WeRateDogs® Data Wrangling Project

ACT REPORT

By Naomi Ogbechi

Introduction

This report is a part of the Wrangle and Analyze data project in the Data Analyst Nanodegree offered by Udacity. In this project, the aim is to gather data about WeRateDogs® from multiple sources including; Twitter API, to clean the data, analyze it, and generate insights.

Background of 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. The account has been tweeting since 2015, and has produced more than 5,000 tweets.

Data Sources

The data that was used in this project came primarily from three sources:

- **WeRateDogs** which provided us with a CSV file containing basic tweet data (tweet ID, timestamp, text, etc.) for all 5,000 of their tweets as they stood on August 1, 2017. These tweets were further filtered to only include tweets that contained dog ratings, leaving only 2,356 tweets.

- **Twitter API** also provided us with a JSON file containing the Retweet count, favorite count, as well as any other additional data we found interesting for each tweet.

- **Udacity** finally provided us with a tsv file containing the respective tweets top three image predictions, image url and number of images. What kind of; dog, animal, or object, was present in each tweet according to a neural network fed with twitter data provided by WeRateDogs.

Objectives

In this execution of the analysis in this project, we tried to:

- Establish which tweet had the highest interactions, and what the peculiarities about it were.
- Determine which dog class had tweets with the most interactions.
- Determine which dog breed had the most interactions.
- Determine Which months had the highest number of tweets by WeRateDogs.

Steps Taken

This project was executed in meticulous and stepwise phases, and included:

1. Gathering data from the various sources mentioned above
2. Assessing the data to detect any issues and challenges
3. Cleaning the data to fix the issues detected in the previous phase
4. Storing the cleaned data in a safe and accessible format for analysis
5. Additional Wrangling Efforts to further prepare the data for analysis
6. Analyzing the data to see if we could achieve the goals stated prior
7. Reporting on the findings, as well as the wrangling efforts undertaken during the project.

# Analysis and Visualizations

Table 1.1

tweet_id	timestamp	text	expanded_urls	rating_numerator	rating_denominator	dog_name	dog_class	tweet_month	tweet_year	retweet_count	favorite_count	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3	p3_conf	p3_dog	interact_count
74423479936002	2016-06-18 18:26:18+00:00	Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad) https://t.co/7wE9LTEXC4	https://twitter.com/dog_rates/status/744234799...	13	10	None	doggo	June	2016	70258	144035	https://pbs.twimg.com/ext_tw_video_thumb/744234667679821824/pu/img/1GaWmtJtdqzZV7jy.jpg	1	Labrador_retriever	0.82533	TRUE	ice_bear	0.04468	FALSE	whippet	0.018442	TRUE	214293

The table above shows the tweet with the highest interaction terms of retweets and favorites count, it was retweeted 70,258 times and favorited 144,035 times, combining to form a total of 214,293 interactions. The tweet is predicted to contain the picture of a Labrador Retriever which is a doggo.

Image 1.1
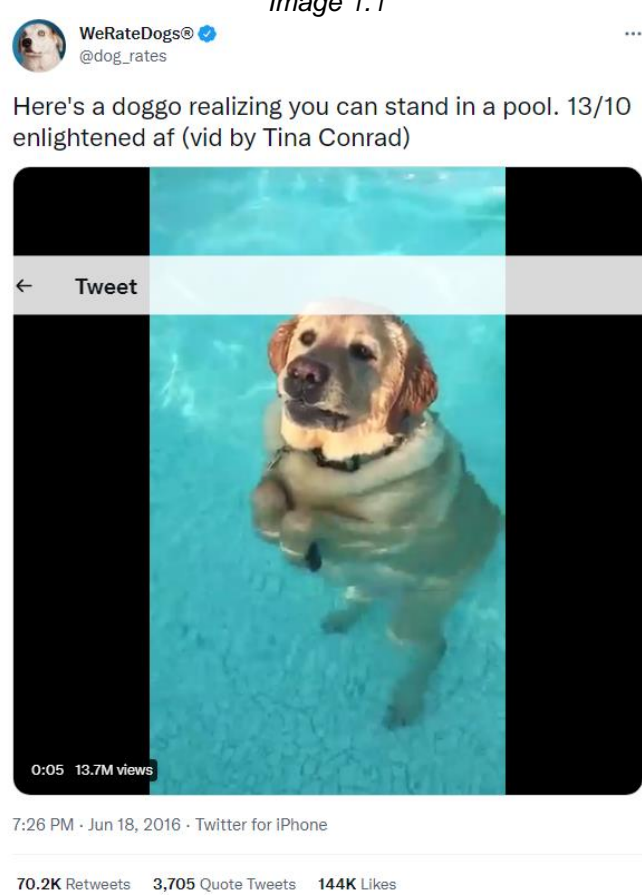
The image above shows a screenshot of WeRateDog's most engaged tweet.

