Analysis of We Rate Dogs Tweets

Introduction

This document captures the results and visualizations of Exploratory Data Analysis of a popular @WeRateDogs twitter archive. 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."

Dataset

In this project I gathered datasets from multiple sources, assessed them and cleaned them into a single master dataset. This wrangled dataset is used for analysis and visualizations in the subsequent sections.

Below is the information of this master dataset

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2116 entries, 0 to 2115
Data columns (total 27 columns):
tweet_id                    2116 non-null int64
in_reply_to_status_id       2116 non-null object
in_reply_to_user_id         2116 non-null object
timestamp                   2116 non-null datetime64[ns]
source                      2116 non-null object
text                        2116 non-null object
retweeted_status_id         2116 non-null object
retweeted_status_user_id    2116 non-null object
retweeted_status_timestamp  0 non-null datetime64[ns]
expanded_urls               2116 non-null object
rating_numerator            2116 non-null float64
rating_denominator          2116 non-null int64
name                        2116 non-null object
dog_stage                   2116 non-null object
jpg_url                     1993 non-null object
img_num                     1993 non-null float64
prediction1                 1993 non-null object
prediction1_confidence      1993 non-null float64
prediction1_dog             1993 non-null object
prediction2                 1993 non-null object
prediction2_confidence      1993 non-null float64
prediction2_dog             1993 non-null object
prediction3                 1993 non-null object
prediction3_confidence      1993 non-null float64
prediction3_dog             1993 non-null object
favorite_count              2116 non-null int64
retweet_count               2116 non-null int64
dtypes: datetime64[ns](2), float64(5), int64(4), object(16)
memory usage: 462.9+ KB
```

Analysis

Descriptive Statistics
I get the statistical summary of the master dataframe to analyze the numerical data.

