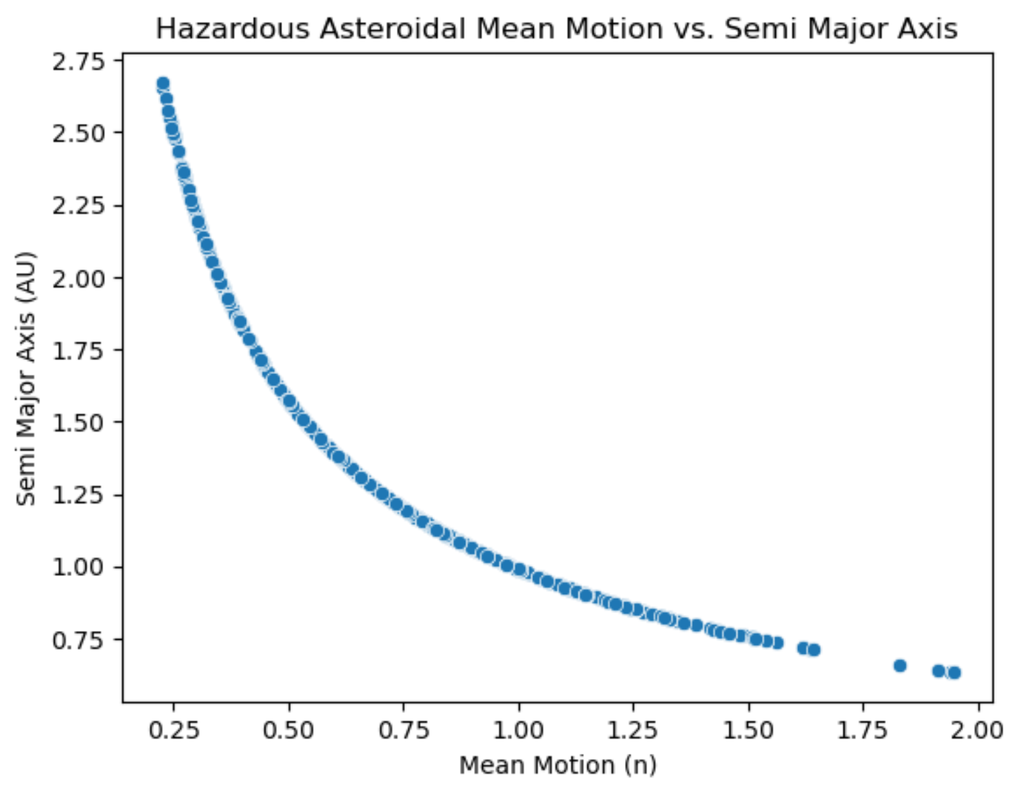
Results

NEO GRAPHIC BEHAVIOR



PHO COMETS

The smallest Semi Major Axis is within the range of around 2 AU, meaning it definitely crosses Earth's orbit at least once

