Wrangling and analyzing data using tweepy API for weratedogs account tweets

Introduction:

This report is written to clarify the useful insights we got after doing the hard work of data wrangling on datasets extracted from twitter API tweepy from twitter account we rate dogs.

After doing the hard part which is Gathering and Assessing the data then cleaning it, we get to the fun and useful part of the data analysis which is working with what we gathered and getting some useful insights and info.

First, I did some basic analyzation that I always was curious about when scrolling through social media like what's the most retweeted tweet and got this video with 79515 retweets
https://twitter.com/dog_rates/status/744234799360020481

Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad)

Then, I queried the dataset for the most favorited(loved) tweet and got this pic with 132810 favorites
https://twitter.com/dog_rates/status/822872901745569793/photo/1

Also, a fun one I wanted to explore who got the highest nonrealistic rating by querying for the highest rated numerator with 1776/10 rating.
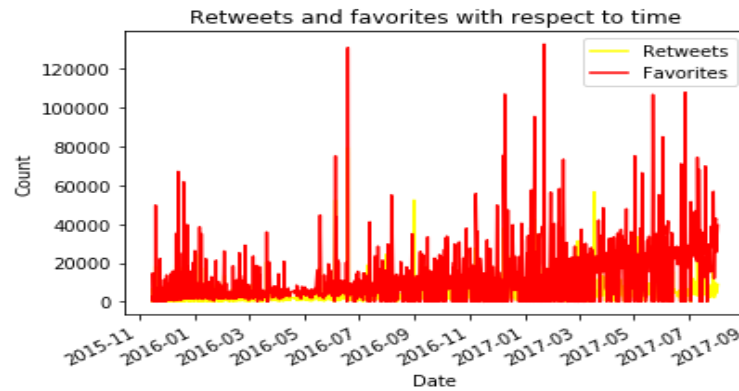
https://twitter.com/dog_rates/status/749981277374128128

Then I asked myself what else can I pull from this dataset that can be insightful and useful, so I plotted a graph that correlates between retweets count and favorite count with respect to time by using matplotlib library which is shown below.

Retweets and favorites with respect to time

As we can see here obviously the favorites count is much higher than the retweets count possibly because it requires less effort to favorite than to retweet, Also the count does not seem to be stable but at certain points in time it explodes with people favoriting the tweets