# Report: Act_Report


**Fig.1** @dog_rates dog rating tweets as at April 16 (Image via Boston Magazine)

# Table of Contents



# 1. Introduction

This report is aimed at communicating the insights into the analysis of **WeRateDogs** data as documented in the wrangle_act.ipynb file. **WeRateDogs** is a Twitter account with the handle @dog_rates that rates people's dogs with a humorous comment about the dog. An example of tweets from this handle is shown in Fig. 1. 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.

# 2. Data wrangling

To start with, I acquired twitter_archive_enhanced.csv, image_predictions.tsv data files and additional data of @dog_rates programmatically with lines of python codes. This implies that the process can be replicated by running these codes if download urls for data files remain valid and personal API keys for elevated access using tweepy are imputed. Reading all gathered data into pandas dataframe made it easier to manipulate data as much as possible.

In line with best practices, I documented issue detected before cleaning them. There were 16 issues detected and 13 of them were cleaned. The last of them was about dataset fragmentation into 3 dataframes which was resolved by a merger of all 3 into twitter_archive_master.csv. Another like it is the messy data issue presented by the spread of dog-stage data over several columns. These columns were merged into one column named dogtionary- a collection of terms to classify dog stages as shown in Fig 2.

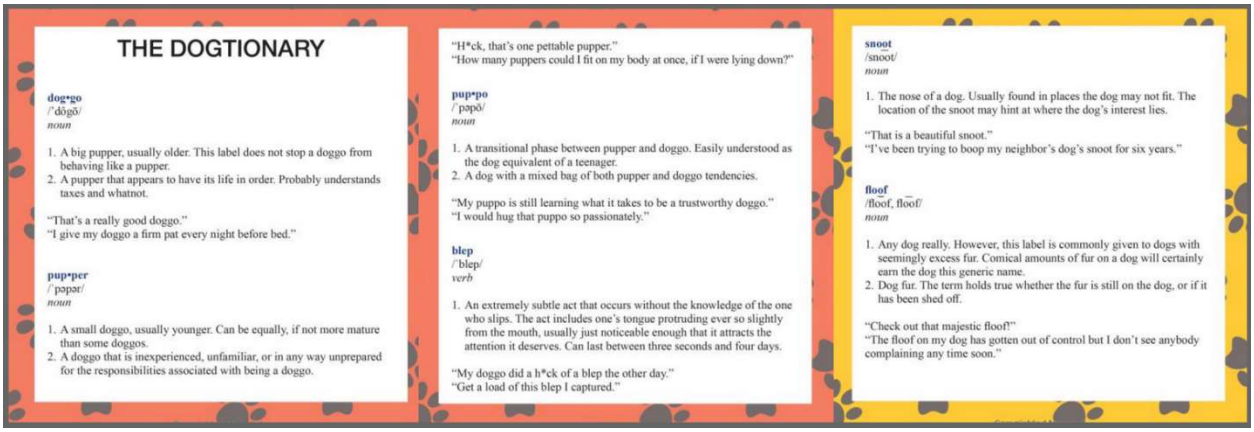


**Fig 2.** The Dogtionary explains the various stages of dog: doggo, pupper, puppo, and floof(er) (via the #WeRateDogs book on Amazon)

Let's read twitter_archive_master.csv into a df_master dataframe for exploratory analysis to establish trends and pattern in wrangled data. In addition, Visualisations would help communicate these insights better.

# 3. Analysis and visualisations

Exploration of data and visualisations were guided with questions and documented as insights. These insights are documented in the following subcategories.

## 3.1 Insight 1

How do categories of dogtionary affect ratings?