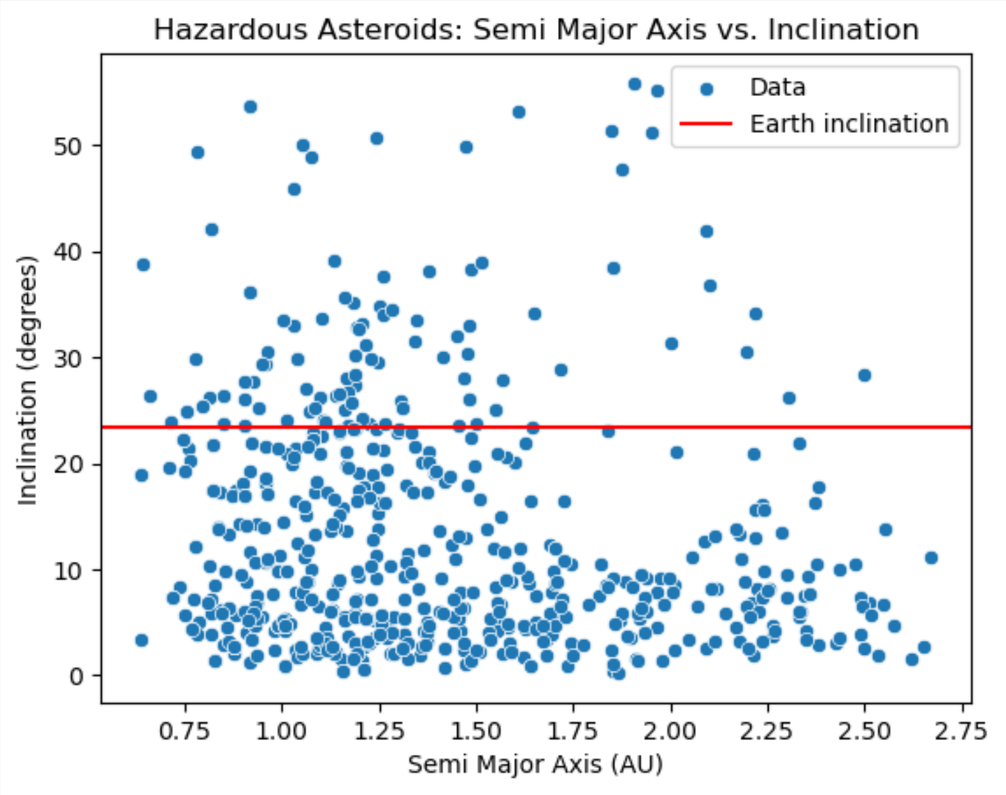


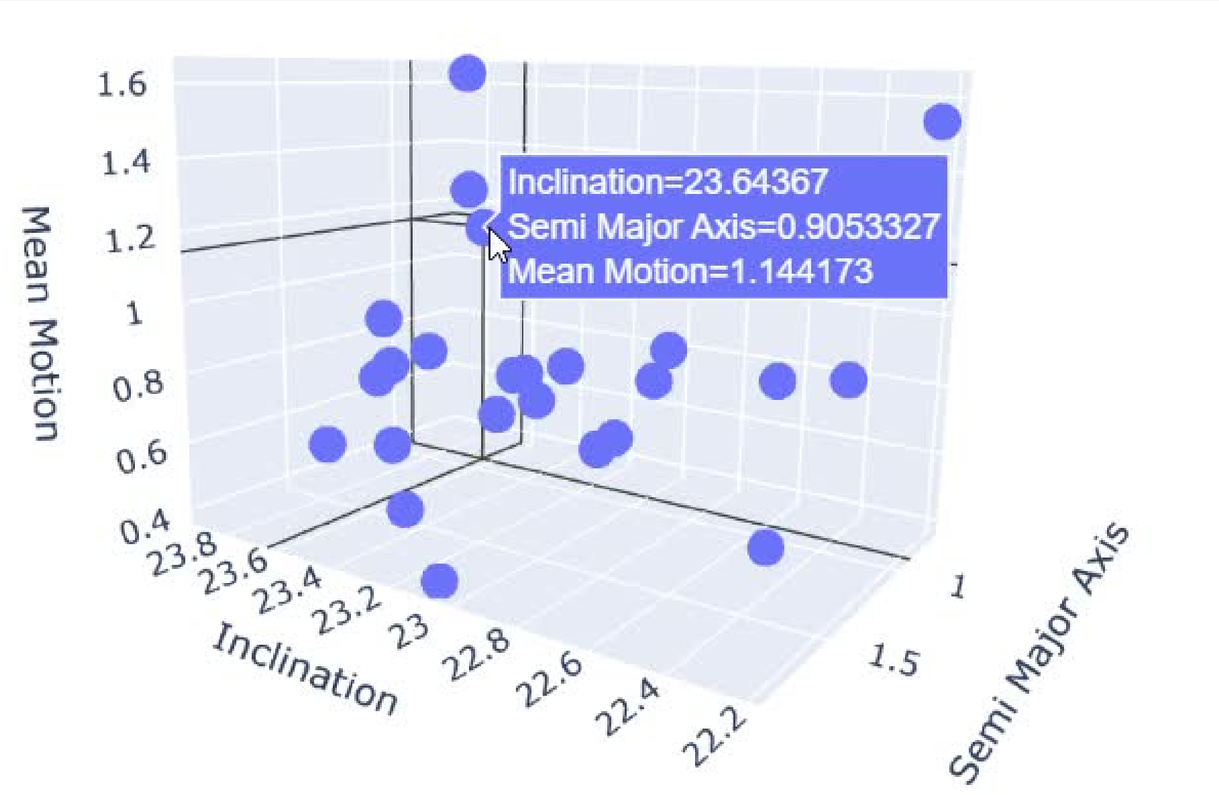
PHO ASTEROIDS

The smallest Semi Major Axis is less than 0.75 AU and there are also significantly more data points within the 2 AU range

Results

ASTEROID GRAPHIC BEHAVIOR





PHO ASTEROID GRAPHS

Earth's inclination is 23.4 degrees. There are multiple asteroids that cross Earth's orbital path in the 2D graph. All of their Semi-Major Axes are less that 3 AU. We can look at the table to prove this.

