SECTION-C  GROUP-09

# UNDERSTANDING EXOPLANET DISCOVERY PATTERNS FOR OPTIMIZED SPACE RESEARCH

## ASTRONOMY / SPACE SCIENCE ANALYTICS


Krishna: 2401010237

Neeraj Singh: 2401010297

Satyam Singh: 2401010430

Bineet Keshari: 2401010130

Antik Mondal: 2401010084

Rudraksh Rathod: 2401010396

# CONTEXT & PROBLEM

## SECTOR CONTEXT

Over the last few decades, thousands of exoplanets have been discovered using multiple detection methods. However, space observation resources such as telescope time, funding, and research manpower are limited. Understanding which discovery methods and planetary characteristics yield the most reliable results is essential for efficient future missions.

## PROBLEM STATEMENT

Despite a growing number of discovered exoplanets, not all discovery methods and planetary conditions contribute equally to reliable detections. There is a lack of clarity on which factors most strongly influence successful discoveries.

## OBJECTIVE

To analyze historical exoplanet data to identify key discovery patterns and influential planetary and stellar characteristics, enabling better decision-making for future space exploration strategies.

# DATA ENGINEERING

**DATA SOURCE**

- DATASET: EXOPLANETS (PLANETS OUTSIDE OUR SOLAR SYSTEM)
- SOURCE: KAGGLE
- SIZE: ~4,800 RECORDS | 10+ ATTRIBUTES
- TIME PERIOD: 1988 – 2009+

## KEY DATA QUALITY ISSUES IDENTIFIED

- Scientific uncertainty values in numeric fields (e.g., ± ranges, error bounds)
- Non-numeric characters in quantitative columns (commas, symbols, citations)
- Inconsistent categorical labels for discovery methods
- High missing-value concentration in mass, radius, and temperature
- Potential duplicate planet records

## CLEANING & STANDARDIZATION ACTIONS

- Extracted best-estimate numeric values for analysis
- Removed scientific notation, uncertainty ranges, and citations
- Standardized numeric formats and units
- Unified discovery method categories
- Ensured unique planet-level records
- Documented all transformations in a **Data Dictionary**

**Outcome**

✓ Dataset converted from scientific raw format → dashboard-ready analytical format
✓ Enabled reliable aggregation, comparison, and visualization

