NEAR EARTH OBJECTS

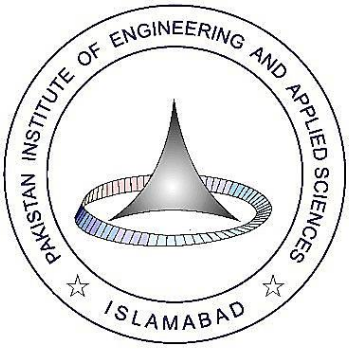
“Racing through space, asteroids are the shooting stars that never make a wish”

STATISTICAL ANALYSIS REPORT OF

NEAR EARTH OBJECTS

By:

Syed Haider Ali



DEPARTMENT OF PHYSICS & APPLIED MATHEMATICS
PAKISTAN INSTITUTE OF ENGINEERING & APPLIED SCIENCES,
NILORE, ISLAMABAD 45650

Executive Summary

This report presents a statistical analysis of Near Earth Objects (NEOs) with a focus on asteroids. The objectives of the report are to identify the types of asteroids that are most hazardous and analyze how different asteroid parameters affect the hazard potential. The data was cleaned and organized using Excel's power query, and exploratory data analysis (EDA) was performed using Python notebooks. The EDA included individual variable analysis and correlation analysis. The statistical analysis focused on the relationship between asteroid size and hazard potential, orbital characteristics and asteroid hazardous potential, asteroid orbit class type and their sizes and distances, asteroid types and time series analysis, and Sentry objects characteristics. Logistic regression modeling was also applied to the data. The limitations of the study include the use of only logistic regression and the need for better models. Overall, this report provides insights into the potential hazards of NEOs and highlights the importance of continued research and monitoring of these Objects.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

Ever since the discovery of objects in orbits that cross the orbit of the Earth, it has been recognized that such objects might collide with the Earth from time to time, and raises the obvious questions, how many of them are there, and how often do they collide with the Earth?

"NEO" stands for "near-Earth object", and includes any object, asteroid or comet, with perihelion distance less than 1.3 AU. Our analysis is of only the asteroid component of that population, called near-Earth asteroids (NEAs). These NEAs are separated by orbit class types under "Amors", "Apollos", "Atens" and "interior-to-earth".

1.2 Problem Statement

The complete statistical analysis of Near-Earth objects along with a implemented machine learning model is discussed in this report.

1.3 Objectives

- To conduct a complete statistical analysis of near-Earth objects and identify the types of asteroids that are most hazardous.
- To analyze how different asteroid parameters affect the hazard potential.
- To investigate the relationship between asteroid size and hazard potential.
- To examine the orbital characteristics of near-Earth asteroids and their association with hazard potential.
- To develop a machine learning model for predicting the hazard potential of near-Earth asteroids using logistic regression.

1.4 Scope

The scope of this report is to describe identify the types of asteroids that are most hazardous and analyze how different asteroid parameters affect the hazard potential.

1.5 Organization

The report is organized into five main sections. Section 1 provides an introduction to the report, including the background, problem statement, objectives, scope, organization, and limitations. Section 2 focuses on the Extract, Transform, Load (ETL) process, which is used to prepare the data for analysis. Section 3 covers the Exploratory Data Analysis (EDA), including the initial inspection of the data, individual variable analysis, and correlation analysis. Section 4 presents the statistical analysis, including the relationship between asteroid size and hazard potential, orbital characteristics, asteroid orbit class types, and time series analysis. Section 5 covers the logistic regression modeling used in the study.

1.6 Limitations

The machine learning model used on the target variable is limited to Logistic regression for the scope of this report. Better models can be built to achieve a better fitting model to hazard potential.

2 Extract Transform Load

We transformed the raw data into an organized form using excel's power query. We went along by removing the following:

- Null values
- Unnecessary columns
- Duplicate records

Table 1: Raw data

id	neo_refer	name	name_lim	designatic	absolute_	is_potenti	is_sentry_	kilometer
2001981	2001981	1981 Mida	Midas	1981	15.22	TRUE	FALSE	2.4019
2002059	2002059	2059 Babo	Baboquiva	2059	15.97	FALSE	FALSE	1.700415
2002061	2002061	2061 Anza	Anza	2061	16.36	FALSE	FALSE	1.420872
2002062	2002062	2062 Aten	Aten	2062	17.1	FALSE	FALSE	1.010543
2002063	2002063	2063 Bacc	Bacchus	2063	17.28	FALSE	FALSE	0.930154
2002100	2002100	2100 Ra-Sh	Ra-Shalon	2100	16.23	FALSE	FALSE	1.508534
2002101	2002101	2101 Ador	Adonis	2101	18.64	TRUE	FALSE	0.497227
2002102	2002102	2102 Tanta	Tantalus	2102	16	TRUE	FALSE	1.677085
2002135	2002135	2135 Arist	Aristaeus	2135	18.02	TRUE	FALSE	0.661538
2002201	2002201	2201 Oljat	Oljato	2201	15.25	TRUE	FALSE	2.368945
2002202	2002202	2202 Pele	Pele	2202	17.15	FALSE	FALSE	0.987541

Table 2: Cleaned data

absolute_m	is_potentially_l	is_sentry_object	mean_esitmate	orbit_c
15.22	TRUE	FALSE	3.886356325	APO
15.97	FALSE	FALSE	2.751329577	AMO
16.36	FALSE	FALSE	2.299019299	AMO
17.1	FALSE	FALSE	1.635093593	ATE
17.28	FALSE	FALSE	1.505021198	APO
16.23	FALSE	FALSE	2.440858575	ATE
18.64	TRUE	FALSE	0.804530692	APO
16	TRUE	FALSE	2.71357992	APO
18.02	TRUE	FALSE	1.070391414	APO
15.25	TRUE	FALSE	3.833033517	APO
17.15	FALSE	FALSE	1.597874319	AMO

3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. It is a good practice to understand the data first and try to gather as many insights from it. EDA is all about making sense of data in hand, before getting to work with it.

Tool used: Python Notebooks

3.1 Initial inspection of the data

Data was loaded into pandas' data frame object. Shape of the data frame and data types of its columns were studied. Several data types were observed including object, float, Boolean and integer. Speaking of data shape it contained 23984 observations and 12 variables.

```
Data columns (total 12 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   Name                              23984 non-null  object
 1   absolute_magnitude_h              23984 non-null  float64
 2   is_potentially_hazardous_asteroid 23984 non-null  bool
 3   is_sentry_object                  23984 non-null  bool
 4   kilometers_estimated_diameter_min 23984 non-null  float64
 5   kilometers_estimated_diameter_max 23984 non-null  float64
 6   mean_esitmated_diameter(Km)       23984 non-null  float64
 7   orbit_class_type                  23984 non-null  object
 8   perihelion_distance               23984 non-null  float64
 9   aphelion_distance                 23984 non-null  float64
 10  first_observation_year            23984 non-null  int64
 11  last_observation_year             23984 non-null  int64
dtypes: bool(2), float64(6), int64(2), object(2)
```

Figure 1: Data types of features

3.2 Individual variable analysis

Descriptive stats were applied on individual variables:

