

WeRateDogs Udacity Project

A data wrangling project by Theresa Sunday

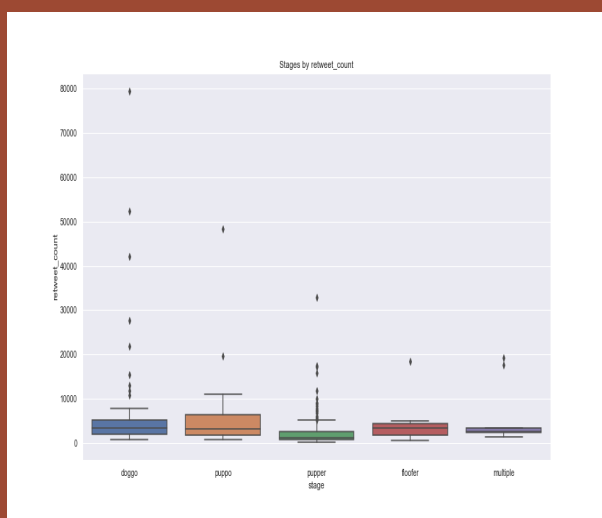
WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The twitter account has garnered over 9 million followers and has been active since November 2015.

With over 2000+ tweets from the Twitter account, we have answered popular questions from what makes a tweet get high likes to what breed gets the most liked!!

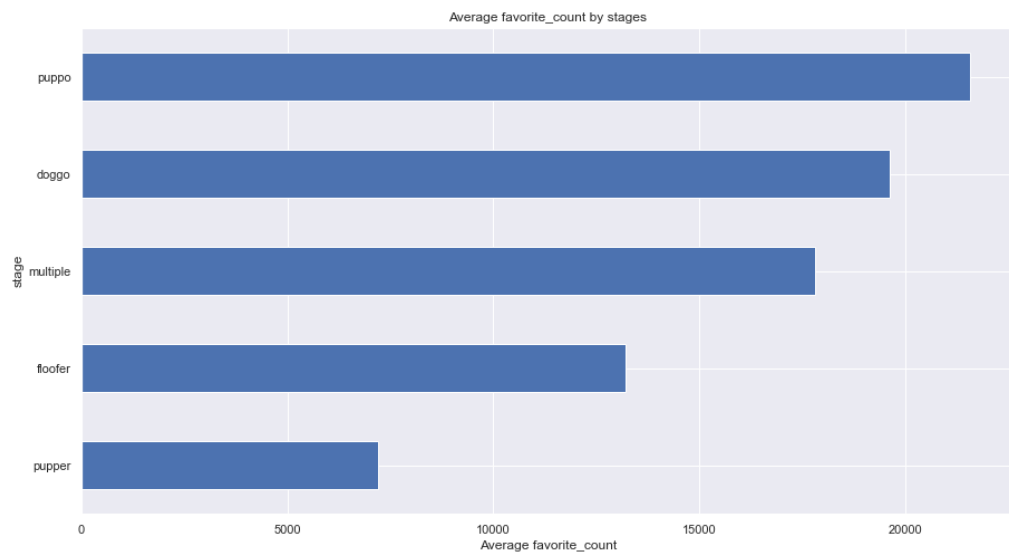
After gathering the data from various sources and cleaning, we have the visualizations to help you answer these burning questions.



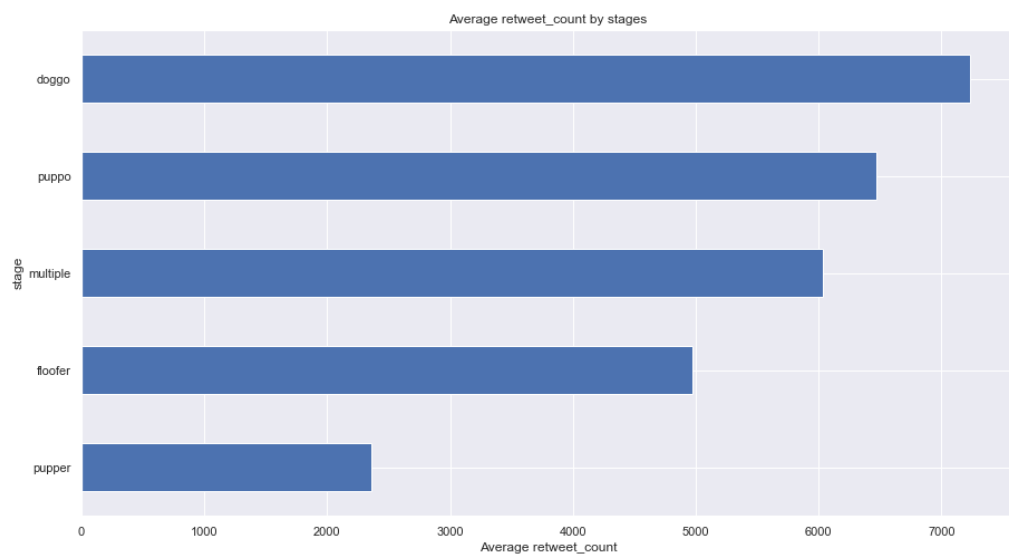
Question 1: How do different dog stages compare in likes and retweets?

