WeRateDogs Data analysis

The main goal of this project is to achieve a complete data analysis process (gather, access, and clean) of twitter WeRateDogs account. In addition, to analyzing this data and provide some valuable answers from it.

WeRateDogs account is an account that rate dogs with some humorous comment, which have more than 8 million followers. Usually, their rating depends on dog type, stage, features and mostly pictures.

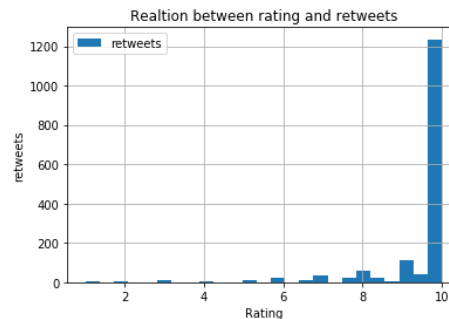*Figure 1: WeRateDogs twitter account*

- **Data Storing:**
  In data wrangling step the three data sets were merged in one data frame (twitter_archive_master) and saved as CSV file.

Done by: Ohoud Almdani

- **Data Analysis and visualization:**
  Three questioned were meant to be answered from this analysis:

  1- The impact of rating on retweet?

  Dogs who have higher rating have high number of retweets too. Which means there is a positive correlation between rating and gaining retweets and revers (gaining more retweets means having high rating).

  As an example, hugo who has a rating of 13 (Yes, in WeRateDog some dogs have a rating of more than 10 because common they are dogs) have less retweet and favorites than Boris and Doug who have rating of 14. (Though all of them are so adorable dogs)
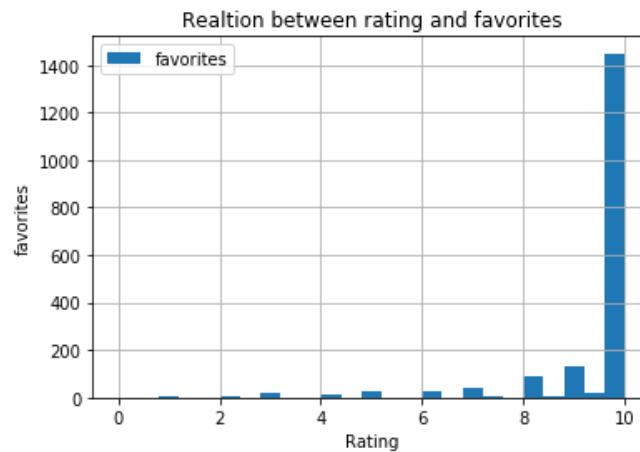

Figure 2: Hugo

Figure 3: Boris and Doug

Done by: Ohoud Almdani

2- The impact of rating on favorite?

Same with retweets, dogs who have higher rating have high number of favorite counts too. Which means there is a positive correlation between rating and gaining favorite and revers (gaining more favorite count means having high rating). You can go back to check favorites counts in Hugo, Boris and Doug pictures as an example for this, too.

Or, in case you want to check on more adorable dogs, this is another example:
Louie who has rating of 13 have more favorite counts than Ivy who has rating of 12.

Figure 4: Louie

Figure 5: Ivy

Done by: Ohoud Almdani

3- The impact of tweet's source on dog's rating?

Tweets from iPhone have higher rating, while tweets from web or TweetDeck have low rating. Which means there is a correlation between source and rating. Though more investigation could be done in this insight to check the impact of other factor such as (camera quality, ease of access which mean more participation, etc) on this correlation.

Relation between tweet source and dog's rating