Retweet and Favorite Count Trends

Looking at Figure 1 and Figure 2 below, we can see that there is a rise in retweet and favortie count over time. The rise in retweet count (Figure 1) is slower, however.

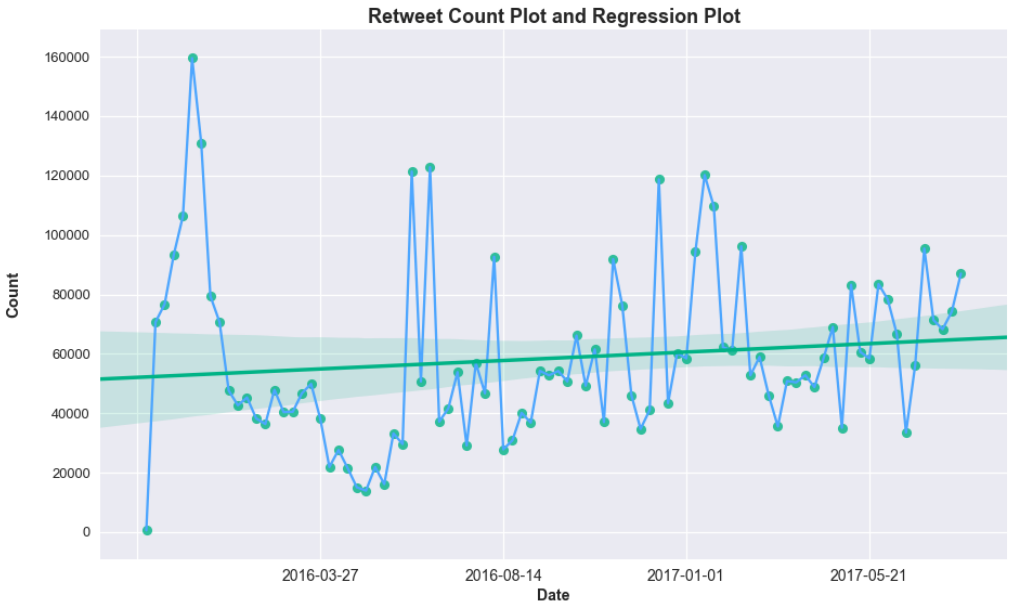


Figure 1

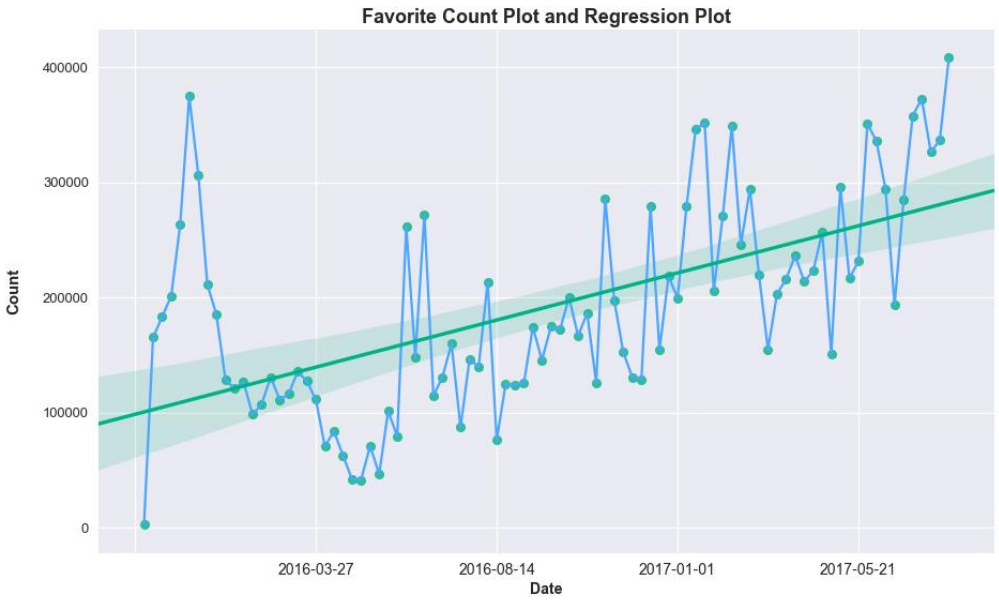


Figure 2

Figure 3

In Figure 3, I superimposed both the retweet and favorite count data with seaborn.regplot. Again there is a general rise in retweet and favorite count overtime. However, retweet count needs to be zoomed in for us to see the rising trend.

