

UDACITY Data Analyst Nanodegree Program

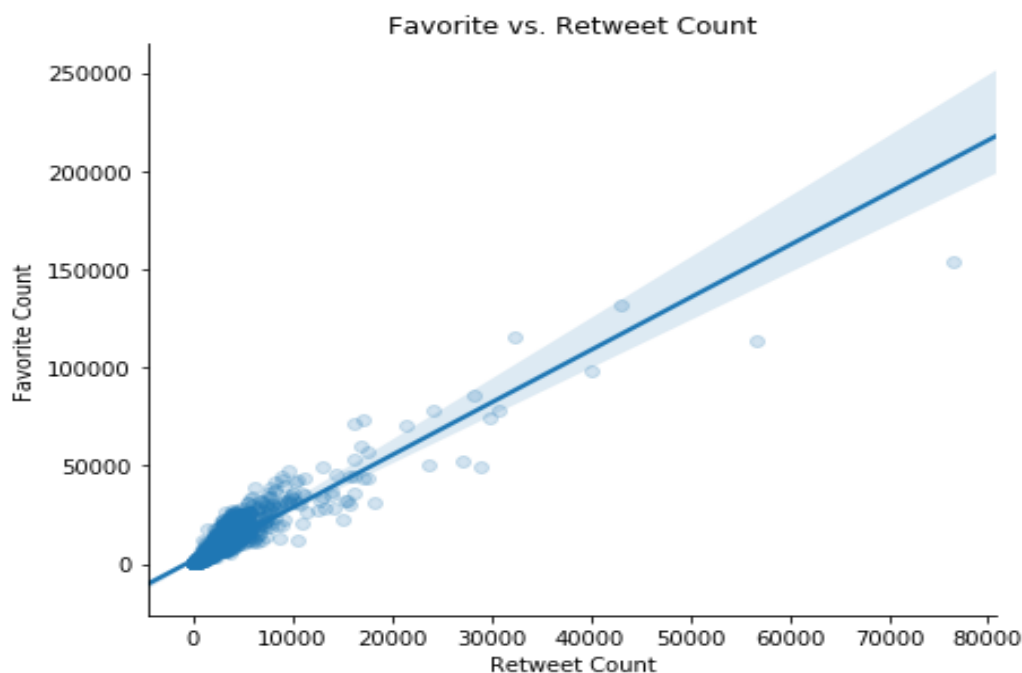
PROJECT DETAILS: Wrangle and Analyze Data

INTRODUCTION:

The Wrangle and Analyze Data project is a part of Udacity's Data Analyst Nanodegree Program. This project involves wrangling of data from various sources associated with tweets from the Twitter user @dog_rates, also known as WeRateDogs. The WeRateDogs rate's pictures of people's dog in a humorous manner, most often by giving them ratings higher than 10/10. After scraping together the data, quality and tidiness issues were accessed and then cleaned. Finally two visualizations were created and its insights are as follows:

1.Favourite dog v/s Retweet count:

WeRateDogs has over 4 million followers in twitter, therefore their tweets are likely to get many favourites and retweets. In addition there may be some tweets that are extremely popular and they become part of the international news and go viral. In the figure below it can be seen that the favourite and retweet count are highly co-related. The majority of the data falls below 40000 favourites and 10000 retweets.

