COURSE NAME :- Data Visualization and Analysis

Pivot Table Creation Methodology

Project Report Section :- C Group :- 09

This project follows a structured, audit-friendly pivot table approach to transform raw exoplanet data into actionable analytical insights. All pivot tables were created only after data cleaning and standardization to ensure accuracy and consistency.1.1 Data Preparation Before Pivoting

Before creating pivot tables, the following steps were performed:

- Removed records with missing or invalid critical values (mass, radius, discovery year, distance).
- Converted numeric fields stored as text into proper numeric formats.
- Standardized units (Mass in MJ, Radius in RJ, Distance in light years, Temperature in Kelvin).
- Created derived categorical fields to enable meaningful grouping:
    - Mass Category (Small, Medium, Giant)
    - Radius Category (Small, Medium, Large)
    - Period Category (Long, Short, Ultra Short)
    - Distance Category (Nearby, Mid-Range, Far)
    - Host Star Temperature Category (Cool, Sun-like, Hot)

These derived categories allow trend-based analysis rather than raw-value noise.1.2 Pivot Table Design Strategy

Each pivot table was designed with a single analytical question in mind. We avoided unnecessary aggregation and ensured that non-additive metrics were never summed.

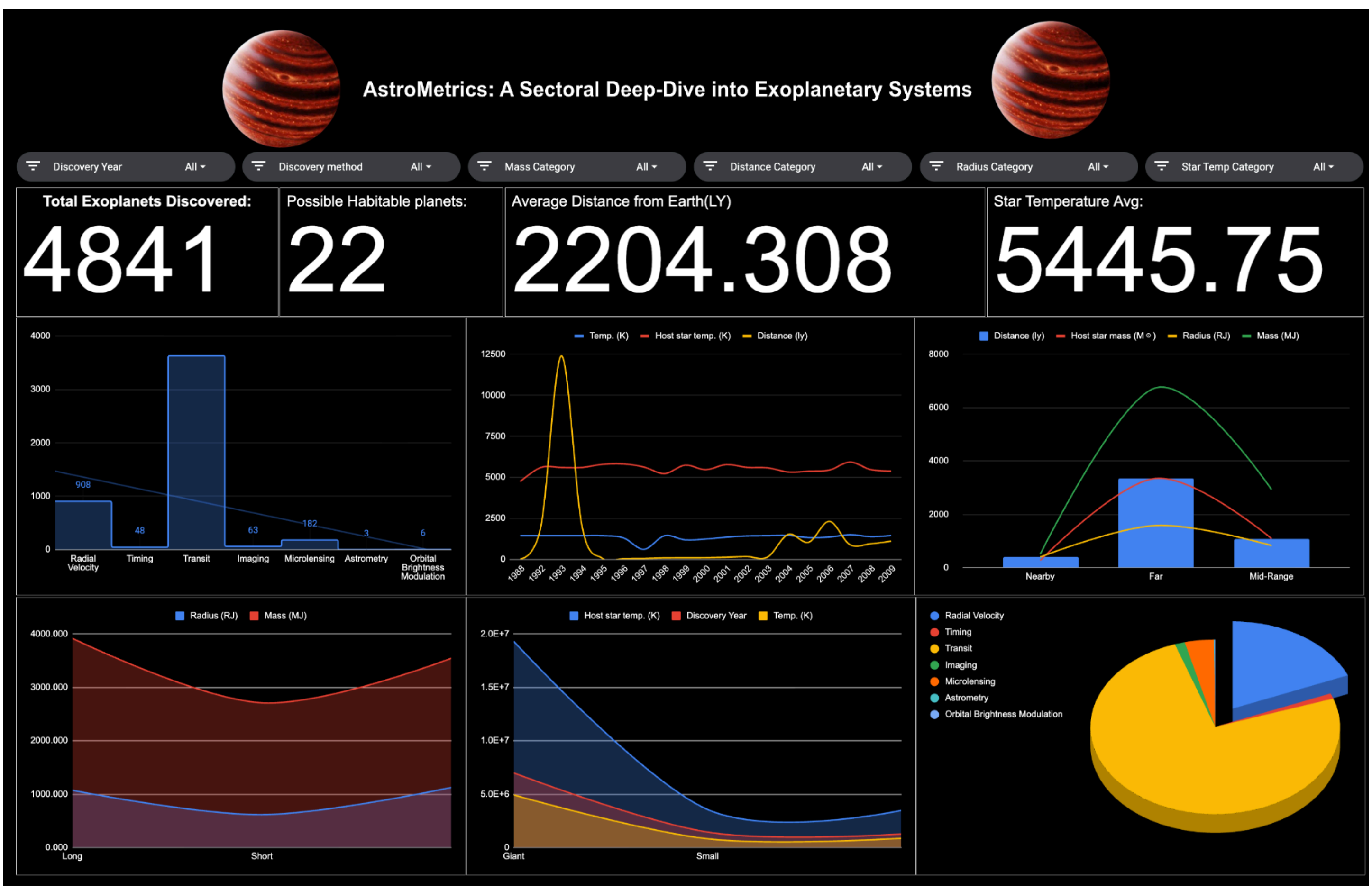
Pivot Design Principles Used:

- **COUNT** → for frequency and dominance analysis
- **AVERAGE** → for physical properties (mass, radius, temperature, distance)
- **Binning** → for continuous variables to improve interpretability
- **KPIs** added outside pivots to preserve pivot integrity

2. Key Pivot Tables and Their Purpose

Pivot Table	Pivot Structure	Purpose	Why this pivot matters
2.1 Discovery Method Dominance	Rows: Discovery Method, Values: Count of Exoplanets	To identify which discovery techniques contribute most to detected exoplanets.	It establishes the baseline detection bias of the dataset.
2.2 Average Planet Mass by Discovery Method	Rows: Discovery Method, Values: Average Planet Mass (MJ)	To compare the type of planets each discovery method is most sensitive to.	Planet mass is a physical property and must not be aggregated.
2.3 Exoplanet Mass Distribution by Orbital Period Category	Rows: Period Category, Columns: Mass Category, Values: Count of Exoplanets	To understand how orbital period influences the detectability of different mass planets.	
2.4 Exoplanet Size Distribution by Host Star Temperature	Rows: Host Star Temperature Category, Columns: Radius Category,	To analyze how detected planet sizes vary with host star	

	Values: Count of Exoplanets	temperature.	
2.5 Discovery Methods Across Distance Categories	Rows: Distance Category, Columns: Discovery Method, Values: Count of Exoplanets	To assess how discovery effectiveness changes with distance from Earth.	
2.6 Year-wise Discovery Trend	Rows: Discovery Year, Values: Count of Exoplanets, Additional KPI: Year-over-Year Growth %	To identify discovery acceleration phases and technology-driven spikes.	



3. Dashboard Design Logic

The dashboard was designed using a top-down analytical flow:

- **Top KPI Section:**
  - Total Exoplanets Discovered
  - Possible Habitable Planets
  - Average Distance from Earth
  - Average Host Star Temperature

    (These KPIs give instant context and scale.)

- **Middle Section – Comparative Analysis:**
  - Discovery Method Dominance
  - Average Mass by Discovery Method
  - Mass Distribution by Orbital Period
  - Size Distribution by Host Star Temperature

    (This layer explains what types of planets we detect and why.)

- **Bottom Section – Trend & Bias Analysis:**
  - Discovery Evolution Over Time
  - Distance vs Detection Intensity
  - Heatmap of Distance × Discovery Method

    (This layer answers how technology and distance shape discoveries.)

4. Key Insights and Conclusions4.1 Discovery Method Bias Is Strong

- **Observation:** Transit and Radial Velocity methods dominate detections. Transit alone accounts for the majority of discovered exoplanets. Less common methods (Astrometry, Timing) contribute marginally.
- **Conclusion:** The observed exoplanet population is shaped more by detection techniques than by true planetary abundance.

4.2 Large and Close-In Planets Are Over-Represented

- **Observation:** Giant planets dominate the mass distribution. Ultra-short and short orbital periods have significantly more detections. Long-period planets are under-represented.
- **Conclusion:** Close-in, massive planets are easier to detect, leading to systematic over-representation.

4.3 Discovery Method Determines Planet Characteristics

- **Observation:** Imaging and Astrometry detect higher-mass planets on average. Transit methods detect lower-mass and smaller-radius planets. Radial velocity sits between the two extremes.
- **Conclusion:** Each discovery method reveals a different subset of the exoplanet population.4.4 Host Star Temperature Influences Detection Volume
- **Observation:** Sun-like stars host the highest number of detected planets. Cool and hot stars show fewer detections across all size categories.
- **Important Note:** This does not imply Sun-like stars form more planets — it reflects observational focus and detectability.

4.5 Distance Strongly Limits Detection

- **Observation:** Most detections occur in far and mid-range categories only for transit. Detection diversity decreases rapidly with distance. Heatmap shows sharp intensity drop for non-transit methods at large distances.
- **Conclusion:** Exoplanet detection is distance-limited, and detection diversity decreases with distance.

4.6 Discovery Growth Is Technology-Driven

- **Observation:** Sharp growth spikes appear in specific years. Growth is not linear or uniform. Year-over-year KPIs show bursts rather than steady increase.
- **Conclusion:** Discovery rates reflect technological advancement, not changes in planetary formation.

5. Final Project Conclusion

This project demonstrates that modern exoplanet datasets primarily reflect observational bias and technological constraints rather than the true distribution of planets in the universe. While physical relationships such as mass, radius, and orbital trends remain valid, any population-level conclusions must account for discovery method, distance, and time-based biases.6. Why This Analysis Is Reliable

- Transparent pivot-based methodology.
- No misuse of aggregation functions.
- Clear separation between data, KPIs, and visuals.
- Bias explicitly acknowledged.