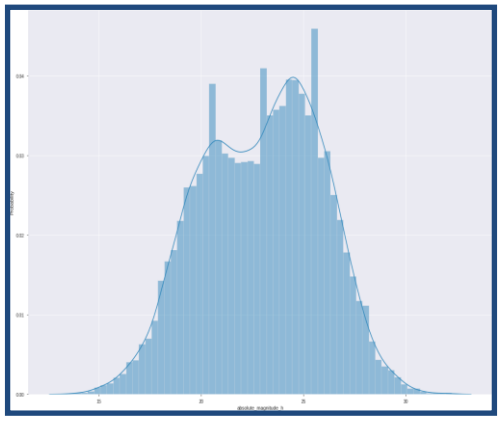
Table 3: Descriptive measures

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Name	23984	13162		2754	NaN	NaN	NaN	NaN	NaN	NaN	NaN
absolute_magnitude_h	23984.0	NaN	NaN	NaN	22.942412	2.953052	12.58	20.65	23.2	25.2	33.2
is_potentially_hazardous_asteroid	23984	2	False	21891	NaN	NaN	NaN	NaN	NaN	NaN	NaN
is_sentry_object	23984	2	False	22918	NaN	NaN	NaN	NaN	NaN	NaN	NaN
kilometers_estimated_diameter_min	23984.0	NaN	NaN	NaN	0.16776	0.293036	0.000609	0.024241	0.060891	0.196135	8.101305
kilometers_estimated_diameter_max	23984.0	NaN	NaN	NaN	0.375124	0.655248	0.001362	0.054205	0.136157	0.438571	18.115068
mean_esitmated_diameter(Km)	23984.0	NaN	NaN	NaN	0.271442	0.474142	0.000985	0.039223	0.098524	0.317353	13.108187
orbit_class_type	23984	4	APO	13239	NaN	NaN	NaN	NaN	NaN	NaN	NaN
perihelion_distance	23984.0	NaN	NaN	NaN	0.915189	0.232182	0.070431	0.785118	0.965252	1.068939	1.299988
aphelion_distance	23984.0	NaN	NaN	NaN	2.653448	4.469635	0.653754	1.706561	2.480455	3.398143	631.895456
first_observation_year	23984.0	NaN	NaN	NaN	2011.831179	8.195892	1931.0	2008.0	2014.0	2018.0	2020.0
last_observation_year	23984.0	NaN	NaN	NaN	2016.19217	4.839755	1979.0	2014.0	2018.0	2020.0	2022.0

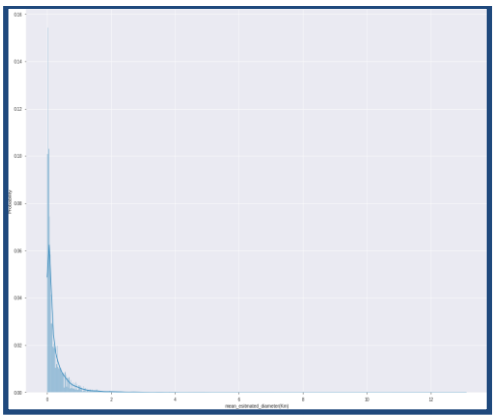
Drawn conclusions.

- Absolute Magnitude has mean and median close, and a very little difference between Q3 and max, which suggests resemblance with bell shape distribution.
- Skewness is present and therefore it confirms the presence of outliers in diameter parameters.
- Aphelion distance has mean and median close enough but Q3 and Max values are wildly different suggesting extreme skewness and presence of outliers.

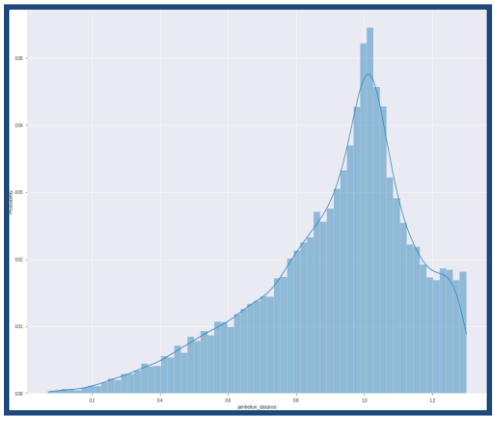
A: Absolute Magnitude



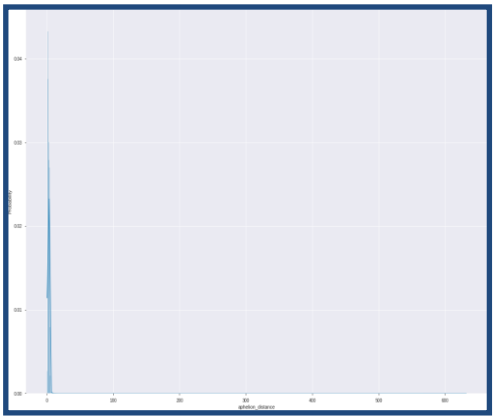
B: Mean Estimated Diameter (Km)



C: Perihelion Distance (AU)



D: Aphelion Distance (AU)



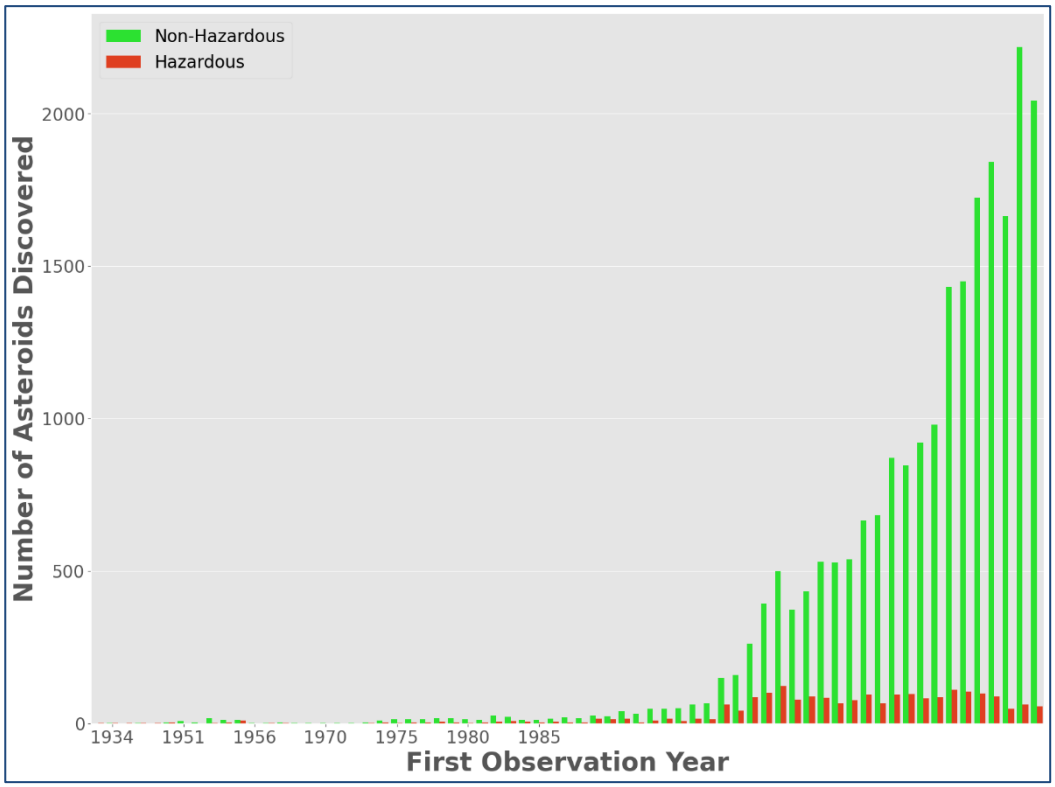


Figure 3: Asteroids discovered (a)

