# "The Stellar Project: Searching for Earth's Next Home"


**Indian Institute of Information Technology ,Nagpur**

**Department of Computer Science and Engineering**

Team name-IIITian

Sumit Kumar
(9509214739)

Rakshit Modanwal
(9793572633)

Sujal Chondhekar
(7620648377)


## 14th February, 2025

Google Colab link: Stellar_analytics.ipynb

# INDEX TABLE

# Problem Statement/Objective

By the year **2224**, Earth has become uninhabitable due to severe environmental degradation, resource depletion, and climate collapse. With no viable solutions left on Earth, humanity must look beyond its home planet for survival. In response, the **Stellar Coalition**, a global alliance of nations, has launched the **Stellar Project**—a mission to identify exoplanets that could serve as a new home. While **hundreds of exoplanets** have been detected, not all possess the necessary conditions to sustain life. A systematic method is required to evaluate and rank these planets based on their habitability potential. This study aims to analyze and preprocess exoplanetary data, identify key planetary characteristics such as atmospheric retention and stability, and develop a **scientifically driven Habitability Index (HI)** to assess and classify exoplanets based on their suitability for sustaining life. The goal is to provide structured insights that can guide future space exploration and decision-making.

# Solution Approach (Statistical)

### Reason for Choosing a Statistical Approach Over Machine Learning

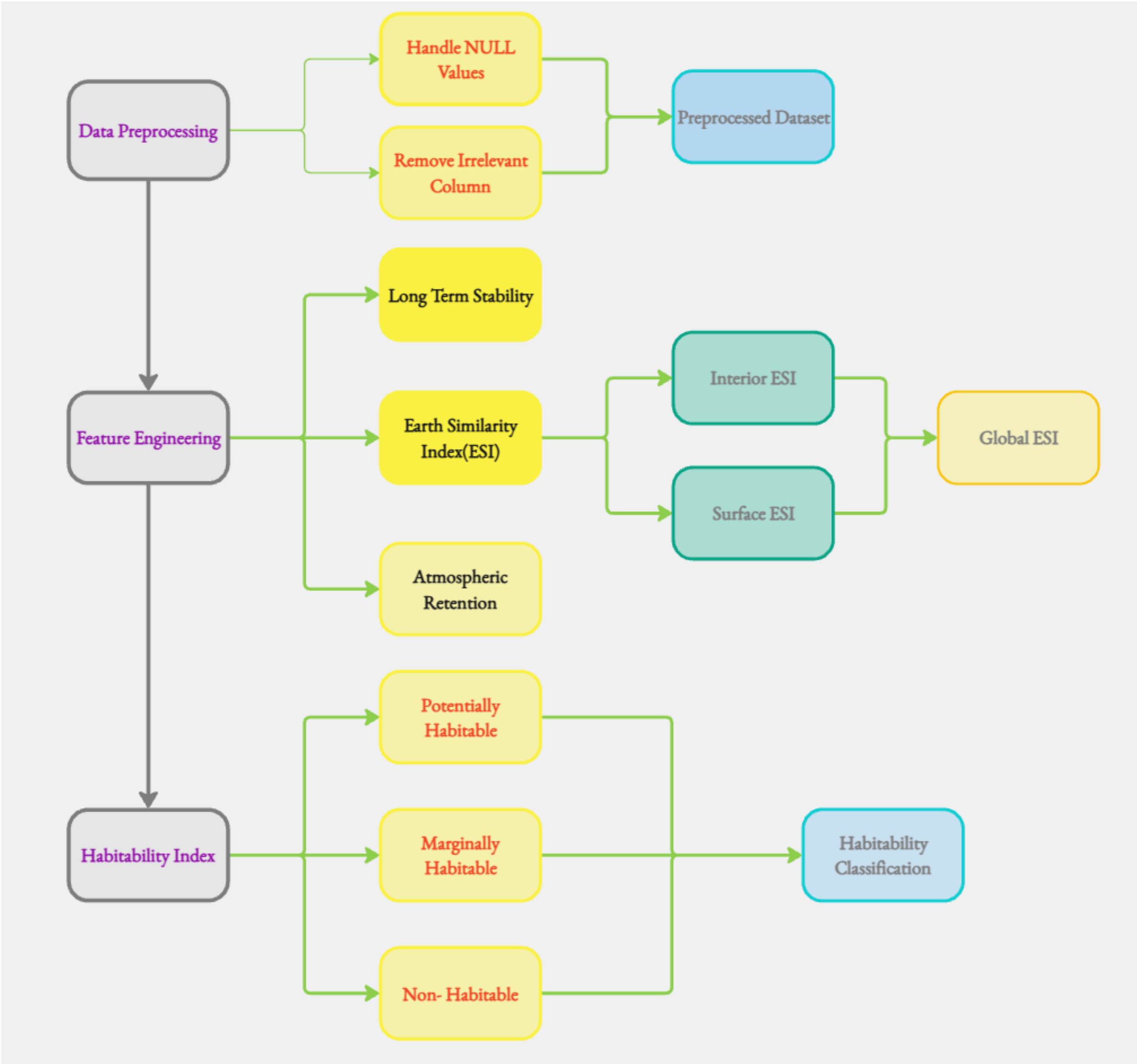
**A statistical approach was preferred over machine learning (ML) for evaluating exoplanet habitability due to the following reasons:**

1. **Simplicity & Interpretability** – The Habitability Index (HI) and other metrics are based on well-established scientific formulas, making the results easy to explain. ML models, in contrast, often act as "black boxes."
2. **Limited Dataset Size** – With only ~5400 data points, statistical methods are more reliable than ML, which requires large datasets to avoid overfitting.

3. **Domain-Specific Knowledge** – Metrics like HI and ESI are derived from astrophysical research, ensuring meaningful insights without the need for complex model training.
4. **Focus on Physical Parameters** – The approach directly incorporates measurable planetary features like escape velocity and orbital stability, ensuring scientifically sound calculations.
5. Transparency & Reproducibility – Statistical formulas are easily shared and verified, unlike ML models, which require hyperparameter tuning and computational power.
6. **Avoiding Overfitting** – ML models can struggle with noisy or incomplete exoplanet data, while a statistical approach ensures robustness.
7. **Scientific Insights Over Prediction** – The goal was to analyze habitability factors rather than just classify planets, making a statistical method more appropriate.

