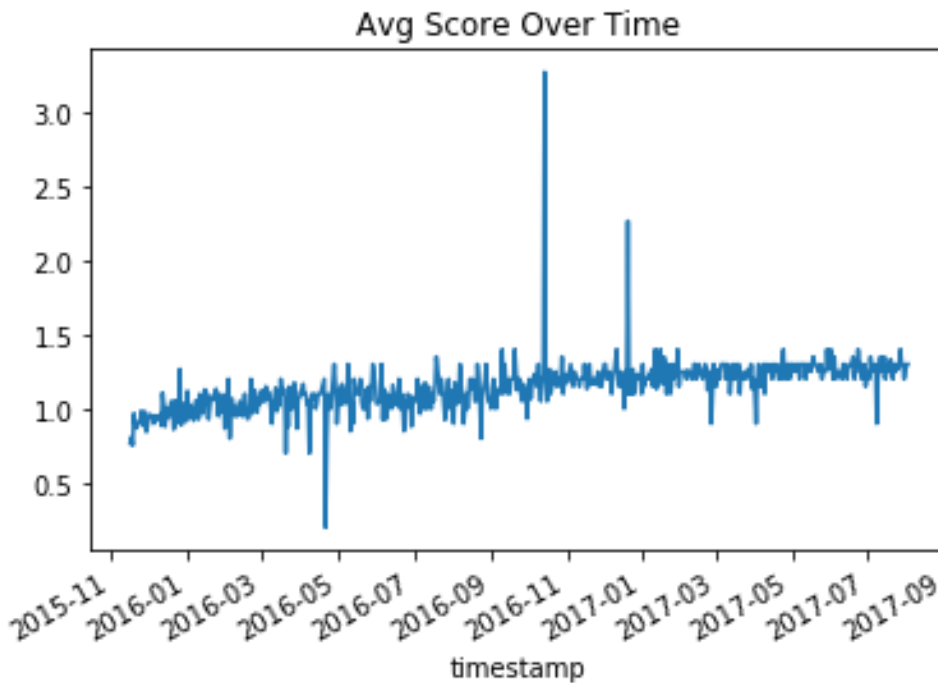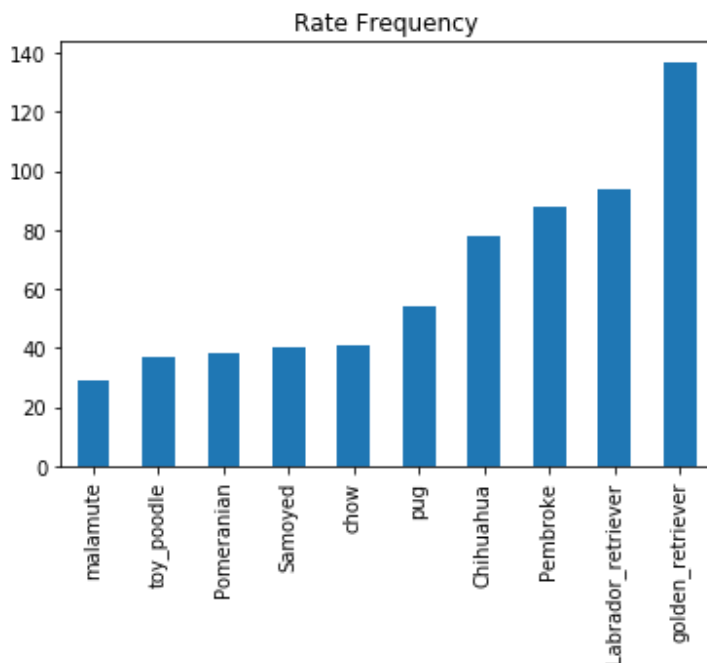# Insights and Visualizations
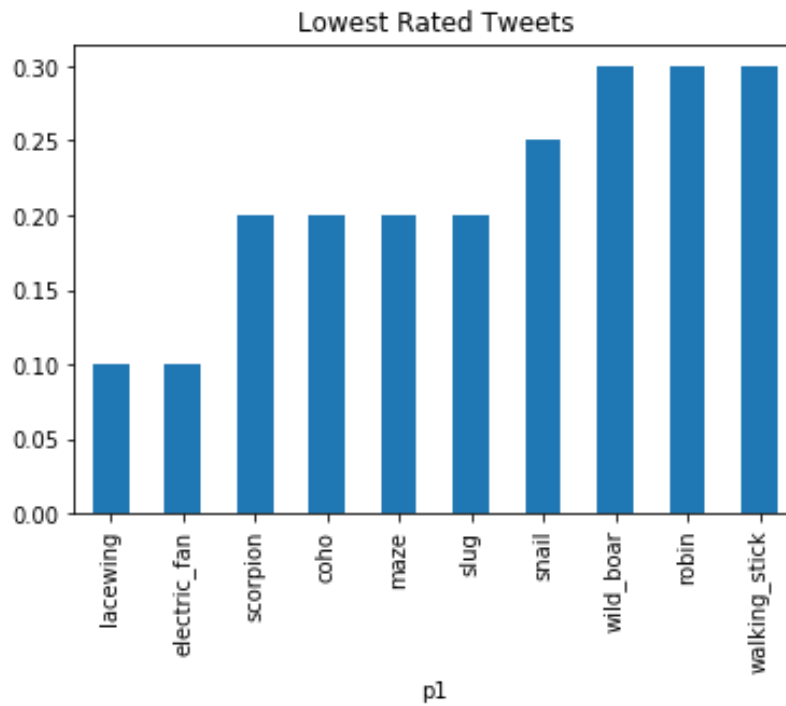
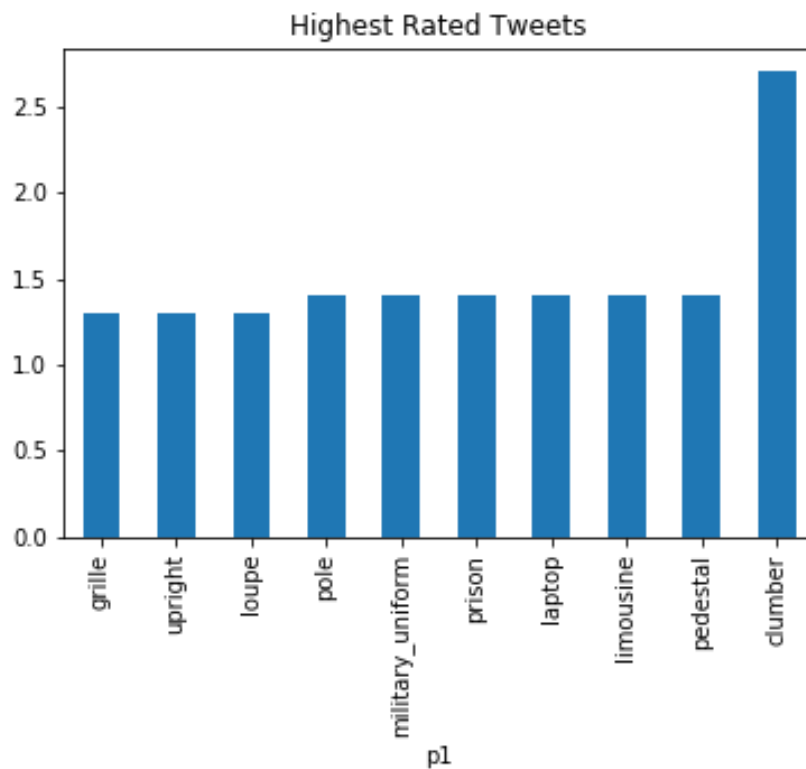The chart below shows that the score over time has been steadily increasing.

