**Level 1 (Regular Data Science Questions):**

1. Calculate the range of the column OrbitDuration

3980

1. Calculate the standard deviation of the column PlanetMass

11.4216265

1. Find the mean and median of the column SolarFlare

136.02, 136

**Level 2 (Multiple Step Hard Data Science Questions):**

1. For each category in the column "StarType", calculate the skewness and kurtosis of the column "PlanetMass". Which category has the most positively skewed Plant Masses?

Yellow Dwarf

1. Calculate a “Habitability Score” for each planet, using a weighted combination of WaterCoverage (40%), NitrogenConcentration (30%), and PlanetaryCoreTemp (10%). For the top 5 planets by this score, sum up their PlanetMass. What is this total?

192.0

1. For each unique StarType, calculate the correlation between StarAge and SolarFlareFrequency. Which StarType shows the weakest correlation?

Red Dwarf

1. Divide the dataset into three categories based on ProximityToBlackHole (Near, Moderate, Distant). For each category, compute the average SatelliteCount. Which category has the highest average.

Moderate and Distant both have 50 average SatelliteCount

1. Perform a rolling 5-planet average on TectonicActivity and identify the PlanetID where this rolling average peaked. For that PlanetID, what was the corresponding VolcanicActivity?

Planet ID: GC1\_P13 and VolcanicActivity: 6

1. Identify planets that have OrbitDuration within 10% of the median OrbitDuration. For these planets, what’s the average Planetmass and median NeighboringPlanetDistances.

Average PlanetMass: 19.40 and Median NeighboringPlanetDistances is 292.50

**Level 3 (Multistep Data Analysis, Aspects of data science/ML):**

1. By treating planets with a very high ArtificialObjectsCount as anomalies, can you use algorithms like Isolation Forest to detect these outlier planets? What is the most common StarType among the anomalies.

Red Dwarf and Yellow Dwarf are both 2

1. The MagneticFieldStrength column follows a pattern. Figure out that pattern and predict the next 5 values for the MagneticFieldStrength column. Output the pattern found as well as the 5th generated value.

The values increase by 10 for each consecutive planet. The 5th generated value is the MagneticFieldStrength of 1050.

1. The SatelliteCount column follows a pattern. Figure out that pattern and predict the next 5 values for the SatelliteCount column. Output the pattern found as well as the 3rd generated value.

The Satellite count maxes out at 50. The 3rd generated value is 50.

1. Using the features WaterCoverage, MagneticFieldStrength, NitrogenConcentration, and TectonicActivity, train a random forest classifier to predict the PlanetStatus (Exploded, Intact). What is the F1-score of the model?

F1-score: 0.13