The **Stellar Project** follows a structured, step-by-step methodology to evaluate exoplanet habitability systematically. The approach begins with **data preprocessing**, ensuring clean and reliable data by handling missing values, removing redundant features, and standardizing planetary attributes. Next, key **habitability-related features** such as the **Earth Similarity Index (ESI)**, **long-term orbital stability**, and **atmospheric retention** are derived to enhance evaluation accuracy. These features play a crucial role in determining a planet's potential to support life. Finally, a **custom Habitability Index (HI)** is formulated, integrating these factors to rank exoplanets based on their habitability potential. This structured evaluation ensures a scientific, data-driven classification of exoplanets, guiding future exploration efforts.

# Flow chart



Data Preprocessing
Handle NULL Values
Remove Irrelevant Column
Preprocessed Dataset
Long Term Stability
Feature Engineering
Earth Similarity Index(ESI)
Interior ESI
Surface ESI
Global ESI
Atmospheric Retention
Potentially Habitable
Habitability Index
Marginally Habitable
Habitability Classification
Non- Habitable

# Data Collection & Preprocessing

## 3.1 Data Sources

The dataset used for this project is sourced from the Exoplanet Archive, which offers a wealth of detailed information about exoplanetary characteristics. The Exoplanet Archive is an excellent resource for accessing reliable data regarding various planetary attributes such as mass, radius, temperature, and orbital details. You can access the dataset directly at **Exoplanet Dataset**. For more details on the specific column headings, refer to the **Exoplanet Dataset Documentation**.

Additionally, to enhance the understanding of exoplanetary habitability, several supplementary resources provide valuable context:

- **Planetary Habitability**: Explore in-depth articles discussing the conditions necessary for life to exist on other planets, such as the atmospheric composition and climate stability.
- **How Scientists Search for Habitable Planets**: This resource delves into the methods used by researchers to identify planets in the habitable zone where life may thrive.
- **Earth Similarity Index (ESI)**: The Earth Similarity Index evaluates how similar a planet is to Earth, considering factors like size, temperature, and atmosphere. More information on the Earth Similarity Index can be found at the **Earth Similarity Index**.

The **Exoplanet Archive** remains the primary source of our data, ensuring that the foundation of the study is scientifically sound and comprehensive.

## 3.2 Handling Missing Values - Detailed Explanation

Handling missing values is a crucial part of data preprocessing, especially in datasets related to exoplanets. Missing data, if not handled correctly, can result in inaccurate analyses and lead to faulty conclusions. In this section, we will explain how missing values were dealt with, why certain columns were excluded, and how the missing data was imputed using scientific methods.

### Distribution of Missing Values

To efficiently handle the missing data, we categorized the columns based on the percentage of missing values:

- Columns with Missing Values Above 60%
  These columns were excluded due to a high proportion of missing data:
  *S_NAME_HD, S_NAME_HIP, P_OMEGA, S_TYPE.*
- Columns with Missing Values Between 10% and 60%
  These columns were considered for imputation to preserve valuable information:
  *P_TEMP_SURF, P_INCLINATION, S_AGE, P_ECCENTRICITY.*
- Columns with Missing Values Below 10%
  Minor gaps were handled using imputation techniques:
  *S_METALLICITY, P_PERIOD, S_LOG_G, S_LOG_LUM, P_TYPE_TEMP, P_FLUX, P_TEMP_EQUIL, S_LUMINOSITY, S_SNOW_LINE, S_RADIUS.*
- Columns with No Missing Values
  These columns had no missing data and were left unchanged:
  *P_HABZONE_CON, P_HABZONE_OPT, P_HABITABLE, S_CONSTELLATION, S_DEC_TXT.*

To better illustrate the distribution of missing values across different columns, refer to the figure 1.1 below. This visual representation highlights the extent of missing data, aiding in understanding the dataset's overall completeness.

| | Column | Missing Values | Percentage |
|---|---|---|---|
| S_NAME_HD | S_NAME_HD | 4628 | 82.657617 |
| S_NAME_HIP | S_NAME_HIP | 4579 | 81.782461 |
| P_OMEGA | P_OMEGA | 3940 | 70.369709 |
| S_TYPE | S_TYPE | 3578 | 63.904269 |
| P_TEMP_SURF | P_TEMP_SURF | 3158 | 56.402929 |
| P_INCLINATION | P_INCLINATION | 1311 | 23.414896 |
| S_AGE | S_AGE | 1207 | 21.557421 |
| P_ECCENTRICITY | P_ECCENTRICITY | 777 | 13.877478 |
| S_METALLICITY | S_METALLICITY | 433 | 7.733524 |
| P_PERIOD | P_PERIOD | 249 | 4.447223 |
| S_LOG_G | S_LOG_G | 246 | 4.393642 |
| S_LOG_LUM | S_LOG_LUM | 235 | 4.197178 |
| P_TYPE_TEMP | P_TYPE_TEMP | 234 | 4.179318 |
| P_FLUX | P_FLUX | 234 | 4.179318 |
| P_TEMP_EQUIL | P_TEMP_EQUIL | 234 | 4.179318 |
| S_LUMINOSITY | S_LUMINOSITY | 233 | 4.161457 |
| S_SNOW_LINE | S_SNOW_LINE | 233 | 4.161457 |
| S_RADIUS | S_RADIUS | 232 | 4.143597 |
| S_ABIO_ZONE | S_ABIO_ZONE | 223 | 3.982854 |
| S_MAG | S_MAG | 219 | 3.911413 |
| S_TEMPERATURE | S_TEMPERATURE | 219 | 3.911413 |
| S_TYPE_TEMP | S_TYPE_TEMP | 194 | 3.464904 |
| S_DISTANCE | S_DISTANCE | 21 | 0.375067 |
| P_HILL_SPHERE | P_HILL_SPHERE | 12 | 0.214324 |
| P_MASS | P_MASS | 7 | 0.125022 |
| P_TYPE | P_TYPE | 7 | 0.125022 |
| P_GRAVITY | P_GRAVITY | 7 | 0.125022 |
| P_POTENTIAL | P_POTENTIAL | 7 | 0.125022 |
| P_ESCAPE | P_ESCAPE | 7 | 0.125022 |
| P_DENSITY | P_DENSITY | 7 | 0.125022 |
| P_RADIUS | P_RADIUS | 7 | 0.125022 |
| P_SEMI_MAJOR_AXIS | P_SEMI_MAJOR_AXIS | 4 | 0.071441 |
| S_TIDAL_LOCK | S_TIDAL_LOCK | 4 | 0.071441 |

Figure 1.1

## Categorization of Relevant and Irrelevant Columns for Habitability Index

To assess exoplanet habitability, we need to distinguish between relevant and irrelevant columns. Relevant columns contribute to determining a planet's potential for habitability, while irrelevant columns do not provide meaningful insights for this purpose.

