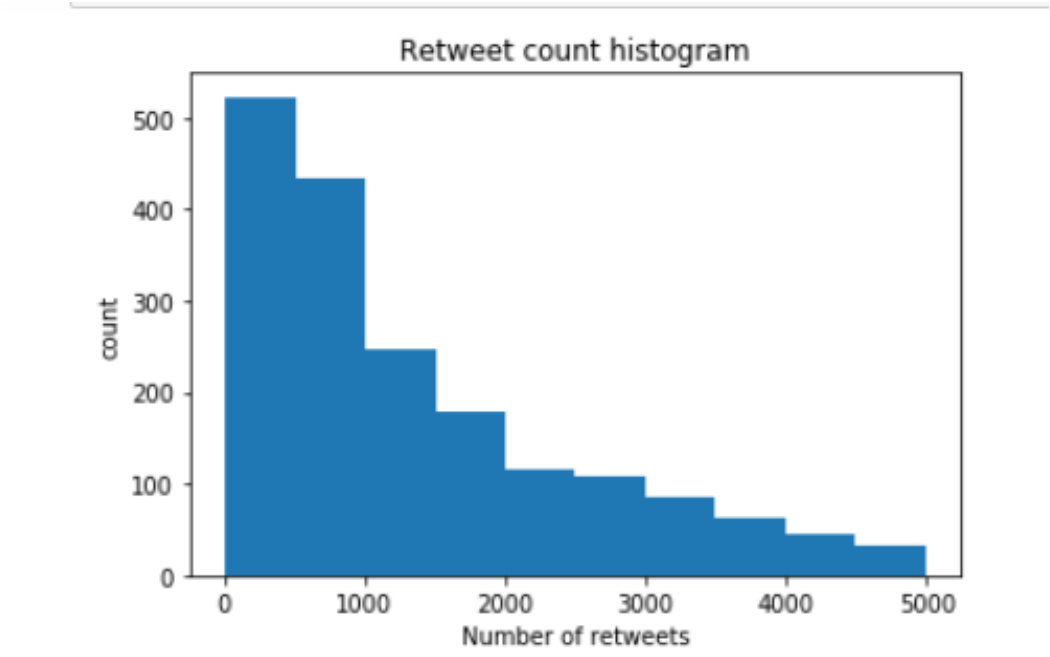
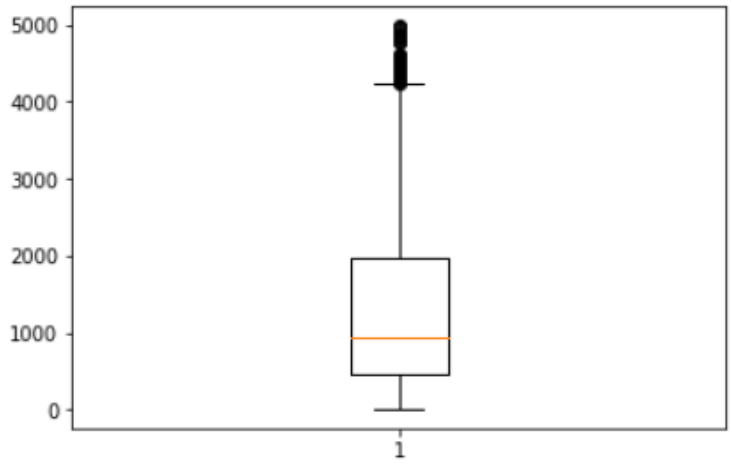
WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The account was started in 2015 by college student Matt Nelson, and has received international media coverage and applause for its cute dogs ratings and love.

We were given this tweeter archive data from the account to clean, analyze and visualize it. In this report i will give some of my insights and findings.

https://en.wikipedia.org/wiki/WeRateDogs some of the information quoted here

The twitter page having so many interactions I first sorted and cleaned the data. Some of the columns that I felt like I did not need, I.e Timestamp of the tweet, were done away with.
On data visualisation and analysis I focused on the Rating numerator and retweet count. I wanted to see if there was a correlation of the rate a certain dog is given and the number of retweets it got afterwards.

