

Act Report

For the analysis and visualizations part of the project, I chose to highlight a few of the key indicators in the data.

I included a general descriptive that provided a full highlight of the different metrics in the data. I used `describe()` to show the mean, min, max, std, 25%, 50% and 75% of all numerical fields.

The second descriptive I included was the max value of the retweet column.

The third descriptive I included was a grouping of the top 10 rating_numerator and prediction1 fields. This shows the top 10 rating_numerators grouped by the prediction1 field. This gives insight into which predictions had the highest rating_numerators. I only analyzed the first prediction field in this descriptive set, which indicated that "Labrador Retriever" and "Golden Retriever" were two of the highest rated predictions, which are also indicated in the next visual comparing the prediction fields against the top retweets.

For the visuals, I first chose to compare the prediction fields (breed), against the retweet field. The visuals show all three predictions along with the 10 ten retweets for each one. These three visuals show that there is a strong indicator towards a particular two breed predictions that get more retweets than others, Labrador Retriever and Golden Retriever.

For my last visual I choose to plot the source field to show which sources are more utilized than others. The plot indicated that "Twitter for iPhone" is the most utilized platform of all that were recorded.

In []: