#WeRateDogs
Analyzing @dog_rates Twitter Archive

by Sara Vicente

Real-world data rarely comes clean. Using Python and its libraries, I gathered data from Twitter's API, assessed its quality and tidiness, then cleaned it.
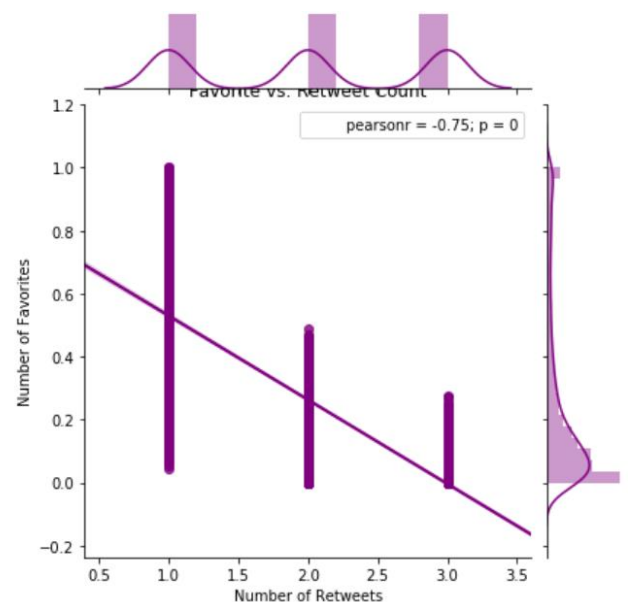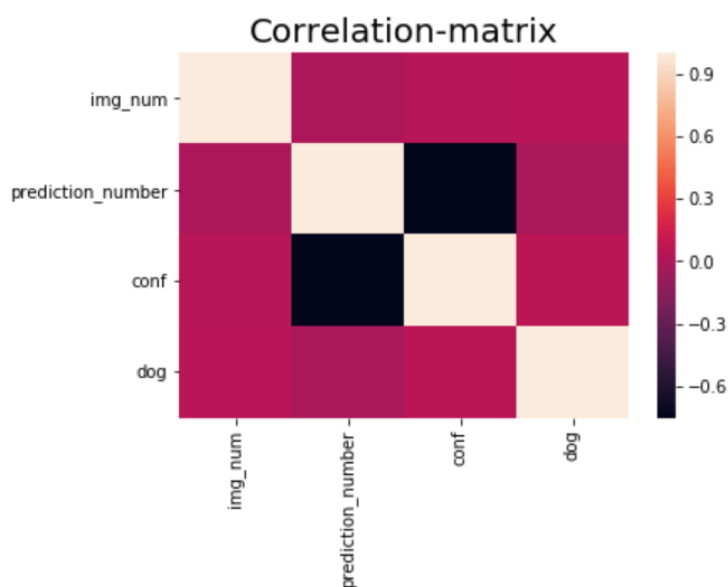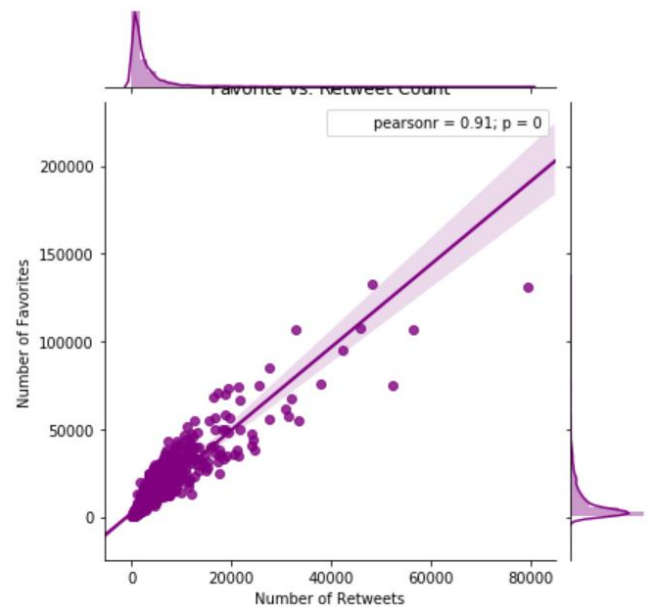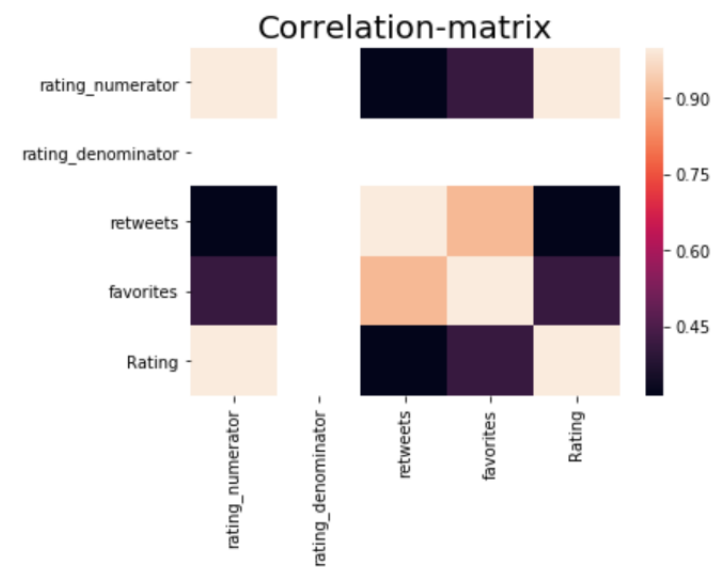
The dataset that I wrangled, analyzed and visualized is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for this project, the archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017. Nevertheless, retweet count and favorite count are two of the notable column omissions. Fortunately, this additional data has been gathered from Twitter's API.