The chart below shows that the golden retriever (using 1st prediction) is the top-rated dog by **volume** of tweets.
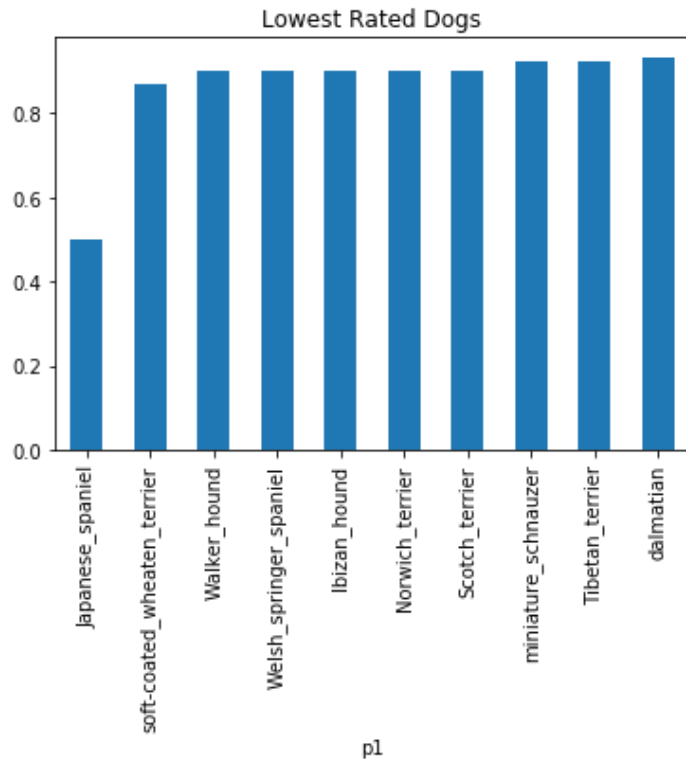
The lowest rated tweets appear to be tweets that are not dogs at all.