Relationship between Retweet/Favorite Count and Rating Numerator

From Figure 4 below, one can't tell if there is a relationship between Retweet/Favorite Count and rating_numerator.

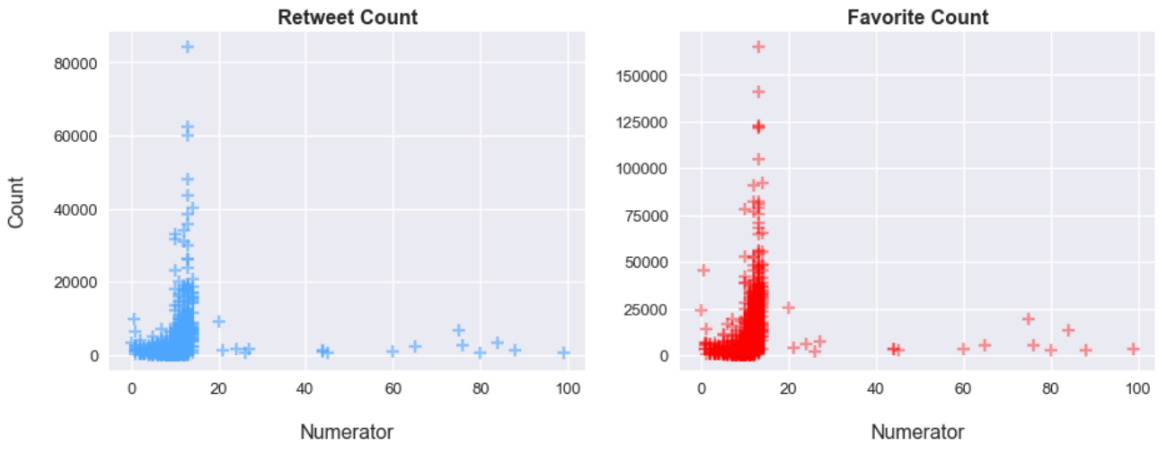


Figure 4

We need to transform the Retweet/Favorite Count into log scale. The reasoning is given in wrangle_act.ipynb.

