

Act Report

**Udacity Data Analyst NanoDegree: Wrangle and
Analyze Data Project**

By: Bijay Shrestha

Introduction

This dataset consists of 2356 tweets in total that range from Nov 2015 to Aug 2017. WeRateDogs is a twitter account that has received a huge amount of international media attention as users rate dogs with a humorous comment. Before creating visualization, the data had to be gathered, assessed and wrangled after which it was stored in a csv file.

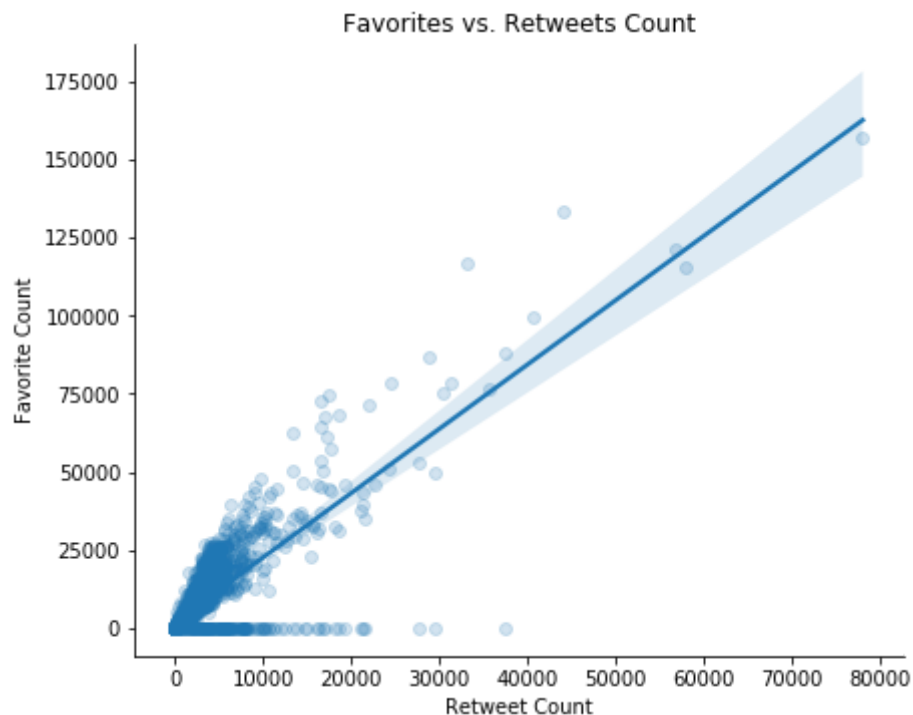
Wrangling Data

The data was collected from 3 different sources. The data was then assessed and cleaned using various methodologies available in python libraries. The data was cleaned programmatically in three stages: define, code and test. Some of the examples of cleaned data are deleting unnecessary columns, converting data types into correct ones, combining four different columns into one, replacing the inconsistent values in a specific column, converting the text of a specific column into readable format etc. They were then combined together based on the tweet_id and loaded into a csv file.

Visualizing Data

Pictures speak louder than words and that applies very much to data analytics. It is easier to understand the meaning of data through visuals. Below are the answers to the key questions through visuals.

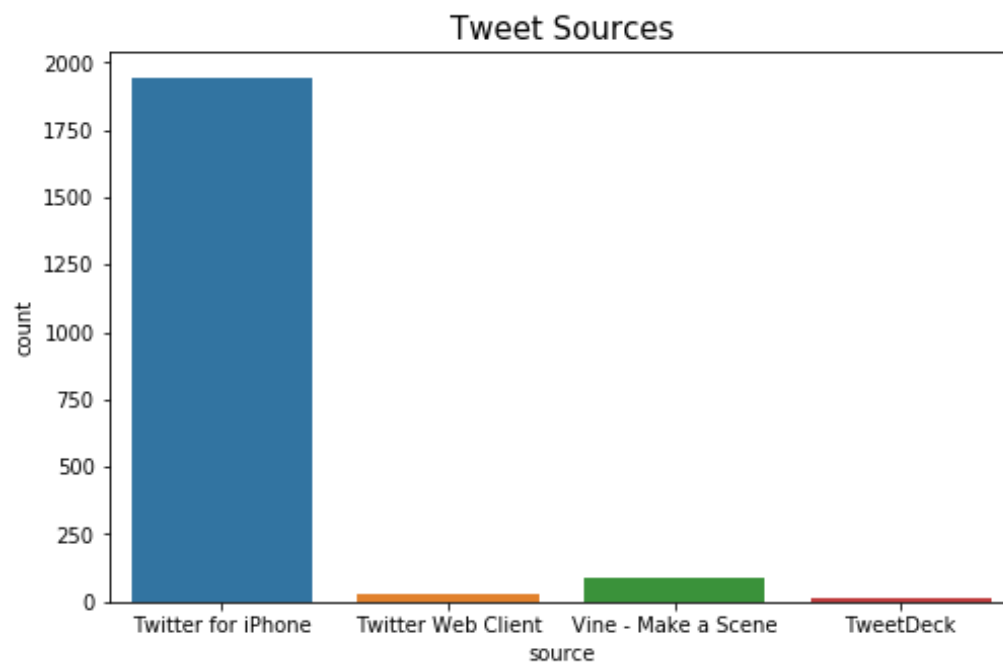
What is the correlation between favorites and retweets?



