Act Report

The data master_df obtained by wrangling the three datasets obtained from the twitter archive show some interesting insights on Exploratory analysis. In exploring the data, I asked certain questions which I needed to further help the exploration.

Questions:

1. **Which dog_stages had the highest ratings?**
2. **What has been the followers_count for the werate_dog account from 2015 till 2017?**
3. **Which posts had the highest retweet_counts?**

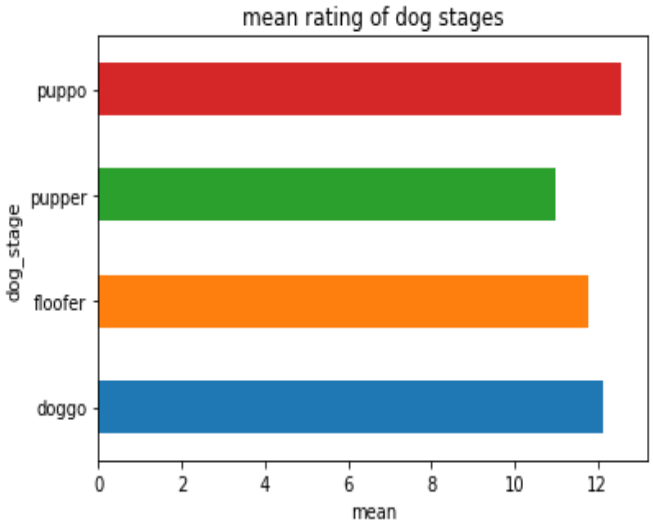
Some of the insights are described below:

Insight 1:

Dog_stages according the dogtionary on the https://www.amazon.com/WeRateDogs-Most-Hilarious-Adorable-Youve/dp/1510717145, the four stages of dogs are defined as follows:

- **Doggo**: A big pupper, usually older.
- **Pupper:** A small doggo, usually younger. Can be equally, if not mature than some doggos
- **Puppo:** A transitional phase between pupper and doggo
- **Floof:** Any dog really. However, this label is commonly given to dogs with seemingly excess fur.

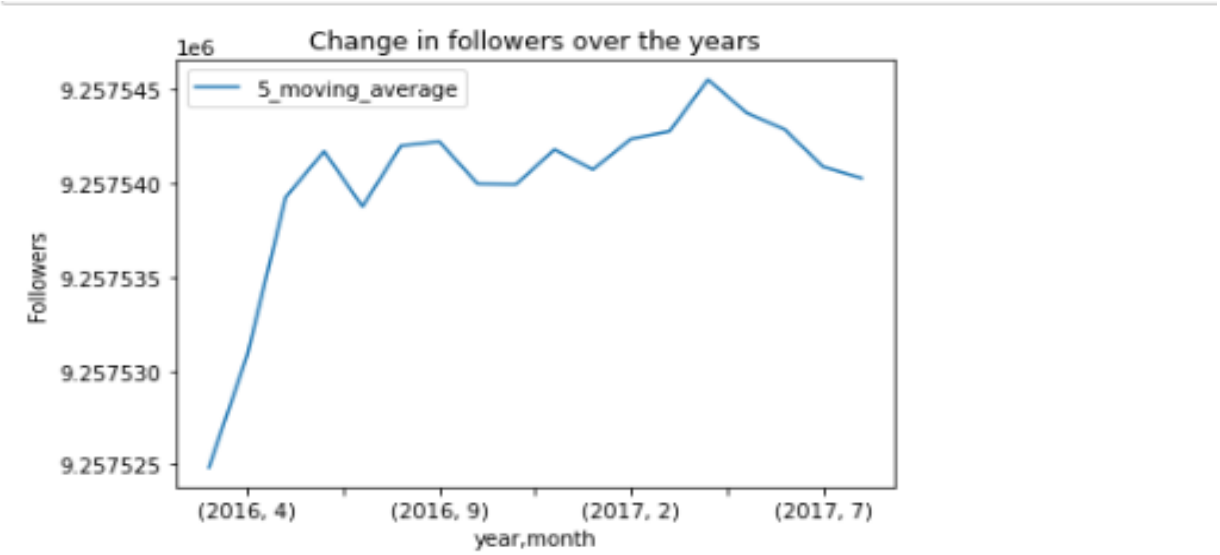
It can be observed from the data that the average rating for the 'stated dog-stages' is roughly the same. However, puppo are rated slightly higher than the others, followed by doggo



Insight 2:

WeRateDogs currently has 9.2 million followers in their page. From November, 2015 till July, 2017 (span of time the datasets covered), their followers have remained constant with just slight increase in 10 (tens).

The change in followers is quite small, as it is measured in 10s. However, there seems to be an increase in the followers over the months, but a slight drop in 2017



Insight 3:

it can be observed that retweet count and favorite count had the strongest correlation with value of 0.926577 suggesting that a positive linear relationship.

	rating_numerator	retweet_count	favorite_count	followers_count
rating_numerator	1.000000	0.344515	0.440946	0.147577
retweet_count	0.344515	1.000000	0.926577	0.073013
favorite_count	0.440946	0.926577	1.000000	0.102508
followers_count	0.147577	0.073013	0.102508	1.000000

The relationship between retweet_count and rating_numerator didn't have such a strong correlation as expected (0.3445). However, posts that had high retweet_count also had high favorite_count.