

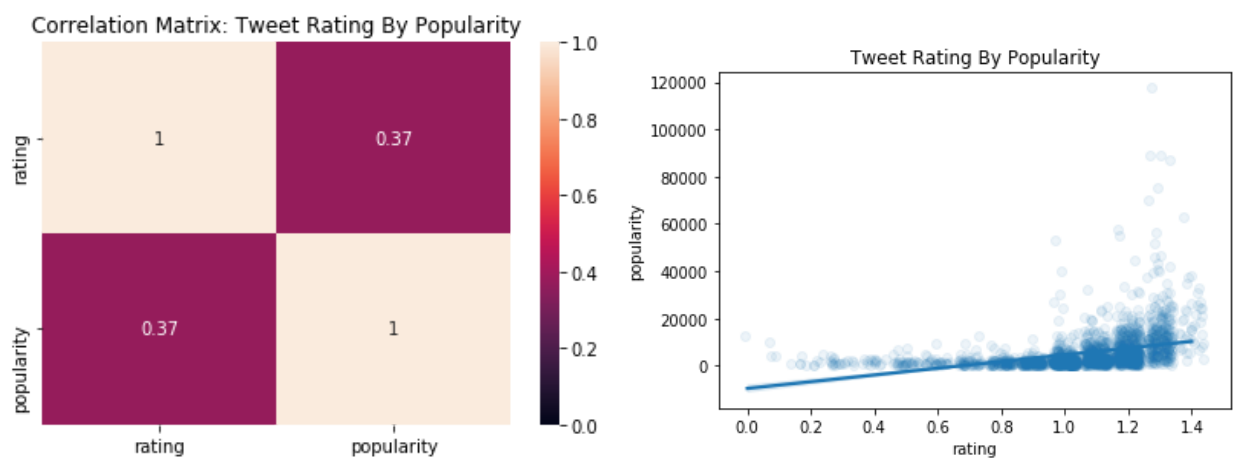
Analysis Report

This analysis of the WeRateDogs twitter account aims to accomplish two goals: to understand the key metrics to predict tweet popularity and to understand a basic data story of the textual tweets via data visualization and sentiment analysis.

Popularity Analysis

Popularity of tweets is measured by likes and retweets. Due to these two measures having a strong positive correlation, popularity is benchmarked by taking the average of the two values. This allows for simpler reporting and data visualizations.

It was predicted that the ratings of the dogs had no correlation with popularity. That is, they are just used for entertainment purposes. This proved to be unlikely to be true. Note here that this wasn't tested rigorously and is only descriptive at this point in time.



Notice that the correlation matrix shows a positive correlation between rating and popularity. This is supported visually by the scatterplot.

The tweets within the range of 0.0 and 0.8 are the least popular. Between 1.2 and 1.3 they seem to be the most popular. 1.0, 1.1, and 1.4 seem to be moderately popular.

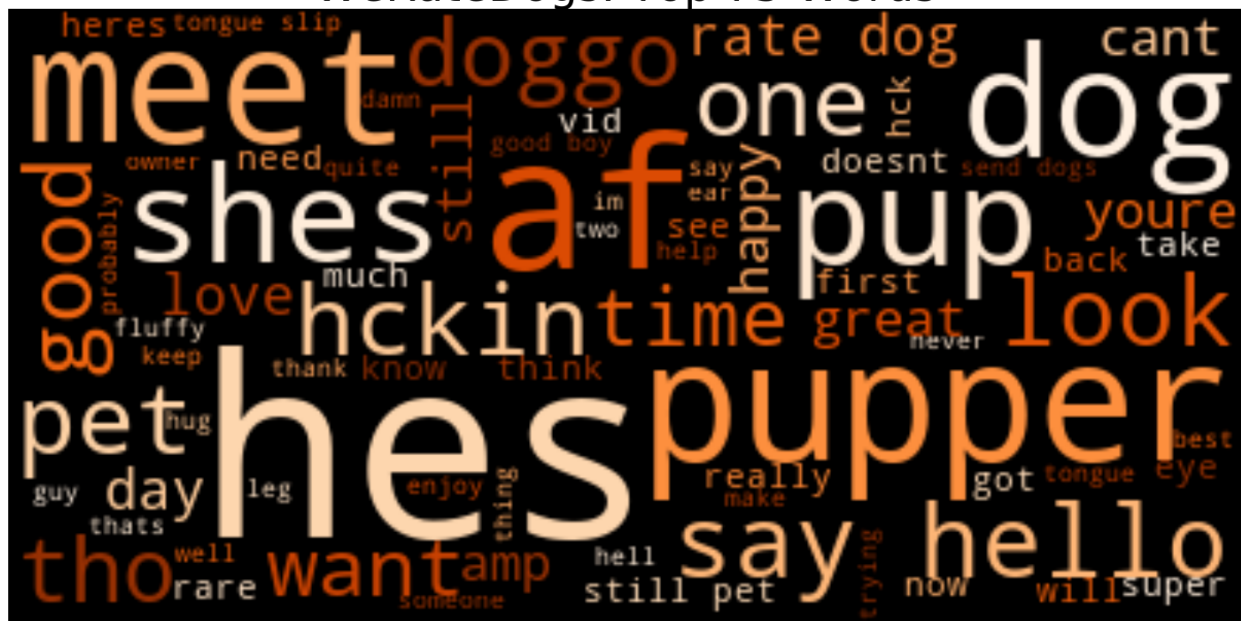
For context, the ratings of the user WeRateDogs are made for comedic purposes and aren't based on anything objective. This explains why ratings of 1.4 are typically less popular than ratings of 1.2 and 1.3. This leads to the question of why there is any correlation between higher ratings and higher popularity if the ratings are given without any objective measurement.

For example, perhaps the user WeRateDogs uses the rating of 0.0-0.9 typically for non dog photos. This would be like giving a turtle a low rating and calling it a dog for comedic purposes. The 1.2 to 1.3 ratings could be given primarily to puppies. The 1.4 ratings could be given to inspirational stories.

Note that the patterns investigative are tentative and haven't been tested for statistical significance. It could be due to chance.

A textual analysis is also performed for the purposes of telling a data story of the users tweets.

WeRateDogs: Top 75 Words



Words like `pup`, `doggo`, `pupper`, `hckin` `hck`, `af`, `good boy`, and many more are seen that are largely associated with the account. It shows patterns in the comedic elements of the user's posts.

A sentiment analysis is also performed amongst the top 75 words and the corpus of tweets.

Sentiment Analysis	Polarity	Subjectivity
Top 75 Words	0.486	0.577
Corpus	0.159	0.588

The polarity and the subjectivity of the top 75 words compared to the entire corpus of tweets is analyzed. The subjectivity, measuring between 0 and 1, shows the posts trend toward subjectivity rather than objectivity for both the top 75 words and the entire corpus.

The top 75 words and the corpus trend in the positive direction for polarity. Interestingly, the top 75 words are 3 times as positive in polarity to the entire corpus (0.16 V 0.48). This could be the result of WeRateDogs tendency to repeat positive words frequently (like hug) yet having posts that are comically negative (like telling a puppy to get his life together).

Further Investigation

Further investigation on using timestamps to predict tweet popularity by time posted and across time may also yield predictive results.

Also breed by popularity was analyzed but proved to be not very useful, because the popularity scores are ultra similar to each other and there are too many breeds to chart. However, by pulling various metrics about the different dog breeds, scatterplots can be generated on dog metrics and popularity. For example, light weight dog breeds may be more popular then heavy breeds.

Lastly, plotting the sentiments to popularity may lead to predictive results, seeing if polarity or subjectivity correlates to popularity.