

Wrangling and Analyzing Data Project - WeRateDogs

Name: Michael Owusu Agyebeng

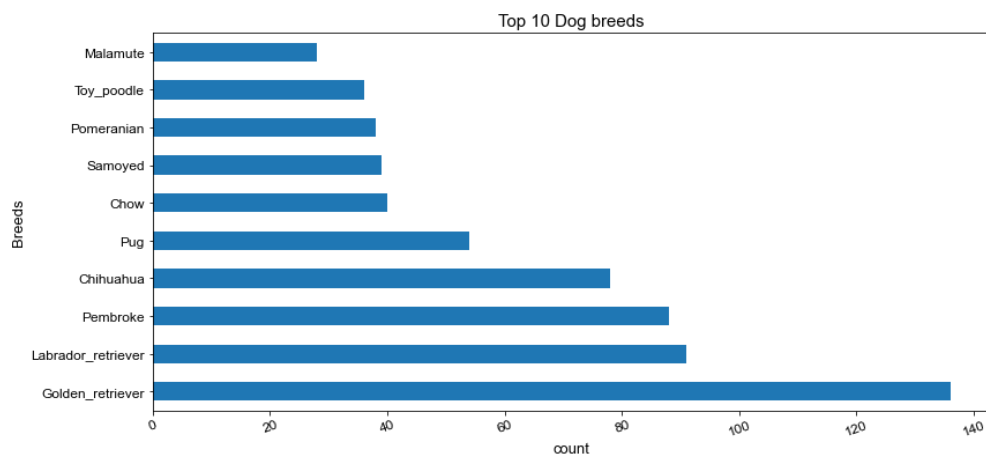
Description: Udacity Student

The Data Gathering process included downloading a csv file, downloading a file from URL, and gathering data from the Twitter API. I was required to request for permission from twitter to grant me developer status and permission to gather data from their platform.

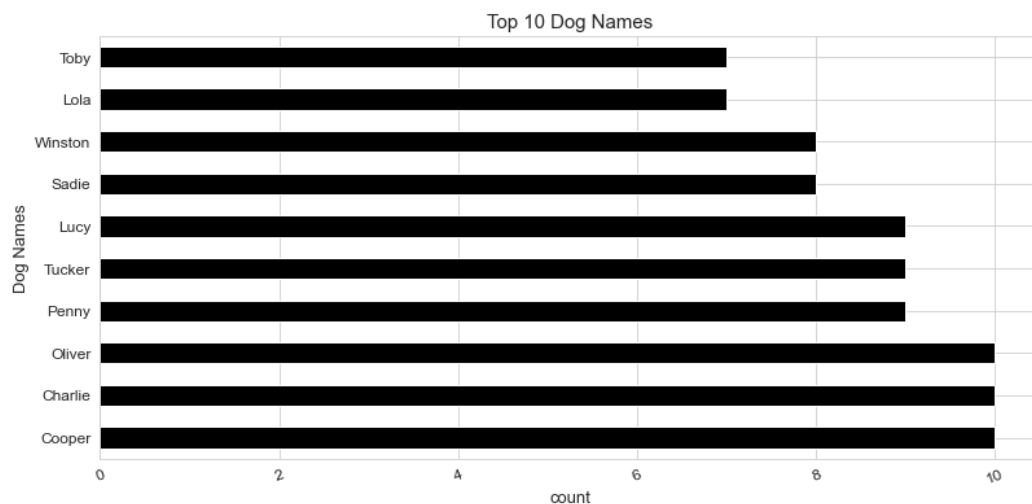
After gathering the data, I found 12 issues which included ten (10) quality issues and two (2) tidiness issues. I cleaned then and merged them together to form one dataset which I stored and named "twitter_archive_master.csv".

After assessing the data, I analyzed and visualized it and found some insights in it. These include;

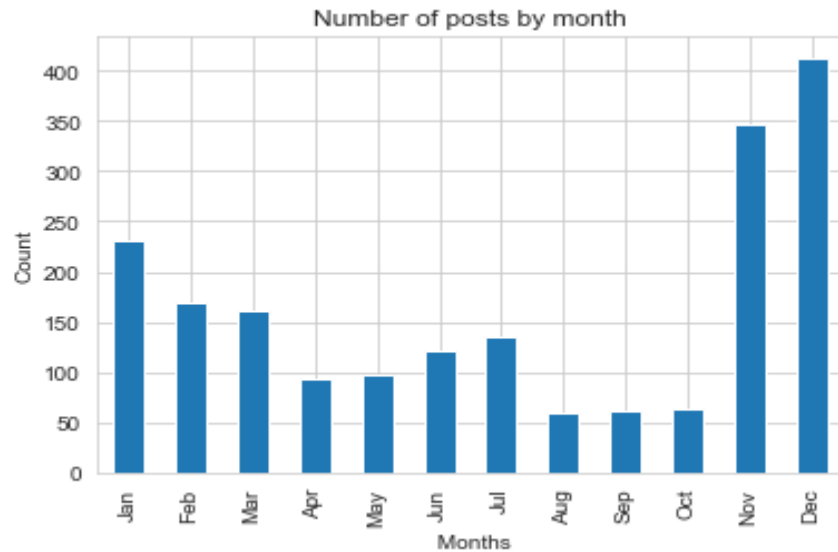
1. Golden Retriever was the most common breed of dog found in the data with over 130 mentions



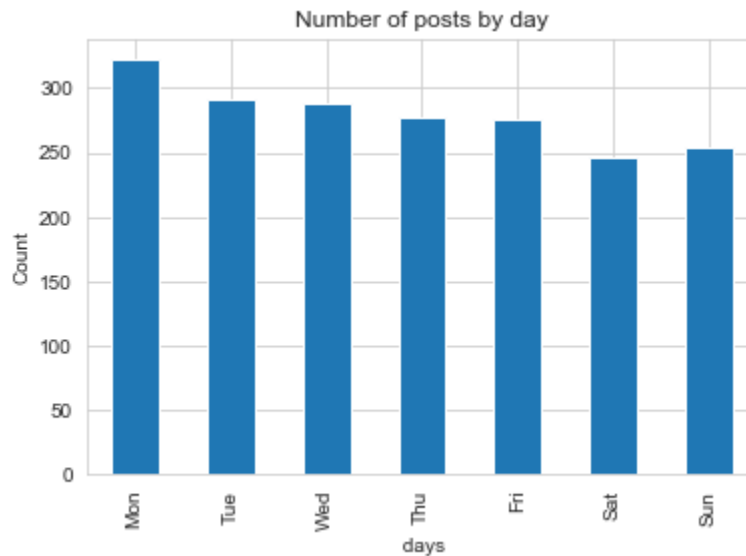
2. Oliver, Charlie, and Cooper were the most common dog names giving a tie of 10 counts each



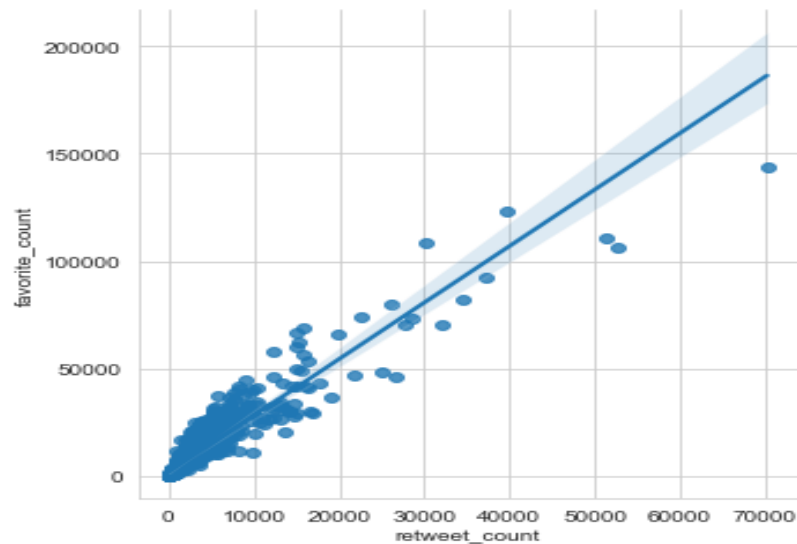
3. Most dog tweets were posted in December and may be because of the festive season and holidays hence most people spent time at home with their dogs. The month of August had the least number of posts.