The figure above shows that favorites and retweets are highly positively correlated. For about every 4 favorites, there is 1 retweet. Majority of the data falls under 40,000 favorites and 12,000 retweets.

What is the breakdown of the source?

Twitter for iPhone	1943
Vine - Make a Scene	91
Twitter Web Client	31
TweetDeck	10

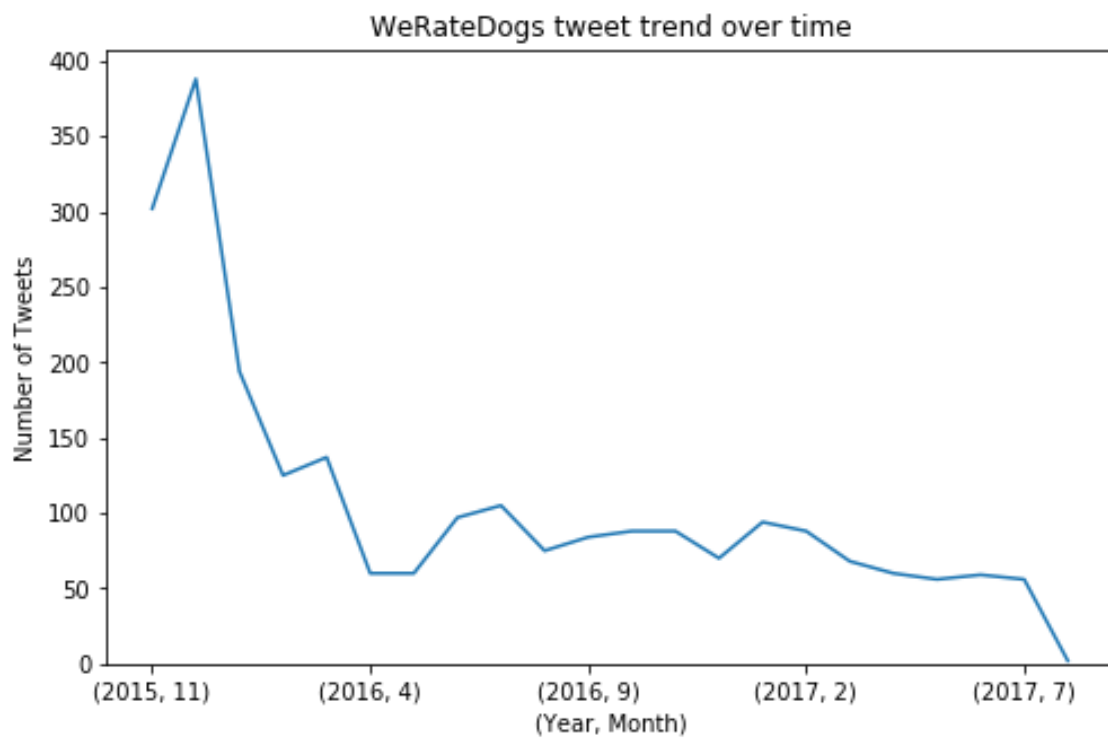


The figure above shows that the most of the tweets come from iphones, followed by vine-make a scene, and twitter web client. The least number of tweets come from tweetdeck.

What is the monthly number of tweets?

