

Adding Inferential Statistics to my Capstone Project Data

The first bit of Inferential Statistics I used in my data were 3 separate heat maps. These Heatmaps gave me a better understanding of a few of my data points which were closely correlated. 'Energy' & Loudness' were the closest correlated variables in my data. The correlation was also not very high, just a bit above average.

The tests I used to measure the level of correlation are the Pearsons, Spearman, and Kendall tests. For these tests, I used the 'Energy' & 'Loudness' columns as they had shown the closest correlation during a few Heatmap plots I did. After analyzing these tests, the Spearman Test had the closest correlation significance with an R-Value of 0.778 & P-Value = 1.828. Since our P-Value is so large, it signifies that there it is a non-significant result.

I also applied some List-wise deletion & Pair-wise deletion but found no significant discoveries in my data using these tests. The results were actually very comparable to the Heatmaps I had previously tested.

The final statistical testing I tried was the T-Test. Our T-Score was pretty low, coming in at 4.1224. A small t-score tells you that the groups are similar. After comparing the t statistic with the critical t value (computed internally) we get a low p value of 0.0013. A low P-Value is good as they indicate your data did not occur by chance. Thus it proves that the mean of the two distributions are different and statistically significant.

After close examination of my data I was not able to find a Dependant Variable. More specifically, there is not one variable that determines if a song will make it in the Top 100 songs of the year chart. There are multiple variables that determine this. Since my data contains more than 2 Independent Variables, this is a multiple logistic regression model.