Results

ASTEROID GRAPHIC BEHAVIOR

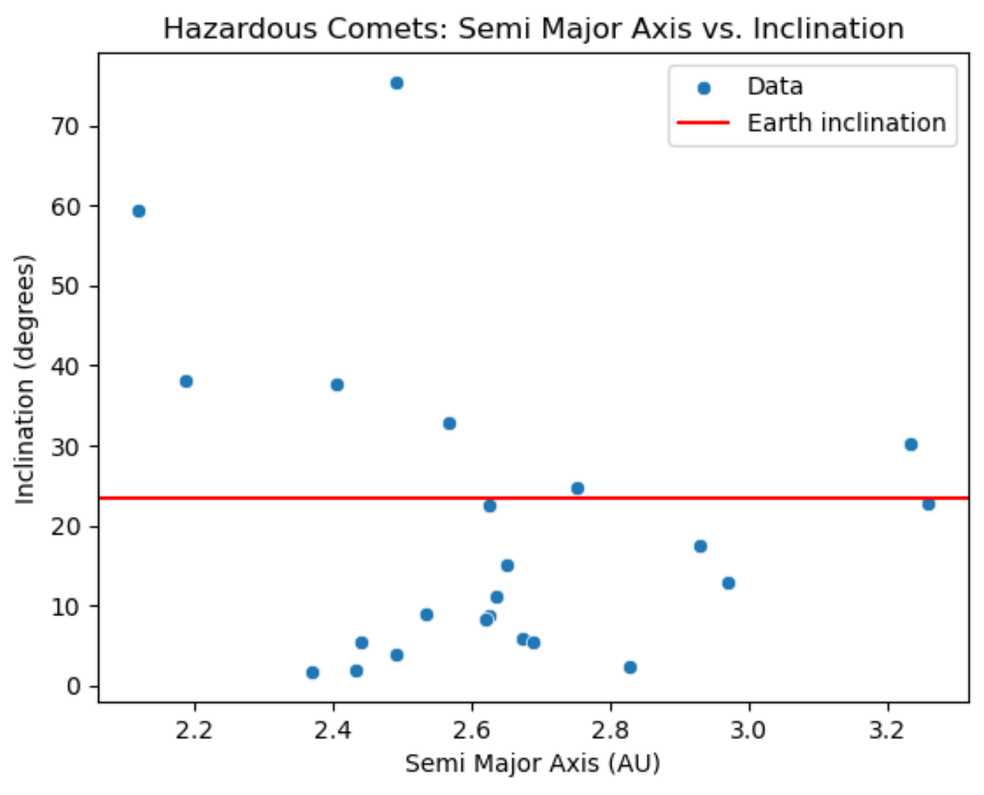
	Inclination	Semi Major Axis	Close Approach Date	Mean Motion
0	23.766741	1.225054	1995-03-22	0.726894
1	23.277205	1.182149	1995-04-22	0.766824
2	23.861945	0.714212	1996-01-22	1.632913
3	22.858798	1.145920	1996-11-08	0.803475
4	22.503754	1.103617	1997-03-22	0.850113
5	23.309957	1.190598	1997-05-22	0.758676
6	23.745884	0.850927	1998-11-22	1.255642
7	23.746184	0.850943	1998-11-22	1.255605
8	23.277205	1.182149	2000-07-08	0.766824
9	23.897140	1.110193	2001-09-15	0.842570
10	22.893293	1.297289	2002-09-08	0.667035

