WeRateDogs Tweet Insights

By Esther Goneh

Introduction

This project aims to wrangle, analyze, visualize and gain insight into WeRateDogs tweet data. WeRateDogs is a Twitter account that rates people's dogs with humorous comments about the dog.

An example of a WeRateDogs tweet:

(Source: https://twitter.com/dog_rates/status/883482846933004288)

The data gathering

The dataset for this analysis was gathered from the following sources:

1. The enhanced WeRateDogs Twitter archive as it stood on August 1, 2017. It contains basic tweet data like the tweet text, timestamp, tweet source, tweet URL, dog's name, rating denominator, and numerator e.t.c.

2. The tweet image prediction dataset is a TSV dataset that results from a neural network classifying the images in Twitter archive data.

3. Supplementary Twitter JSON dataset, this dataset contains supplementary data -like retweet counts, and favorite counts- for WeRateDogs tweet's.

Assessing data

The data was assessed to ascertain if it's quality meets six dimensions of data quality: accuracy, completeness, consistency, timeliness, validity, and uniqueness, and the three dimensions of tidy data:

- Each variable forms a column and contains values
- Each observation forms a row
- Each type of observational unit forms a table.

Cleaning data

Once the quality and tidiness issues have been documented the data is cleaned using the following steps:

- Define: Document exactly what needs to be done.
- Code: Programmatically clean the data.
- Test: Evaluate the dataset to ensure that it was cleaned.

Analysis and visualization

The following insight was gained by analysing the data:

1. The Golden Retriever is the most tweeted breed by WeRateDogs, it is also the most retweeted and has more favorite count than any other breed for the period under analysis.