```
archive_clean.describe()
```

	tweet_id	rating_numerator	rating_denominator	img_num	prediction1_confidence	prediction2_confidence	prediction3_confidence	favorite_count
count	2.116000e+03	2116.000000	2116.0	1993.000000	1993.000000	1.993000e+03	1.993000e+03	2116.000000
mean	7.362910e+17	11.638601	10.0	1.203211	0.593802	1.344685e-01	6.026575e-02	8902.163516
std	6.706037e+16	39.452319	0.0	0.560899	0.271951	1.006821e-01	5.089760e-02	12514.794954
min	6.660209e+17	0.000000	10.0	1.000000	0.044333	1.011300e-08	1.740170e-10	79.000000
25%	6.766166e+17	10.000000	10.0	1.000000	0.362835	5.405530e-02	1.619070e-02	2000.500000
50%	7.093173e+17	11.000000	10.0	1.000000	0.587507	1.175080e-01	4.952370e-02	4092.500000
75%	7.870529e+17	12.000000	10.0	1.000000	0.845256	1.952180e-01	9.160200e-02	11177.250000
max	8.924206e+17	1776.000000	10.0	4.000000	1.000000	4.880140e-01	2.734190e-01	143578.000000

Observations:

- The mean rating is 11.63. The highest is 1776 which corresponds to a humorous tweet.

- The machine learning predictions seemed to perform better in Prediction one with a higher mean of confidence rate (0.598)

- The maximum number of times a tweet was favorited was 143578 times

- The maximum number of retweets are 77476

High Rated Tweets