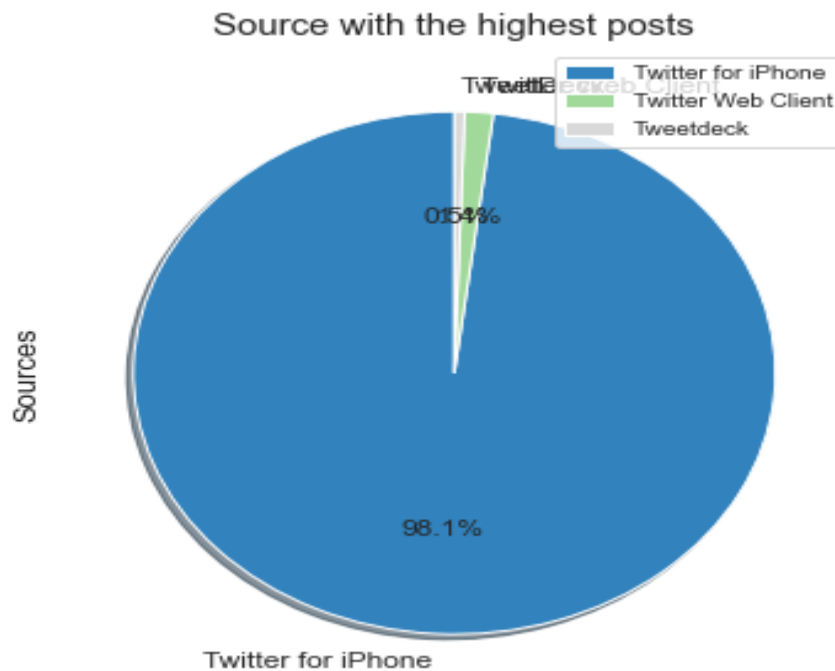
4. Monday recorded the most tweets with over 300 dog tweets signifying it is the day most of the tweets were posted.



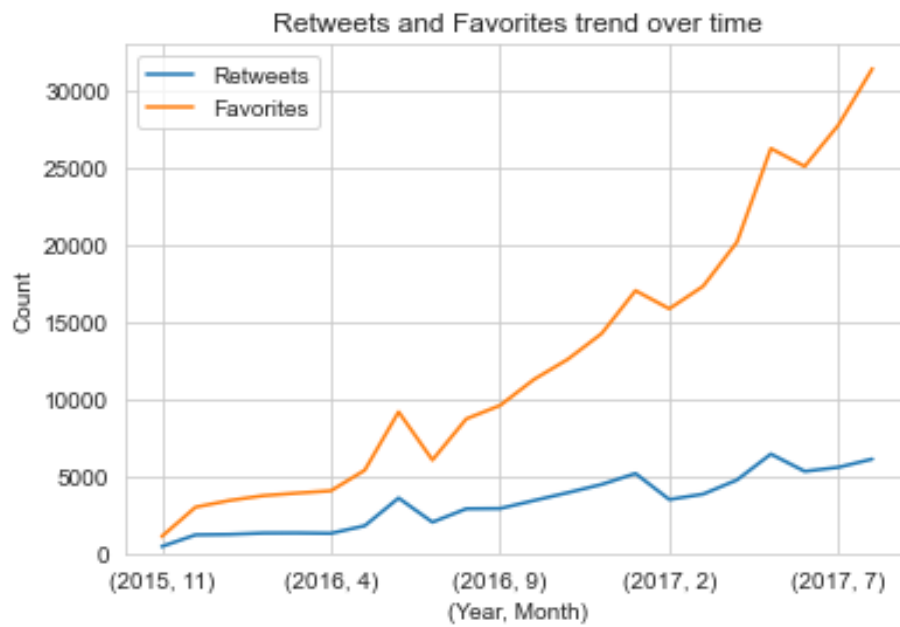
5. There is also a very strong positive relationship between retweet count and favorite count of the WeRateDogs data. This signifies an association between retweets and favorites. Hence the more retweets a post gets, the more favorites it gets.



6. Almost all of our data was gotten from iPhone users with them giving a 98.1% of our data.



7. When retweet and favorite count was put viewed over time, it was recognized that favorite counts were always more than retweets but as time went by, while retweet count increased approximately linear, favorite count seemed to grow approximately exponentially. Which indicates that as at November 2015 there were fewer engagements (a sum of less than 5000 favorites and retweets) on WeRateDogs posts, but it had popularity over time and racked up more than 30000 favorites alone in July 2017.



8. Putting the average retweet and favorite count side by side, it is evident that favorite counts were averagely more than retweet counts and doublestage had the most favorites while pupper had the least favorite count.