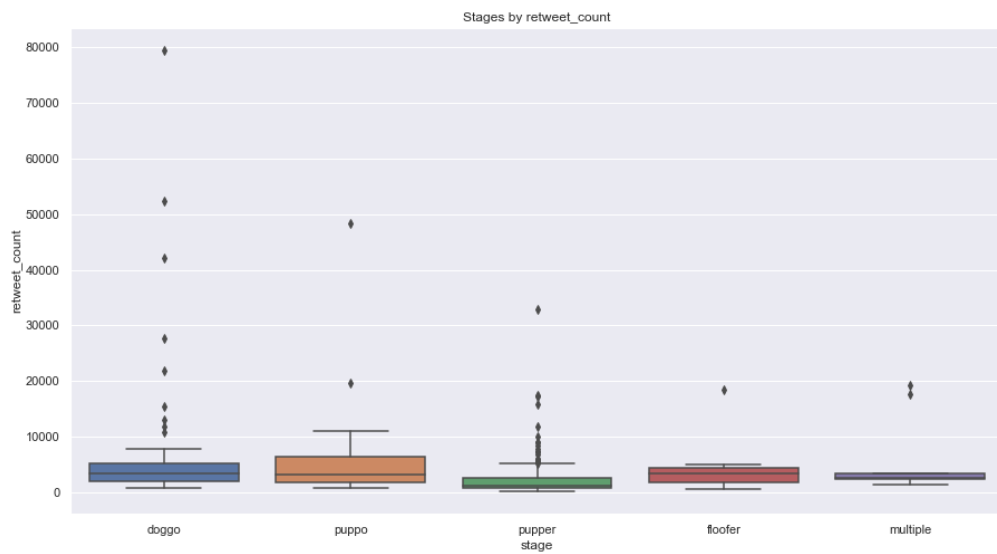


This shows a box plot distribution of the different dog stages and their retweet_count. Doggo shows a retweet_count of almost 8000+ which is the highest based on the dataset.



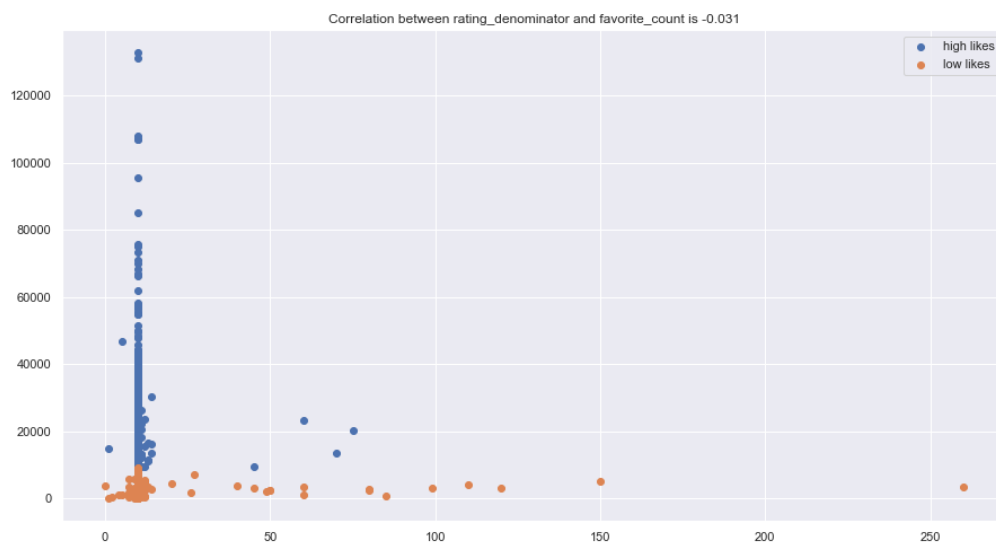
The above chart is the dog stages by average likes. Puppo has the highest average likes per tweet than any other dog stage.