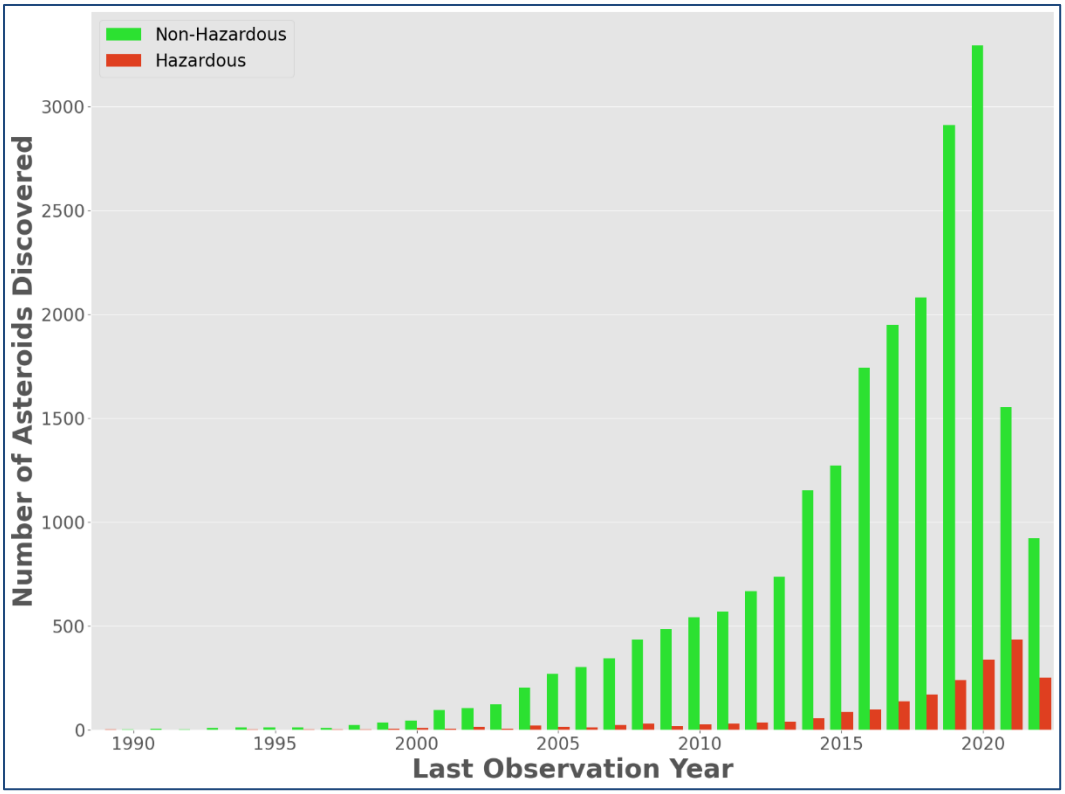


Figure 4: Asteroids discovered (b)

Sentry Objects **Orbit Class Type**

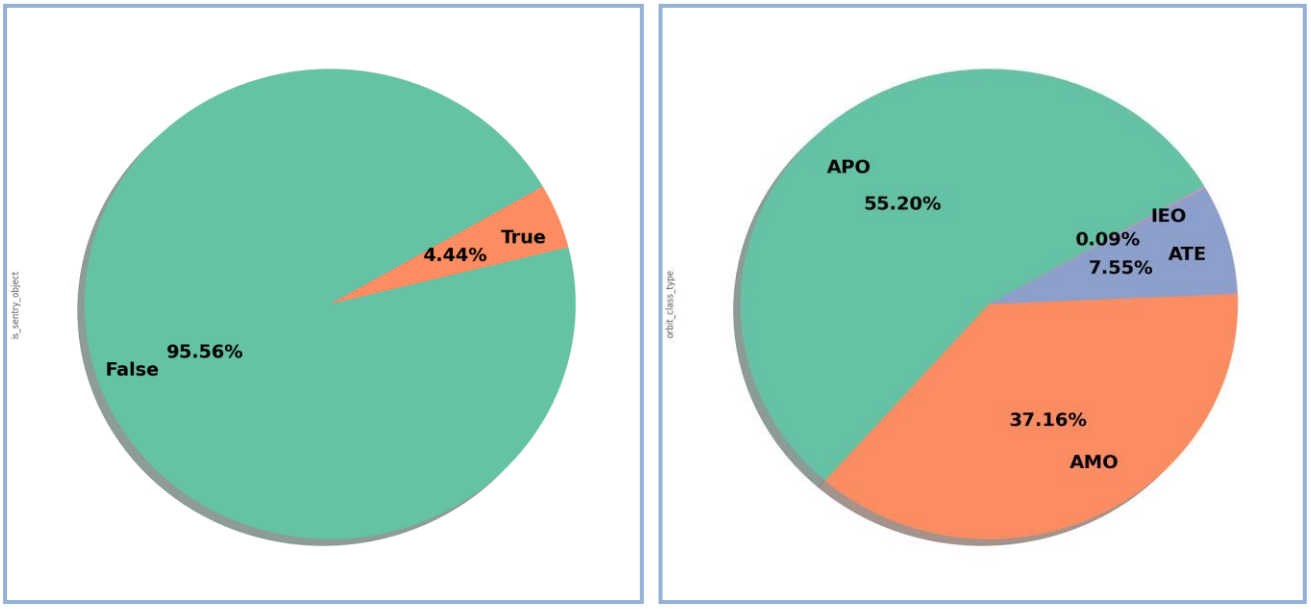


Figure 5: Pie charts of categorical variable

3.3 Correlation analysis

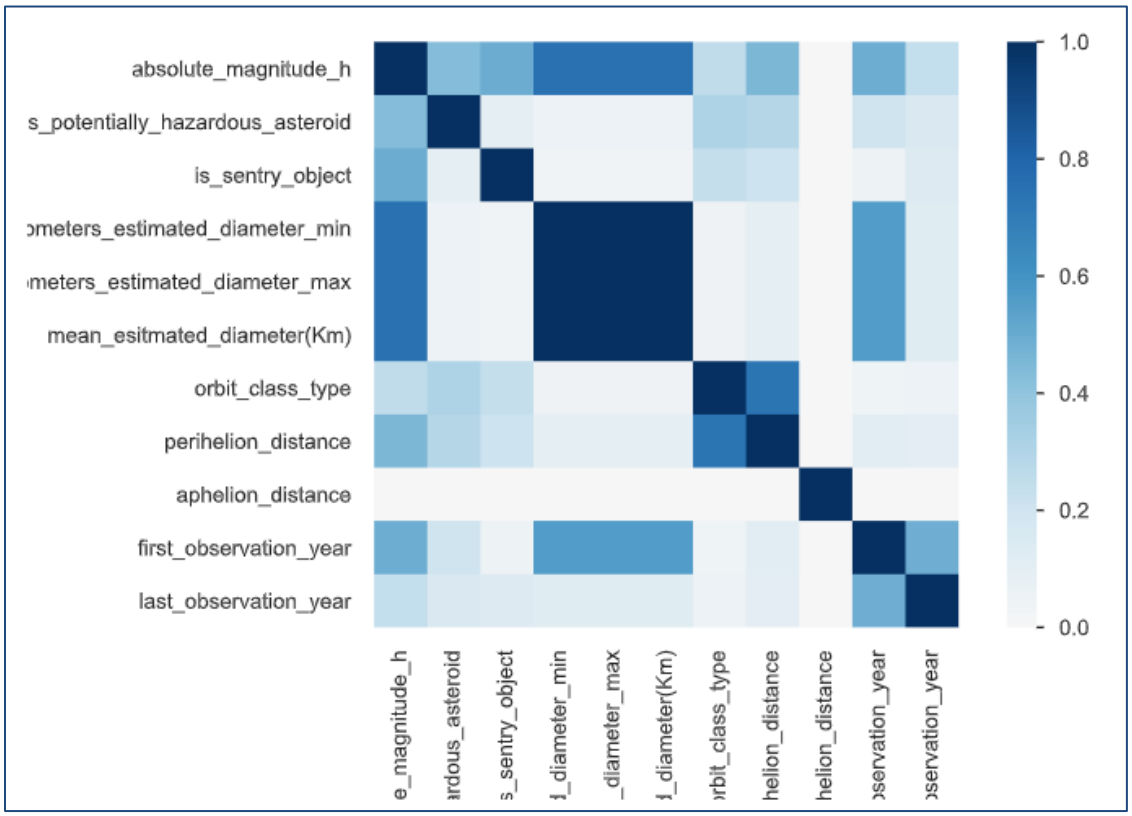


Figure 6: Phik (φk) correlation matrix

Phik (φk) is a practical correlation coefficient that works consistently between categorical, ordinal and interval variables, captures non-linear dependency and reverts to the Pearson correlation coefficient. These are useful features when studying the correlation matrix of variables with mixed types.