Chart 1.1

Total Interactions Count by Dog Classes
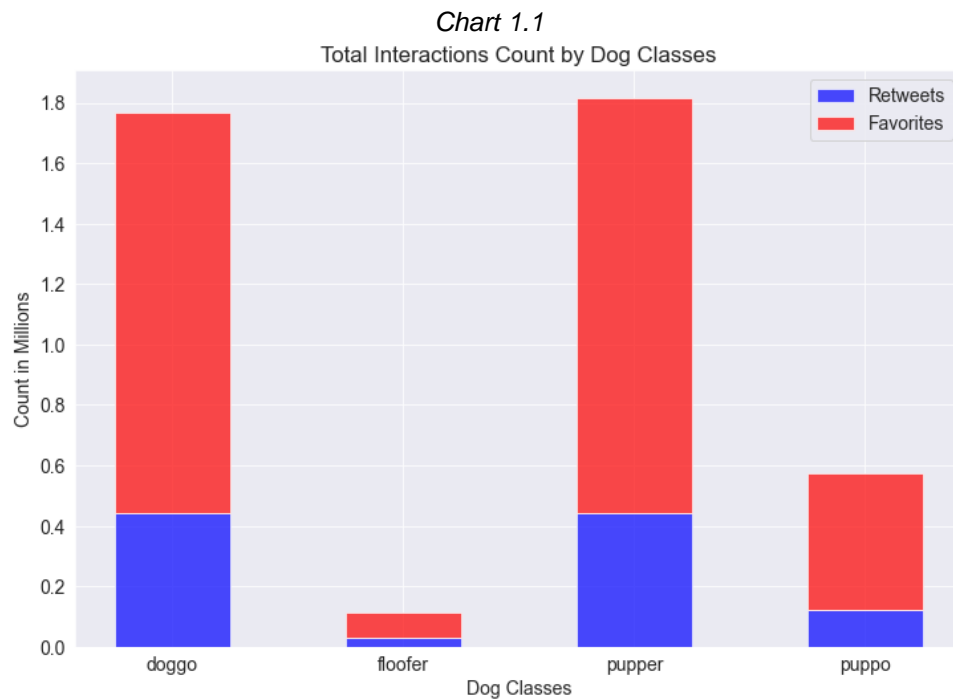
The chart above shows the total interactions of tweets, categorized by dog class. It is obvious that pupper and doggo tweets drive the most interaction in terms of retweets and favorites, with pupper slightly getting the edge.