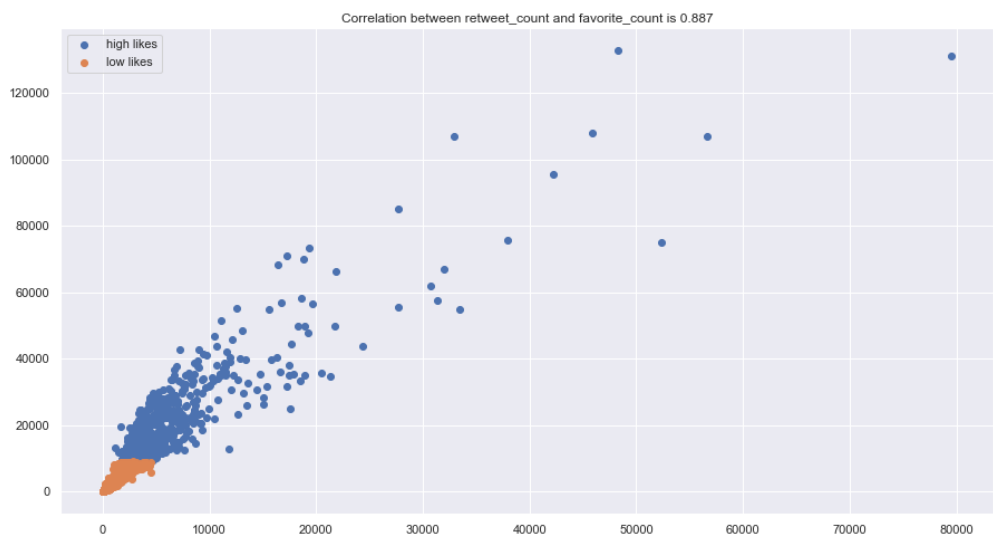
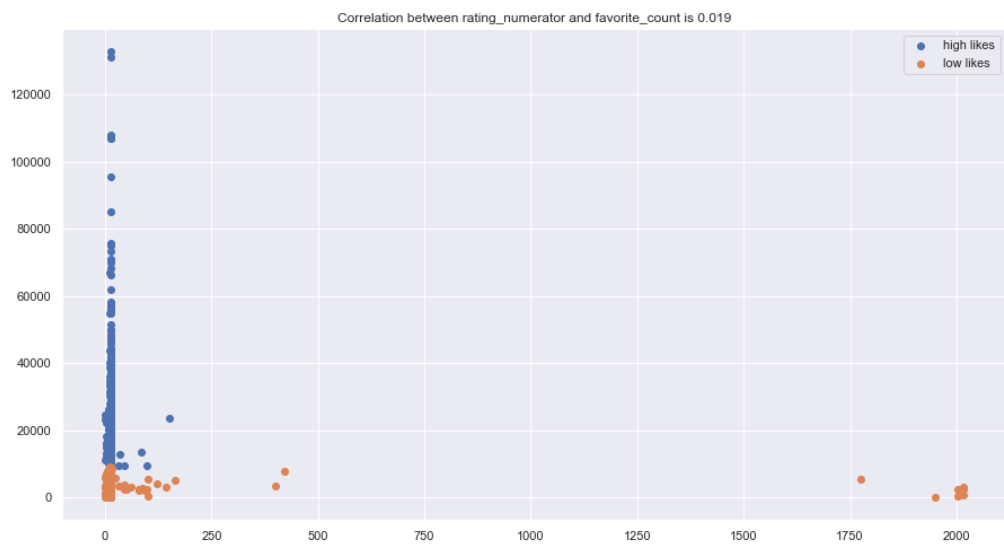