This correlation matrix signifies a strong correlation between estimated diameter of the asteroid and absolute magnitude of the asteroid. There also exist a strong correlation between the perihelion distance and the orbital class type.

4 Statistical Analysis

Statistical measures are usually performed in the following manner.

- Descriptive Analysis
- Hypothesis Testing
- Correlation Analysis
- Regression Modeling

Upon performing the EDA, inferences are likely to be drawn from the data.

4.1 Relationship between the asteroid size and hazard potential

The data extracted from the original data frame and separated by the hazardous Boolean key. The first five records are displayed.

	Estimated Diameter(Km)	Hazardous
0	3.886356	True
1	0.804531	True
2	2.713580	True
3	1.070391	True
4	3.833034	True

	Estimated Diameter(Km)	Hazardous
0	2.751330	False
1	2.299019	False
2	1.635094	False
3	1.505021	False
4	2.440859	False

Table 4: Asteroid size & hazard potential

For Hazardous Class:
Mean: 0.539886 Km
Variance: 0.309123 Km^2
n = 2093

For Non-Hazardous Class:
Mean: 0.245776 Km
Variance: 0.209214 Km^2
m = 21891

The mean diameter and the variability of sizes for hazardous asteroids is c clearly greater than nonhazardous asteroids. To get the statistical evidence and compare population means to significant confidence levels. A **right tailed t-test** was applied for unequal

variances to test if the average diameter of the asteroids that are dangerous is significantly greater than those who are not dangerous.

Group A: Hazardous Asteroids Diameters

Group B: Non-Hazardous Asteroids Diameters

In right tailed t-test for unequal variances:

$$H_0 : \mu_A \leq \mu_B$$

$$H_1 : \mu_A > \mu_B$$

$$T - statistic = \frac{\overline{x}_A - \overline{x}_B}{\sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{m}}}$$

$$D.F = \frac{\left(\frac{S_A^2}{n} + \frac{S_B^2}{m}\right)^2}{\left(\frac{S_A^4}{n^2(n-1)} + \frac{S_B^4}{m^2(m-1)}\right)}$$

$\overline{x} = Sample\ Mean$

$S = Sample\ Standard\ dev$

$n = Sample\ size\ for\ group\ A$

$m = Sample\ size\ for\ group\ B$

Result:

T-Statistic = 23.4538

P-value = 7.1835×10^{-110}

Since the p-value << 0.05, the null hypothesis was rejected and therefore there is a statistical proof that **Hazardous asteroids tend to be bigger in size than non-hazardous ones.**

4.2 Orbital characteristics & asteroid hazardous potential

The data extracted from the original data frame and separated by the hazardous Boolean key. The first five records are displayed:

Table 5: Orbit Class type vs. hazard potential

	orbit_class_type	is_potentially_hazardous_asteroid
0	APO	True
1	AMO	False
2	AMO	False
3	ATE	False
4	APO	False
...	...	...

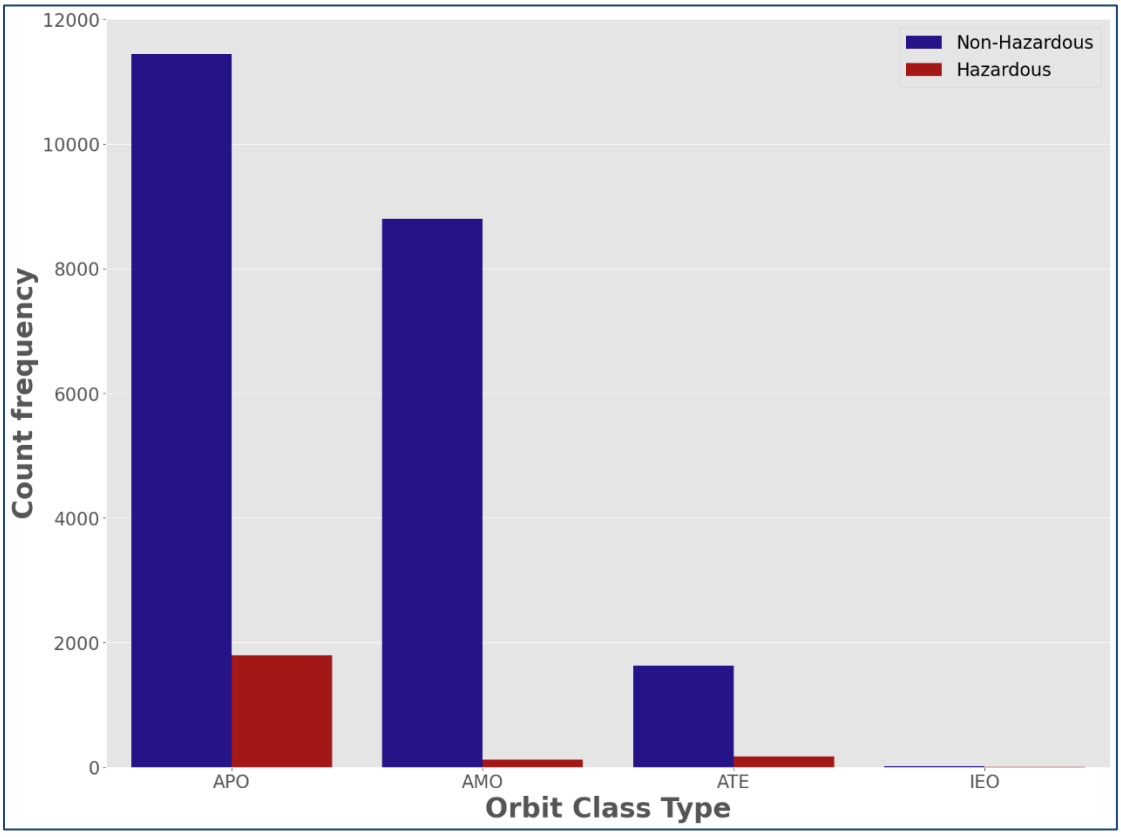


Figure 7: Pareto chart of orbit class type

Upon inspecting the count plot, it can be deduced that.

- APO constitutes the greatest number of near-earth asteroids.
- Majority of the NEA's are non-hazardous.
- Most of the hazardous are of the APO type followed by ATE, AMO and IEO

The following cross table shows this:

Table 6: Hazard potential vs. asteroid type

is_potentially_hazardous_asteroid	False	True
orbit_class_type		
AMO	8795	118
APO	11445	1794
ATE	1635	175
IEO	16	6

To check for a significant association between the orbit class and the hazard status of near-earth asteroids. **Chi-squared contingency test** was applied.

H_0 : The orbit class type of NEOs is independent of their hazard potential.

H_1:The orbit class type of NEOs is associated with their hazard potential.

$$\chi^2 = \frac{\sum(O_i - E_i)^2}{E_i}$$

$$E_i = \frac{(row\ i\ total \times column\ i\ total)}{n}$$

$\chi^2 = test\ statistic$

$O_i =$ the observed frequency for cell i

$E_i =$ the expected frequency for cell i

n = Total Sample size

Test Results:

Chi-test statistic = 1011.56

p-value = 5.5732×10^{-219}

Since the p-value << 0.05, Based on the results of the chi-square contingency test, there is strong evidence of an association between the orbit class type and hazard potential of Near-Earth Objects (NEOs).