Similar to the lowest rated tweets the predictor for high rated tweets are not dogs either based on score

We will now filter using the p1_dog field to pull lowest rated dogs. The Japanese spaniel is lowest rated on average.

Lowest Rated Dogs

We will now filter using the p1_dog field to pull highest rated dogs. The dumber is highest rated on average.

Highest Rated Dogs
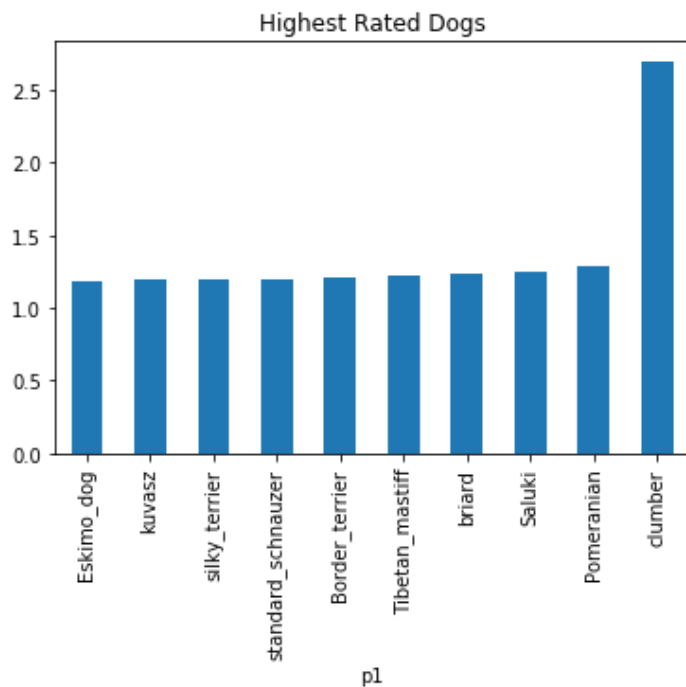
Now adding more dimensions like the number of likes, retweets, etc we will begin to use the seaborns library.