WeRateDogs almost always rates dogs above a 10 with the maximum being 14. However, on analysis it is found that few tweets have much higher values which I initially suspected to be a wrangling oversight but turned out to be a valid humorous tweet. Below are two such tweets.

Also, since these are humorous posts, I was curious to see how the machine learning predictions were for these images.

WeRateDogs™ ✔
@dog_rates
Follow ˅

This is Atticus. He's quite simply America af. 1776/10

8:00 AM - 4 Jul 2016

2,712 Retweets 5,513 Likes

Tweet ID: 749981277374128128

This tweet was posted on July 4, 2016.

It was rated 1776/10 and was one of the popular tweets.

Predictions: bowtie, sunglasses, sunglass.

Tweet ID: 670842764863651840

Although, not truly a dog rating. This tweet posted in November, 2015 is one of the most popular and humorous tweets of We Rate Dogs.
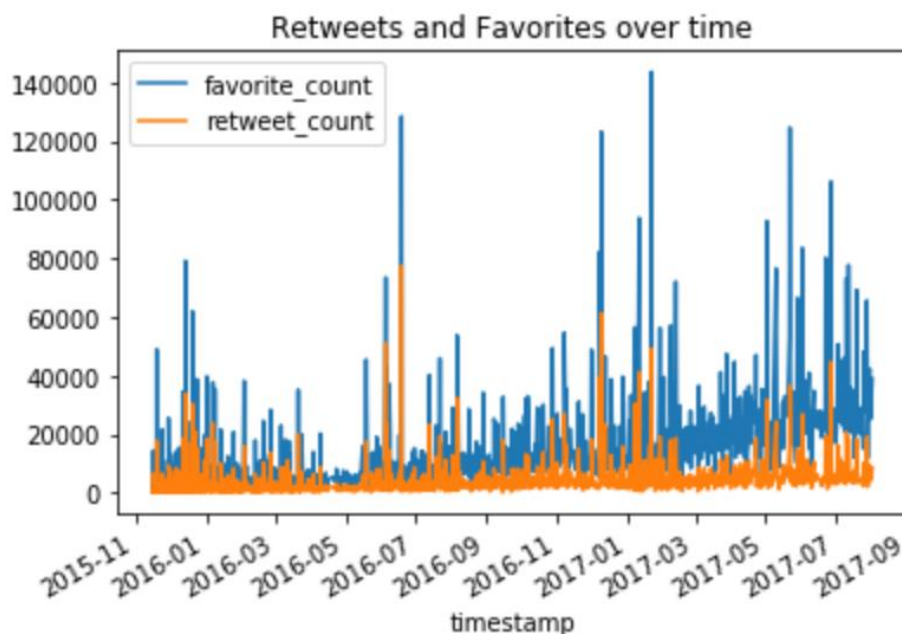
Predictions: Microphone, accordion, drumsticks.

Visualizations

Retweets and Favorites

As WeRateDogs have gained popularity over period of time, I have analysed the retweet and favorite count over month/year since 2015. There are few spikes during funnty posts and vacations such as July 4 but overall the retweets/favorites are on an upward trend.

I also analyze the correlation between retweets and favorites. It can be clearly seen that there is a linear and a strong positive correlation between retweets and favorites.
