

# Dataset Description: NASA TESS Confirmed Planets

Name: Hemanth

Source: NASA Exoplanet Archive (Caltech/IPAC)

Date Retrieved: January 2, 2026

Subject: Confirmed Exoplanets discovered by the TESS Mission.

## 1. Overview

This dataset contains the best consensus parameters for all confirmed exoplanets discovered by the **Transiting Exoplanet Survey Satellite (TESS)**. It is a subset of the NASA Planetary Systems Composite Table, filtered to remove duplicate observations and non-TESS discoveries.

## 2. Dataset Statistics

- **Total Records:** 6,065 confirmed planets.(after removing duplicate rows)
- **Total Attributes:** 289 columns
- **Key Filters Applied:**
  - `disc_facility like '%TESS%'`: Restricted to TESS discoveries.
  - `default_flag = 1`: Selected the single definitive set of parameters for each planet.

## 3. Key Attributes

For habitability prediction, the following columns are the primary features:

|          |           |   |                 |
|----------|-----------|---|-----------------|
| Identity | pl_name   | Official Planet Name (e.g., TOI-700 d)    | string          |
| Geometry | pl_rade   | Planet Radius                             | Earth Radii(R+) |
| Geometry | pl_orbper | Orbital Period (Year length)              | days            |
| Energy   | pl_insol  | Insolation Flux (Stellar energy received) | Earth Flux (S+) |

|        |        |                         |           |
|--------|--------|-------------------------|-----------|
| Energy | pl_eqt | Equilibrium Temperature | Kelvin(K) |
|--------|--------|-------------------------|-----------|

|          |                 |  |                  |
|----------|-----------------|--|------------------|
| Star     | st_teff         | Stellar Effective Temperature          | Kelvin (K)       |
| Star     | st_rad          | Stellar Radius                         | Solar Radii (R.) |
| Metadata | discoverymethod | Method of detection (mostly "Transit") | Categorical      |

#### 4. Data Quality & Preprocessing Notes

- **Missing Data:** A significant portion of **pl\_bmasse** (Planet Mass) and **pl\_insol** (Insolation) is missing because TESS uses the Transit method, which measures radius rather than mass.
- **Imputation Strategy:** Missing masses may need to be imputed using Mass-Radius relations (e.g., Chen & Kipping, 2017).
- **Class Imbalance:** "Habitable" planets represent <1% of the dataset.

#### 5. Target Variable (Habitability)

This dataset does not contain a "Habitable" label. For this project, a binary target variable **is\_habitable** was engineered based on the **Conservative Habitable Zone** criteria:

1. **Rocky Composition:** Radius  $\leq 2.0 R_{\oplus}$
2. **Habitable Zone Flux:** Insolation between 0.2 and 1.5  $S_{\oplus}$