

Final Project – DSC530

This project is to analyze the Exoplanets database from *exoplanet.eu*; Exoplanets are beyond our solar system.

I did the dataset search on *exoplanet.eu*. This is an astronomy website founded in Paris with an Extrasolar Planets database.

The database has 129 columns or variables and this is a database that constantly is being updated with new observations.

I used the PyVO database. PyVO lets find and retrieve astronomical data available from archives that support standard IVOA virtual observatory service protocols.

To connect to the *exoplanet.eu* database, I have to use an API. The API was developed by the Virtual Observatory (VO). After connecting to the database, I will be able to extract the data with SQL instructions.

- **Statistical/Hypothetical Question**

Are the ellipses perimeters influenced by the exoplanet detection methods?

- **Outcome of your EDA**

See *Aldana – Final Project.pdf*

- **What do you feel was missed during the analysis?**

The *exoplanet.eu* database has many variables (129) with which I can make more analysis. My focus was related to Exoplanets with the following variables:

1. mass: planetary mass.
2. period: orbital period around the star.
3. detection type: method used to discover the exoplanet
4. semi major axis: the semi-major axis is half of the longest diameter of an ellipse.
5. semi minor axis: [derived variable] minor semi-axis is a line segment that is at right angles with the semi-major axis and has one end at the center.
6. eccentricity: the eccentricity is a measure of how much an ellipse is squashed. This is the shape of the ellipse.

7. star distance: distance to a host star.

8. ellipse: [derived variable] total ellipse perimeter.

○ **Were there any variables you felt could have helped in the analysis?**

There are many variables of host stars. These variables can contribute more to my analysis.

○ **Were there any assumptions made you felt were incorrect?**

I feel that p-values should be analyzed better. I found results that can be influenced by outliers. The dimensions in the universe are extremely large and there are quite differences between exoplanets.

○ **What challenges did you face, what did you not fully understand?**

The most challenging was to calculate the ellipse perimeter variable. Because there are several variables and formulas involved. The ellipse variable is the most important variable in my analysis. This variable indicates the total ellipse perimeter.

I have to calculate the ellipse perimeter with semi_major_axis and semi_minor_axis variables, but the semi_minor_axis variable does not exist in the database. Although, the eccentricity variable exists in the database.

I can calculate semi_minor_axis using eccentricity and semi_major_axis. This is the formula to calculate semi_minor_axis:

$$\text{semi_minor_axis} = \text{semi_major_axis} [a] * \text{SQRT}(1 - \text{POWER}(\text{eccentricity}, 2))$$

I used *Ramanujan's formula* which is more accurate:

$$h = \text{POWER}(\text{semi_major_axis} - \text{semi_minor_axis}, 2) / \text{POWER}(\text{semi_major_axis} + \text{semi_minor_axis}, 2)$$

$$\text{ellipse} = \text{math.pi} * (\text{semi_major_axis} + \text{semi_minor_axis}) * (1 + ((3 * h) / (10 + \text{SQRT}(4 - (3 * h))))))$$

○ Submit a link to your repository to the assignment link during the final week of class.

<https://github.com/moaldana/exoplanets>