KOI Exploration Analysis: Kepler's Object Of Interest

Analysis by Perla Lopez Flores

Abstract

The Kepler Objects of Interest (KOI) data focuses on exoplanets and their host stars. I noticed an increase in false positives in the exoplanet search, and with many variables to work with, I decided to focus on a specific group of variables to determine their influence on an exoplanet's confirmation or false positive. This led to my research question: What variables play an important role in false positives? With the expansion of exoplanet searches, there are multiple benefits such as finding other life along with society as a whole being able to better understand the process of the formation of planets by looking at other planets not in our solar system. My ultimate goal was to find out the influence that the variables "transit epoch," "transit duration," "transit depth," and "transit signal-to-noise" has in exoplanet false positives. I created a smaller dataset from the original dataset using only these variables and the disposition column (column stating whether an exoplanet was confirmed or a false positive). The finalized dataset included 2179 potential exoplanet findings in which I performed a generalized linear model (glm). I found that my model had a 0.58 accuracy, meaning that there were false outcomes after testing the influence that the four different variables had on disposition. This is important because with these results, we can see that the four variables resulted in more false outcomes in the confirmed exoplanets than the false positives.

Background

The KOI dataset is a collection of exoplanet candidates, confirmations, and false positives along with more specific information about each exoplanet. It includes information including:

- KeplD, KOI name, Kepler name
- disposition, orbital periods, planetary radius, impact parameter
- transit epoch, transit duration, transit depth, transit signal-to-noise

Research Question

What variables play an important role in false positives?

Hypotheses

Null: There is a significant role that the variables transit epoch, transit duration, transit depth, and transit signal-to-noise plays in exoplanet false positives.
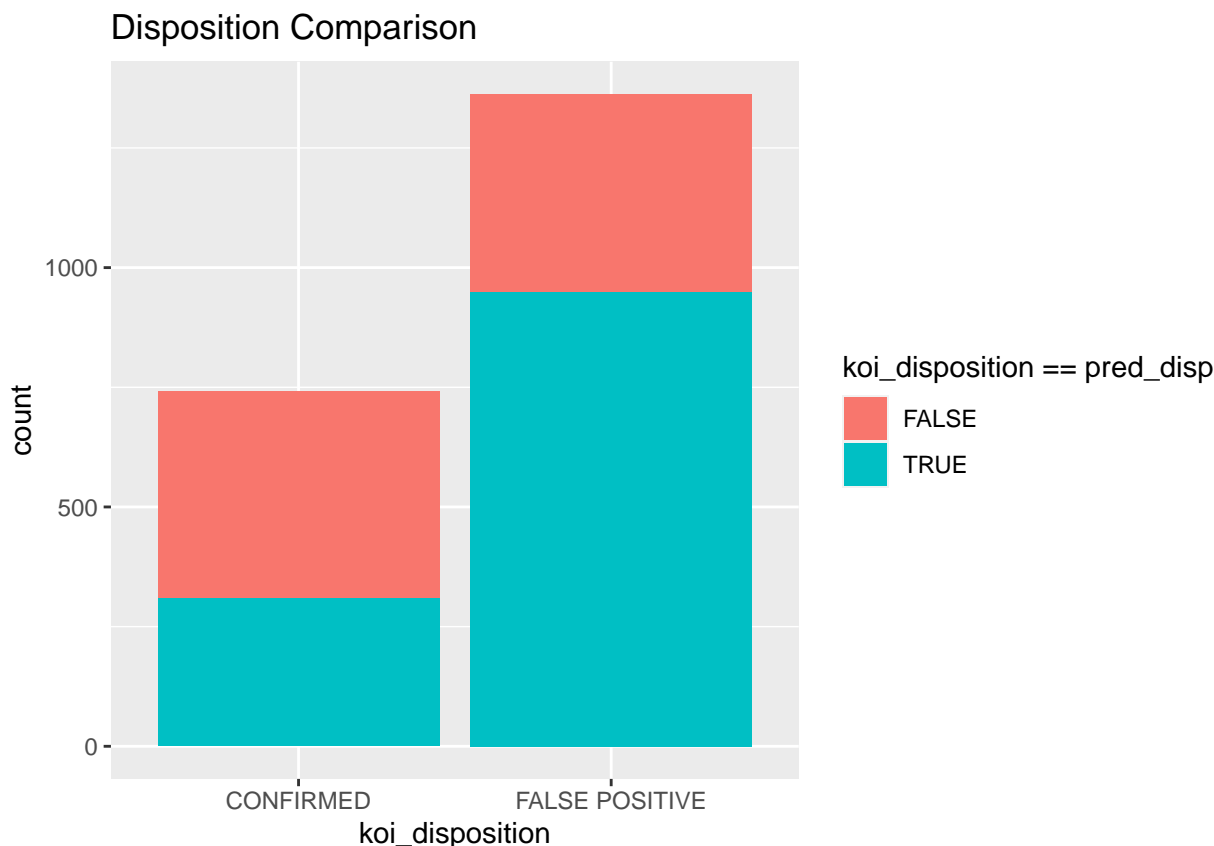
Alternate: There is no significant role that the variables transit epoch, transit duration, transit depth, and transit signal-to-noise plays in exoplanet false positives.