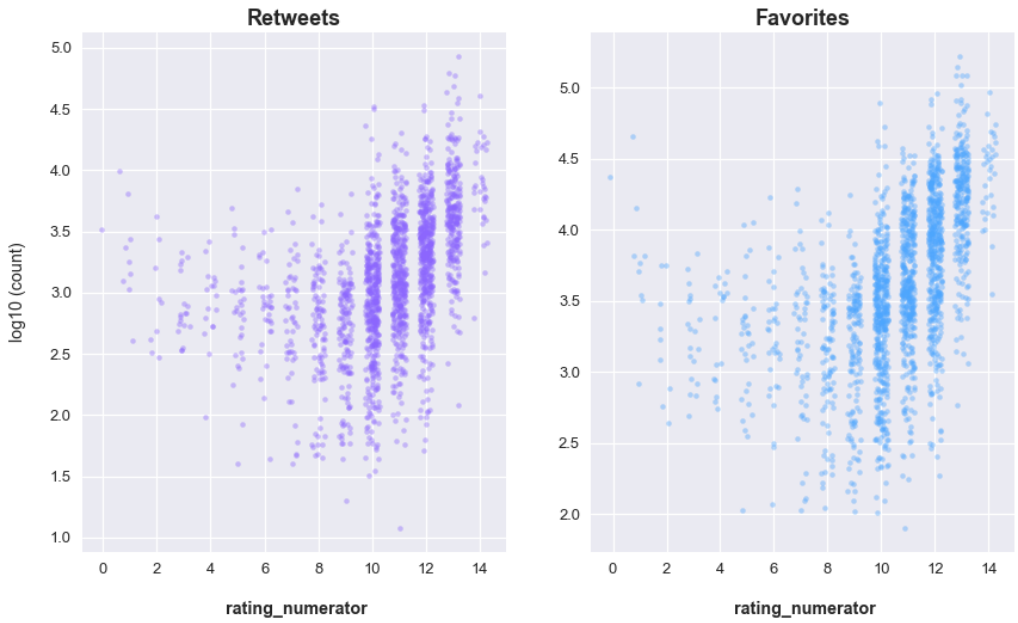


Figure 5

Looking at the plot in Fig.5, we can see that there are more retweet and favorite counts as rating-numerator increases.

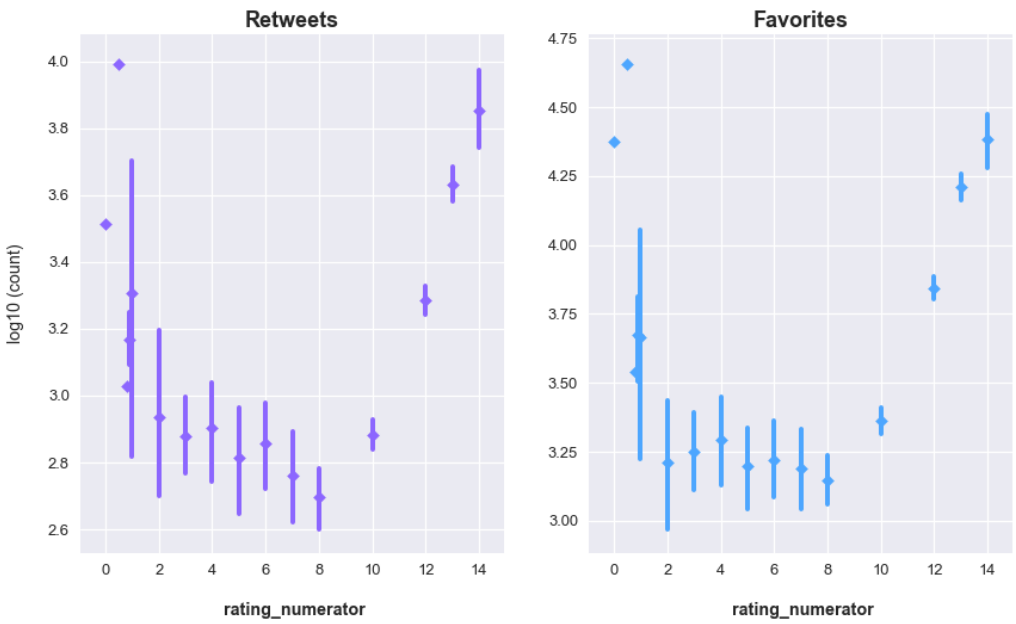


Figure 6

- In Fig.6, we take the mean of retweet and favorite count of each numerator corridor (discrete variable).

- The fitted line is a positive one. This means that as rating_numerator increases, retweet and favorite counts also increase.

