

EDA

- At first I drew pairgrid and correlation matrix to get the general idea about the data. In some subplots of the pairgrid especially on the diagonal we can notice some outliers that need to be investigated further. In the correlation matrix we can see that `nufnu` and `nufnu_lower` are highly correlated(as expected).
- Then I tried a quintile outlier detection technique and plotted boxes and histograms to visually identify the outliers, but they didn't work well, although visually there are outliers. I still need a professional opinion from an astrophysicist whether to consider them as outliers or not. As you can see on my codes I've found the outliers, but when I try to look at the samples with outlier values they seem pretty normal for me. Anyway I dropped 339 samples containing outliers.
- After cleaning the outliers the correlation between `nufnu` and `nufnu_lower` became 1 which means we may not need `nufnu_lower`. For this I have done one more experiment and plotted scatter plot of `nufnu-nufnu_err` and `nufnu_lower`, which showed a simple rule: `nufnu_lower` is 0 when `nufnu-nufnu_err ≤ 0` and `nufnu-nufnu_err` when `nufnu-nufnu_err > 0` so we can safely drop the `nufnu_lower` column.
- After cleaning the data the correlation between `nufnu` and `nufnu_err` became 0.75, but I won't do anything with them as I need both to revive information.
- I drew a pie plot of the `flag` column which shows that 87.8% of the observations are determined.
- Then I got an interesting 3D plot by combining `right_ascension`, `declination` and `flag`. The plot shows a slice of 3D unit ball which shows the directions in which the observations were taken, green dots are determined observations and the red ones are upper limits.(I did this without dropping outliers to get full info about observation positions)

Conclusion: After cleaning data from (possible) outliers I could answer one of the open questions about the need of a `nufnu_lower` column(we can safely drop it). I also showed the columns relation with pairgrid and correlation matrix. And finally I could show the observation range on a unit ball.