In this paper I document my findings from the data analysis about the WeRateDogs Twitter Archive.
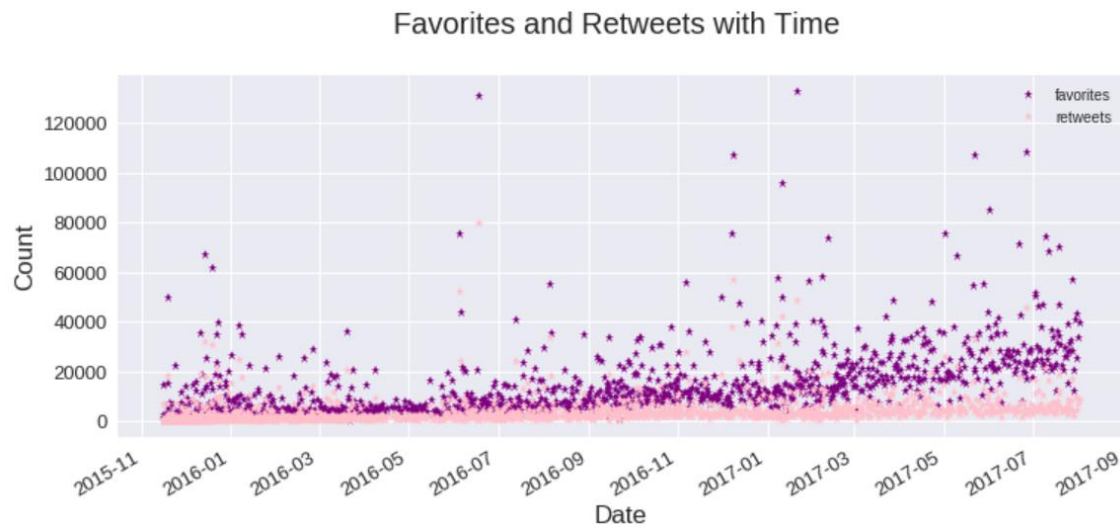
QUESTION 1: ARE THERE ANY VARIABLES CORRELATED WITH EACH OTHER?