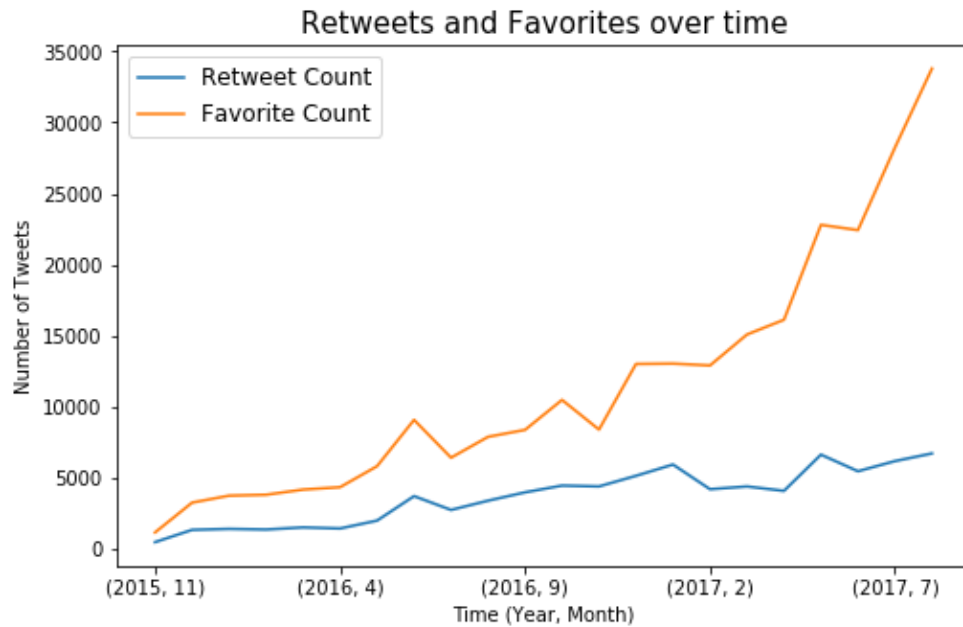
2015-11	302
2015-12	388
2016-01	194
2016-02	125
2016-03	137
2016-04	60
2016-05	60

2016-06	97
2016-07	105
2016-08	75
2016-09	84
2016-10	88
2016-11	88
2016-12	70
2017-01	94
2017-02	88
2017-03	68
2017-04	60
2017-05	56
2017-06	59
2017-07	56
2017-08	2



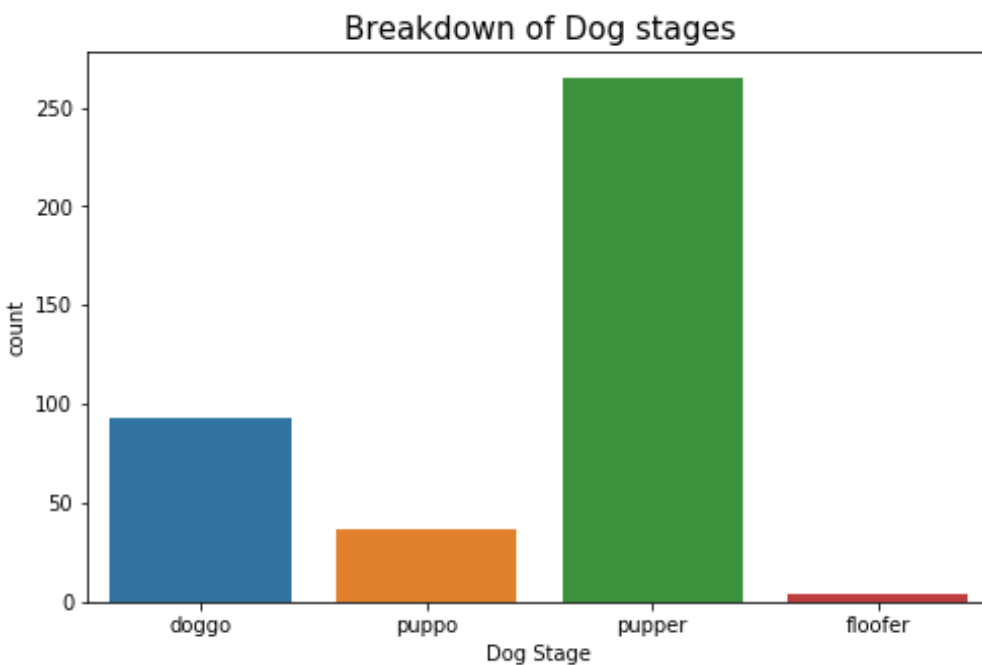
The above line graph shows that the tweets were higher in the beginning and slowed down rapidly after. December 2015 had the highest number of tweets after which the number of tweets started to decline rapidly until April 2016. The tweets frequency started to pick up slightly after that but then started to decline at a slower rate. Overall the number of tweets have a downward trend with a steep decline until April 2016 and slower decline after that.

What are the retweets and favorites counts over time?



The above graph shows that both the number of favorites and retweets have grown over the time. The favorite counts are higher than that of retweet counts. The growth rate of retweets is gradual over the years. However, the number of favorites have grown heavily after a steep decline in mid 2016.

What is the breakdown of the dog_stage?



The above bar shows that amongst the four dog stages, pupper was the highest reported with 265 counts, followed by doggo with 93 counts and puppo with 37 counts. Floofer was the least reported with 4 counts.