# KPI & METRICS FRAMEWORK

*Before analyzing, what did we decide to measure, and why?*

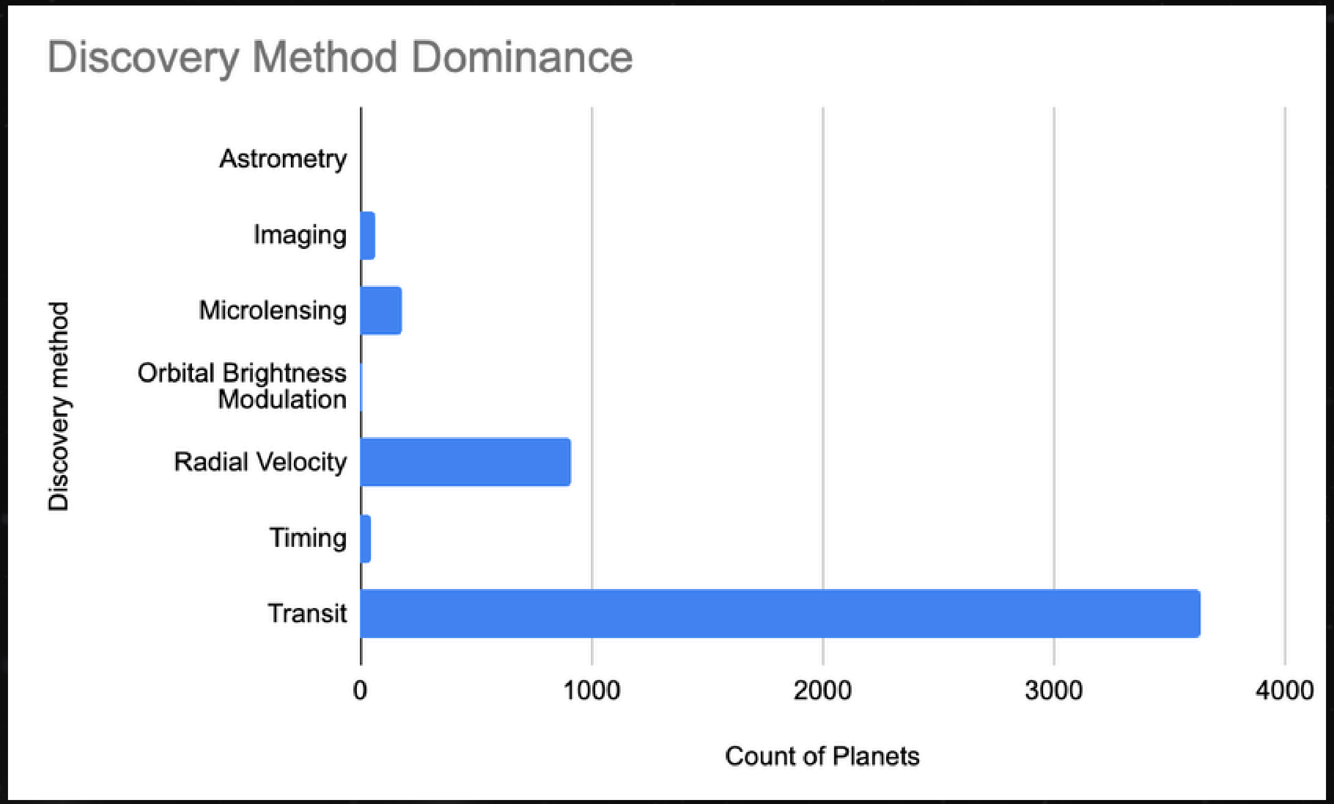
| KPI / Metric | Why it matters |
| --- | --- |
| 1. Discovery Count by Method | Measures which detection techniques yield the highest number of confirmed exoplanets. |
| 2. Discovery Trend Over Time | Tracks how exoplanet detection rates evolved with technological advancements. |
| 3. Planet Mass Distribution | Identifies observational bias toward detecting larger planets. |
| 4. Orbital Period Category Distribution | Evaluates detectability based on planetary orbital characteristics. |
| 5. Distance vs Discovery Method | Assesses how detection effectiveness changes with distance from Earth. |

**Why These KPIs Matter:**

These KPIs collectively evaluate discovery efficiency, bias, and scalability, ensuring that subsequent analysis directly supports strategic decision-making for future exoplanet observation missions.