As expected, there is no strong correlation between two different variables except for *Prediction_number* and Confidence, as the negative correlation was already a norm in the dataset construction.
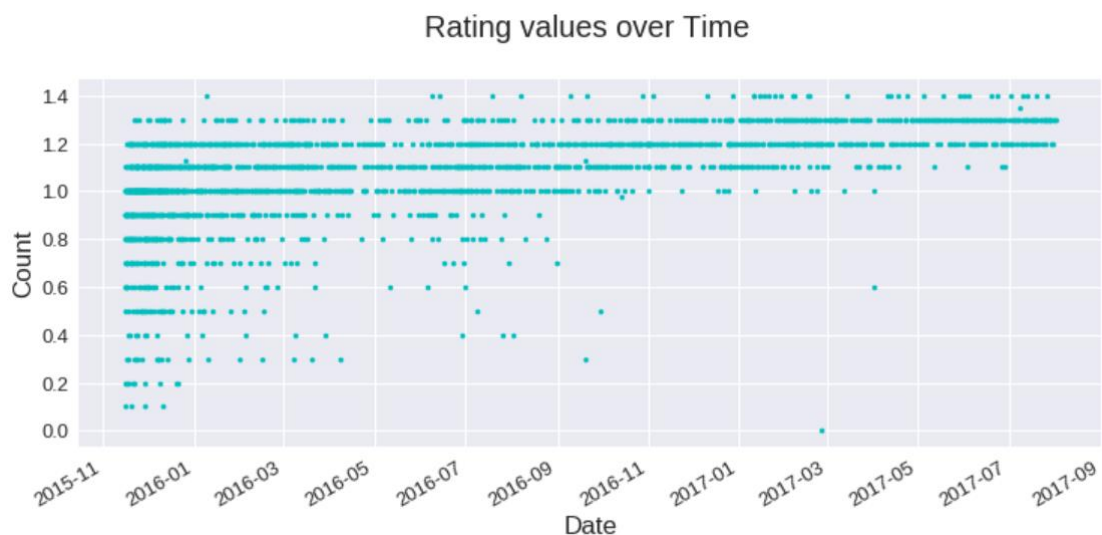
QUESTION 2: DISTRIBUTION OF FAVORITES, RETWEETS AND RATING OVER TIME?

Favorites and Retweets with Time

As can be seen in the plot, the number of favorites increases over time, with a moderate number of favorites in November 2015 and growing progressively over the years until September 2017.

However, there is no progressive growth in the number of retweets, which seems to decrease its density in the second semester of the year 2017.
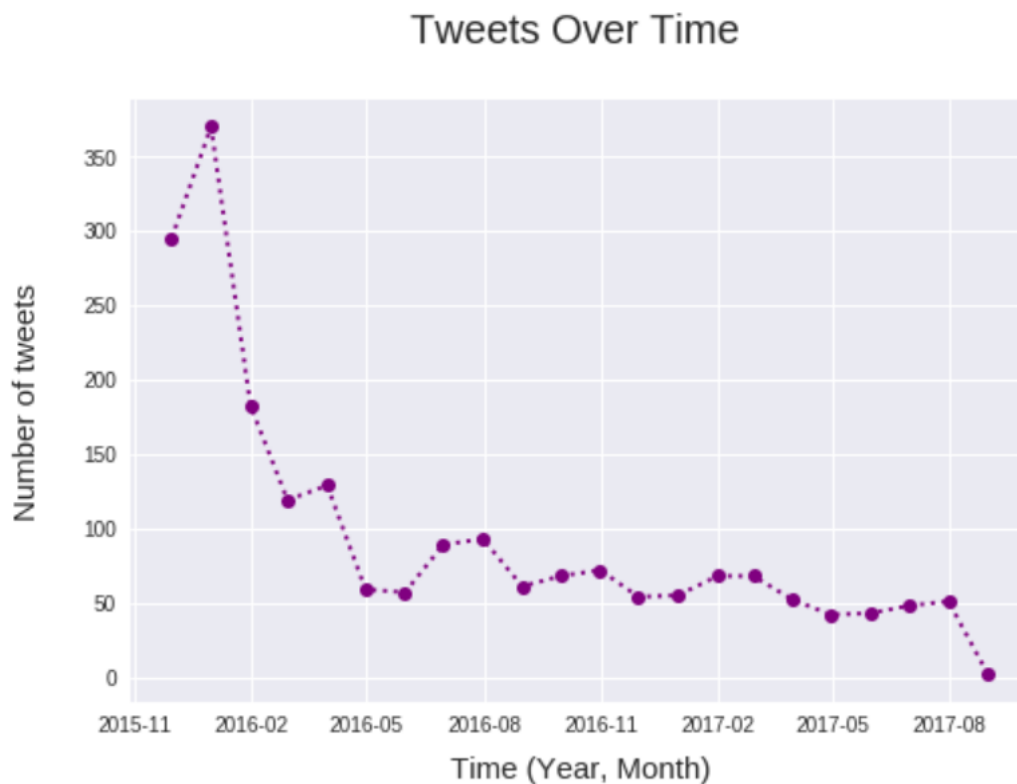
Rating values over Time

