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.

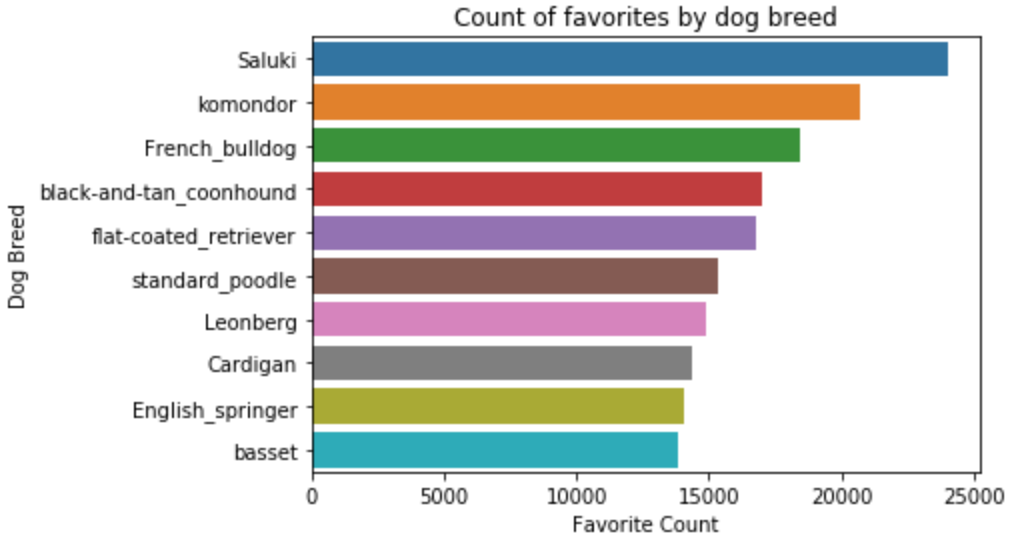

Figure 1; Count of favorites by dog breed

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

Figure 2: Saluki Dog

Figure 3: Komondor

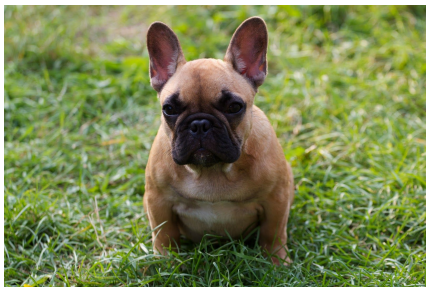

Figure 4: French Bulldog

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.

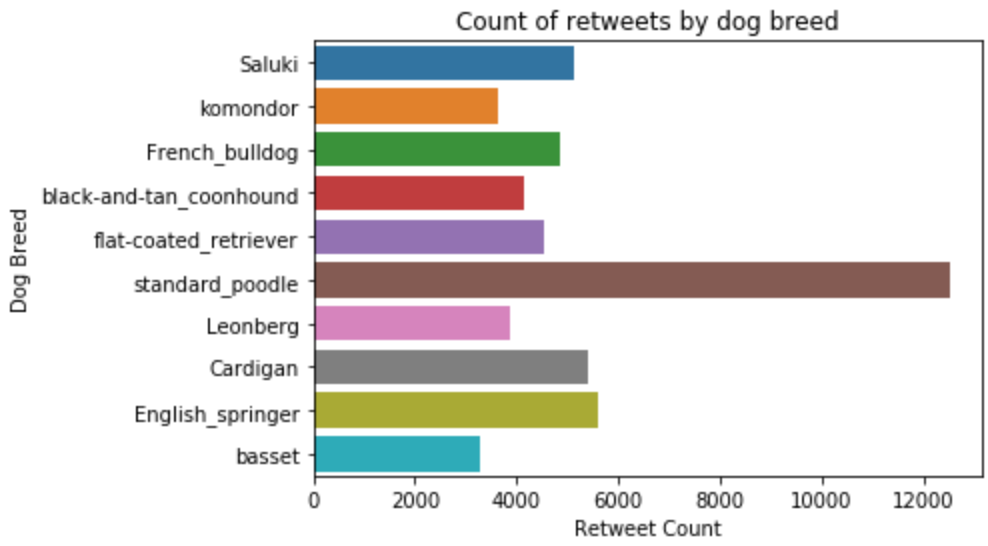

Figure 5: Count of retweets by dog breed

As seen in the count of retweets, the most retweeted dog breed was actually the standard poodle. My assumption is this type of dog had a more “catchy” text and hence was more widely shared than our previous top 3 dogs.

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

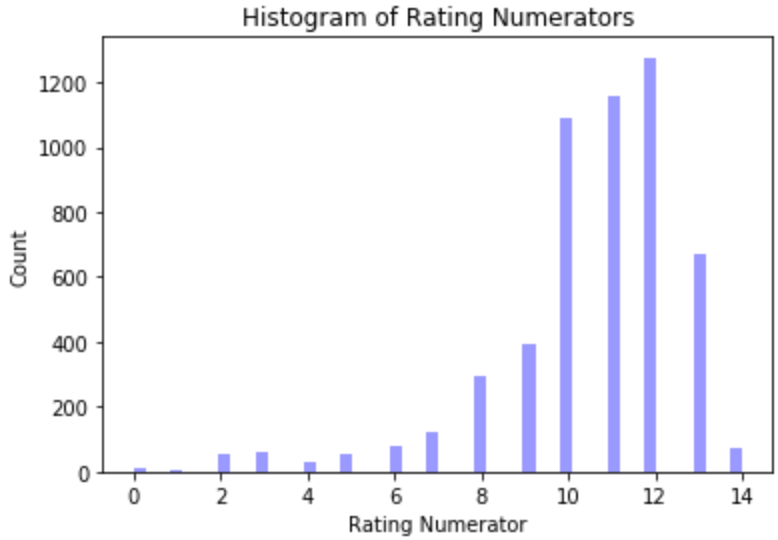

Figure 6: Histogram of rating numerators

It seems that we have a right skewed distribution with the majority of ratings falling amongst 10 - 13, let's compute some statistics on the rating numerator. We have a mean of 10.5 with a standard

deviation of 2.2 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.