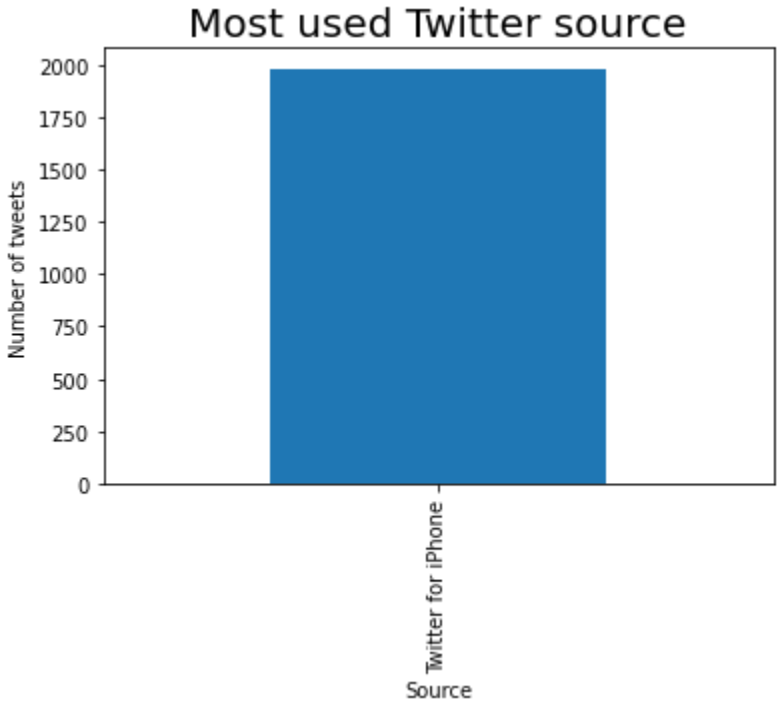
Act report

Having access to data about tweets of “WeRateDogs” account could be very interesting in order to understand more about the dogs and the people owning them. With that said, I performed a modest analysis that could easily be improved over time. So , I encourage everyone with access to that data to practice more methods and functions on it, and do not hesitate to iterate over the wrangling process because I am sure the dataset is not fully clean.

Our main focus was on the tweets, the source from which the tweets are sent, their progress over time alongside with the retweets and the relationship that could possibly be between the number of the retweets with the number of the favorites. The analysis went as follows:

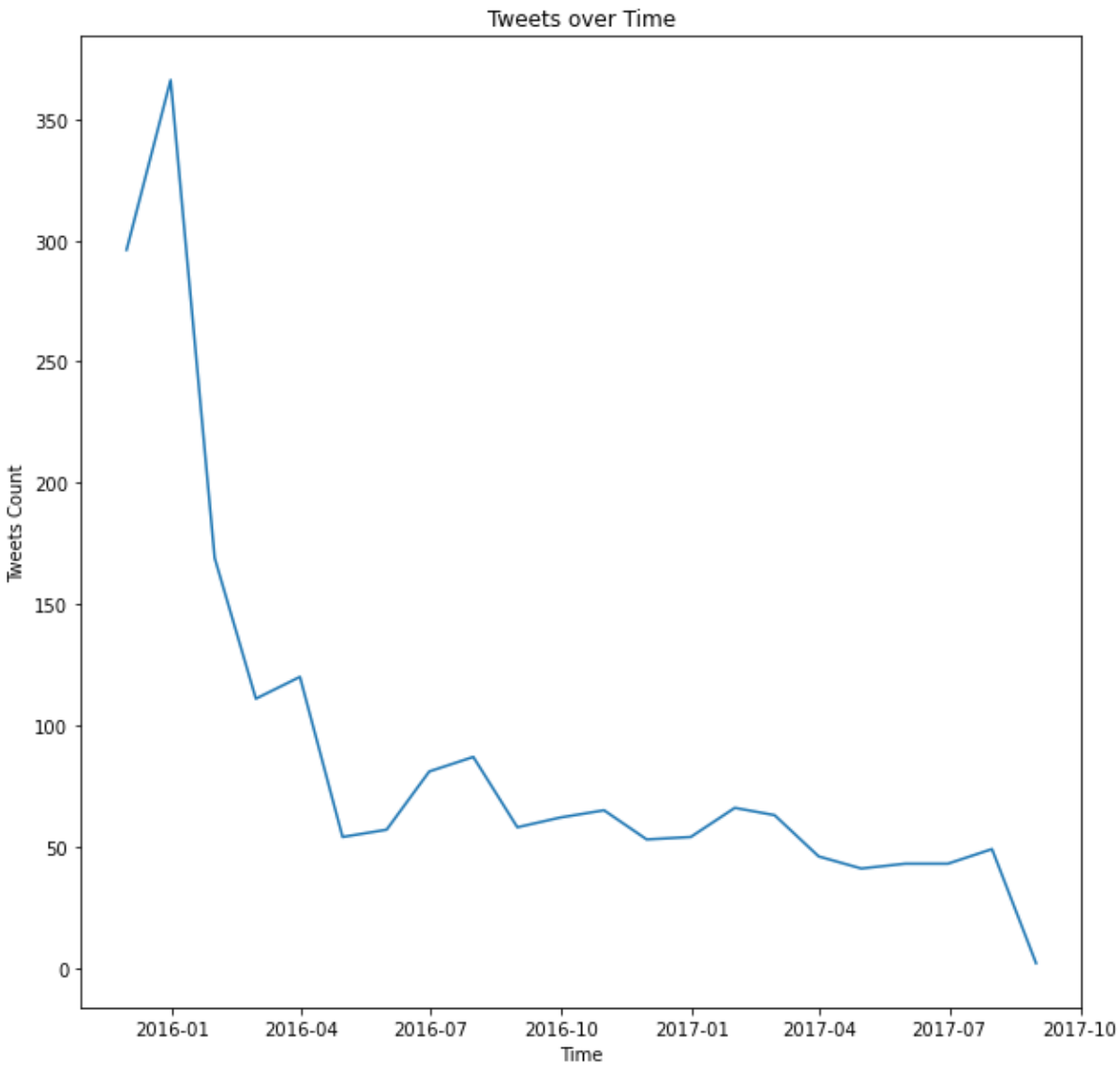
1. Source of twitter traffic:

Before the cleaning, we had the majority connected from their iphones, but some others are connected from the web client, twitterDeck (which is “your personal browser for staying in touch with what's happening now.” according to its developers) or vine.co which used to be an entertainment network but is in an archived state now. After performing the cleaning, we found out that 100% of tweets were sent from iphone users (see figure: Most used Twitter source)



