To learn data analysis is not easy but Udacity in their *Data Analyst Nanodegree Program* helps a lot by giving amazing data. This time I've got to gather, assess, clean and analyze the WeRateDog twitter account. The author rates dogs sent by other users and tries his best to be objective and funny.

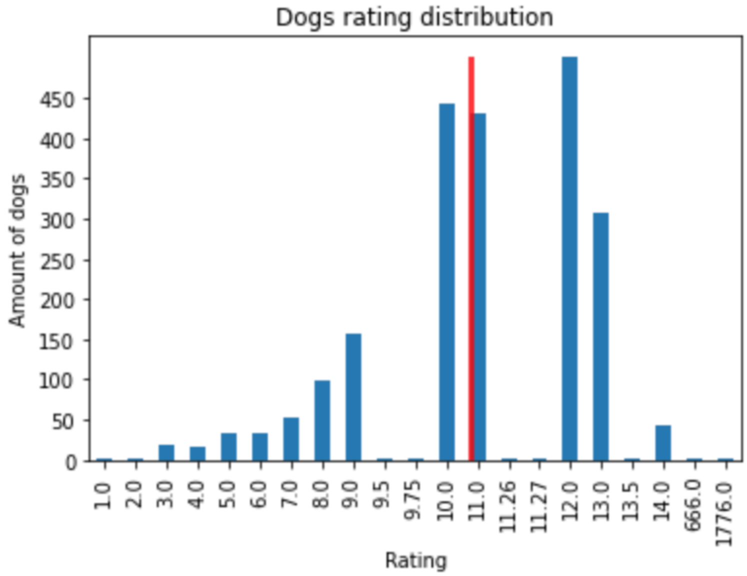
In the beginning I've got only part of the data, the rest I had to gather myself.

Udacity provided a csv file 'twitter_archive_enhanced.csv' (basic information about 2356 tweets) for manual downloading. Next part was 'image_predictions.tsv' that was available for programmatic download. It contains 2075 predictions made by a neural network that can classify breeds of dogs. The third piece of information I gathered using Twitter API and Python's Tweepy library. There is information about the amount of likes and retweets each tweet received for 2327 tweets in the file 'tweet_json.txt'.

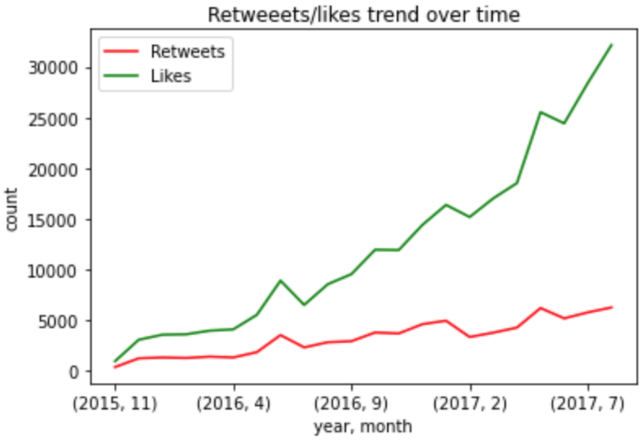
During assessing the data I found 10 quality issues and 3 tidiness issues. I used variety of pandas methods to clean them up.

Here is some insides I've got:

- Most of the dogs received ratings 10-12, few outliers are mostly jokes and not really bad opinions about dogs:



- Account WeRateDog have got a lot of attention over time:



- Rating of a dog does not affect the amount of likes and retweets it gets.
- Puppo and pupper (which both are puppy) are the most favorable stages of dogs, they have the highest ratings and the most likes/retweets.
- The most popular breeds judging by rating/likes/retweets: saluki, bouvier des flandres, samoyed, bedlington terrier, french bulldog, afghan hound and whippet.