Chart 1.2

Proportion by Tweet Interaction and Dog Classes
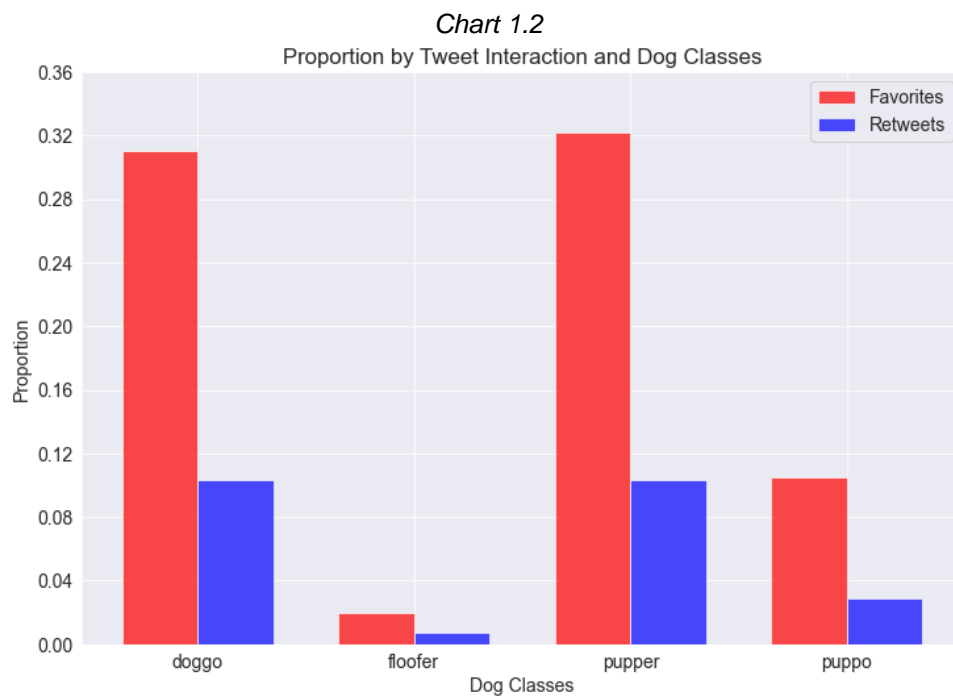
The chart above shows the total proportion by tweet interaction and dog class. This goes deeper to show that pupper tweets favorites count is what gives it the edge over doggo tweets, despite a very similar retweet count.