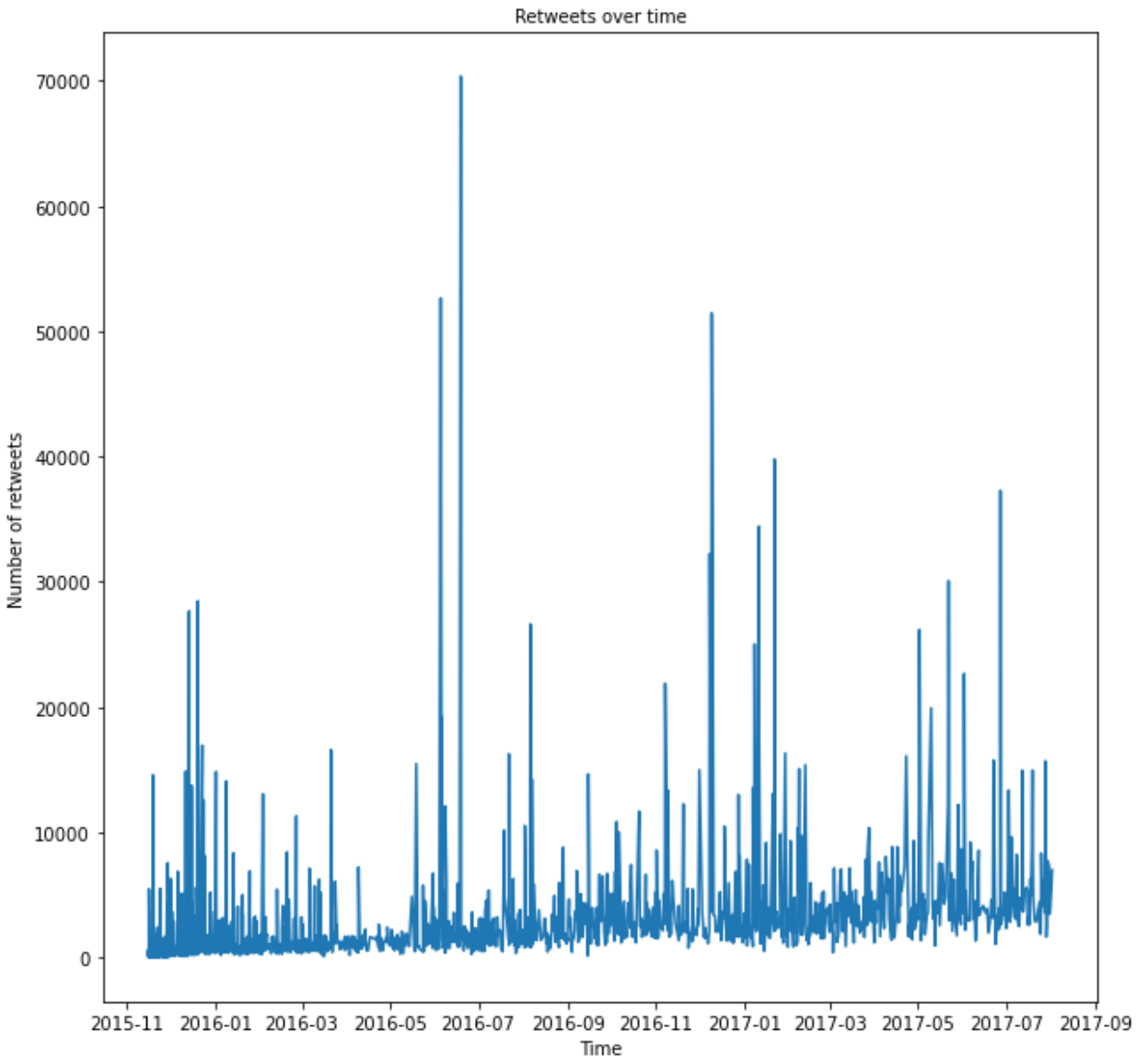
2. Number of tweets over time:

The tweets we have about those iphone users revolved around the period of time from the end of 2015 up until August 2017. The month with the highest number of tweets received by WeRateDog twitter account was by far January of 2016. (see figure below). Then, the number started, generally, decreasing.