4.3 Asteroid orbit class type and their sizes comparison

The data extracted from the original data frame and separated by the hazardous Boolean key. The first five records are displayed:

Table 7: Orbit Class type vs. asteroid sizes

	mean_esitmated_diameter(Km)	orbit_class_type
0	3.886356	APO
1	2.751330	AMO
2	2.299019	AMO
3	1.635094	ATE
4	1.505021	APO
...	...	...

Orbit Class Type	AMO	APO	ATE	IEO
Median Diameter(Km)	0.154008	0.07826	0.056695	0.744587

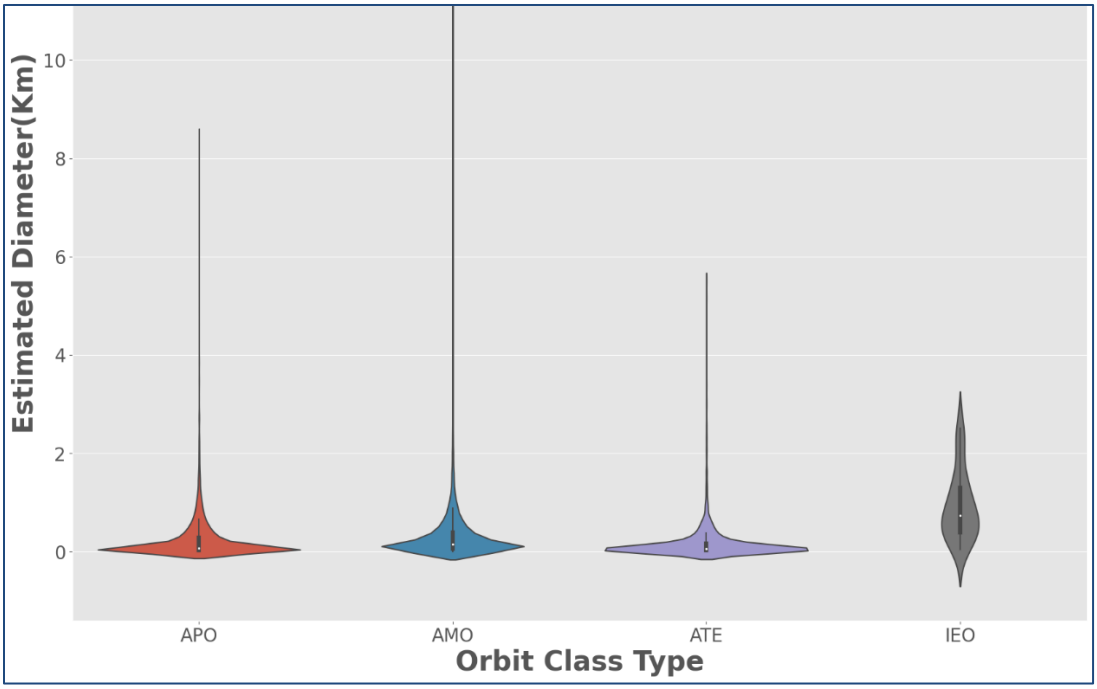


Figure 8: Violin plots of asteroid size and type

To check if the median of the asteroid group differs significantly from each other. We apply **Kruskal-Wallis test** on median asteroid diameter across different orbit classes.

$$H_0 : M_A = M_B = M_C = M_D$$

$$H_1 : M_A \neq M_B \neq M_C \neq M_D$$

Test Results:

Test statistics = 1245.94

p-value= 7.90×10^{-270}

Since the p-value << 0.05, the null hypothesis was rejected and therefore there is a significant difference in Median Sizes of Asteroid types.

4.4 Asteroid orbit class type and their distance comparison

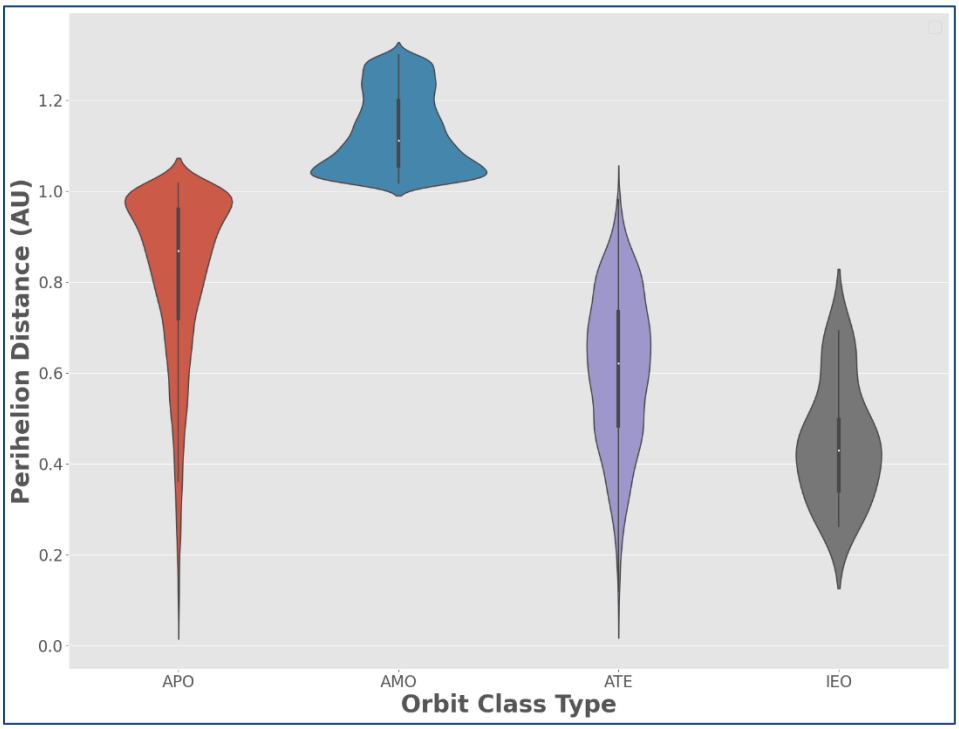


Figure 9: Violin plot of perihelion distance and asteroid type

The perihelion distance is the greatest for the AMOR type asteroids existing between 1AU to 1.3AU. The perihelion distance of the asteroid types compares as AMO > APO > ATE > IEO.

Table 8: Perihelion distance and orbit class type

	perihelion_distance	orbit_class_type
0	0.621512	APO
1	1.238537	AMO
2	1.050403	AMO
3	0.790185	ATE
4	0.701397	APO

Orbit Class Type	AMO	APO	ATE	IEO
Mean Perihelion Distance	1.129295	0.813852	0.607773	0.44698
Median Perihelion Distance	1.111453	0.869005	0.620669	0.42959

Over here we can observe how the means and the medians of different orbital types vary. Using one way ANOVA and Kruskal-Wallis test we can easily verify that the means and the medians of each group differ significantly respectively.

Test results:

Kruskal statistics = 17540.72

p-value = 0

One way ANOVA test statistic = 10058.99

p-value = 0

From the above results we deduce that the null hypotheses are rejected and therefore,

$$\boldsymbol{\mu_A \neq \mu_B \neq \mu_c \neq \mu_D}$$

$$\boldsymbol{M_A \neq M_B \neq M_C \neq M_D}$$

The pattern of the perihelion distances observed above holds.

