act_report

June 5, 2019

Investigation of Twitter archive of `WeRateDogs`

0.1 Table of Contents

0.1.1 About the Dataset

The dataset that is being wrangled (and 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.

This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017. More on this soon.

0.1.2 Inspiration

Is it possible to find the best rated dogs based on the tweets
  ### Questions - - Ratio of people missing appointments - Absence of people based on gender - Appointments based on hour of the day - Appointments based on the days of the week - Appointments based on month - Most important factor which leads to their absence
  Home
  ## Exploratory Data Analysis
  Home ### Research Question 1 (Replace this header name!)

0.1.3 Basic plots based on the available columns

0.1.4 Plot 1

Average Rating of dogs based on their category   The above plot shows that `normal` category dogs have got highest rating. Also their rating has highest deviation. `pupper` category dogs have got the least rating with the lest amoun of deviation.

1

0.1.5 Plot 2

No of tweets based on month   The above plot shows that `December` and `November` are the months having he most no of tweets. `August` has the least no of tweets.

0.1.6 Plot 3

No of tweets based on the day of the week   The above plot shows that `Monday` is having the highest no of tweets. Also we can see that the no of tweets gradually decrease along the week from `Monday` to `Saturday`.

0.1.7 Plot 4

Average Rating of doga based on the day of week   The above plot shows that `Monday` and `Saturday` are having the highest rating among all the days of the week. Also we can see that both of them are having higest amount of deviation. The statistics for the other days are almost the same.

0.1.8 Plot 5

No of tweets based on hour of the Day   The above plot shows that the most no of tweets are done in the early morning time and during the evening around 5. The time duration between 5AM and 2PM is having the least number of tweets.

0.1.9 Plot 6

Retweets based on time difference   The above plot shows that most of the retweets are done within 1 day of the tweeted time. This number of tweets gradully decrease as the time difference increase. After the ertain period which looks near 1 year, the no of retweets are approximately 0.

0.1.10 Plot 7

Heatmap   The above heatmap shows that the `rating` is mostly dependent on the `numerator` of the rating rather than the `denominator`. `Hour` is least correlated with the `rating` as it has value of `0.02`. We can also see some negative correlation between `retweetTime`, `re-Hour` and `denominator`, `re-Hour`. This means that the retweet hour is important in determining the rating of the dogs.

0.1.11 Plot 8

Top 10 most common names of Dogs   From the above plot we can see that the most common names for dogs are - `Charlie`, `Cooper`, `Lucy`, `Oliver`, `Tucker`, `Lola`, `Penny`, `Bo`, `Winston` and `Sadie`

0.1.12 Plot 9

Top 10 most rated dogs   From the above plot we can see that the top 10 highest rated dogs are `Atticus`, `Sam`, `Doobert`, `one`, `Laika`, `Ollie`, `Gary`, `Bo` and `Sunny`

Above all `Atticus` is the highest rated dog with rating of `177.6`

0.1.13   Some of the happy followers of `WeRateDogs`

The End
   Home