# KEY INSIGHTS (EXPLORATORY DATA ANALYSIS)

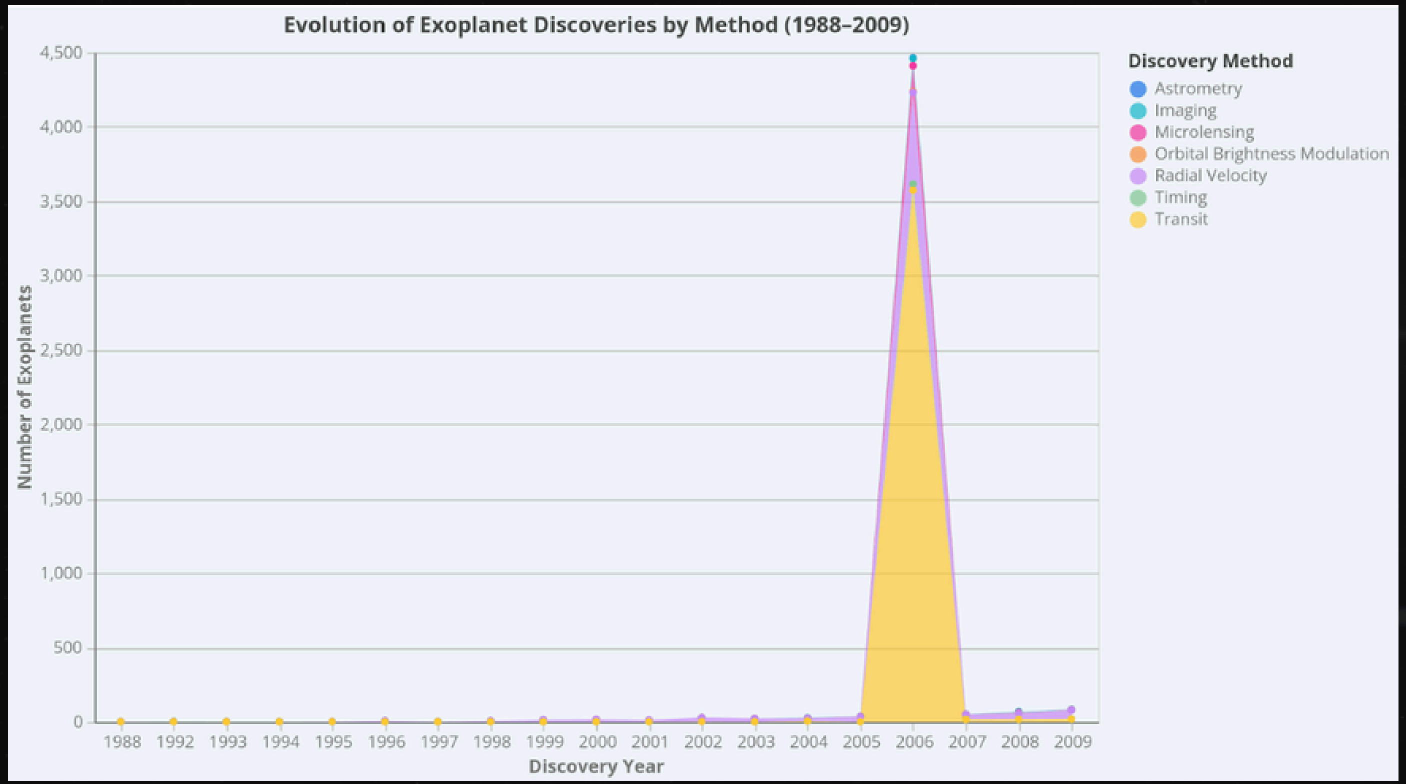
## INSIGHT 1: DISCOVERY METHOD EFFICIENCY




The **transit method dominates exoplanet discoveries,** accounting for most confirmed detections.
This indicates that **scalable photometric techniques offer higher observational efficiency** than resource-intensive methods such as radial velocity or imaging.