In the figure above, it is clearly seen how there are not ratings with values less than or equal to 0.8 (except for very specific outliers) after September 2016.

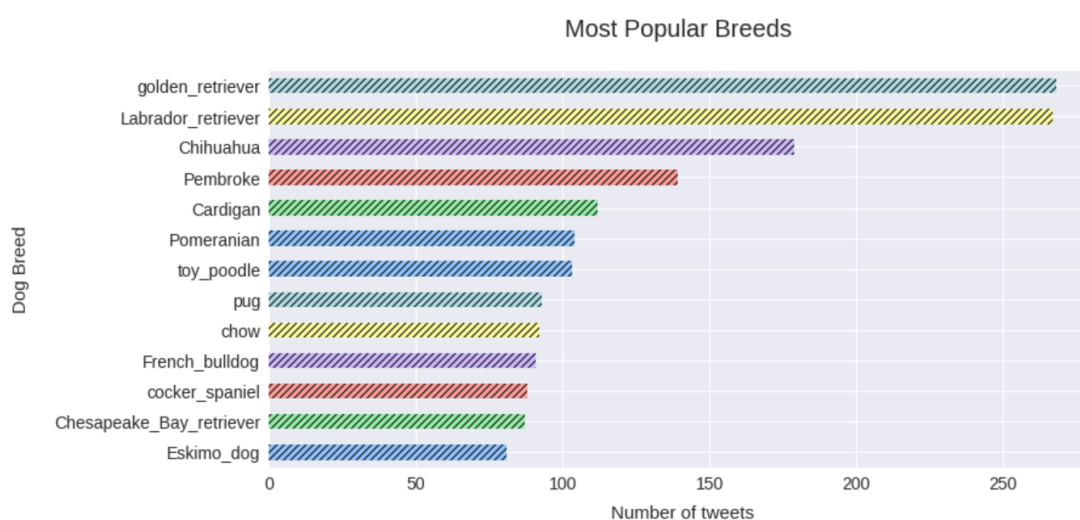The high ratings have maintained its great popularity over time.

QUESTION 3: WHAT IS THE DISTRIBUTION OF THE NUMBER OF TWEETS OVER TIME?

Tweets Over Time

The number of tweets decreases over time, showing a sharp drop in December 2016.