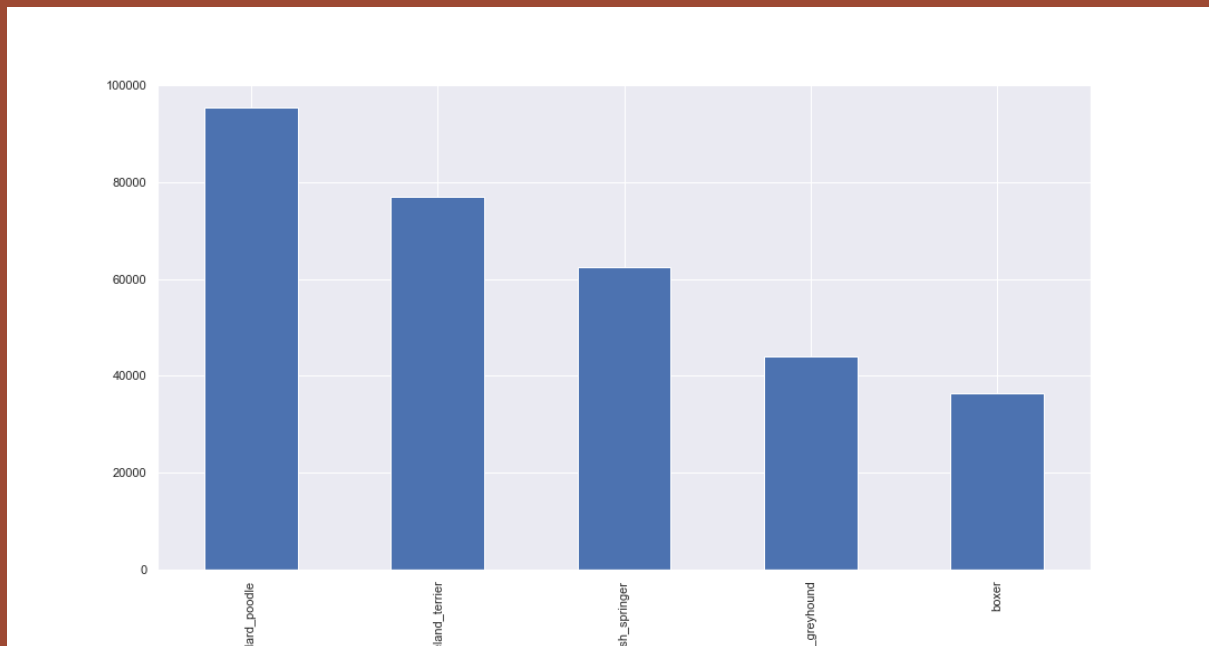


Similarly, the two charts above show a bar chart and box plot respectively. Each chart showing relationship between the various dog stages and retweets. Doggo had the highest average retweets.

Question 2: What are the characteristics associated with tweets with high likes?