- (The mean value decreases between rating_numerator 0 and 8. The mean value increases when rating_numerator > 8. But in general, the fitted line is positive.)

```
low_ratings_df = df_archive_c.query('rating_numerator < 2').copy()
```

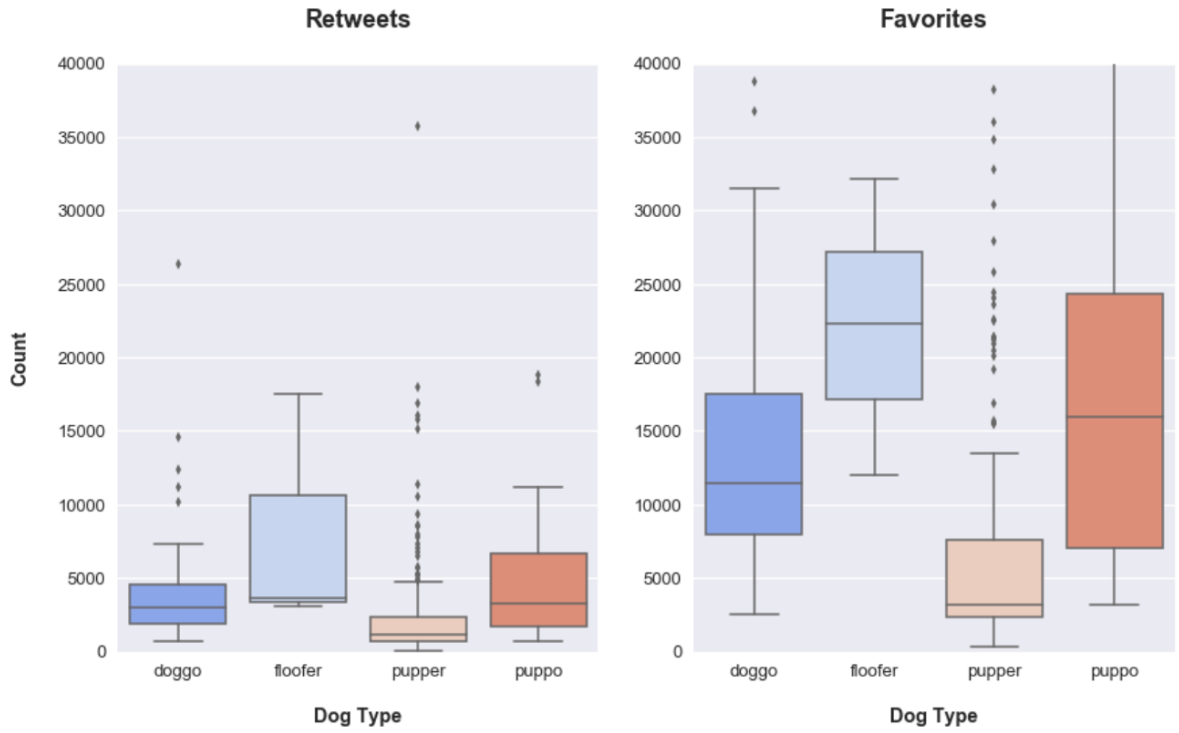
```
selected_cols = ['rating_numerator','retweet_count','favorite_count']
low_ratings_df[selected_cols]
```

	rating_numerator	retweet_count	favorite_count
39	0.5	9826.0	45343.0
231	0.0	3265.0	23713.0
753	0.9	1777.0	6523.0
1060	0.9	1437.0	5074.0
1119	0.8	1069.0	3464.0
1265	0.9	1236.0	3216.0
1497	1.0	2748.0	6507.0
1709	1.0	408.0	834.0
1873	1.0	2335.0	5883.0
1949	1.0	6493.0	14173.0

Figure B3_table

Note: Although from Fig.5 and Fig.6, retweet and favorite counts increase as rating_numerator increases, this doesn't mean that dogs having low rating numerator receive 'no love'. From Fig.B3_table above, retweet and favorite counts are still relatively high! (although rating numerator range from 0 to 1).

Plotting Retweets and Favorites given dogtype (conditional probability)



A few observations on seaborn.boxplot above

- Floofer has the most retweet and favorite counts (by median value).
- Also note that floofer has only 3 data points. This explains why it has no outliers.
- Floofer has the longest boxplot in RetweetCount (as observed previously from statistics above).
- Puppo has the longest boxplot in FavoriteCount while pupper has the shortest (discussed previously).

There are more observations in wrangle_act.ipynb. Please refer to it.