📊 Dataset Description: NASA Exoplanet Archive (Planetary Systems Dataset)

1. Overview

This dataset is obtained from the **NASA Exoplanet Archive** and contains detailed information about confirmed exoplanets and their host star systems. It provides scientifically validated data related to planetary discovery, orbital characteristics, physical properties, and observational methods.

The dataset is mainly used for **astronomical research, data analysis, and machine learning applications** related to exoplanet studies.

---

2. Source of Data

- **Provider:** NASA Exoplanet Archive
- **Website:** https://exoplanetarchive.ipac.caltech.edu
- **Generated Date:** February 13, 2026
- **Selection Criteria:** Default planetary systems detected mainly using the **TESS** mission

---

3. Dataset Size

- **Number of Records (Rows):** 39,386
- **Number of Attributes (Columns):** 289

Each row represents an **exoplanet entry** associated with a star system.

---

4. Data Format

- **File Type:** CSV (Comma-Separated Values)
- **Encoding:** Text-based
- **Structure:** Tabular format with headers
- **Comments:** Lines starting with # contain metadata and constraints

---

5. Main Components of the Dataset

The dataset is divided into multiple logical sections:

A. Planet Identification Information

These columns uniquely identify each planet.

Column	Description
rowid	Unique row identifier
pl_name	Planet name
hostname	Host star name
pl_letter	Planet designation letter
hd_name	Henry Draper catalog name
hip_name	Hipparcos catalog name
tic_id	TESS catalog ID

B. Stellar (Host Star) Properties

These attributes describe the star around which the planet revolves.

Column	Description
st_mass	Mass of the star (Solar mass)
st_rad	Radius of the star (Solar radius)
st_teff	Effective temperature (Kelvin)
st_lum	Stellar luminosity
st_age	Age of the star (billion years)
st_met	Metallicity

C. Planetary Physical Characteristics

These columns describe the physical properties of exoplanets.

Column	Description
pl_massj	Planet mass (Jupiter mass)
pl_radj	Planet radius (Jupiter radius)
pl_dens	Planet density

Column	Description
pl_eqt	Equilibrium temperature
pl_grav	Surface gravity

D. Orbital Parameters

These features define the orbit of each planet.

Column	Description
pl_orbper	Orbital period (days)
pl_orbsmax	Semi-major axis (AU)
pl_orbeccen	Orbital eccentricity
pl_orbincl	Orbital inclination
pl_orblper	Longitude of periastron

E. Discovery and Observation Data

These columns describe how and when planets were discovered.

Column	Description
discoverymethod	Method used for discovery
disc_year	Discovery year
disc_facility	Discovery facility
pl_pubdate	Publication date
releasedate	Release date

F. Data Quality and Validation

These fields help identify reliable records.

Column	Description
default_flag	Indicates best-quality record (1 = default)

Column	Description
pl_nnotes	Number of notes
rowupdate	Last update date

G. Observation Count Information

These columns indicate how many observations were made.

Column	Description
st_nphot	Number of photometric observations
st_nrvc	Number of radial velocity measurements
st_nspec	Number of spectra
pl_ntranspec	Transmission spectroscopy count

---

6. Missing Values

- Some columns contain **null (NaN) values**.
- Missing values occur because:
  - Certain measurements are not available.
  - Some planets have incomplete observations.
- These must be handled during data preprocessing.

---

7. Data Characteristics

- **Type:** Structured scientific dataset
- **Nature:** Observational and experimental
- **Time Span:** Multiple years (from early discoveries to 2026)
- **Update Frequency:** Periodically updated
- **Reliability:** Peer-reviewed and verified

---

8. Possible Applications

This dataset can be used for:

✓ Exoplanet classification
✓ Habitability analysis
✓ Machine learning prediction models
✓ Statistical analysis
✓ Astronomical research
✓ Discovery trend analysis

---

9. Limitations

- Some parameters are estimated.
- Observational bias may exist.
- Not all planets have complete data.
- Multiple records may exist for the same planet (with different sources).

---

10. Conclusion

The NASA Exoplanet Archive Planetary Systems dataset is a comprehensive and reliable collection of exoplanetary data. With 39,000+ records and 289 attributes, it provides in-depth information about planets, stars, and discovery methods. It is suitable for advanced research, data mining, and predictive modeling in astronomy.