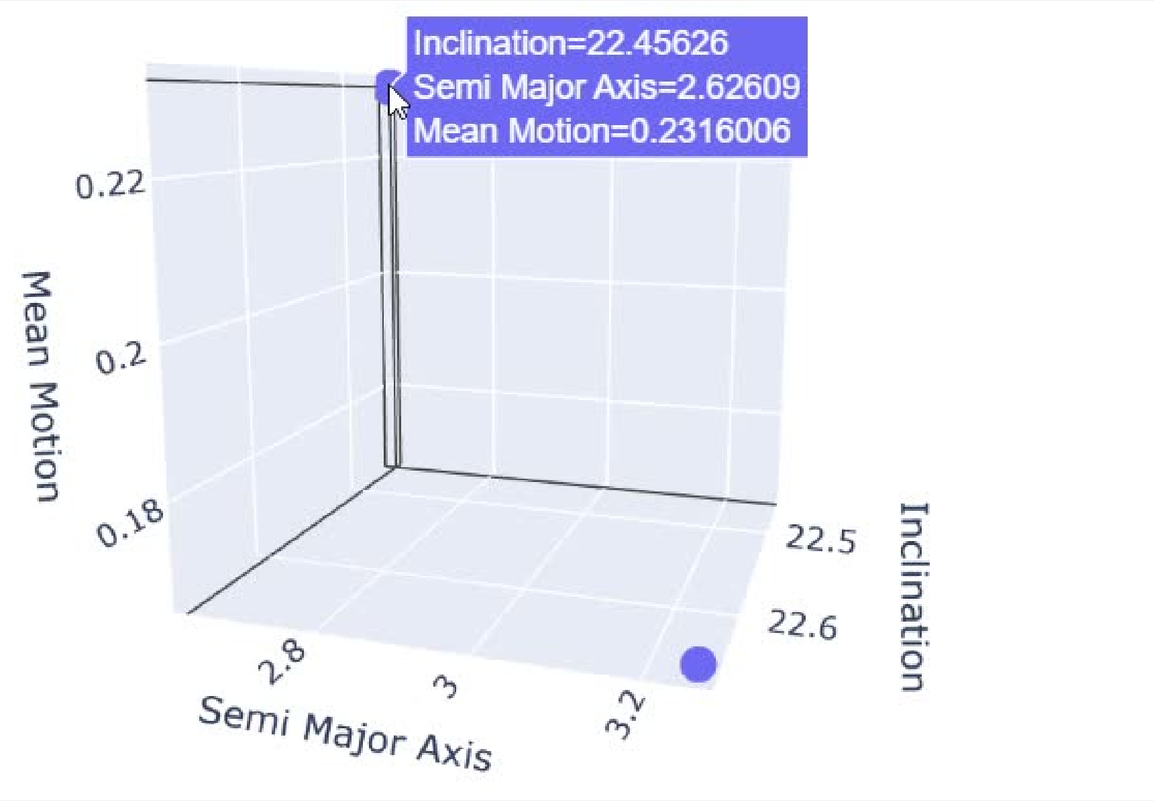
PHO ASTEROIDS

From the table, there are 36 instances of asteroids crossing Earth's orbital path. From the years 1995 to 2016, there are only 4 years with no asteroid passings. That is 1999, 2004, 2008, and 2016. Every other year there is at least 1 asteroid passing.

Results

COMET GRAPHIC BEHAVIOR





PHO COMET GRAPHS

Earth's inclination is 23.4 degrees. Looking at this graph, it looks like 2 comets cross Earth's orbital plane. We can look at the table to prove this.

Results

COMET GRAPHIC BEHAVIOR

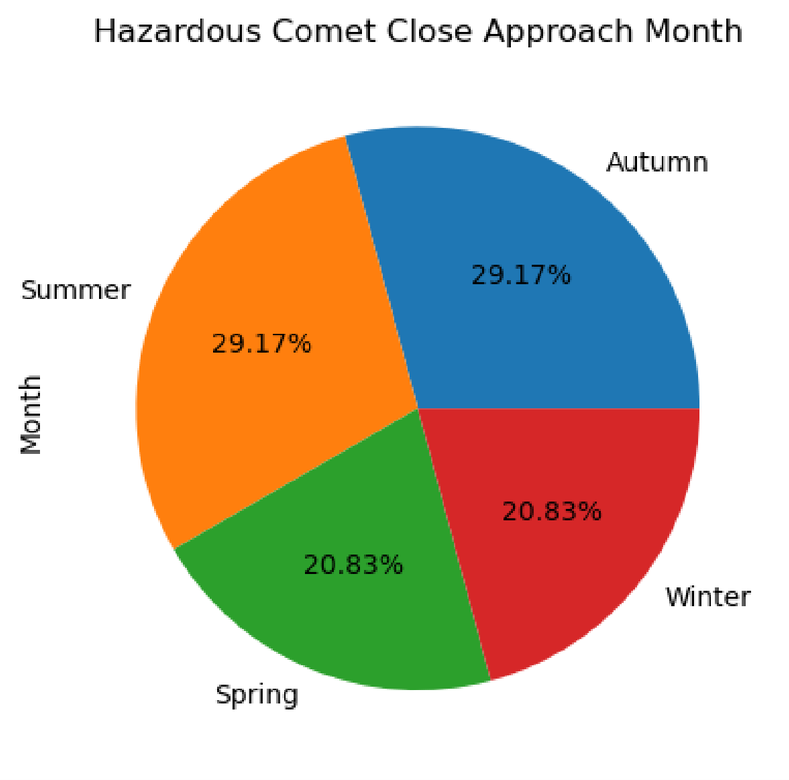
	Inclination	Semi Major Axis	Close Approach Date	Mean Motion
0	22.456259	2.626090	1999-04-15	0.231601
1	22.659883	3.259846	2010-02-08	0.167459