**Why this happens:** Transit surveys can monitor thousands of stars simultaneously.

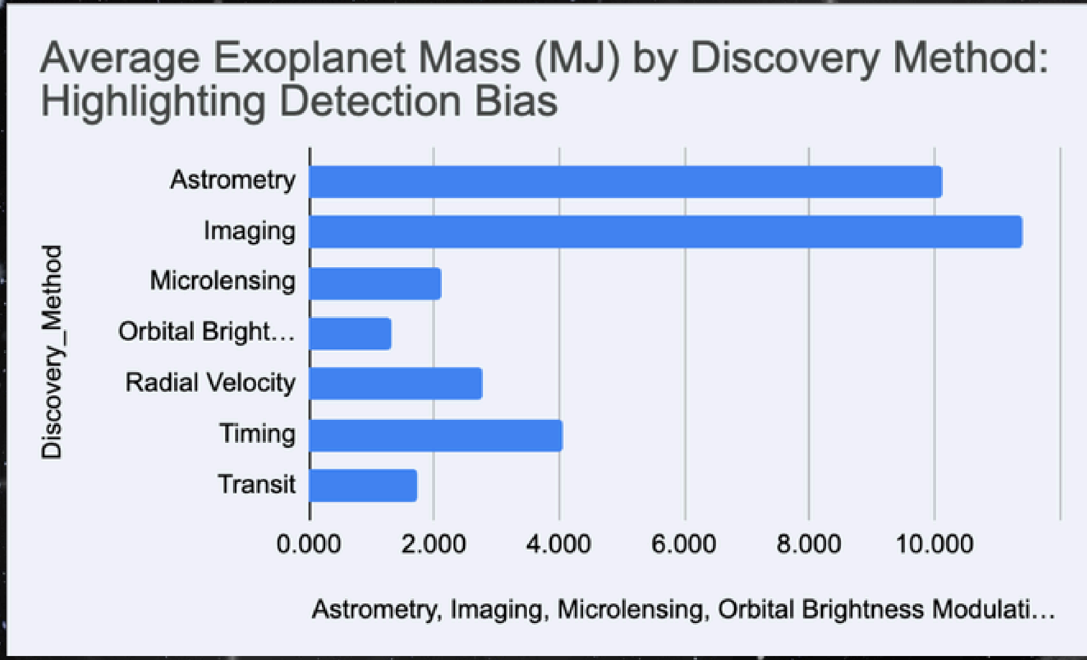
## INSIGHT 2: TECHNOLOGICAL IMPACT ON DISCOVERIES

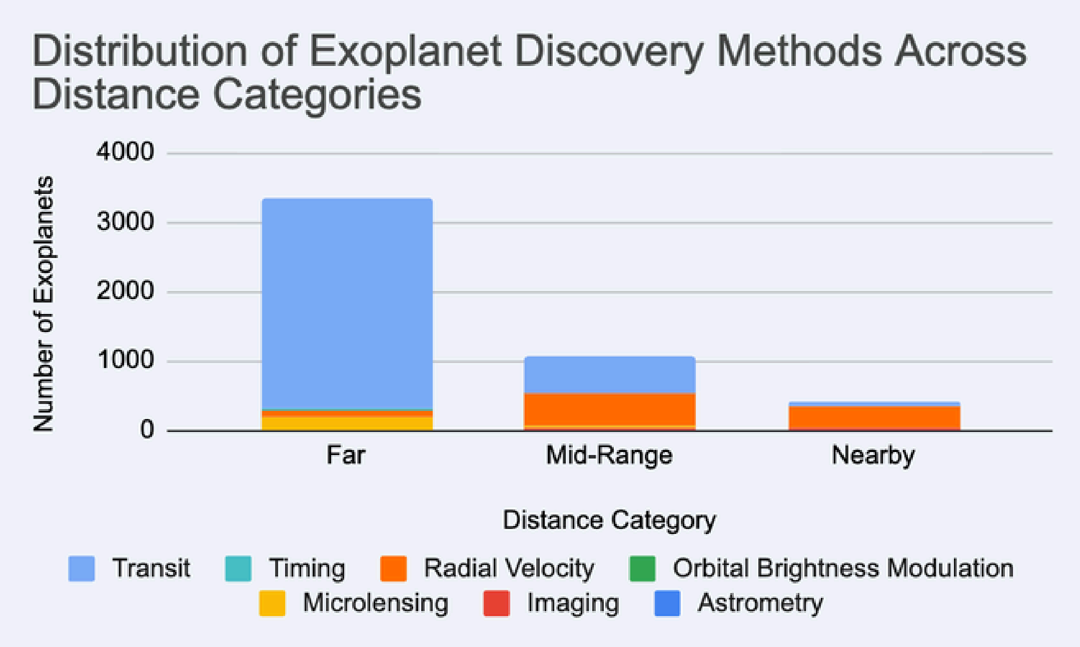



Exoplanet discovery frequency **increased sharply after 2005,** reflecting advances in telescope sensitivity and large-scale survey missions.
This shows that **technology, not planetary scarcity**, was the main limitation in earlier years.

**Why this matters:** Better instruments directly increase discovery output.

# ADVANCED ANALYSIS (ROOT-CAUSE)




## ANALYTICAL FOCUS

Root-cause analysis of discovery efficiency differences across exoplanet detection methods.