4.5 Asteroid types and time series analysis

Table 9: Time series observation of asteroids

class_type	AMO	APO	ATE	IEO
first_observation_year				
1931	0.0	1.0	0.0	0.0
1934	1.0	1.0	0.0	0.0
1936	0.0	1.0	0.0	0.0
1937	1.0	2.0	0.0	0.0
1949	0.0	1.0	0.0	0.0
...	...	...	...	...
2016	679.0	1004.0	140.0	1.0
2017	642.0	1125.0	161.0	1.0

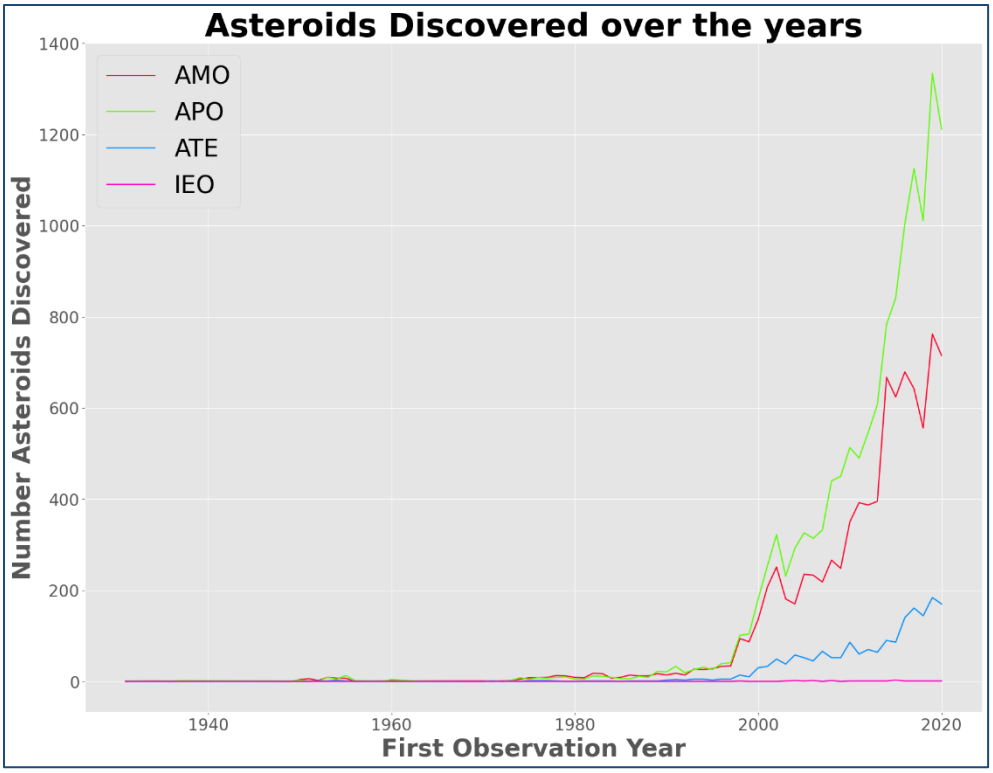


Figure 10: Time series plot of first observation year

Analyzing the time series, it is evident that over the recent years the number of asteroids discovered has increased exponentially. The most discovered asteroid types are in the order APO > AMO > ATE > IEO.

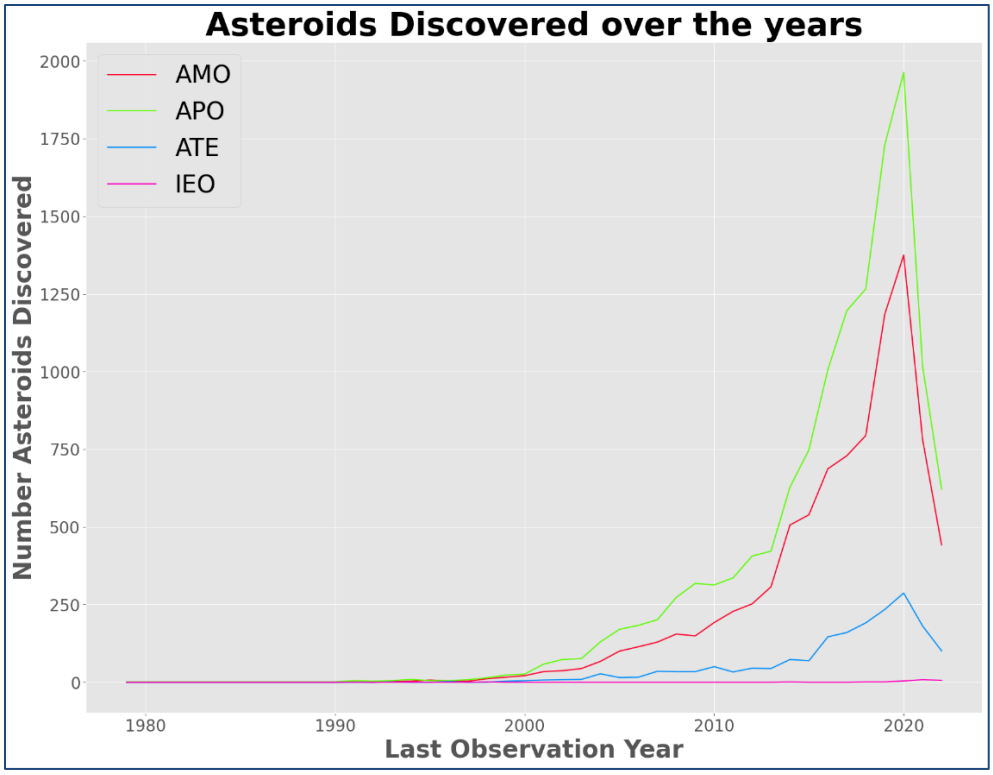


Figure 11: Time series plot of last observation year

The same trend follows for last observation year however it can be observed that the discoveries of asteroids peaked in 2020.

Comparing hazard potential with observation period:

Observation period is defined as the difference between the last observation year and first observation year. Separating out the observation duration using hazard potential is displayed using the following violin plot.

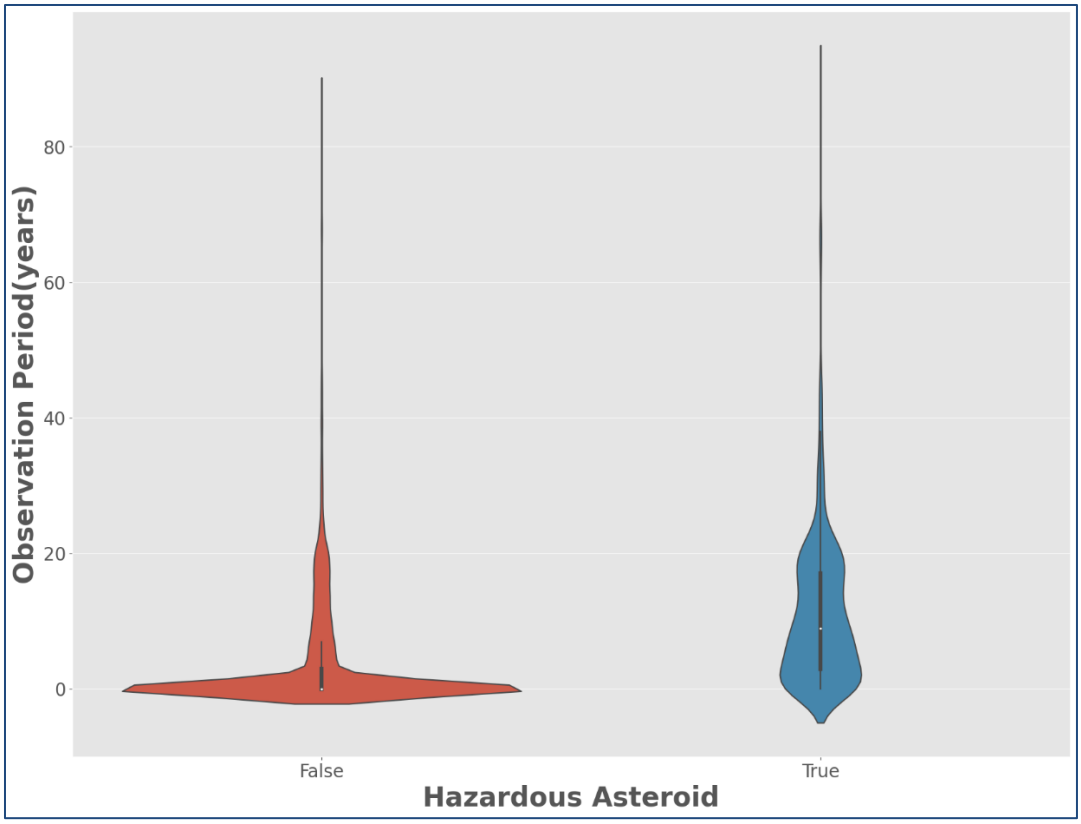


Figure 12: Violin plots of observation periods

The hazardous asteroids have a mean of 11.22 years and median of 9.0 years. Whereas the nonhazardous asteroids have a mean of 3.70 years and median of 0.0 years. Therefore, it's apparent that hazardous tend to have a higher average observation period than nonhazardous one's. To get significant proof of this we will run a independent right tailed t-test of unequal variances.