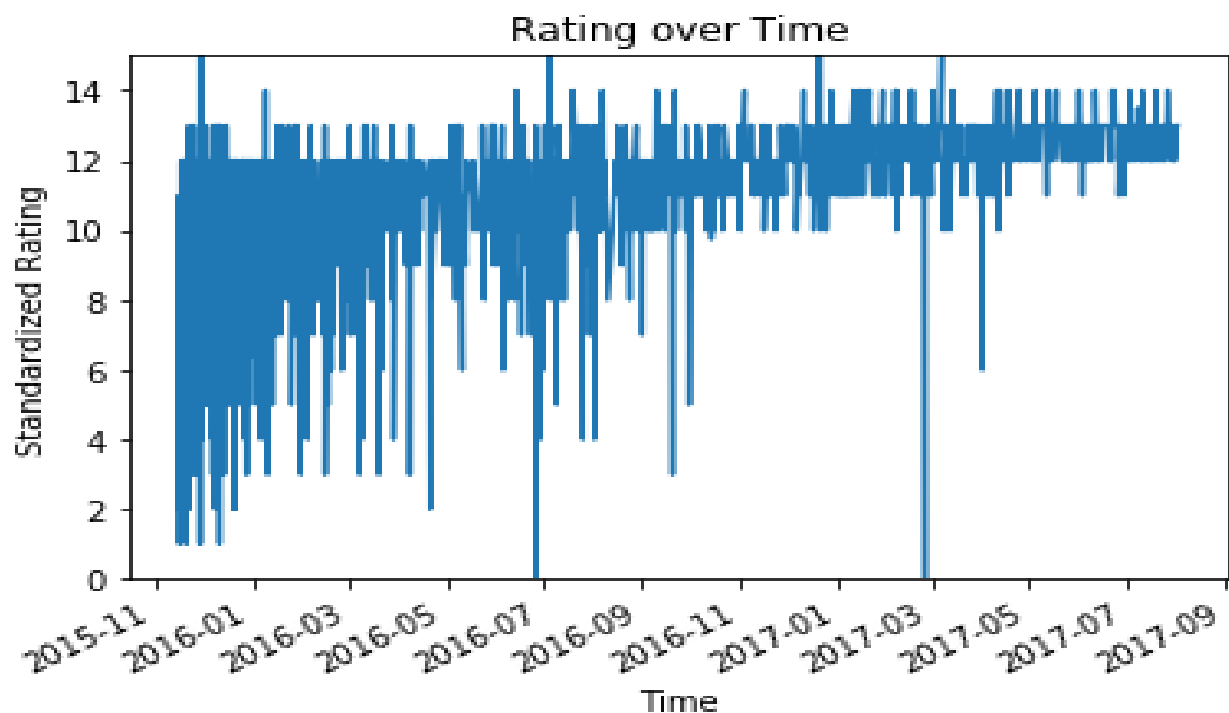


2. Standardized ratings over time:

The idea behind the WeRateDogs account is to ask people to send them photos of their dogs and they will rate them on a scale of 1-10 with humorous comments, however they are often given rating higher than 10. I assumed that almost all dogs were given a rating higher than 10 but I was surprised to see that there were also ratings lower than 10. I was curious to find if overtime, as the account became more and more popular

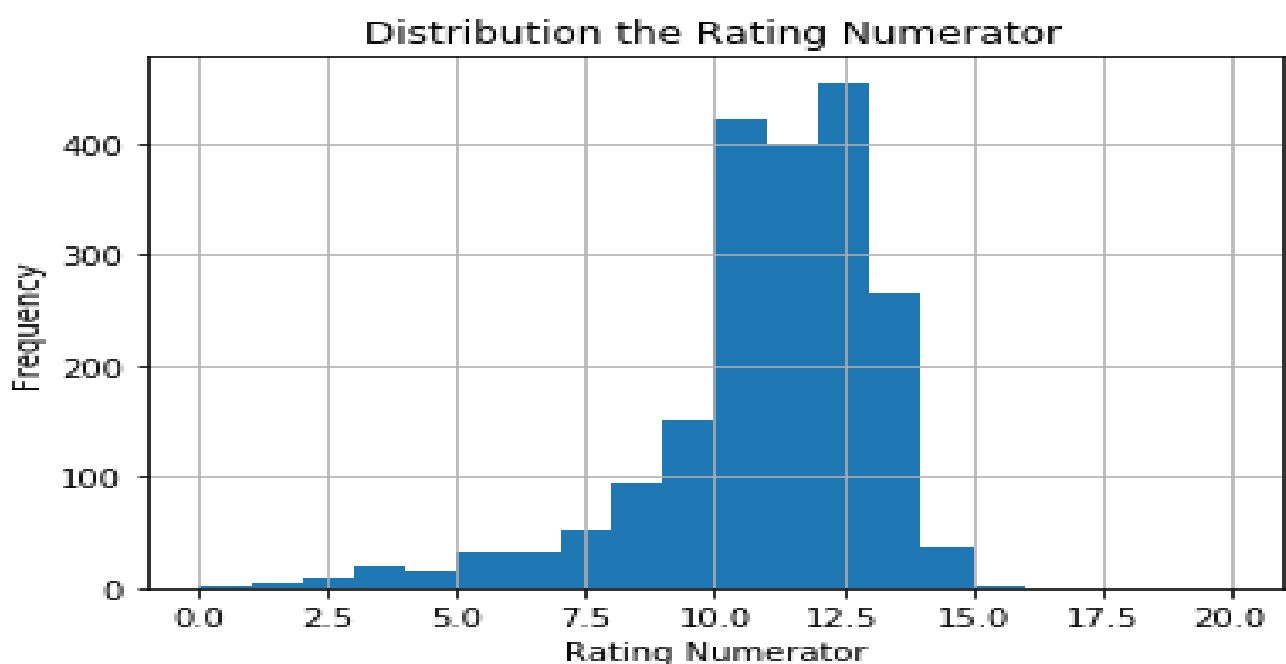
and people associated the above 10/10 raatings with being funny,that the higher ratings would become more prevalent.

In the figure below it can be seen that the rating below 10 decreases overtime. Before 2016-11 there were many ratings lower than 10, while after 2016 there were barely any.



3.Majority of votes ranging between 10-12:

It is found by assessment that the majority of votes lies in the range starting from 10 and ending at 12. It appears that the twitter users had some criteria for voting and upon reaching the criteria they give the suitable vote depending on how many criteria has been fully satisfied.



4. Twitter is very popular among iPhone users:

By assessing the master dataset it is observed that most of the tweets that has been recorded comes from the Twitter-for-iPhone app. This indicates that twitter is very popular among the iPhone users. Also it could mean that twitter's Iphone app is very user-friendly and easy to use with very attractive UI design.