### Key Findings

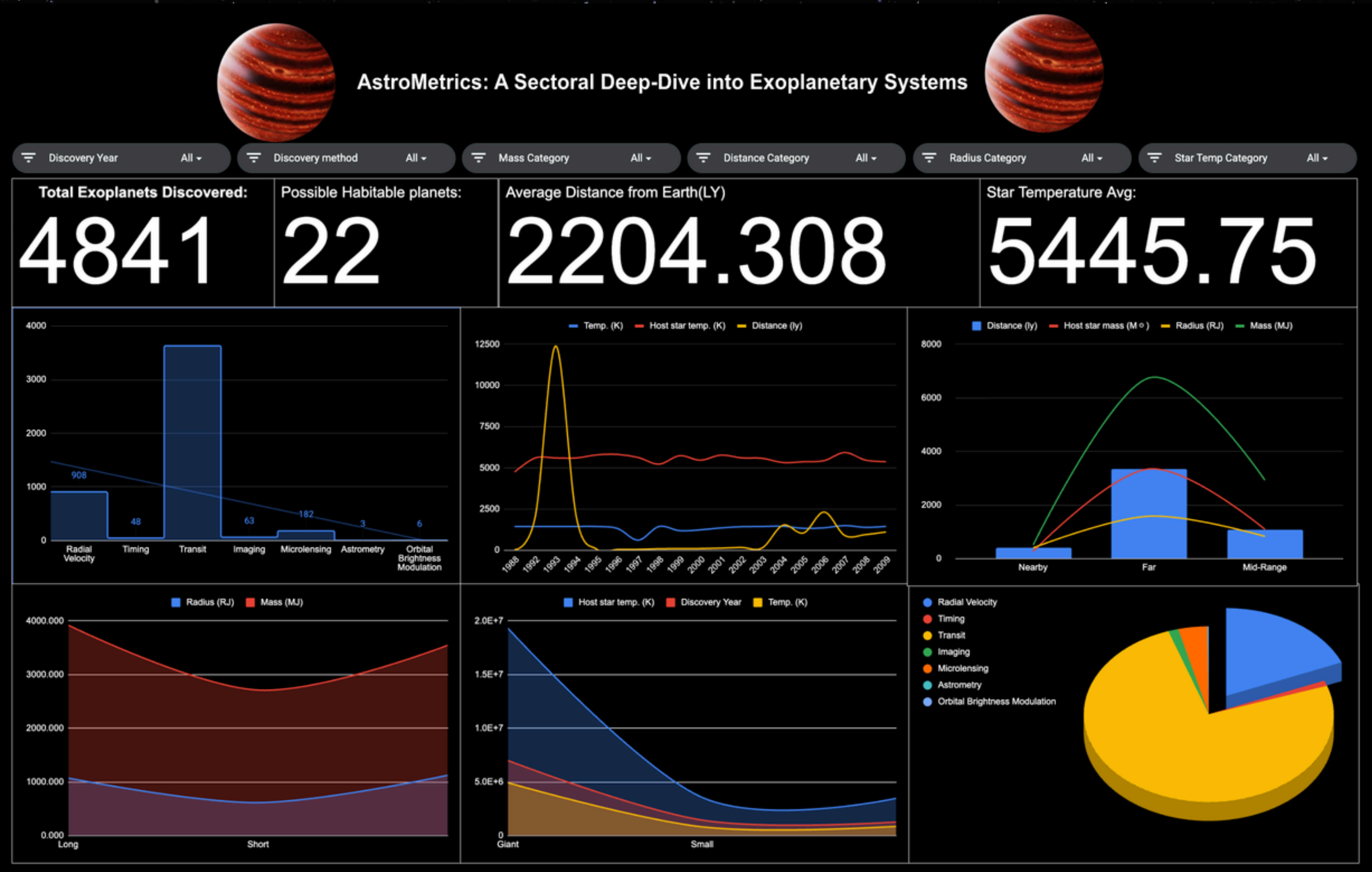
- Transit-based methods dominate discoveries because they are highly scalable and sensitive to small, close-orbit planets.
- Radial velocity techniques provide reliable detections but require long observation time per star, limiting large-scale discovery.
- Imaging and astrometry methods detect fewer planets overall but are effective at identifying massive and distant exoplanets that are inaccessible to other techniques.

### New Understanding

Exoplanet discovery outcomes are driven primarily by method feasibility, observational constraints, and survey scalability, rather than the true underlying distribution of planets in the universe.

# DASHBOARD OBJECTIVE

To provide a single, interactive view that summarizes exoplanet discovery patterns while allowing detailed drill-down across discovery methods, time, distance, and planetary characteristics.




### Executive View (Top Section)

- Presents high-level KPIs including total exoplanets discovered, possible habitable planets, average distance from Earth, and average host star temperature.
- Designed for quick decision-making, enabling stakeholders to immediately understand the overall scale and scope of exoplanet discoveries.
- Filters allow users to dynamically adjust the view by discovery year, method, mass, distance, and stellar properties.

### Operational View (Middle & Bottom Sections)

- Visualizes discovery method dominance, clearly showing the efficiency differences across detection techniques.
- Highlights temporal trends, demonstrating how technological advancements impacted discovery volume over time.
- Shows detection bias across distance, mass, radius, and orbital categories.
- Enables analysts to drill down and compare how different methods perform under varying observational conditions.

### Key Value of the Dashboard

- Bridges the gap between raw astronomical data and actionable insights.
- Supports both strategic overview and detailed exploratory analysis.
- Helps identify why certain planets are discovered more frequently, not just how many.