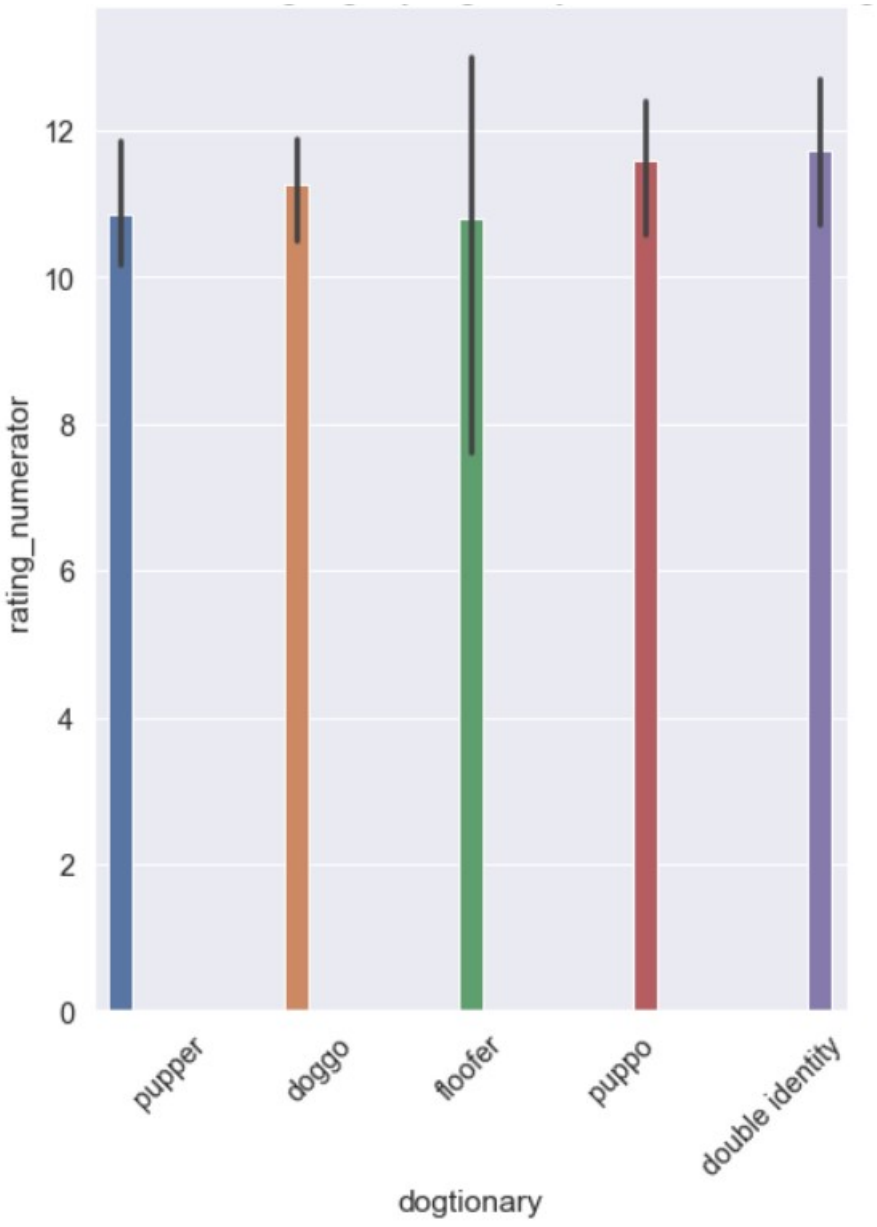


**Fig. 4.** Bar chart showing dogs by dogtionary terms and their ratings

Fig. 3 shows a bar chart of dogtionary by terms for dogs and their ratings. There were only 5 with double identity, they all rated 10 and higher implying that dogs with double identity tend to rate better than others.

## 3.2 Insight 2

For dogs under review, is there a relationship between favorite and retweets counts?