Relevant Columns (Contribute to Habitability Assessment)

- P_TEMP_SURF (Surface Temperature) – Crucial for determining if the planet falls within the habitable temperature range.
- P_TEMP_EQUIL (Equilibrium Temperature) – Helps estimate the planet's potential for supporting liquid water.
- P_FLUX (Incident Flux) – Determines the energy received from the host star, impacting surface conditions.
- P_ECCENTRICITY (Orbital Eccentricity) – Affects climate stability and seasonal variations.
- P_PERIOD (Orbital Period) – Helps understand the length of a planet's year, influencing climate cycles.
- P_SEMI_MAJOR_AXIS (Semi-Major Axis) – Indicates the planet's distance from its star, affecting temperature and radiation exposure.
- S_LUMINOSITY (Star Luminosity) – Determines the energy output of the host star, influencing planetary climate.
- S_RADIUS (Star Radius) – Affects stellar classification, impacting radiation and habitable zone estimation.
- S_TEMPERATURE (Star Temperature) – Determines the type of radiation emitted and habitable zone boundaries.
- S_AGE (Star Age) – Older stars may provide stable environments for life.
- S_DISTANCE (Star Distance from Earth) – Useful for observational feasibility but does not directly affect habitability.
- S_METALLICITY (Star Metallicity) – Higher metallicity can indicate better chances of forming rocky planets.
- S_SNOW_LINE (Snow Line Distance) – Determines the boundary where water and other volatiles can condense, affecting planetary formation.
- P_HABZONE_CON (Conservative Habitable Zone) – Directly related to habitability classification.
- P_HABZONE_OPT (Optimistic Habitable Zone) – Provides a broader estimate of habitability potential.
- P_HABITABLE (Habitability Classification) – Indicates whether a planet is officially classified as habitable.

Irrelevant Columns (Do Not Contribute to Habitability Assessment)

- **Identifiers & Names** – *S_NAME_HD, S_NAME_HIP, S_NAME, P_NAME, S_CONSTELLATION, S_CONSTELLATION_ABR, S_CONSTELLATION_ENG* (These are labels and do not affect habitability).
- **Positional Data** – *S_RA, S_DEC, S_RA_TXT, S_DEC_TXT, S_RA_STR, S_DEC_STR* (Used for locating celestial objects, not habitability).
- **Discovery Information** – *P_YEAR, P_DISCOVERY_FACILITY, P_DETECTION, P_UPDATE* (Relevant for historical tracking, not habitability).
- **Gravitational & Orbital Parameters** – *P_OMEGA, P_INCLINATION, P_TIDAL_LOCK, P_PERIASTRON, P_APASTRON, P_DISTANCE_EFF, P_HILL_SPHERE* (While important for orbital dynamics, they do not directly indicate habitability).
- **Gravitational Properties** – *P_GRAVITY, P_DENSITY, P_POTENTIAL, P_ESCAPE* (Impact planetary physics but are not primary habitability indicators).
- **Logarithmic & Abstract Star Properties** – *S_LOG_G, S_LOG_LUM* (More relevant for stellar classification than habitability).
- **Planetary Classification & Miscellaneous** – *P_TYPE, P_TYPE_TEMP, P_MASS_ORIGIN* (These categorize planets but do not assess habitability directly).

## Handling Missing Values and Removing Irrelevant Columns

After categorizing the dataset into relevant and irrelevant columns for habitability assessment, the next step involves:

- **Handling Missing Values** – Using scientific formulas instead of arbitrary imputation to ensure data accuracy.
- **Removing Irrelevant Columns** – Eliminating unnecessary features that do not contribute to habitability analysis.

These parameters were carefully selected to capture the essential characteristics that influence habitability, ensuring the **habitability index** reflects the most accurate and scientifically consistent conditions for sustaining life on exoplanets. Missing values were not arbitrarily replaced with basic statistical methods; rather, **astrophysical equations and empirical relationships** were used to ensure that the imputed values maintain physical consistency with established planetary and stellar models. This approach not only enhances the accuracy of habitability assessments but also maintains the scientific rigor required for meaningful comparisons between different exoplanets.

| Parameter | Missing Values (%) | Formula Used for Missing Values | Reason for Inclusion | Scientific Source |
|---|---|---|---|---|
| P_MASS (Planetary Mass) | 0.12% | $M_p = \frac{S_{MASS}}{S_{RADIUS} - C}$ | Determines if a planet is terrestrial or gaseous. Affects gravity and atmospheric retention. | Chen & Kipping (2017) |
| P_RADIUS (Planetary Radius) | 0.12% | $R_p = C + S_{MASS} \times M_p$ | Determines planet size; used to calculate density and escape velocity. | Chen & Kipping (2017) |
| P_PERIOD (Orbital Period) | 4.45% | $P = \sqrt{\frac{4\pi^2 a^3}{GM_*}}$ | Determines a planet's year length, affecting climate stability. | Kepler's Laws |
| P_SEMI_MAJOR_AXIS (Semi-Major Axis) | 0.07% | $a = \left(\frac{P^2 GM_*}{4\pi^2}\right)^{1/3}$ | Determines the distance from the host star, affecting temperature. | Kepler's Laws |
| P_ECCENTRICITY (Orbital Eccentricity) | 13.88% | $e = 0.29 \times \left(\frac{a}{1AU}\right)^{0.5}$ | Affects climate variations and long-term habitability. | Exoplanet Archive |
| P_ESCAPE (Escape Velocity) | 0.12% | $v_{esc} = \sqrt{\frac{2GM_p}{R_p}}$ | Determines atmosphere retention capability. | Newtonian Mechanics |
| P_POTENTIAL (Gravitational Potential) | 0.12% | $U = -\frac{GM_p}{R_p}$ | Affects atmospheric retention and surface conditions. | NASA Exoplanet Archive |
| P_GRAVITY (Surface Gravity) | 0.12% | $g = \frac{GM_p}{R_p^2}$ | Determines weight and potential habitability. | Newtonian Mechanics |