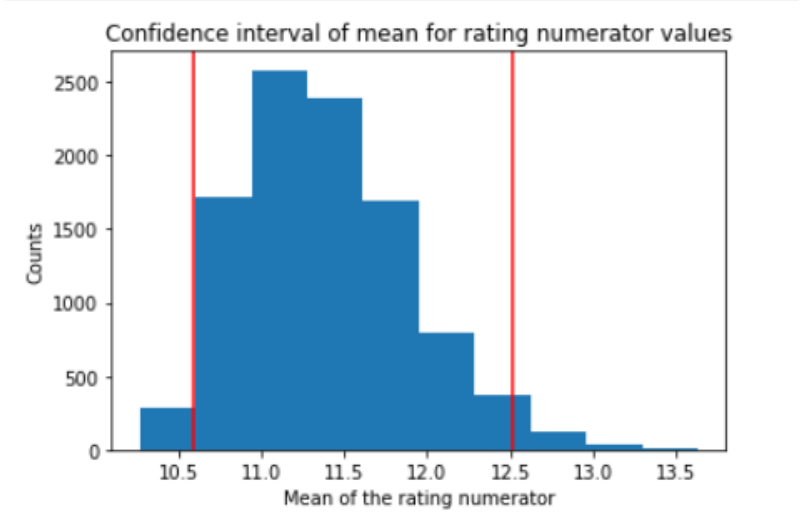




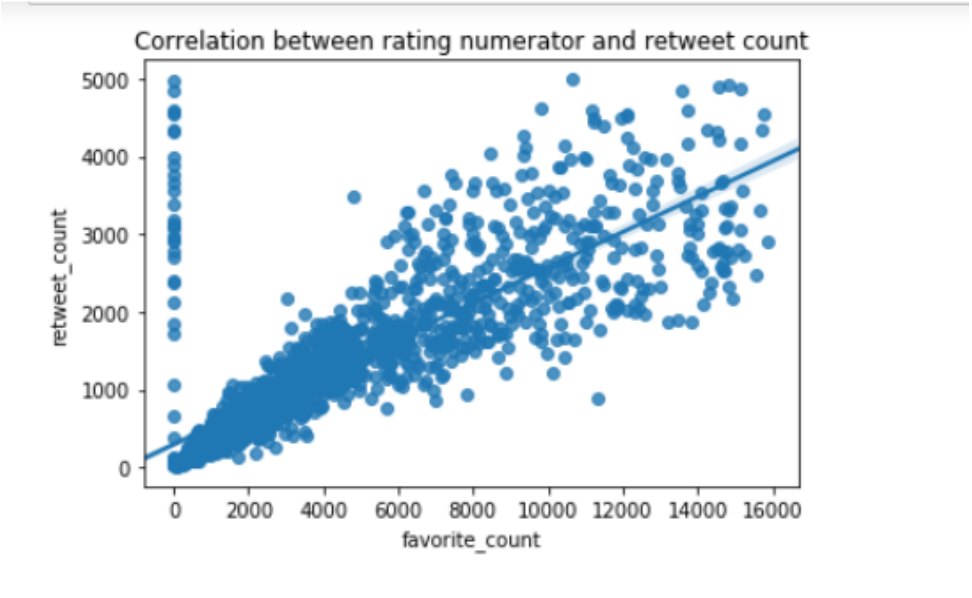
We can see that outliers are above 5000. Plot hist for values less then 5000:

As we can see from the visualizer, We rate dogs has more of its retweets from 0 to 1000 with sime of its retweets getting to 5000. Which is impressive and means there is a big consumption of its content.

The other section I analyzed was the Rating Numerator. And we can see most of the dogs are rated between 10.5 and 12.5



Opposing to the Rating numerator which is skewed to the left. We can see below that the retweet count histogram and the favourite count histogram are skewd to the right and their mean/ standard deviation being mean: 1210.3327536231884, standard deviation: 1037.8435672800558



To sum it up, There is a lot on can do with the project because the page is very popular and alot of information can be extracted from this archive. I hope I have displayed some of the key factors I have been learning In data wrangling and visualization.