Create a 250-word-minimum written report called act_report.pdf or act_report.html that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

As part of the Udacity project, I went on to analyze dog ratings from the twitter account WeRateDogs. At first this was dirty data so I cleaned it and made it ready to gather some insights.

After analyzing the dog rating data we found interesting facts on which were the most rated dogs and the ones that got the most amount of retweets. We also found insights related to the spread of the ratings - and how most dogs received a high rating.

For my first insight, I analyzed the tweeter prediction dataset along with the twitter API data to compare which types of dogs had the most amount of retweets and favorites.

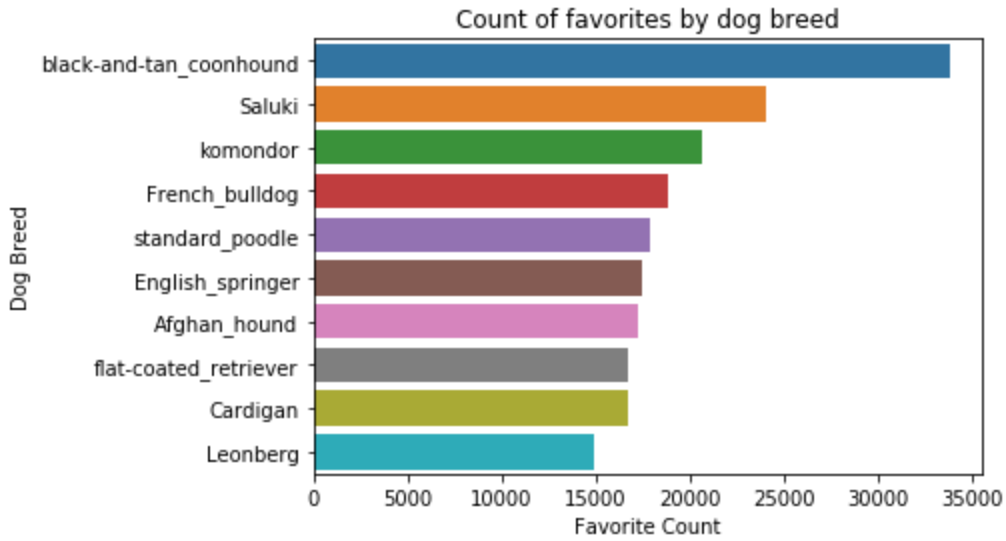

Figure 1; Count of favorites by dog breed

As seen in figure 1, the most favorited dog was the Black & Tan Coonhound followed by the Saluki and the Komondor. Here are some pictures of each dog to show why they were so favorited:

Figure 2: Black & Tan Coonhound

Figure 3: Saluki Dog

Figure 4: Komondor

We can all agree they deserve the top 3 spot. After examining the favorite count I went on to check if this translated to the amount of retweets.

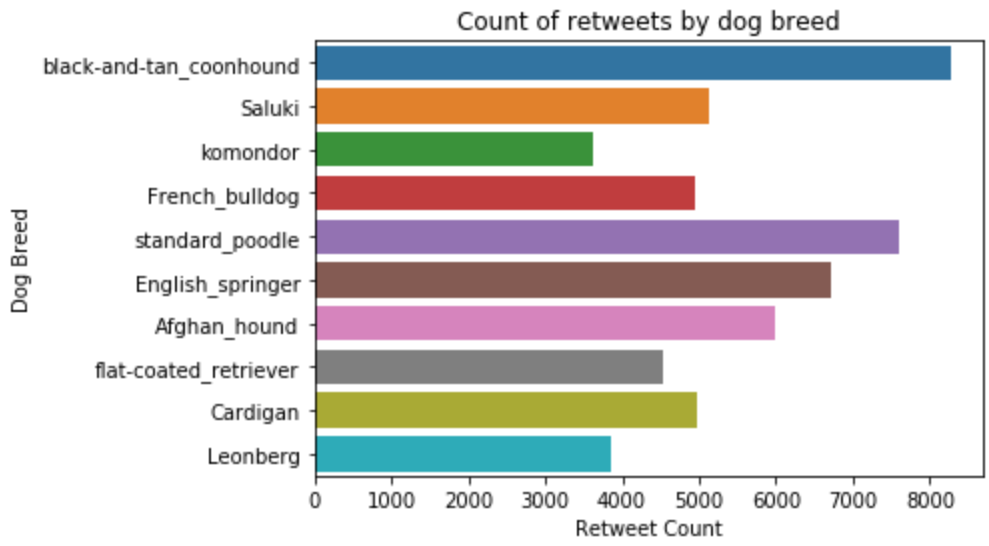

Figure 5: Count of retweets by dog breed

As seen in the count of retweets, the most retweeted dog breed translated to a similar ranking in favorites, however, the Saluki and the Komondor didn't get as many retweets as the other dog breeds. My assumption is that these types of dogs didn't have a text as "catchy" and hence weren't as shared as the other leading dog breeds in the ranking.

Lastly, I checked what is the spread of ratings across the tweets, my goal was to detect outliers or any particular trends in ratings.

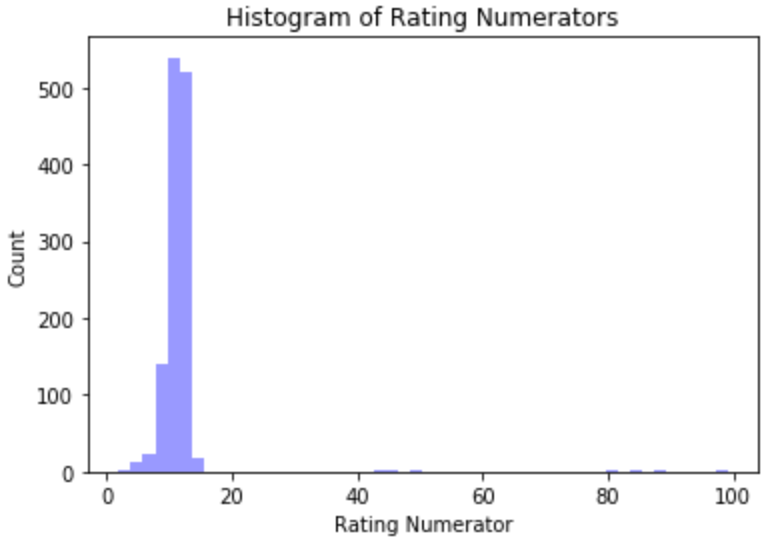

Figure 6: Histogram of rating numerators

It seems that we have a left skewed distribution with the majority of ratings falling amongst 10 - 13 with a few outliers above 40. We have a mean of 11.3 with a standard deviation of 4.9 points, it

seems that the rating itself is not a big factor in how famous the dogs are. From our analysis the retweets and favorite count are a better predictor.