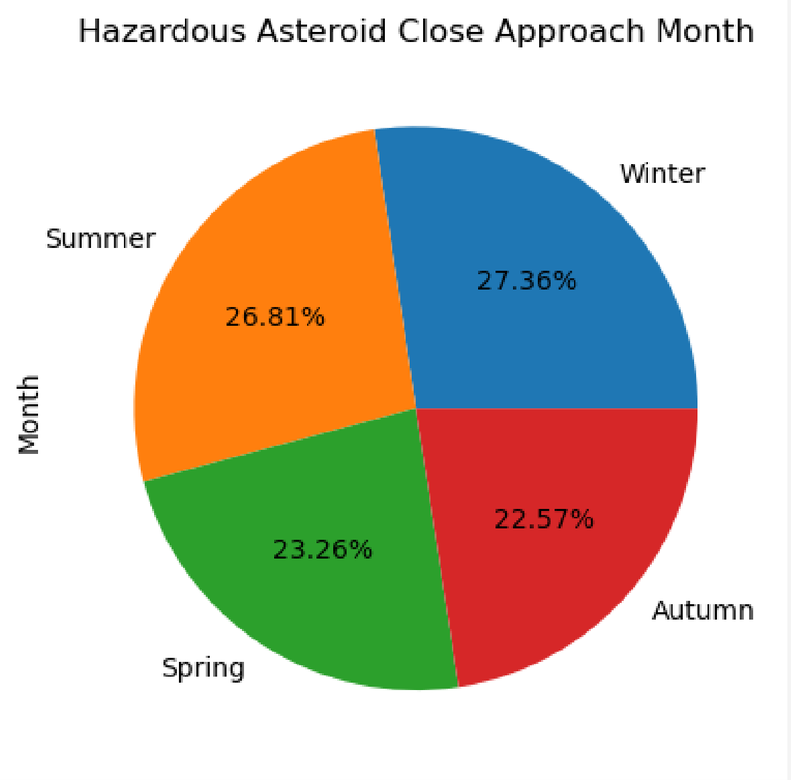
PHO COMETS

From the table, there are 2 instances of comets passing Earth's plane. One is in April 1999 and the other is in February 2010. For the year 1999, though there were no asteroid crossings, there was a comet crossing. Therefore, we can infer that from 1995 to 2016, Earth experienced no asteroid or comets crossing in the years 2004, 2008, and 2016.

Results

SEASONAL BEHAVIOR





PHO COMETS

From 1995 to 2016 and out of 24 instances, Earth experienced the most comet PHOs in Summer and Autumn

PHO ASTEROIDS

From 1995 to 2016 and out of 731 instances, Earth experienced the most asteroid PHOs in Winter