Chart 1.3
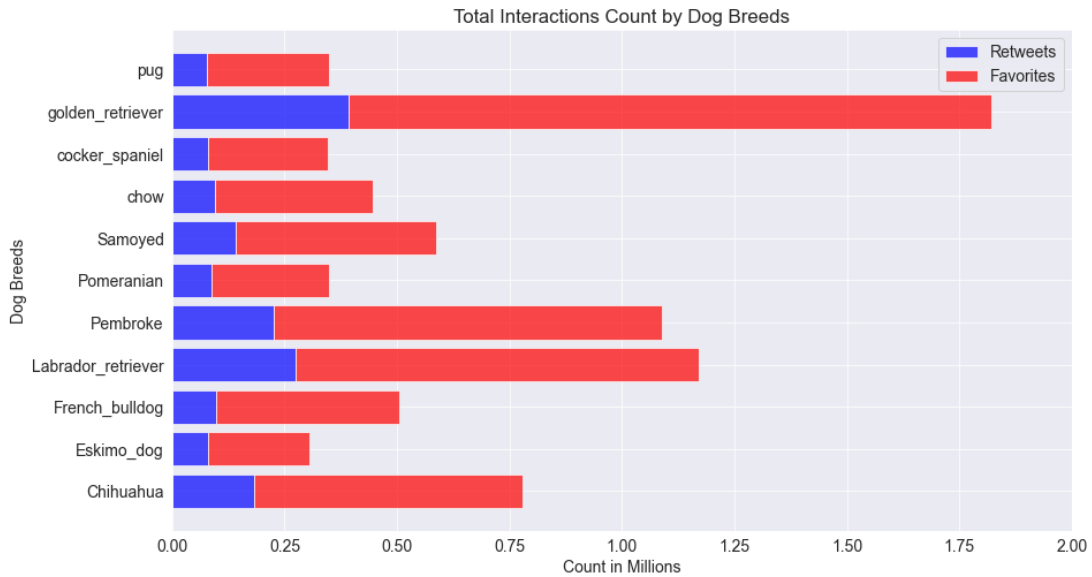
The chart above shows the total interactions of tweets, categorized by dog breed. It appears that tweets containing Golden Retrievers drive far more interactions than tweets containing any other dog breed, both in terms of retweets and favorites.
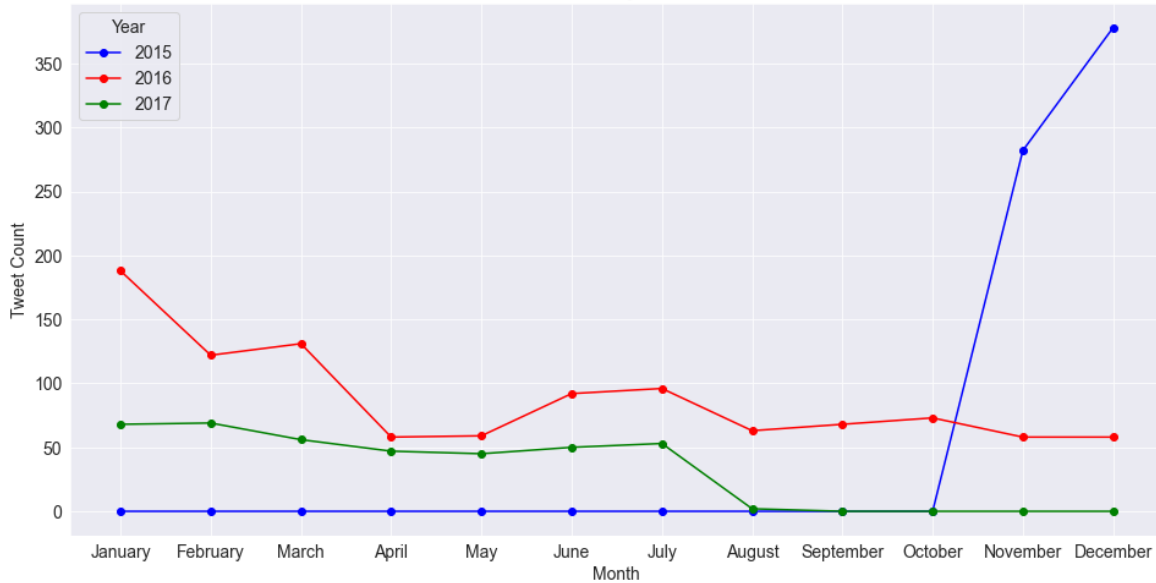
Image 1.1

Here is a picture of a **Golden Retriever**; the most engaged dog breed in WeRateDogs tweets.

Chart 1.4

Total Tweet Count by Month and Year

The chart above shows the total tweet count of WeRateDogs by month and year.

Insight

1. The tweet with the highest number of interactions has 214,397 interactions. It was published on the 18th of June in the year 2016 and it is about a doggo, predicted to be a Labrador retriever at a confidence level of 82%, with a dog rating of 13/10.

2. Tweets about dogs who are classified as doggos have more retweets than floofers, puppers and puppos. Conversely, tweets about floofers are the least retweeted.

3. Tweets that are about puppers are favored more as compared to doggos, floofers, and puppos.

4. Tweets which are about dogs who are classified as floofers are the least retweeted and also the least favored, hence they have the lowest interaction.

5. Tweets about dogs which are puppers have the most interaction in terms of retweets and favorites as compared with doggos, floofers and puppos.

6. WeRateDogs tweeted a lot more in the year, 2016 as compared to 2015 and they significantly tweeted more in the months of November, December and January.

7. Dog that are golden retrievers have the highest retweet count and favorite count; basically the highest interaction.

8. WeRateDogs tweeted most towards the ending of 2015 and the beginning of 2016, after which their tweet frequency steadily declined until their lowest periods in 2017.

Summary and Conclusion

In summary, WeRateDogs tweet volume has steadily reduced over the years, while the same cannot be said for their engagements. For their engagements however, it appears that WeRateDogs' followers prefer content about puppers in general, and golden retrievers specifically, since tweets about these seem to drive the most engagements.