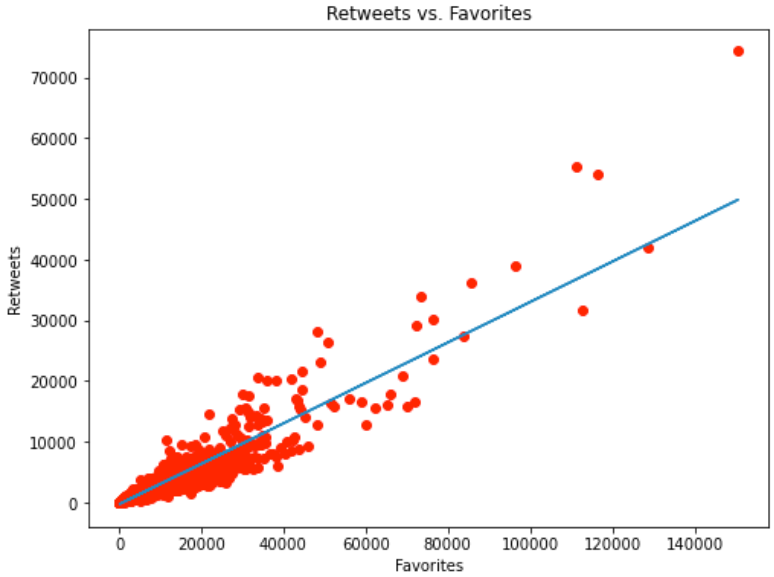
Having just become the owner of an adorable Golden Retriever puppo three weeks ago, I know that Golden Retrievers are the best dogs on the planet. I wanted to use this data, from the WeRateDogs twitter, to find out if my preference for Goldens is backed by statistical evidence or if I am blinded by love. Additionally, I was interested in what stage of dog (puppo, pupper, floofer, and doggo) would garner the most retweets, as well as the relationship between retweets and favorites on twitter.

To begin my statistical analysis I wanted to look at the relationship between retweets and favorites. Not being a twitter user, I had little knowledge of the importance of the two types of ways to show interest in a tweet. I set up a linear regression comparing the two, and found a p-value of 0.000 which shows significance. The correlation coefficient of 0.333 means that retweets occur at ⅓ the rate of favorites. This would make sense because retweets are a stronger endorsement of a tweet. A favorite simply means that a point is added to it's score. Where as a retweet promotes a tweet by attaching one's name to it. In the following graph the relationship between retweets and favorites is obvious, especially with the regression line add in.



I began my investigation by simply inquiring what dog breed is most commonly tweeted about. To no surprise, it is the Golden Retriever. My fear was that there would be very few Goldens in the dataset, but with 158 out of 2152 total posts it gave me hope that a significance test would garner reliable data. I created dummy columns of dog breeds, and set up a linear regression with retweet count as the dependent variable. I chose retweets as the dependent variable because it is a numerical value that relates to user preference. The more retweets a post has, the more people like that post. I set the significance level at $p <= 0.05$. With a p-value of 0.013 we can reject the null hypothesis that Golden Retrievers receive no more retweets than any other breed. The correlation coefficient for Goldens is 857, meaning they correlate with 857 more retweets than the 2307 retweets that any other breed is expected to receive.

Lastly, I looked at the significance of the dog stage name used in the tweets. The four terms are floofer, puppo, pupper, and doggo. When I ran a multiple linear regression of the terms I found that both doggo and puppo resulted in statistically significant increases in retweet count. The correlation coefficient of doggo was 3293 and for puppo it was 3236. This means

that both terms correlate with nearly 3300 more retweets than the average post with no term included.

I found statistical significance, but are these findings practically significant? Knowing the relationship between retweets and favorites are important for understanding the “currency” of Twitter. With such a strong correlation and a correlation coefficient of ⅓ it is a very clear insight into how users respond to the WeRateDogs twitter account. I do not know very much about how retweets affect the visibility of a tweet, but I would assume that 850 more retweets is a significant number when the average of the whole dataset is 2300. Using doggo or puppo would also be practically significant as well, as they receive more than two and a half times as many retweets as those without a dog stage in the tweet.