QUESTION 4: WHICH IS THE MOST POPULAR DOG BREED?
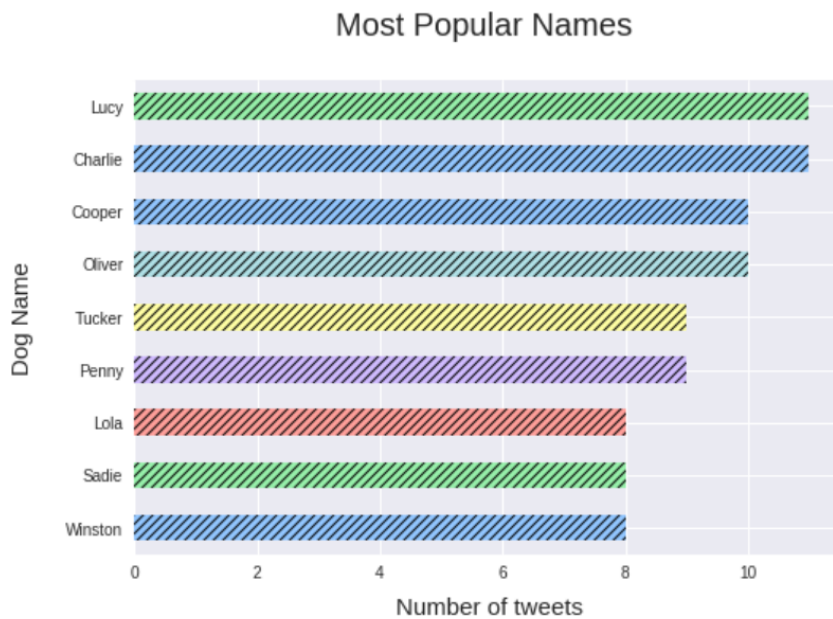

Most Popular Breeds

The most popular dog breeds are Golden Retriever and Labrador Retriever, followed by Chihuahua and Pembroke.

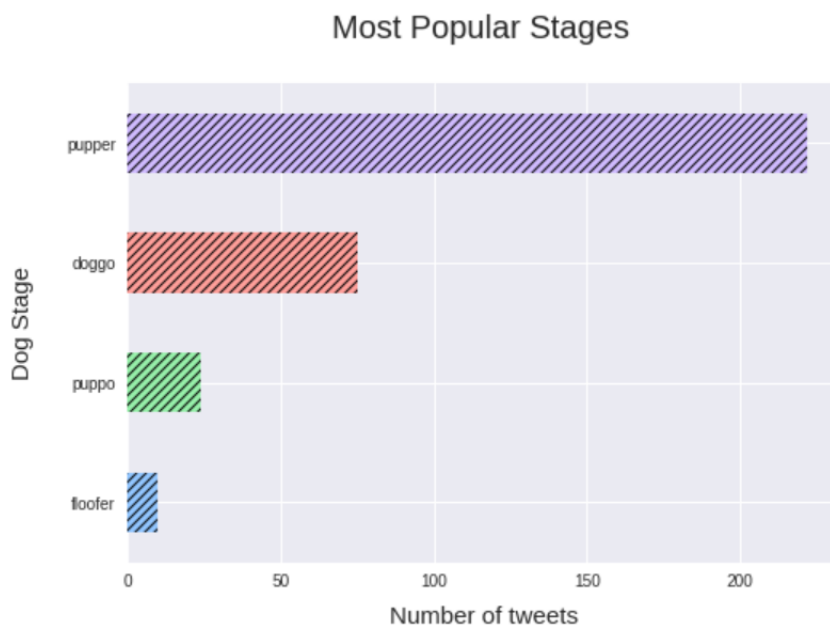## QUESTION 5: WHAT IS THE MOST POPULAR DOG NAME?

Most Popular Names

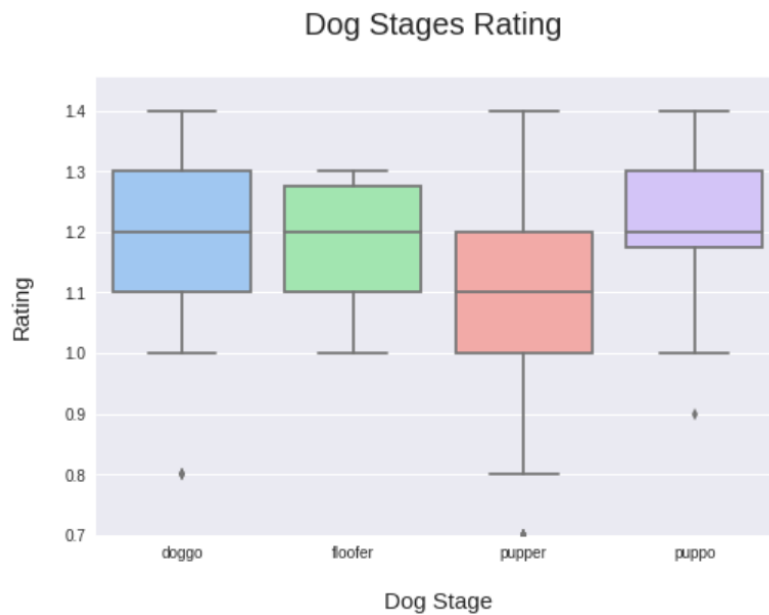The most popular dog names are Charlie and Lucy, followed by Cooper and Oliver.