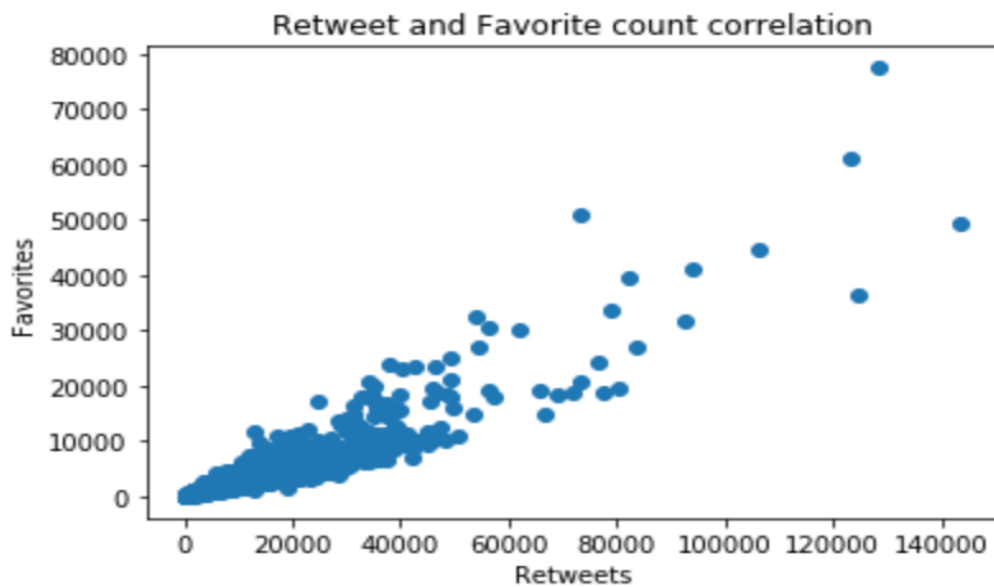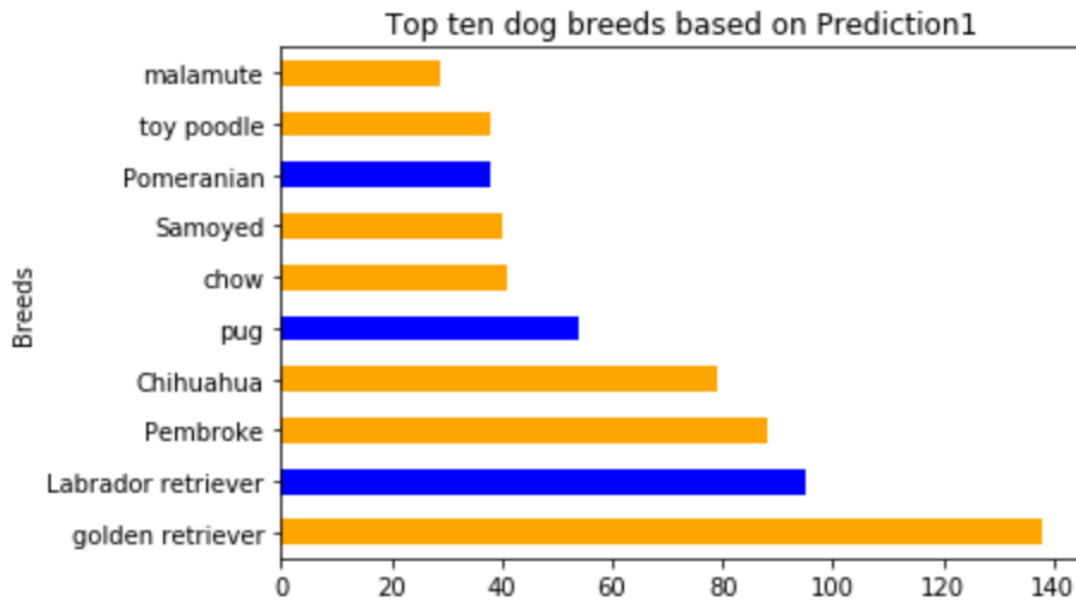

What is trending?

I analyze the dataset and extract the popular dog names and breeds that were popular in 2015-2017. Since prediction 1 has higher confidence rate, I used prediction 1 to determine the popular breeds.

Top 10 names in 2015-2017
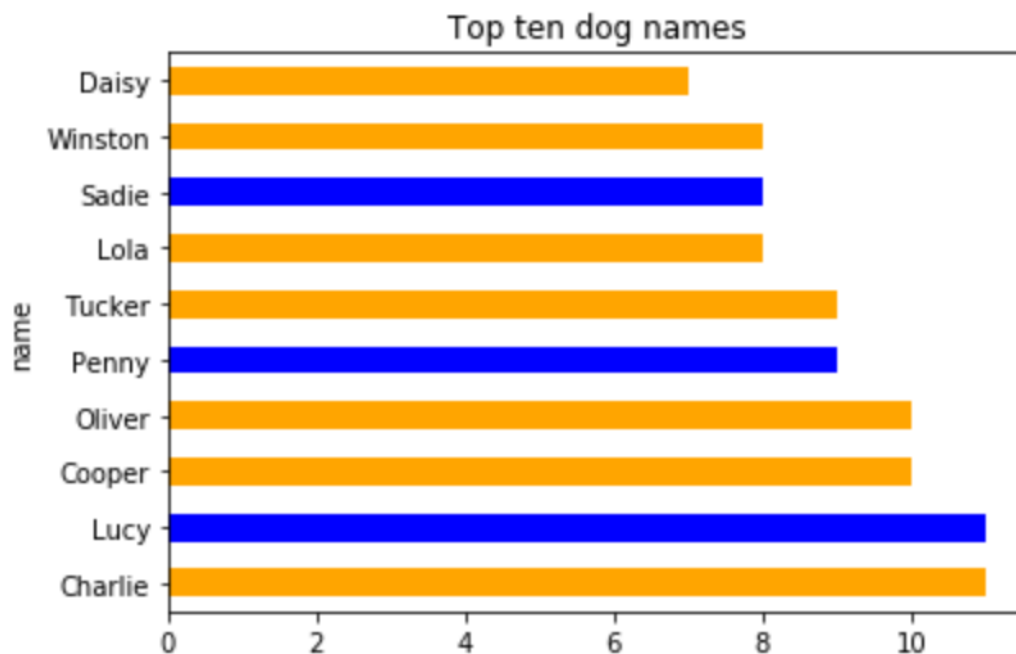
Top 10 breeds in 2015-2017

Top ten dog breeds based on Prediction1

Rating Analysis

WeRateDogs is known for its unique rating system and the analysis would not be complete without a rating analysis. I have plotted the ratings over time to see if there is a specific trend. Overall, most dogs score between a 10 -12. I have also excluded outliers for a more meaningful analysis. It is also interesting to note that the lower scores are more pronounced earlier in the time graph. There are very few low ratings starting 2017.
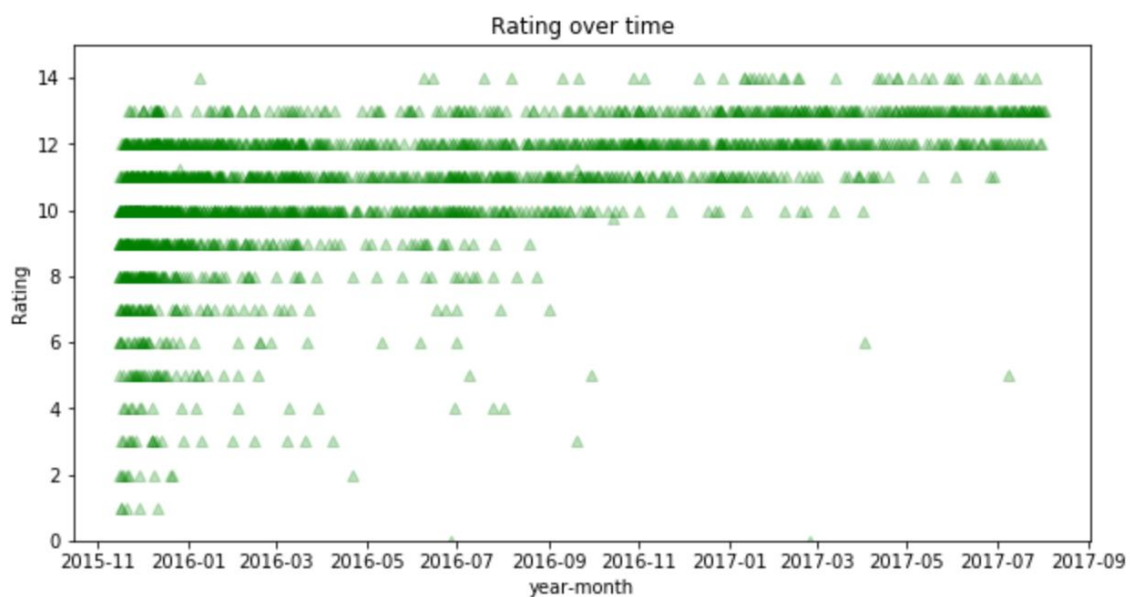
Conclusion

The data anlaysis part of this project was very interesting and gave a glimpse into the unique WeRateDogs twitter account. The time plots and trend analysis provided an insight into how the tweets, retweets and user favorites have evolved over time. It would be interesting to see how the names and breed trends change in the future datasets.