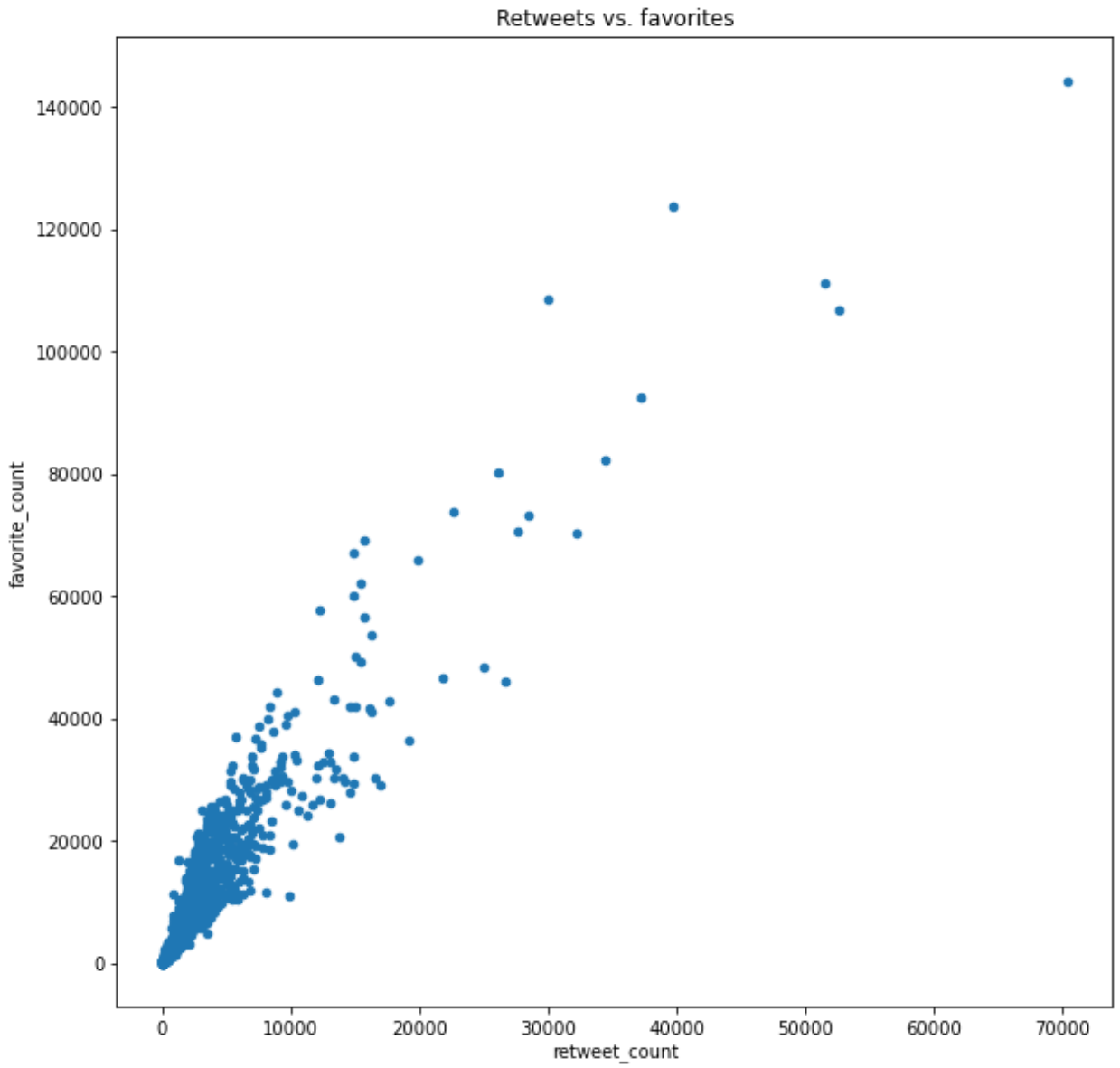
3. Number of retweets over time:

The number of retweets on the other hand, wasn't as high as the number of tweets at that time. But, there were ups and downs all over the same period of time with some obvious peaks (see chart below) that are probably related to some cute dogs that made the tweets go viral (Just an assumption that needs to be verified).



4. Number of favorites vs. number of retweets:

A very interesting measure is the correlation coefficient between the number of retweets and the number of favorites which is equal to almost 0.93 . This is a very high correlation that was represented in the below figure:



Conclusion:

We have had some very interesting insights that motivates us to try and dig deeper and deeper into this dataset in order to acquire more insights and gain more data wrangling skills. We can have more information about the most common dog names for instance, the stages in which the dogs are, most dog breeds represented in this dataset and more. I will continue the adventure and will meet you in part 2 ;)