QUESTION 6: WHAT IS THE MOST POPULAR DOG STAGE?

Most Popular Stages

The most popular dog stage is by far "Pupper", followed by "Doggo", "Puppo" and "Floofer".
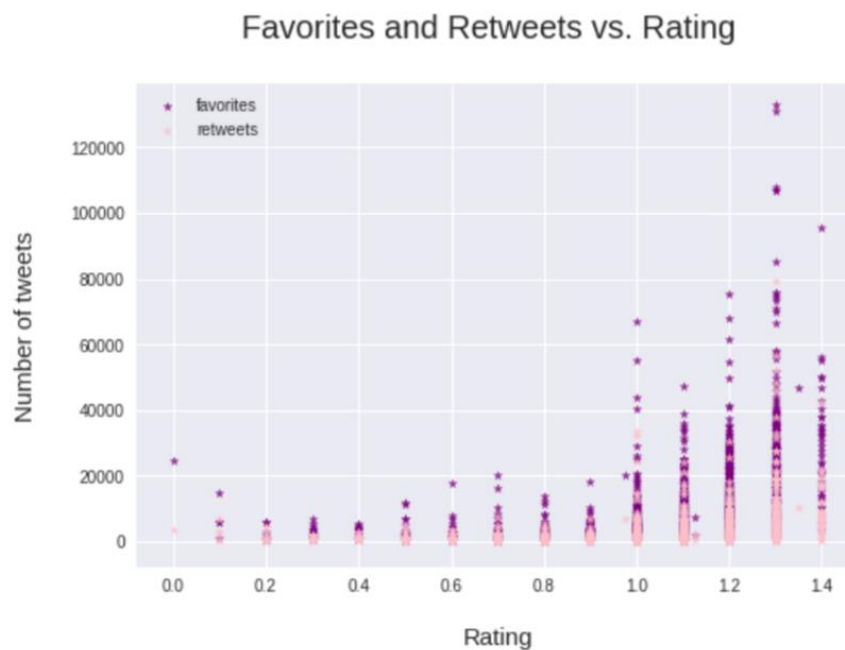
QUESTION 7: WHICH DOG STAGE HAS THE HIGHEST RATING?

Dog Stages Rating

"Puppo" dogs got the highest rating and are the less rated, while "Puppers" got the lowest rating and are the most rated ones.

QUESTION 8: DO TWEETS WITH GREATER RATING GET MORE RETWEETS AND FAVORITES?

Favorites and Retweets vs. Rating

The number of retweets and favorites increases with the Rating value, reaching the bullish peak for a rating of 13.

QUESTION 9: DO TWEETS GET MORE RETWEETS OR FAVORITES IF THEY HAVE A PICTURE?

The average confidence level for the true picture's prediction is: 61.3978574033 %

The number of retweets of tweets WITH a picture is: 4057446
The number of retweets of tweets WITHOUT a picture is: 1845634
The percentage difference is: 37.4687790103 %

The number of favorites of tweets WITH a picture is: 13530488
The number of favorites of tweets WITHOUT a picture is: 5150552
The percentage difference is: 44.857973646 %

Therefore, tweets have more retweets and favorites if they include an image, also, this fact affects the favorites more.

QUESTION 10: IMAGES OF THE GOLDEN RETRIEVERS (MY FAVORITE DOG BREED) BEST RATED AND WITH THE MOST NUMBER OF FAVORITES?

WINNER #1   WINNER #2 (MY FAVORITE<3)

WINNER #3

WINNER #4

WINNER #5

Still not your favorites? 😉