Number of tweets for top 10 dog breeds
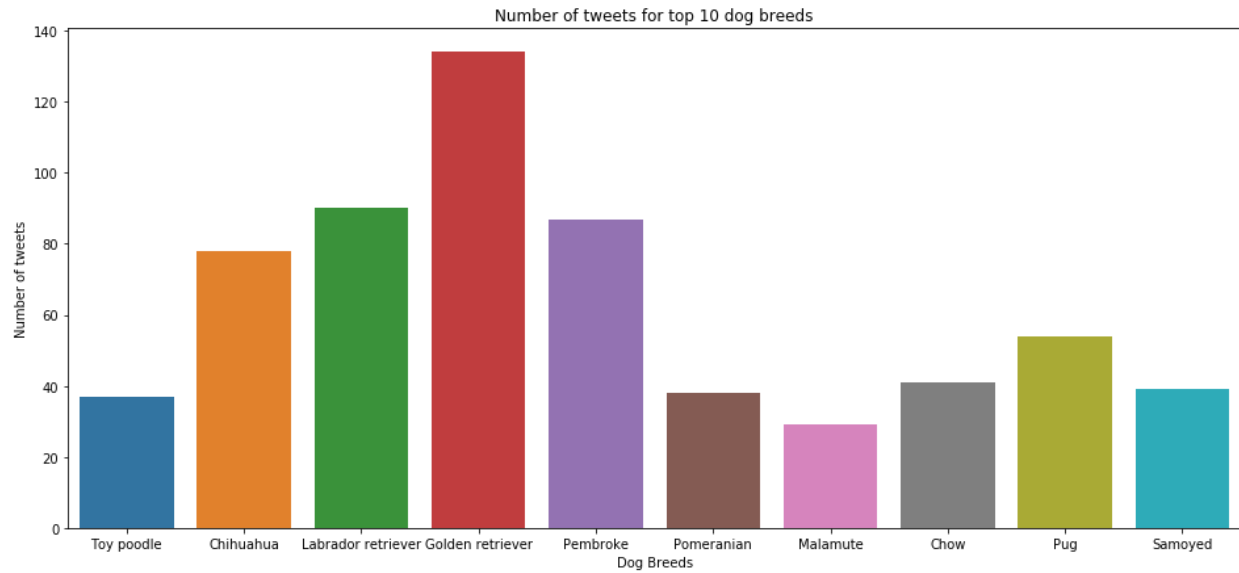
2. Dog ratings with high retweet count also have high favorite count, and dog ratings with low retweet count also have low favorite count.
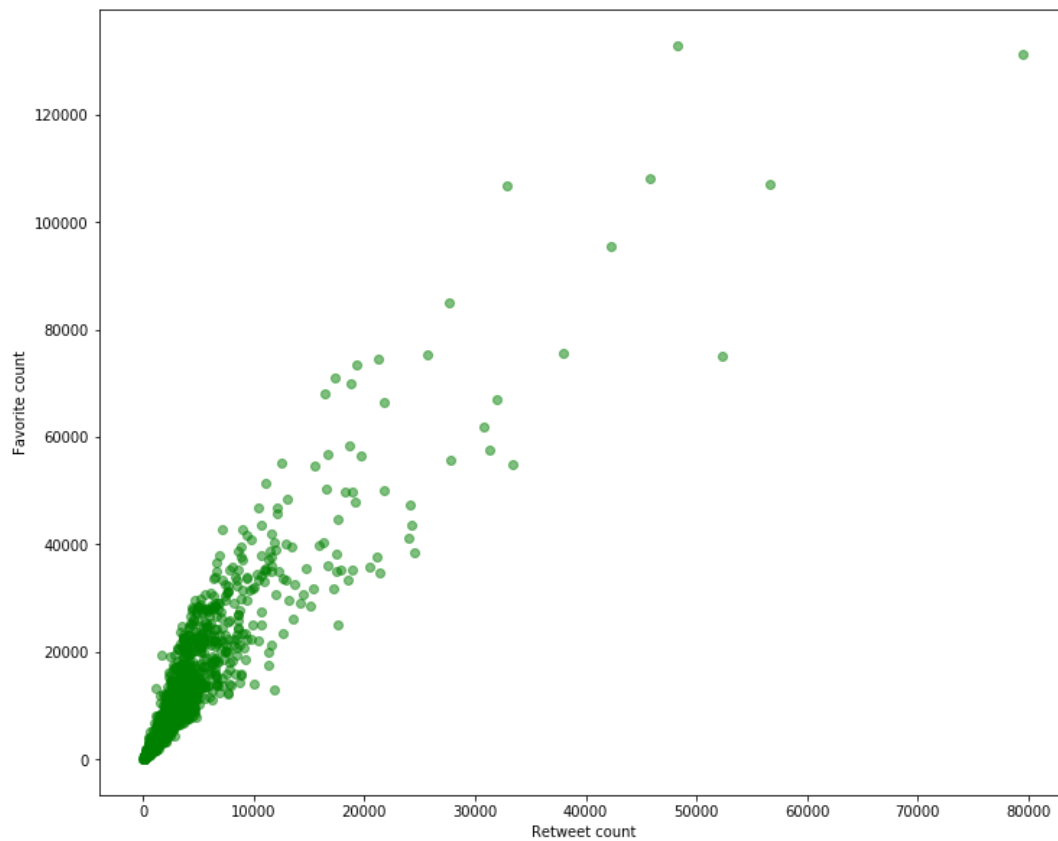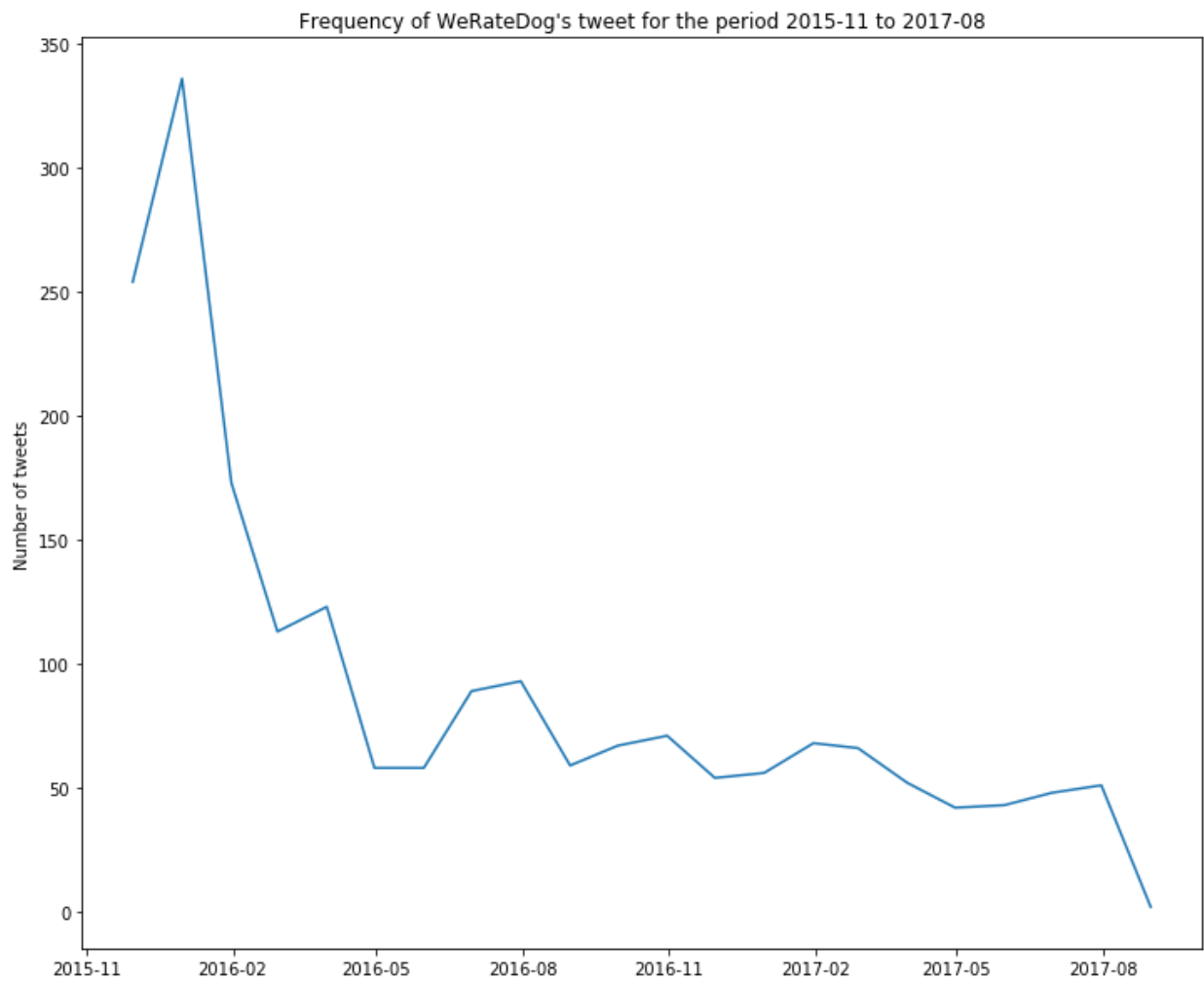
3. The dog, Atticus, holds WeRateDog's highest rating for the period under study,

with a rating of 1776/10. While the dog Crystal, holds WeRateDog's least rating of 2/10 for the period under study.

4. There is a sharp increase of tweets during the holiday period of 2015 (2015-11 to 2015-12). From January 2016 there was a steady decline in the number of tweets till the summer of 2016 where there is a noticeable spike in number of tweets. Followed by a steady decline during the periods 2016-08 - 2016-9. The

rates of tweets largely remained constant with noticeable increase during the holidays and summer period throughout the remainder of months till 2017-08 when there was a steep decline. From this data it can be deduced that the operator of WeRateDogs twitter page is most likely a student during this period. The increase in the amount of tweets coincide with period when students are on break.

Frequency of WeRateDog's tweet for the period 2015-11 to 2017-08

5. WeRateDogs was tweeting 93.5% of the time from an apple iphone.

WeRateDogs Tweet Sources

Twitter for iPhone
Vine
Twitter Web Client
TweetDeck

93.5%

0.6%
1.4%
4.6%