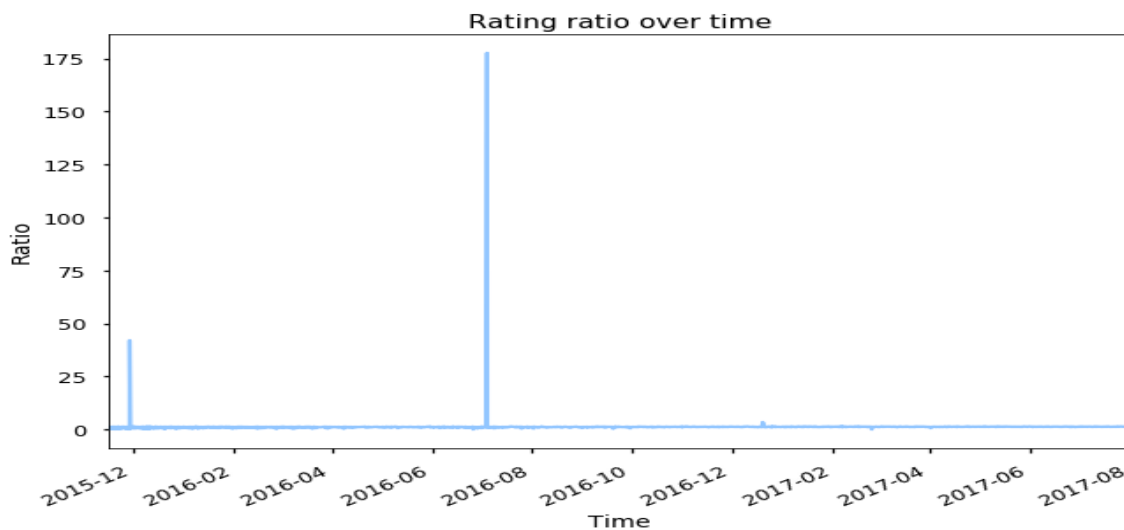


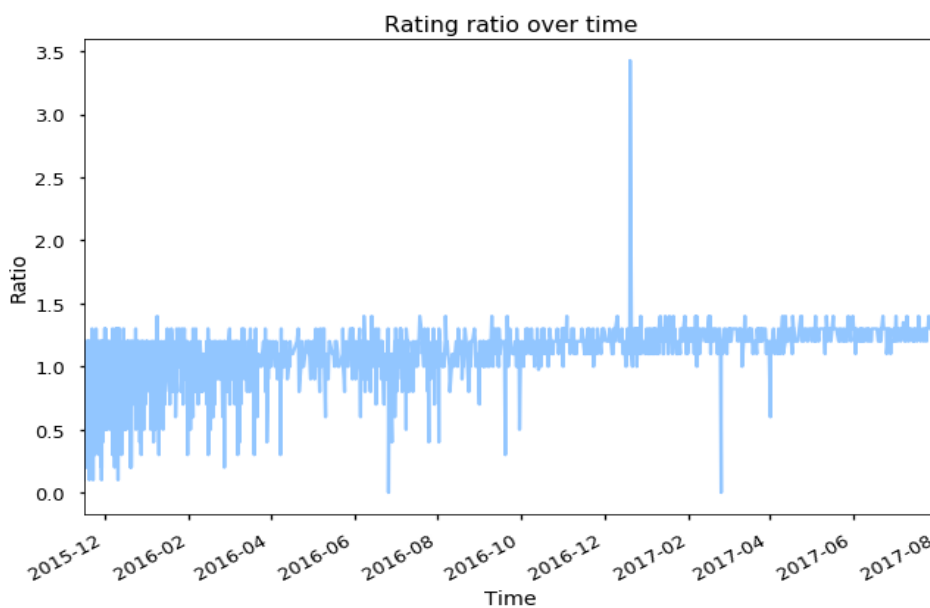
Introduction

The project aims to provide insights of a cleaned twitter data the data used here is from We rate dogs twitter archive. The cleaning efforts are captured in the wrangle_act.ipynb file. Variable like rating and retweets and favourites are visualized to depict trends and patterns.

Analysis

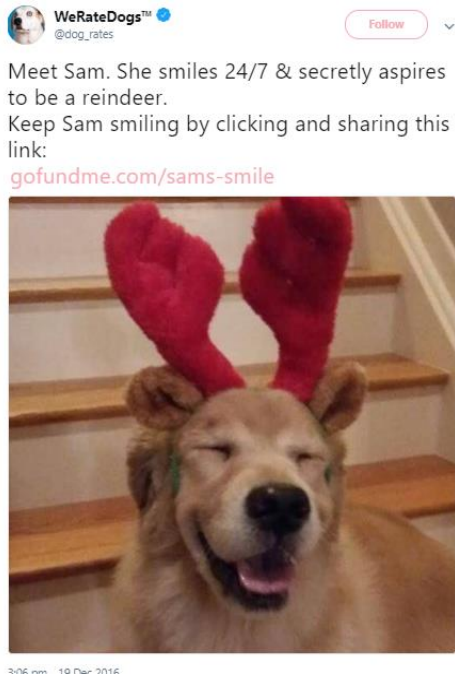


Instead of using rating numerator or denominator a ratio is used in order to normalize the data set then the ratio is plotted against time as shown above. Here nothing can be understood about rating pattern due to the outliers .

