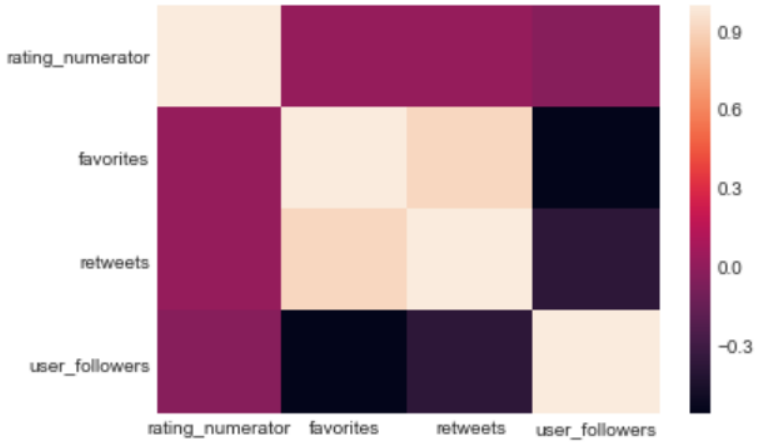
Analysis and Visualization

Mariam Abouzeid

After cleaning our data, it's time for analyzing it and see visualizations

1. Correlation Matrix

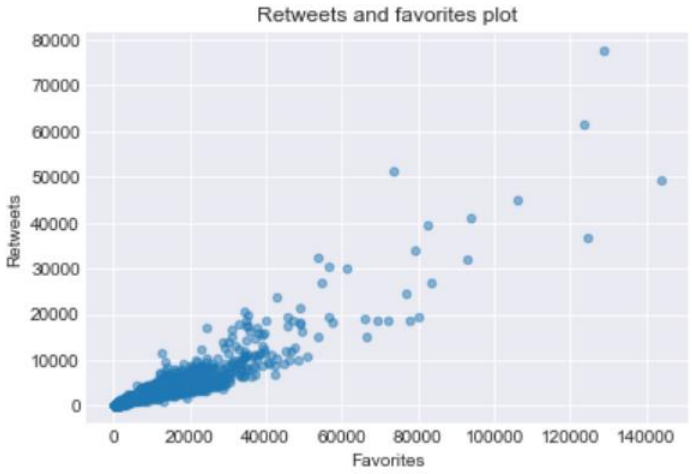
We usually use correlation matrix to know the relation between dataset variables



- We can see the strong correlation between favorites and retweet, this is normal (more favorites mean more retweets)
- User followers and retweet have negative correlation (this seems the opposite of normal prediction)
- Rating doesn't get affected by any attribute

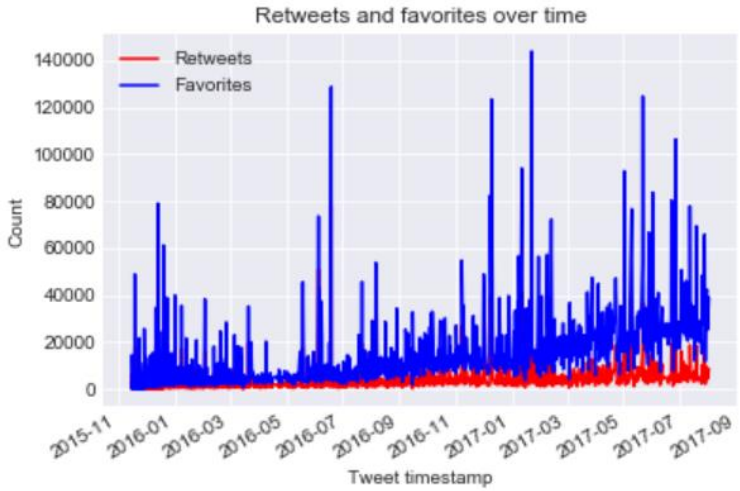
2. Plotting and Analysis

2.1. Retweets VS Favorites



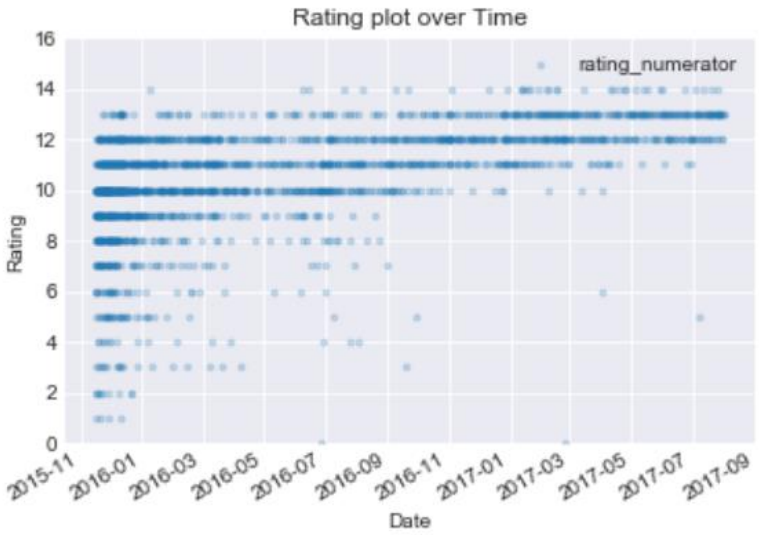
- As we've seen in the matric, there is high correlation between these 2 variables. More favorites more retweets

2.2. Retweets VS Favorites over the time



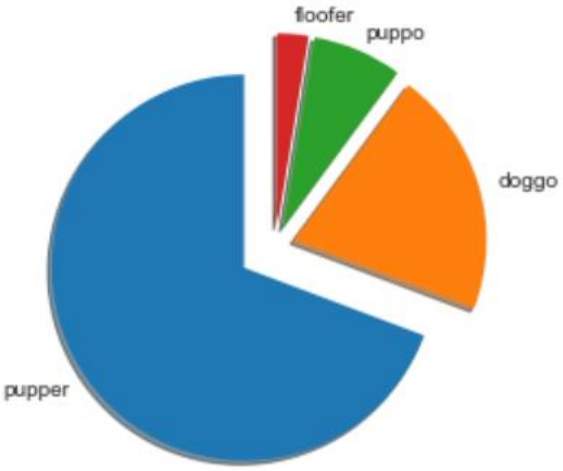
- We can see that favorites number is much higher than retweets across the years
- Retweets are slightly increasing across time

2.3. Rating over the time



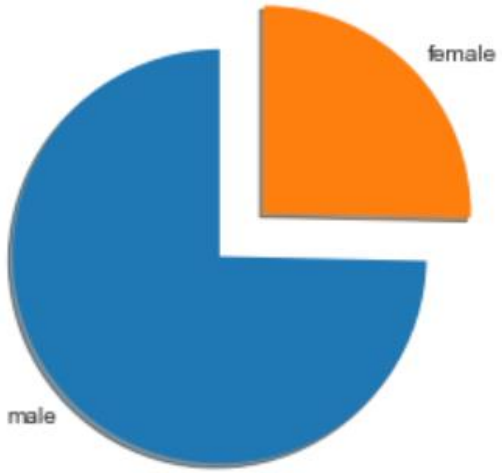
- The users were not used to rating for the past years but over time they started using it more and more frequently

2.4. Dog Stages pie chart



- Pupper is the most owned doge stage, followed by doggo. foofer is the least owned one

2.5. Dog Gender pie chart



- Almost 3/4 of the dogs are males while only 1/4 are females