The number of retweets to favorites is a fairly linear relationship.
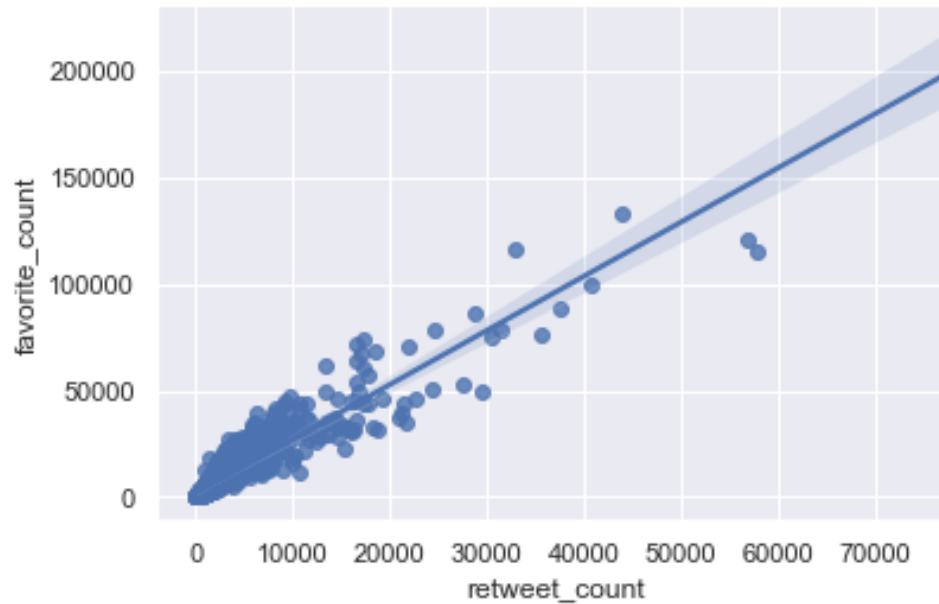
Now getting further into the analysis we can breakdown the above chart by setting the hue to the source of the tweet. The chart shows iPhone sourced tweets being favorited more than retweeted.
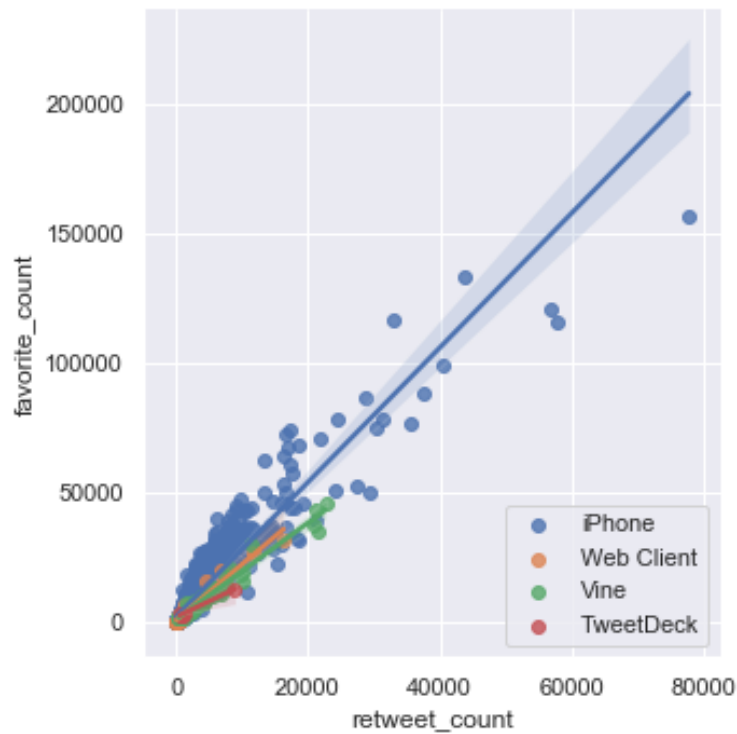
Now filtering for dog only we get the chart below.   Notice how the Vine source is completely removed from the dataset. It looks like the bad predictions are actually coming from vine sources. This makes sense because Vine are either GIFs or videos and probably are not compatible with the algorithm. GIFs at their resting state may not have the dog in view at first. The predictor is probably using a "thumbnail" of the vine to grade the dog species. These of course are my assumptions.