| | | | | |
|---|---|---|---|---|
| P_FLUX (Incident Flux) | 4.18% | $F = \frac{L_*}{4\pi a^2}$ | Determines energy received from the star. | Stefan-Boltzmann Law |
| P_TEMP_EQUIL (Equilibrium Temperature) | 4.18% | $T_e = T_* \times \sqrt{\frac{R_*}{2a}}$ | Determines baseline temperature before atmospheric effects. | Selsis et al. (2007) |
| P_TEMP_SURF (Surface Temperature) | 56.40% | $T_s = 9.650 + 1.096 \times T_e$ | Crucial for liquid water stability. | Schulze-Makuch et al. (2011) |
| P_HABITABLE (Habitability Index) | 0.00% | Derived from planetary properties | Indicates potential for life. | NASA Exoplanet Archive |
| P_DENSITY (Planetary Density) | 0.12% | $\rho = \frac{M_p}{\frac{4}{3}\pi R_p^3}$ | Determines the planet's composition (rocky, icy, or gaseous). | NASA Exoplanet Archive |
| S_TEMPERATURE (Stellar Effective Temperature) | 3.91% | $T_* = \left(\frac{L_*}{4\pi R_*^2 \sigma}\right)^{1/4}$ | Determines spectral type and energy output. | Stefan-Boltzmann Law |
| S_MASS (Stellar Mass) | 0.07% | No formula | Determines stellar lifetime and energy output. | Exoplanet Archive |
| S_RADIUS (Stellar Radius) | 4.14% | No formula | Used in calculating luminosity and habitable zone. | Exoplanet Archive |
| S_LUMINOSITY (Stellar Luminosity) | 4.16% | $L_* = L_\odot \times \left(\frac{S_{MASS}}{M_\odot}\right)^{3.5}$ | Determines energy output, affecting planet temperature. | Stefan-Boltzmann Law |

To maintain scientific integrity, missing values were primarily filled using astrophysical models rather than simple statistical imputation.

**Mean Imputation:** Used for parameters with minimal missing values (S_MASS, S_RADIUS, P_DISTANCE, and P_SEMI_MAJOR_AXIS).
**Physics-Based Models:**

- Orbital Parameters derived using Kepler's Third Law.
- Planetary Mass & Radius estimated based on stellar properties.
- Escape Velocity, Gravity, & Gravitational Potential computed using Newtonian Mechanics.
- Incident Flux, Equilibrium Temperature, and Surface Temperature determined using Stefan-Boltzmann Law.
- Eccentricity estimated using semi-major axis relationships.
- Luminosity & Stellar Temperature: Calculated using mass-luminosity relations and Stefan-Boltzmann equations.

**Several columns were removed because they did not contribute directly to exoplanet habitability:**

| *Column* | *Reason for Removal* |
|---|---|
| *S_NAME_HD, S_NAME_HIP* | Redundant catalog identifiers, unnecessary for habitability analysis. |
| *P_OMEGA* | Argument of periapsis has minimal impact on habitability. |
| *S_TYPE* | Spectral type is already covered by **S_TEMPERATURE** and **S_LUMINOSITY**. |
| *P_INCLINATION* | Does not directly impact habitability. |
| *S_AGE* | Age of the star is not critical for habitability. |
| *S_METALLICITY* | Metal content is important for planet formation but not immediate habitability. |
| *S_LOG_G, S_LOG_LUM* | Logarithmic gravity and luminosity are redundant with **S_MASS** and **S_LUMINOSITY**. |
| *S_MAG* | Apparent magnitude is not needed for habitability calculations. |
| *S_DISTANCE* | Distance from Earth does not affect the planet's habitability. |
| *P_HILL_SPHERE* | Relates to satellite retention, not planetary habitability. |

This analysis focused on selecting key habitability parameters and using astrophysical formulas to accurately handle missing data. By replacing missing values with scientifically derived formulas, the dataset maintains physical consistency. Redundant or irrelevant columns were removed, streamlining the data for habitability assessments. The final dataset, with no missing values, offers a solid foundation for reliable exoplanet habitability analysis.

# Feature Estimation

## 4.1 Derived Features (ESI, Stability, Atmospheric Retention)

## Earth Similarity Index

Reference:-**Schulze-Makuch D. et al., "A two-tiered approach to assess the habitability of exoplanets," Astrobiology, vol. 11, p. 1041, 2011a**

To calculate the Exoplanet Similarity Index (ESI) for exoplanets, we followed a systematic approach using several key planetary parameters. Here's a summary of the ESI calculations we performed:-

1. **Input Parameters:**
   We used four primary planetary characteristics:

- Radius (R)
- Density (ρ)
- Surface Temperature (Ts)
- Escape Velocity (ve)

2. **Normalization:**
   Each parameter was normalized using the following formula:

$$\mathrm{ESI}_x = \left(1 - \left|\frac{x - x_0}{x + x_0}\right|\right)^w$$

Where:

- x is the planetary parameter value
- x0 is the corresponding Earth reference value
- w is the weight assigned to the parameter

3. **Weights and Reference Values:**
   We used the following weights and Earth reference values:

- Radius: w = 0.57, x0 = 1.0 Earth radius
- Density: w = 1.07, x0= 1.0 Earth density
- Surface Temperature: w = 5.58, x0 = 288 K
- Escape Velocity: w = 0.70, x0= 1.0 Earth escape velocity

4. **Interior ESI:**
   We calculated the Interior ESI using radius and density:

$$ESI_I = \sqrt{ESI_R \times ESI_\rho}$$

5. **Surface ESI:**
   We computed the Surface ESI using surface temperature and escape velocity:

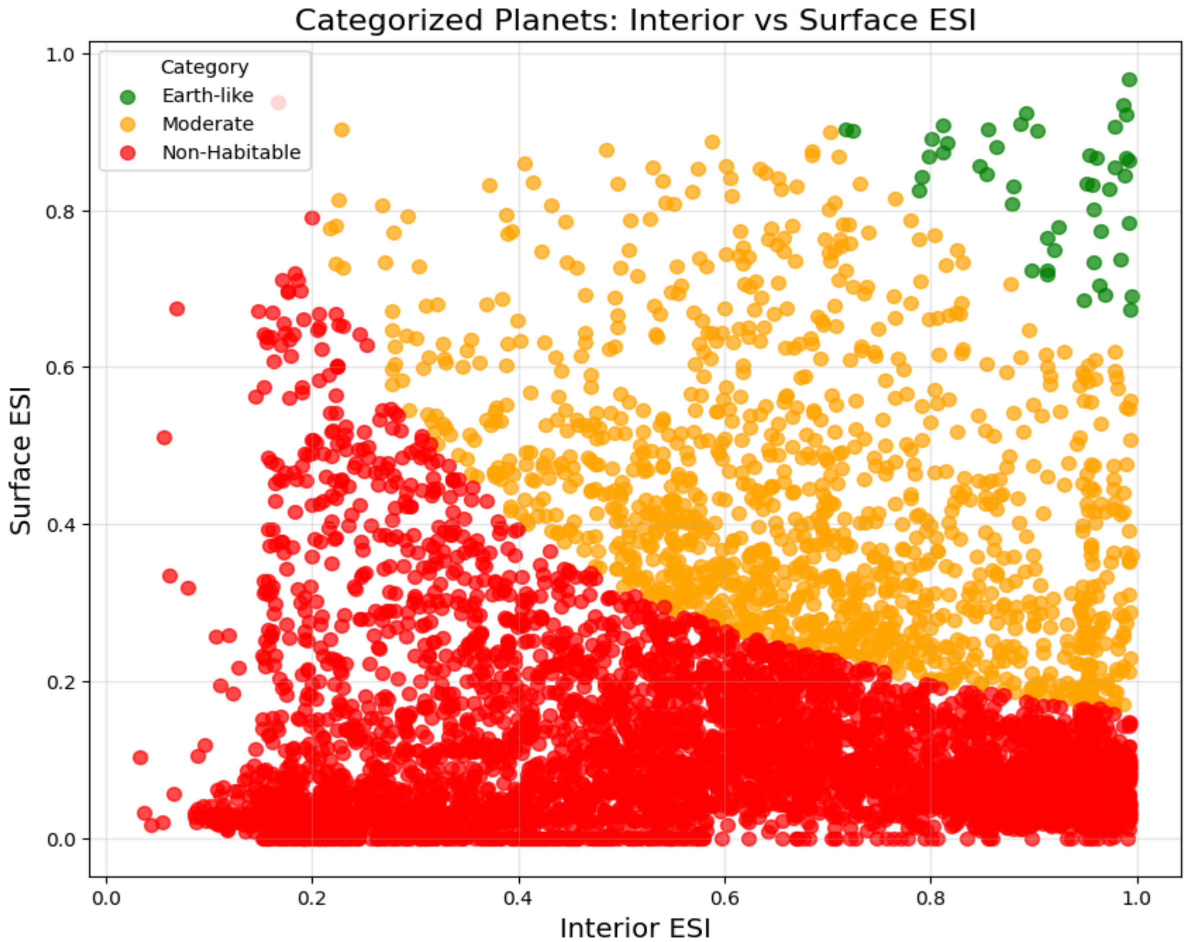
$$ESI_S = \sqrt{ESI_{Ts} \times ESI_{ve}}$$

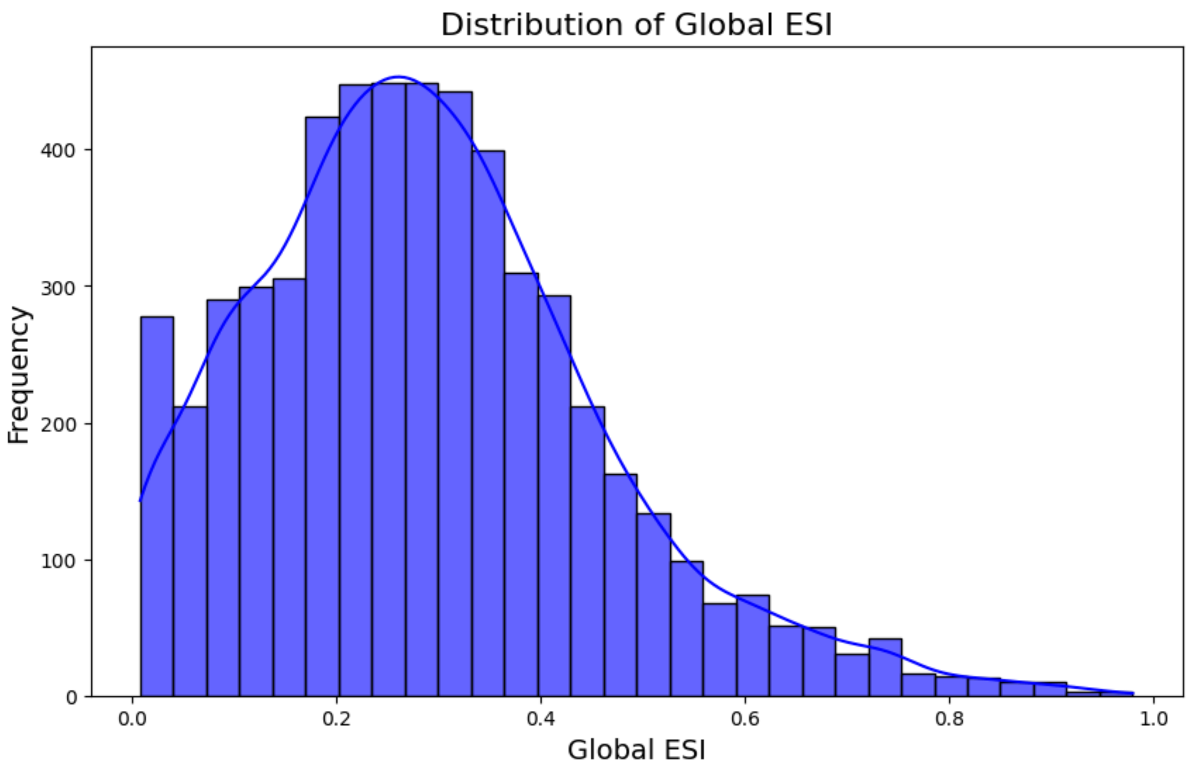
6. **Global ESI:**
   Finally, we calculated the Global ESI by combining Interior and Surface ESIs:

$$ESI = \sqrt{ESI_I \times ESI_S}$$

**A sample results of calculated ESI**

| Names | Radius (EU) | Density (EU) | Surface Temperature (K) | Escape Velocity (EU) | ESI Interior | ESI Surface | ESI Global |
|---|---|---|---|---|---|---|---|
| Teegarden's Star b | 1.05 | 1.002 | 3034 | 0.98 | 0.99 | 0.96 | 0.97 |
| TOI-700 d | 1.073 | 1.011 | 3459 | 0.973 | 0.986 | 0.933 | 0.95 |
| Kepler-1649 c | 1.06 | 1.007 | 3240 | 0.978 | 0.989 | 0.921 | 0.955 |
| TOI-700 e | 0.953 | 0.94 | 3459 | 0.973 | 0.977 | 0.906 | 0.941 |
| Proxima cen b | 1.03 | 0.97 | 2900 | 0.993 | 0.990 | 0.866 | 0.92 |


Categorized Planets: Interior vs Surface ESI
Category
Earth-like
Moderate
Non-Habitable
Surface ESI
Interior ESI
0.0
0.2
0.4
0.6
0.8
1.0