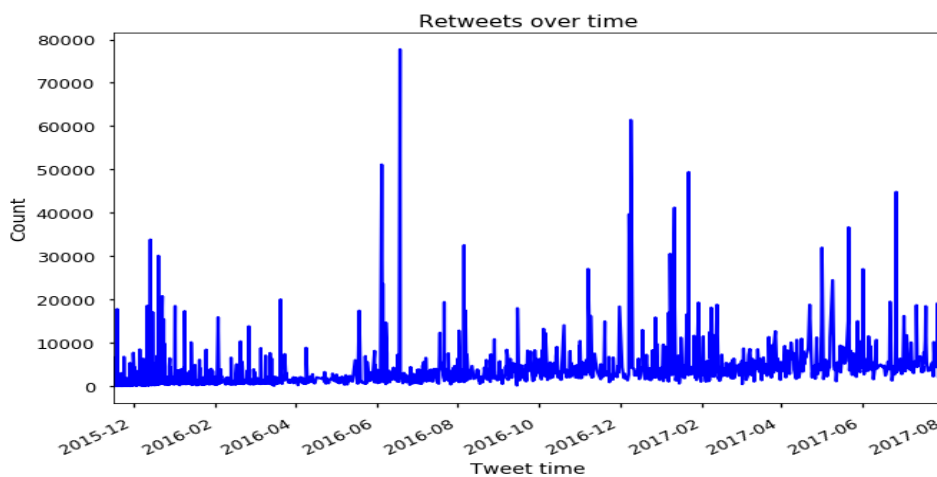


After getting rid of the outliers we get the above graph. In this graph it can be seen that most of the ratios are 1 or around 1 and there are little to no ratios which are zero or less.

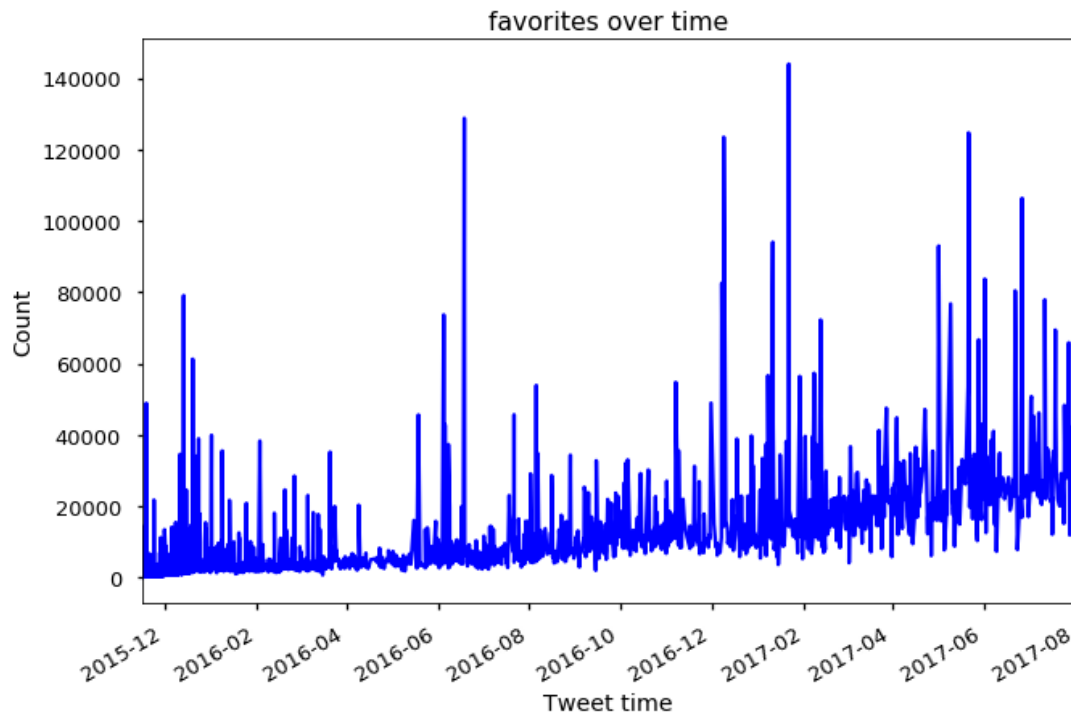
Interestingly a spike can be seen near 3.5 digging further I found out that the numerator and denominator depict something other than ratings .



The post above corresponds to the 3.5 spike ratio seen in ratio vs. time graph here 24/7 is a figure of speech rather than rating, I am not cleaning this because then Sam will have no rating and I would hate to see him sad.



Retweets were very high in 2016 , mostly the count is between 0 to 20,000 . this shows that the account has a consistent fan base



Favourites are increasing over time this depicts that the account is gaining popularity. Many outliers can be seen in the above graph this is because of very popular posts or the so called viral posts other than that Mostly the favourites count is above 0 after 2015.