Group A: Hazardous Asteroids Observation Period

Group B: Non-Hazardous Asteroids Observation Period

In right tailed t-test for unequal variances

$$H_0 : \mu_A \leq \mu_B$$

$$H_1 : \mu_A > \mu_B$$

T-test statistic = 29.36

P-value = 1.896×10^{-161}

Since the p-value << 0.05, the null hypothesis was rejected and therefore there is a statistical proof that **Asteroids with longer observation periods tend to be Hazardous.**

4.6 Sentry objects characteristics

The term "**Sentry Object**" is simply a designation given to near-Earth objects that have been identified and tracked by astronomers. The purpose of tracking these objects is to determine if they have the potential to become hazardous in the future.

While some Sentry Objects have been classified as potentially hazardous asteroids (PHAs), meaning they have the potential to collide with Earth and cause significant damage, not all of them fall into this category. In fact, the majority of Sentry Objects are classified as non-hazardous, as their orbits do not pose a threat to Earth in the foreseeable future.

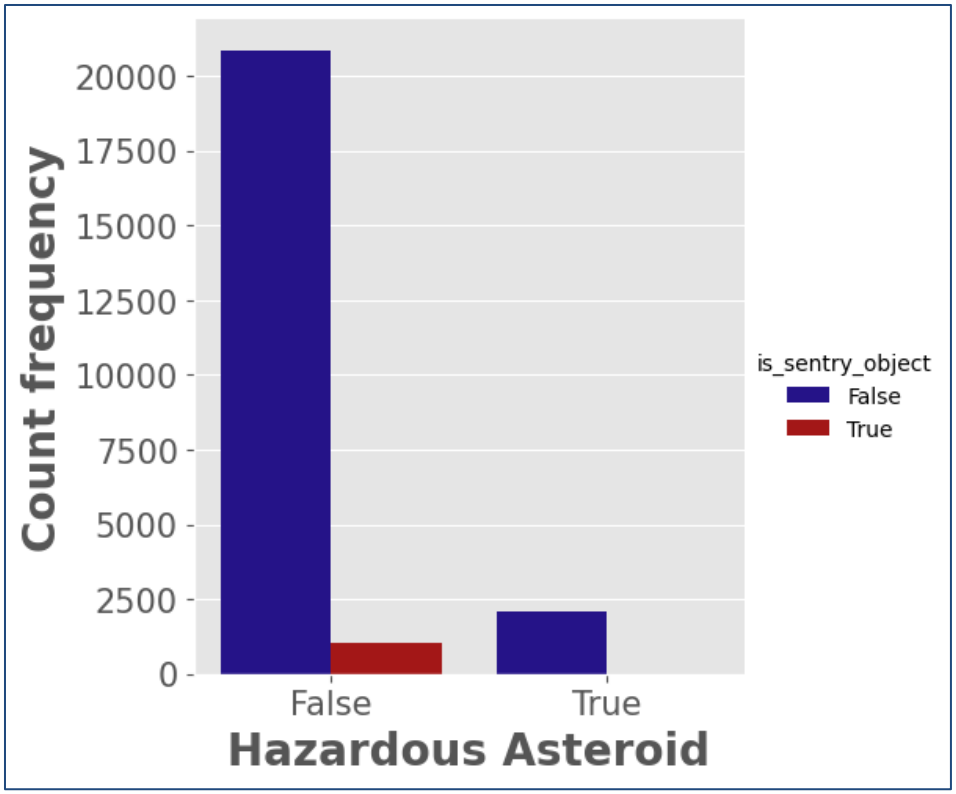


Figure 13: Pareto chart for sentry objects

Hazardous sentry objects are 11 in count while the rest of the sentry object are nonhazardous (count=1055).

Comparing perihelion distances of sentry and non-sentry objects:

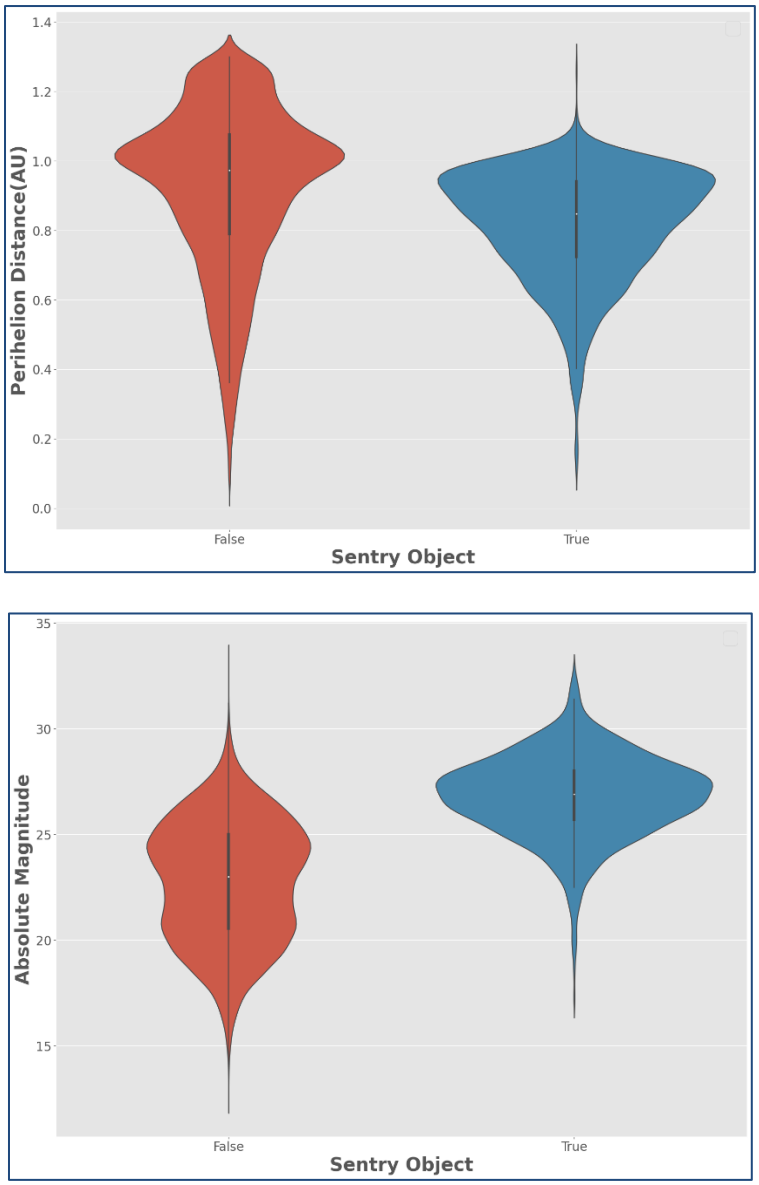


Figure 14: Violin plots of sentry objects and parameters

It is clear from these plots that most of the sentry objects are nonhazardous, have a median perihelion distance and absolute magnitude of 0.84806 AU and 26.9 respectively.

