# Revealing sectrets of planets habitability using the NASA KOI Exoplanet Dataset

# 

## 1. Introduction

The Kepler Object of Interest (KOI) dataset is a vital resource in exoplanet research, providing data about potential exoplanets identified by NASA’s Kepler Space Telescope. This project aims to analyze the KOI dataset to identify patterns and predict habitability using a regression model. The analysis includes data preprocessing, exploratory data analysis (EDA), and predictive modeling.

## 2. Dataset Description

The dataset comprises 10,000 rows and multiple features, including:

* **KOI\_TEFF**: Stellar effective temperature.
* **KOI\_PRAD**: Planetary radius.
* **KOI\_SRAD**: Stellar radius.
* **KOI\_INSOL**: Stellar flux incident on the planet.
* **Disposition**: Classification of objects (e.g., confirmed planets, false positives).

The primary goal was to use these features to predict habitability, represented as a percentage score.

## 3. Data Preprocessing

Key preprocessing steps included:

* **Handling Missing Values**: Rows with missing values in critical features were removed.
* **Normalization**: Features were scaled to ensure uniformity for regression analysis.
* **Filtering Data**: Focused on confirmed planets by filtering rows with Disposition = confirmed.

## 4. Exploratory Data Analysis (EDA)

EDA revealed:

* **Distribution**: Most confirmed planets had KOI\_PRAD values below 10 Earth radii.
* **Correlations**: A significant correlation between KOI\_INSOL and habitability scores.
* **Visualizations**: Bar charts, histograms and pie charts were employed to understand feature relationships and distributions.

## 5. Feature Selection

No explicit feature selection methods were applied. Instead, the analysis focused on using important features identified in the literature and dataset documentation: KOI\_INSOL, KOI\_TEFF, KOI\_PRAD, and KOI\_SRAD.

## 6. Model Building

Machine learning model used for prediction:

* **Linear Regression**: Applied to predict habitability scores based on the selected features.

### Model Performance:

* Linear Regression Performance: Derived from notebook results.

## 7. Findings and Results

The analysis highlighted:

* **Key Factors of Habitability**: KOI\_INSOL and KOI\_TEFF emerged as the strongest predictors of habitability.
* **Model Insights**: Linear regression provided a straightforward approach to predicting habitability percentages.
* **Distribution of Habitability**: Only a small fraction of the celestial bodies analyzed had high habitability scores.

## 8. Conclusion

This project successfully utilized the KOI dataset to predict exoplanet habitability. The results demonstrate the potential of regression modeling in advancing exoplanet research. Future work can involve:

* Expanding the dataset with new observations.
* Employing advanced models for improved predictions.