

Exoplanet Dataset Overview

1. Introduction to the Dataset

This dataset contains detailed information about **exoplanets**, which are planets that exist outside our solar system. The data has been collected from astronomical observations and includes both **planetary** and **stellar** characteristics.

The dataset consists of approximately **39,508 rows** and multiple columns, where each row represents one discovered exoplanet and its associated properties. The dataset is useful for **astronomical analysis, data science exploration, and machine learning tasks**, such as predicting planet habitability.

2. What I Understand from the Dataset

From this dataset, I understand that it aims to describe:

- The **physical properties of exoplanets** (size, mass, orbital period, etc.)
- The **environment around the planet**, including the host star's properties
- Factors that help determine whether a planet could be **habitable**

The dataset can be used to:

- Compare planets with Earth
 - Study planetary systems
 - Build ML models to predict **habitability or classification of exoplanets**
-

3. Description of Main Attributes

- **Planet Radius (pl_rade)**
Shows the size of the planet and helps distinguish rocky planets from gas giants.
- **Planet Mass (pl_bmasse)**
Indicates how heavy the planet is and reflects its gravitational strength and ability to hold an atmosphere.
- **Planet Density (pl_dens)**
Helps determine the planet's composition, such as whether it is rocky or gaseous.
- **Orbital Period (pl_orbper)**
Represents the time taken by the planet to complete one orbit around its host star.

- **Semi-Major Axis (pl_orbsmax)**
Indicates the average distance between the planet and its star, which affects planetary temperature.
 - **Equilibrium Temperature (pl_eqt)**
Estimates the planet's surface temperature based on stellar energy and is a key factor for habitability.
 - **Host Star Temperature (st_teff)**
Shows how hot the star is, influencing the amount of energy received by the planet.
 - **Stellar Luminosity (st_lum)**
Represents the total energy output of the star and helps identify habitable zones.
-

4. Importance of the Dataset

This dataset is important because:

- It provides real astronomical data for analysis
- It can be used in **machine learning projects**
- It helps understand planetary diversity beyond Earth
- It supports research on **potentially habitable planets**