Distribution of Global ESI
Frequency
Global ESI
0
100
200
300
400
0.0
0.2
0.4
0.6
0.8
1.0

## Long Term Stability

Reference:-**Orbital stability analysis of hypothetical Earth-mass and Luna-mass moons in the Sagarmatha (HD 100777) star system**

Long-term stability is a measure of how stable a planet's orbit is over long timescales. This is critical for maintaining stable conditions that may support life. Planets in stable orbits are less likely to experience dramatic climate changes or disruptions from gravitational interactions.

$$\text{stability indicator} = \frac{a(1-e)}{M_p/M_*}$$

Where:

- a is the semi-major axis (P_SEMI_MAJOR_AXIS)
- e is the eccentricity (P_ECCENTRICITY)
- M_p is the planet mass (P_MASS)
- M* is the star mass (S_MASS)

Note:- The original Formulae is too long and complicated.This is simplified by taking some assumptions and approximations but this will not have much effect since its weight while calculating the Habituality Index is less.

## Atmospheric Retention

Atmospheric retention is a planet's ability to hold onto its atmosphere over extended periods. This depends on the planet's ability to counteract the tendency of atmospheric gases to escape into space due to thermal velocity.

The atmospheric retention is calculated using the following formula:

$$\text{Atmospheric Retention} = \text{normalized escape} \times (1 - \text{normalized temp})$$

Where:

- $\text{normalized escape} = \frac{\text{P ESCAPE}}{\max(\text{P ESCAPE})}$
- $\text{normalized temp} = \frac{\text{P TEMP SURF}}{\max(\text{P TEMP SURF})}$

This formula takes into account the planet's escape velocity and surface temperature, both normalized to their respective maximum values in the dataset. It provides a simplified estimate of the planet's ability to retain its atmosphere.

### 4.2 Habitability Index Calculation

## Habitability Index

The Habitability Index (HI) provides a quantitative measure of an exoplanet's potential to support life by combining three critical factors: Earth Similarity Index (ESI), Long-Term Stability, and Atmospheric Retention.

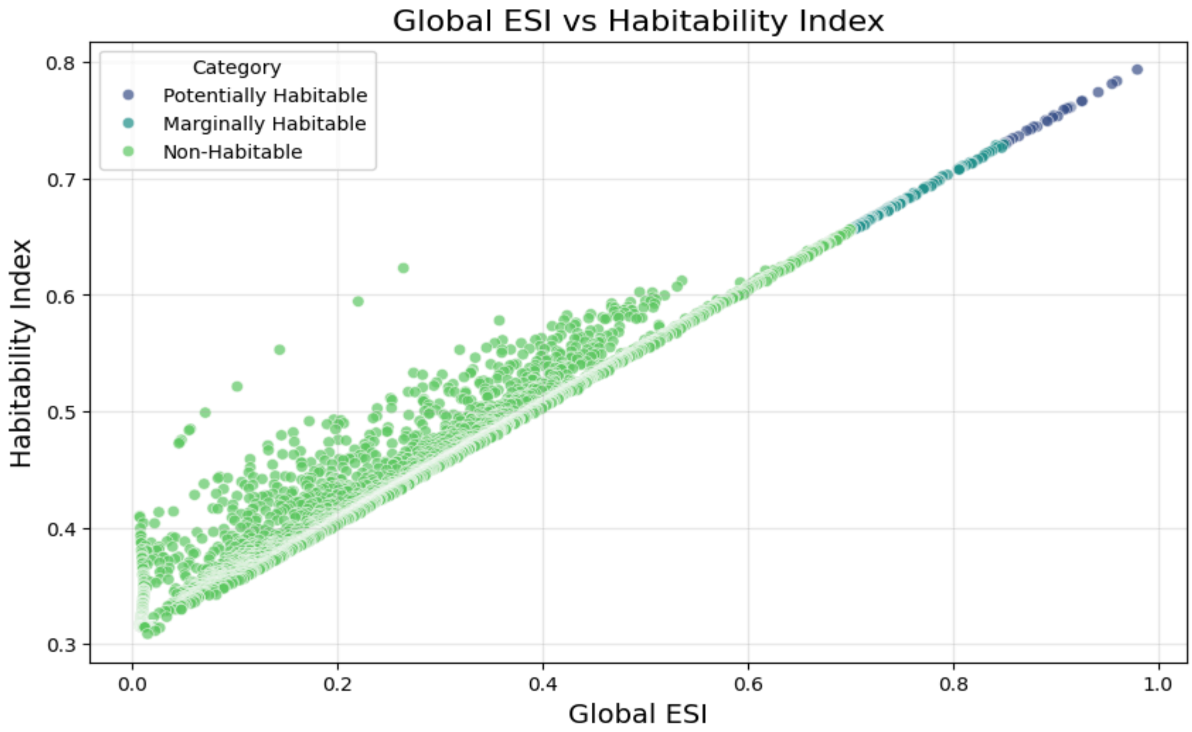
The formula used to calculate HI is:

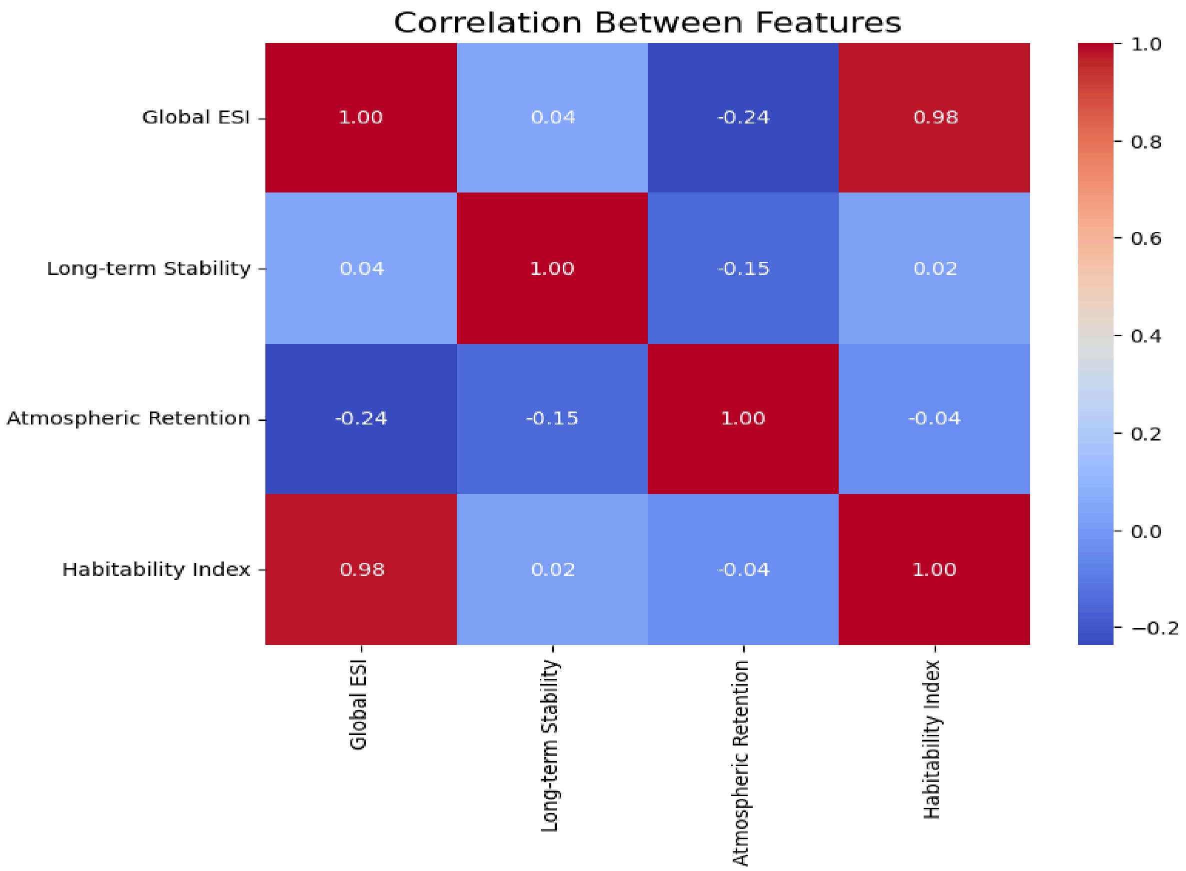
$$HI = (0.5 \cdot \text{esi global}) + (0.3 \cdot \text{long term stability}) + (0.2 \cdot \text{atmospheric retention})$$

In this more weight is given to esi global since most of the research works give importance to ESI and we have authentic formulae for this but for the other two the formulae are approximate formulae and they do not contribute significantly to HI

A Sample Table of Habitability Index Calculation

| Exoplanet | Global ESI | Long-term Stability | Atmospheric Retention | Habitability Index |
|---|---|---|---|---|
| Teegarden's Star b | 0.979 | 1 | 0.019315 | 0.79 |
| TOI-700 d | 0.959 | 0.99 | 0.019528 | 0.783 |
| Kepler-1649 c | 0.955 | 1 | 0.019 | 0.781 |
| TOI-700 e | 0.941 | 0.99 | 0.016 | 0.77 |
| Proxima Cen b | 0.92 | 1 | 0.01 | 0.76 |




This Correlation Matrix clearly shows that Global Esi is the most important feature for the habitability index.

# Results & Analysis

## 5.1 Classification Results

To categorize exoplanets based on their potential habitability, we employed a classification system that considers both the Habitability Index (HI) and the Earth Similarity Index (ESI). This approach allows for a more comprehensive assessment of a planet's potential to support life. The classification criteria are as follows:

1. Potentially Habitable:
   - Habitability Index (HI) ≥ 0.70
   - Earth Similarity Index (ESI) ≥ 0.85

   Planets in this category exhibit characteristics that are most similar to Earth and have the highest potential for habitability based on our current understanding.

2. Marginally Habitable:
   - Habitability Index (HI) ≥ 0.60
   - Earth Similarity Index (ESI) ≥ 0.70

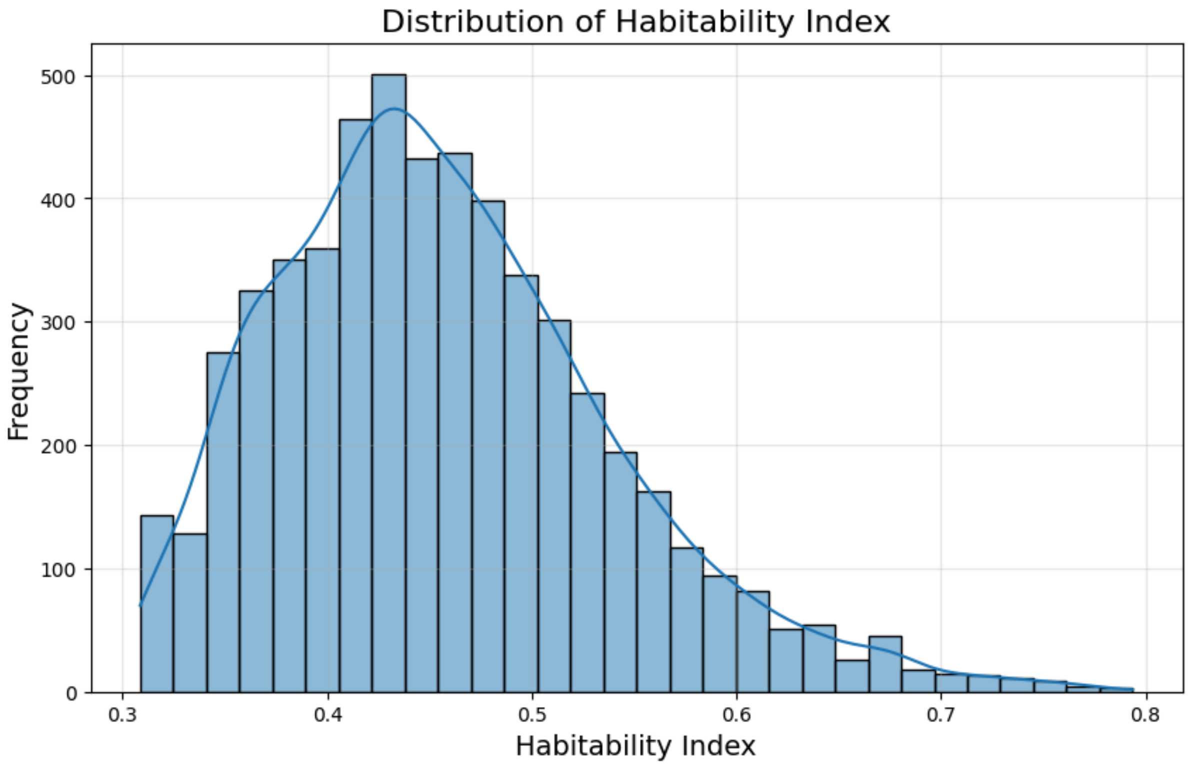
   These planets show promising characteristics for habitability but to a lesser degree than the potentially habitable category. They may require further investigation to determine their true potential for supporting life.