Figure 1: Kepler Exoplanets' Disposition

Exoplanet Disposition
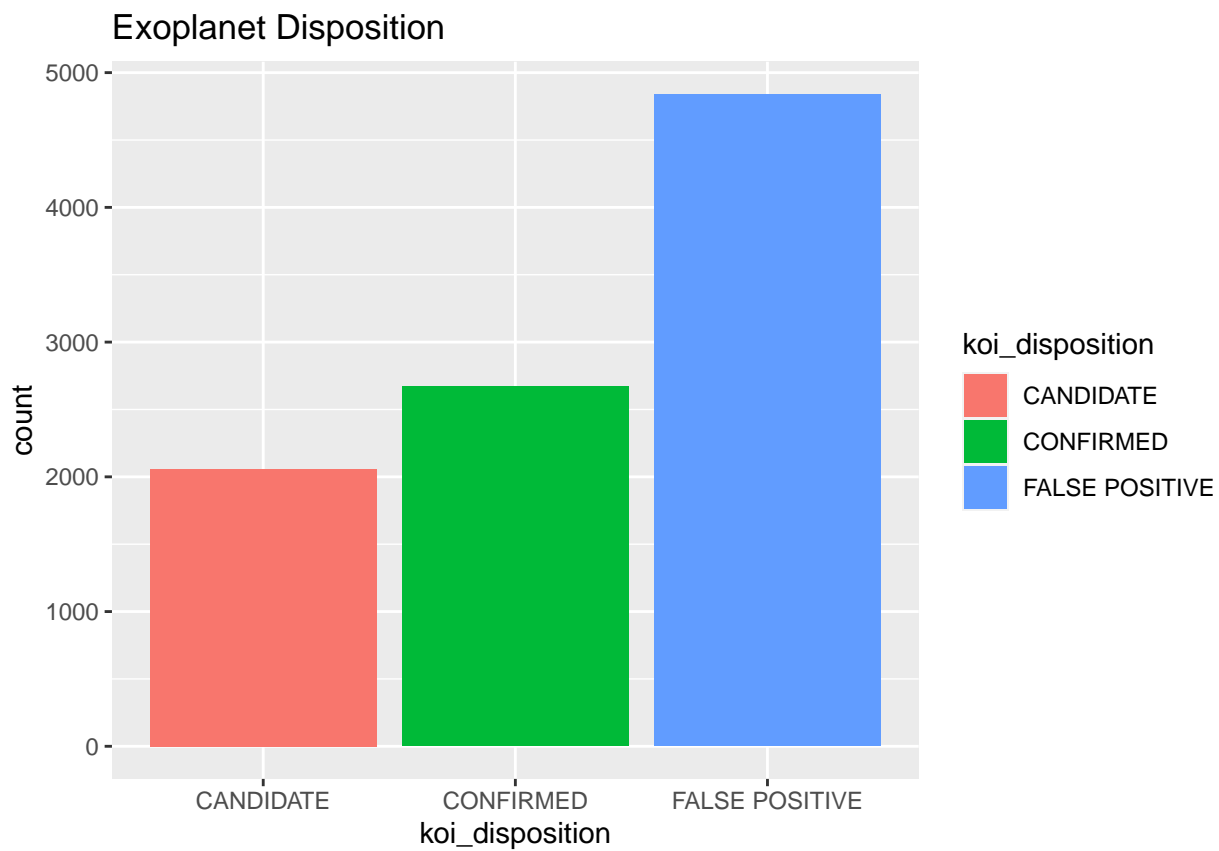
This plot shows the Kepler exoplanets that have been candidates, confirmed, or false positives. There is a significantly higher rate of false positive exoplanets than candidates and confirmed exoplanets.

Figure 2: Comparing Kepler Exoplanets' Dispositions

This plot shows the results of the analysis, comparing the true disposition with the predicted disposition (meaning the amount of exoplanets that were confirmed in comparison to false positives). There were approximately 750 confirmed exoplanets, but more than half were seen as false. There were approximately 1300 false positive exoplanets, and less than half were seen as false.

Discussion

The variables "transit epoch," "transit duration," "transit depth," and "transit signal-to-noise" impacted the confirmed and false positive exoplanets differently. With these variables, there were more false outcomes in the confirmed than the false positives. This can ultimately help others better understand why there are more false positives than confirmed. Before running my analysis, I expected more accurate results as to why there are more false positives, but the outcome I got supported my hypothesis due to there being less false results in the false positive. After this analysis, some next steps would be to compare confirmed and false positives again while looking at different variables. By doing so, I can then further my understanding by comparing past results with new results.

Code Availability and Data Availability

All the code used in these analysis is in the GitHub repository **the-codingschool/KOI-exploration.** My analyses are in the **Perla-Project/Scripts** folder. The R markdown file used to create this report is in that folder under the name **Final Report.RMD**, and that includes all of the code to reproduce these plots.

The raw data used was downloaded from the **NASA Exoplanet Archive**. R objects for these datasets can be found in **KOI-exploration/data.** The raw data was pre-processed to combine transit based variables. These pre-processed datasets can be found in **KOI-exploration/Perla-Project/Scripts/cleansing__and__processing.** More information on KOI datasets **here.**