HYPOTHESIS CONCLUSIONS SO FAR

75%

ACCURACY

I predicted 3/4 hypothesis correctly

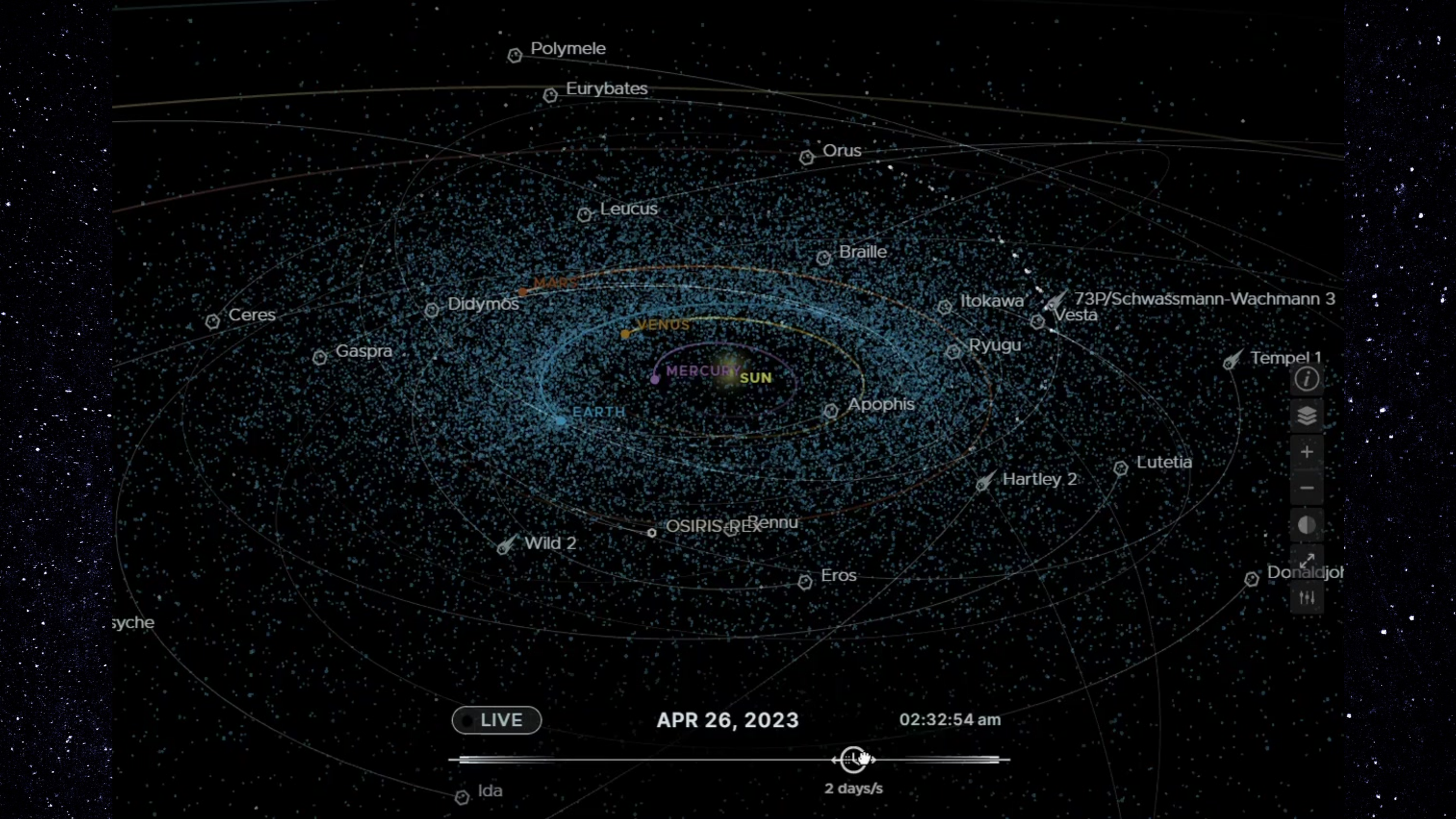
- There are more asteroid NEOs
- There are more asteroid NEOs that are PHOs
- PHOs should have a Semi Major Axis within Earths range

- Earth does not experience the most PHO interactions in just the Fall. It is actually Summer, Fall, and Winter

It's safe to conclude that Earth is in danger of PHO occurrences year round

Future Work

DOWNLOAD	MODEL	PREDICT
NASA's Jet Propulsion Laboratory hosts the Center for Near-Earth Object Studies where they collect hundreds of thousands of samples	Create a model that categorizes all current NEO data into its close approach seasons	Predict when future NEO close approach seasons will be based on the model


Polymele
Eurybates
Orus
Leucus
Braille
MARS
Didymos
Ceres
Gaspra
VENUS
MERCURY
SUN
EARTH
Itokawa
73P/Schwassmann-Wachmann 3
Vesta
Ryugu
Tempel 1
Apophis
Lutetia
Hartley 2
OSIRIS-REx
Bennu
Wild 2
Eros
Donaldjoh
syche
Ida
LIVE
APR 26, 2023
02:32:54 am
2 days/s

DEPARTMENT

Computational Applied Mathematics and Operations Research

EMAIL ADDRESS

kar15@rice.edu

thank you!