3. Non-Habitable:
   - Planets that do not meet the criteria for either of the above categories.

   These planets are considered unlikely to support life as we know it, based on their calculated HI and ESI values.

This classification system provides a nuanced approach to categorizing exoplanets, taking into account both their overall habitability potential (HI) and their similarity to Earth (ESI). It's important to note that these categories are based on our current understanding of habitability and the limitations of available data. Future discoveries and advancements in exoplanet science may lead to refinements in these classification criteria.

## 5.2 Habitability Trends




- The distribution is approximately right-skewed, with most planets having an HI between 0.4 and 0.6.
- Fewer planets have very high HI values (closer to 0.7 or 0.8), indicating that highly habitable planets are rare.
- Most of the exoplanets are Non-Habitable

Top 5 Potentially Habitable Planets:-

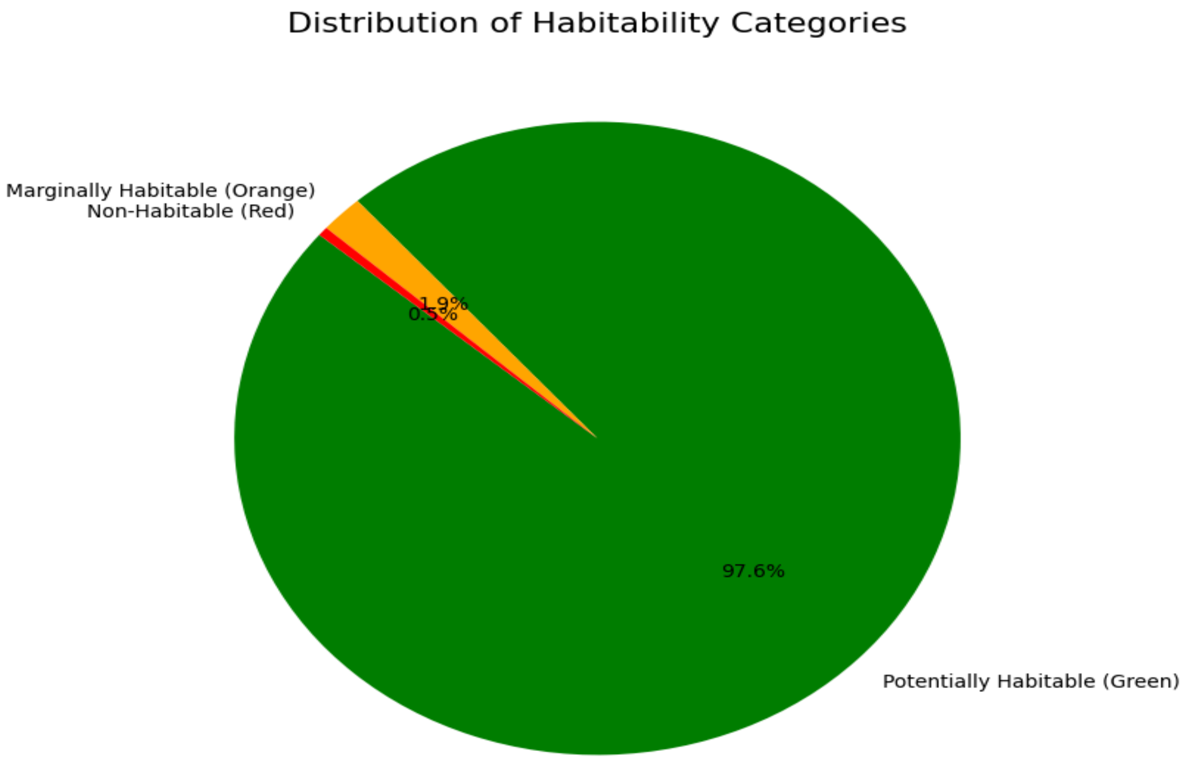
1) Teegarden's Star b
2) TOI-700 d
3) Kepler-1649 c
4) TOI-700 e
5) Proxima Cen b

## 5.3 Key Findings

1. Only **0.5%** planets belongs to Potentially Habitable Category

2. Only **1.9%** planets belongs to Marginally Habitable Category

3. Rest **97.6%** planets belongs to  Non Habitable Category



# Conclusion

**This study aimed to evaluate the habitability potential of exoplanets by using a combination of metrics, including the Habitability Index (HI), Earth Similarity Index (ESI), Atmospheric Retention, and Long-Term Stability. Based on these metrics, planets were classified into three categories: Potentially Habitable, Marginally Habitable, and Non-Habitable.**

The following conclusions can be drawn:

1. Classification Results:
    - The pie chart shows that the majority of planets in the dataset (~97.6%) fall into the Non-Habitable category.
    - Only a small fraction (~1.9%) are classified as Marginally Habitable, and an even smaller fraction (~0.5%) are identified as Potentially Habitable.
    - This highlights the rarity of Earth-like conditions in exoplanetary systems.
2. Key Metrics for Habitability:
    - The Habitability Index (HI) combines multiple factors such as ESI, atmospheric retention, and long-term stability to assess habitability.
    - Feature importance analysis revealed that parameters like surface temperature, escape velocity, and orbital stability play critical roles in determining habitability.
3. Earth-Like Planets Are Rare:
    - The study confirms that planets with high HI and ESI values (indicating Earth-like conditions) are extremely rare in the dataset.
    - Most planets lack the necessary conditions to support life as we know it.
4. Visualization Insights:
    - The pie chart effectively demonstrates the distribution of habitability classifications.
    - Other visualizations (scatter plots, histograms, and heatmaps) provided insights into correlations between planetary characteristics and their influence on habitability.

# References


1. Schulze-Makuch D. et al., “A two-tiered approach to assess the habitability of exoplanets,” Astrobiology, vol. 11, p. 1041, 2011a.

2. Orbital stability analysis of hypothetical Earth-mass and Luna-mass moons in the Sagarmatha (HD 100777) star system

3. Chen, J., & Kipping, D. 2017, ApJ, 834, 17

4. Kepler laws

5. Semi Major Axis

6. The exoplanet eccentricity distribution from Kepler planet candidates

7. Gravitation Potential

8. P_Escape Velocity

9. Equilibrium Temperature

10. Earth Similarity Index and Habitability Studies of Exoplanets