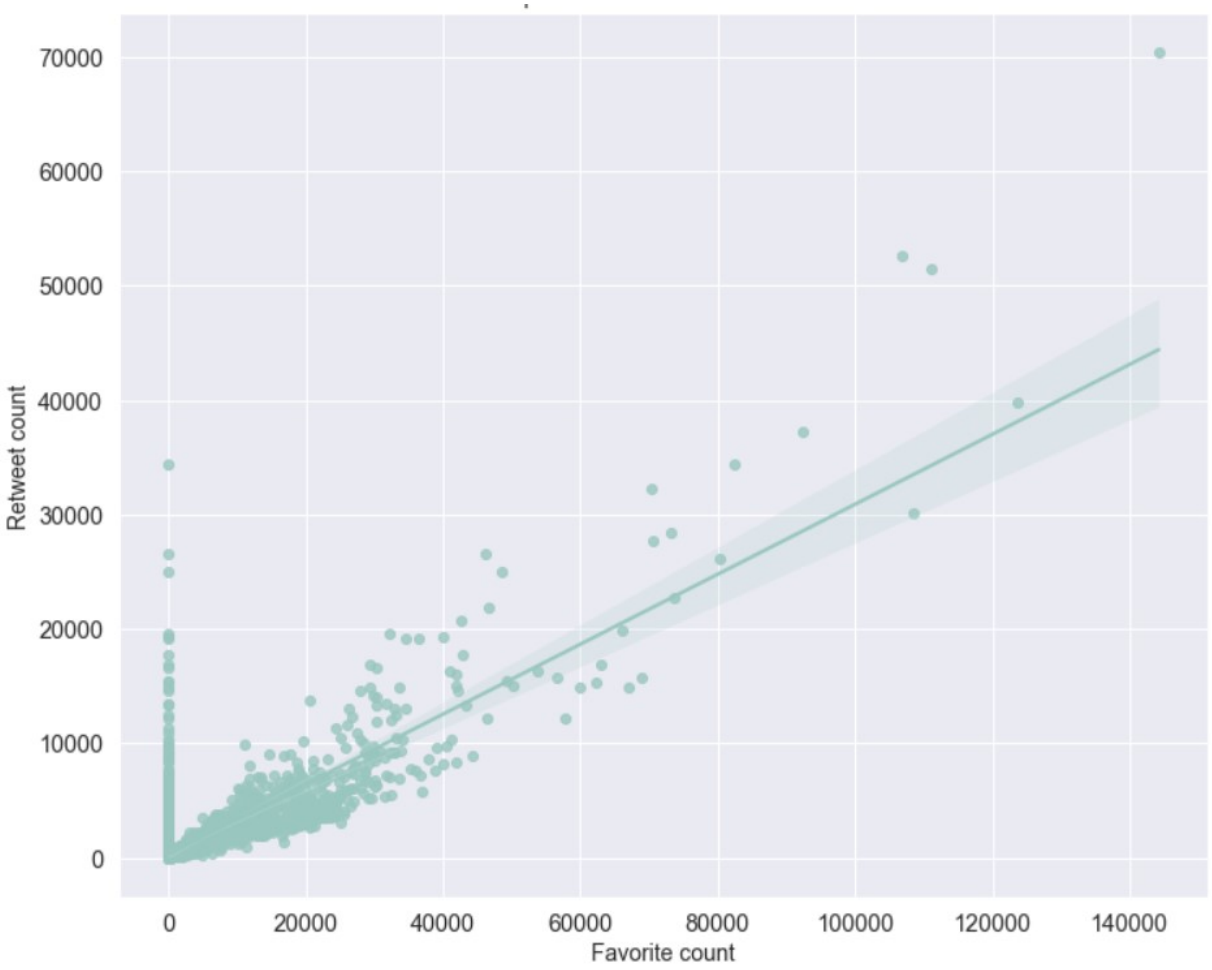


**Fig. 5.** Relationship between favorite and retweet counts

Fig.5 shows the correlation favorite and retweet counts. A line of best fit is drawn with majority of points close by to indicate a strong relationship between both counts. 2327 dogs got retweeted and liked as favorites simultaneously! The average retweet count per dogs (2458 retweets) is one-third the average number of favorite counts (7026 likes). The dog with the maximum likes (144247 favorite counts) got almost double the number of maximum retweets gotten by any dog under review (70334 retweets).

## 3.3 Insight 3

What is the distribution of dogs under review by dogtionary terms?

The largest number of dogs identified by dogtionary are Puppers! Therefore, majority of dogs in the dataset are either small yet, equally or more matured than some doggos. On the contrary, they may be inexperienced or unprepared for responsibilities associated with being a doggo according to definition of dogtionary terms shown by Fig. 3.

# 4. Conclusions

Data analysis and visualisation were carried out on the master archive led to conclusions that dogs with double identities tend to rate higher than others who identify with just one of the original doctionary terms. Also, favorite and retweet counts strongly interrelated. The likelihood of a dog wit retweet getting liked as a favorite is high. Finally, Puppers are the most represented dogtionary category in the dataset.