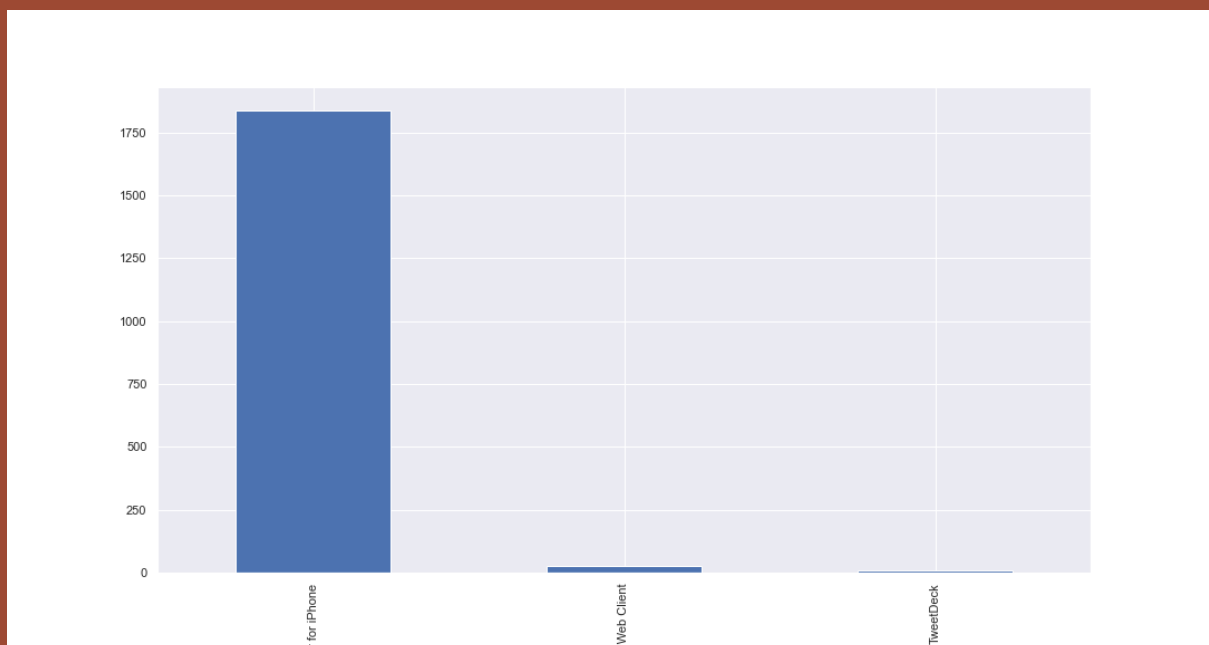






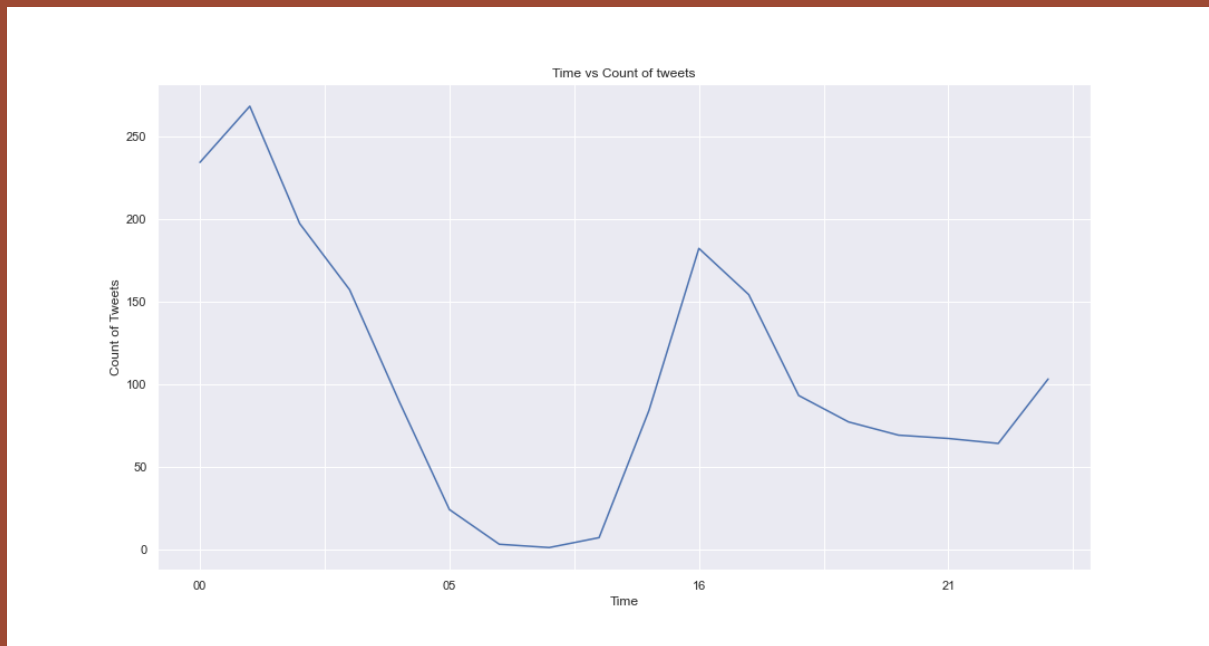
Insight: From the scatter plots, we can see how strongly correlated retweets and likes are. The third scatter plot shows how movies with low likes tended to have lower amounts of retweets and vice versa. The bar chart shows the dog breeds with the top 5 highest likes and it is no surprise that most of them tend to be good family dogs.

Question 3: What device does the WeRateDogs twitter account use in sending tweets?



Almost 80% of the tweets gathered are tweeted from an iPhone with very little tweets using web client and deck.

Question 4: At what time do the WeRateDog account send in tweets?



Interesting to observe that most of the tweets came in after midnight and showed little or no tweet from 5am till afternoon.

LIMITATIONS:

- The dataset latest entry is dated August 2017 which is 5 years ago and a lot of changes might have occurred during that period.
- Due to missing data, some important information might have been lost which would have aided better analysis.