# RECOMMENDATIONS

1. **Prioritise Transit-Based Survey Missions**

**Action:** Allocate more observational resources and funding toward large-scale transit survey missions.
**Rationale:** Transit methods demonstrate the **highest discovery efficiency** due to their ability to monitor thousands of stars simultaneously, as shown by their dominance in confirmed detections.
**Expected Benefit:** Higher discovery rates at **lower cost per planet detected.**

2. **Use Radial Velocity for Targeted Confirmation**

**Action:** Deploy radial velocity techniques selectively to **confirm transit-detected candidates** and refine mass estimates.
**Rationale:** While resource-intensive, radial velocity provides **high-precision measurements** that complement transit surveys.
**Expected Benefit:** Improved data accuracy without compromising scalability.

3. **Reserve Imaging for Specialised Exploration**

**Action:** Apply direct imaging methods primarily for **large, distant exoplanets** that cannot be detected using transit or radial velocity techniques.
**Rationale:** Imaging detects fewer planets overall but is uniquely effective for studying **massive planets at wide orbits.**
**Expected Benefit:** Optimised use of expensive instrumentation for high-value discoveries.

4. **Invest in High-Sensitivity Instruments**

**Action:** Continue investment in next-generation telescopes and improved photometric sensitivity.
**Rationale:** Discovery trends show that **technological advancement directly drives discovery growth**, especially after 2005.
**Expected Benefit:** Expanded detection range and improved discovery of smaller, longer-period planets.

5. **Address Detection Bias Through Hybrid Strategies**

**Action:** Adopt a **hybrid detection strategy** combining transit, radial velocity, and imaging.
**Rationale:** Different methods are biased toward different planetary types; combining them reduces observational blind spots.
**Expected Benefit:** More representative understanding of true planetary distributions.

# IMPACT & VALUE

## Strategic Impact

- Enables **data-driven prioritization** of exoplanet discovery methods.
- Shifts focus from trial-and-error exploration to **evidence-backed survey planning.**
- Improves alignment between **scientific goals and observational resources.**

## Operational Value

- **Reduced Observation Cost:** Prioritizing transit surveys lowers cost per discovery by maximizing star coverage per observation cycle.
- **Time Efficiency:** Targeted use of radial velocity reduces unnecessary long-duration observations.
- **Resource Optimization:** Imaging instruments are reserved for cases where they provide unique scientific value.

## Scientific Value

- Reduces detection bias by combining complementary discovery methods.
- Enables deeper understanding of **planetary diversity and formation mechanisms.**
- Supports more accurate estimation of **habitable planet occurrence rates.**

## Why Stakeholders Should Approve This

- Converts complex astronomical datasets into clear, actionable insights.
- Demonstrates measurable improvements in **efficiency, cost control, and discovery yield.**
- Provides a **scalable analytical framework** applicable to future astronomical missions.

**The "So What?" Answer:** This analysis helps organizations **discover more planets, faster, and at lower cost,** while improving scientific accuracy.

# LIMITATIONS & NEXT STEPS

## Limitations

**1. Detection Bias in the Dataset**

- The dataset is heavily biased toward transit and radial velocity methods.
- Smaller, long-period, and Earth-like exoplanets are under-represented due to current detection constraints.

**2. Incomplete and Missing Data**

- Several parameters (mass, radius, temperature) contain missing or estimated values.
- This limits precision for advanced statistical modeling and habitability analysis.

**3. Limited Time Coverage**

- The dataset primarily spans discoveries up to 2009, excluding newer missions such as Kepler and TESS.
- Recent advancements are not fully reflected.

**4. Simplified Habitability Criteria**

- Habitability estimates are based on basic thresholds and do not account for atmospheric composition, orbital stability, or stellar activity.

🎯 **WHY THIS MATTERS:**
ACKNOWLEDGING LIMITATIONS DEMONSTRATES ANALYTICAL MATURITY AND BUILDS A ROADMAP FOR FUTURE SCIENTIFIC ADVANCEMENT.

## 🔮 NEXT STEPS

**1. Integrate Updated Mission Data**
Incorporate data from modern missions (Kepler, TESS, JWST) to improve completeness and accuracy.

↓

**2. Advanced Predictive Modeling**
Apply machine learning models to predict undiscovered exoplanets and estimate detection probability.

↓

**3. Enhanced Habitability Analysis**
Include atmospheric, orbital, and stellar variability factors for deeper habitability assessment.

↓

**4. Real-Time Dashboard Expansion**
Enable real-time data ingestion and automated discovery trend updates.


THANK YOU!