5 Logistic Regression Modeling

Up till now we have spent the better part of this report performing statistical tests on Hazard potential with other variables to find a relation with them. In order to quantify the mathematical relationship between the predictor variables and our target variable (Hazard Potential), we need some sort of mathematical model. A linear regression model is not suitable since the output variable is **binary.**

In linear regression:

$$y' = \beta_1 x_1 + \beta_2 x_2 + \dots \beta_k x_k$$

To map the output variable y' to a **binary variable** we can apply the transformation:

$$\hat{y} = \frac{e^{y'}}{1 + e^{y'}}$$

The parameters $\beta_1 - \beta_k$ are learned using a method called **Maximum likelihood Estimation (MLE)**

We fitted a binary logistic model to the data with target variable as hazard potential and the predictor variables included: absolute magnitude, mean estimated diameter, sentry object, and orbit class type.

The details of the results are provided in the appendix.

The model gave the following results:

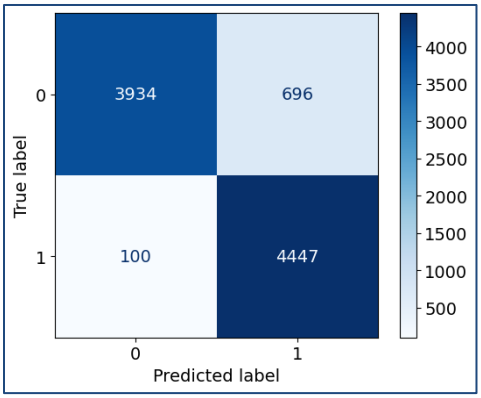


Figure 16: Confusion matrix

```
Classification report:
              precision    recall  f1-score   support

       False       0.98      0.85      0.91      4630
        True       0.86      0.98      0.92      4547

    accuracy                           0.91      9177
   macro avg       0.92      0.91      0.91      9177
weighted avg       0.92      0.91      0.91      9177
```

Figure 15: Classification report

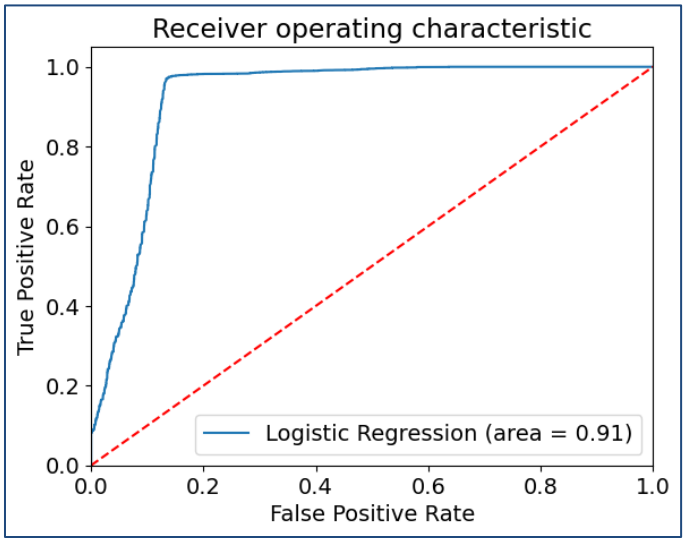


Figure 17: ROC curve

The above metrics signify that the Logistic Model is very strong and a good empirical model on the data. With the Logistic Regression results obtained in Minitab, we were able to obtain a detailed analysis of predictor variables including the Odds Ratio:

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
absolute_magnitude_h	0.3165	(0.3065, 0.3268)
mean_esitmated_diameter(Km)	0.0927	(0.0829, 0.1037)

Odds Ratios for Categorical Predictors

Level A	Level B	Odds Ratio	95% CI
is_sentry_object			
True	False	0.2654	(0.1516, 0.4644)
orbit_class_type			
APO	AMO	24.9591	(22.7680, 27.3611)
ATE	AMO	22.2411	(19.3252, 25.5970)
IEO	AMO	5.5139	(2.6730, 11.3741)
ATE	APO	0.8911	(0.7912, 1.0036)
IEO	APO	0.2209	(0.1074, 0.4546)
IEO	ATE	0.2479	(0.1195, 0.5141)

Odds ratio for level A relative to level B

Table 10: Odds Ratio Table for Logistic Regression

6 Conclusion

In conclusion, our statistical analysis of Near Earth Objects (NEOs) with a focus on asteroids has provided valuable insights into the potential hazards of these objects. Our findings indicate that asteroid size, orbital characteristics, orbit class type, asteroid type, and Sentry objects characteristics all play a significant role in determining the hazard potential of NEOs. We have identified the types of asteroids that are most hazardous and analyzed how different asteroid parameters affect their hazard potential.

Our logistic regression modeling also provided us with a useful tool to predict the hazard potential of asteroids based on their parameters. However, we acknowledge that the limitations of the study include the use of only logistic regression and the need for better

models. Therefore, continued research and monitoring of NEOs are essential to further our understanding of their behavior and potential hazards.

Overall, our analysis emphasizes the importance of preparedness for potential asteroid impacts and the need for ongoing efforts to track and monitor NEOs. Our findings can contribute to the development of effective strategies to mitigate the potential impact of asteroids and ensure the safety of our planet.

‘

7 Glossary

Astronomical Terms:

Near-Earth Object (NEO): Any object, asteroid or comet, with perihelion distance less than 1.3 astronomical units (AU) from the Sun.

Amors (AMO): Near-Earth asteroids (NEAs) with orbits that intersect the orbit of Mars, but not Earth.

Apollos (APO): Near-Earth asteroids (NEAs) with orbits that cross the orbit of the Earth.

Atens (ATE): Near-Earth asteroids (NEAs) with orbits that intersect the orbit of the Earth, but not at a distance less than that of the Earth's aphelion.

Interior-to-earth (IEO): Near-Earth asteroids (NEAs) that have orbits that lie entirely within the orbit of Earth and come within 0.983 astronomical units (AU) of the Sun at their closest approach.

Perihelion distance: The point in the orbit of a planet or other celestial body at which it is closest to the Sun.

Aphelion Distance: The point in an object's orbit around the sun when it is farthest from the sun.

Absolute magnitude: The measure of the intrinsic brightness of a near-Earth object, independent of its distance from Earth. Measured on a logarithmic scale.

Statistical Terms:

A right-tailed t-test for unequal variances is a statistical hypothesis test used to determine if there is a significant difference between the means of two independent groups, where the variances of the two groups are not assumed to be equal

The **p-value** is a statistical measure that represents the probability of obtaining a test statistic at least as extreme as the one calculated from the observed data, assuming the null hypothesis is true. It is used to determine the level of statistical significance of a hypothesis test.

Chi-Square Contingency Test is a statistical hypothesis test that is used to determine whether there is a significant association between two or more categorical variable in a contingency table. It compares the observed frequencies in a contingency table to the expected frequencies under the null hypothesis that there is no association between the variable.

One-way ANOVA (Analysis of Variance) is a statistical test used to determine whether three or more groups have significantly different means. It compares the variance within groups to the variance between groups to assess if there is a significant difference between them.

The Kruskal-Wallis test is a nonparametric statistical test used to determine whether there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable.

Maximum likelihood estimation (MLE) is a statistical method for estimating the parameters of a probability distribution by maximizing the likelihood function. The likelihood function represents the probability of observing a given set of data as a function of the parameters of the distribution

8 Appendix

The Complete statistical analysis along with machine learning models is organized in python Google Collaboratory notebooks:

Access the Notebooks