

Analyzing and Visualizing Data

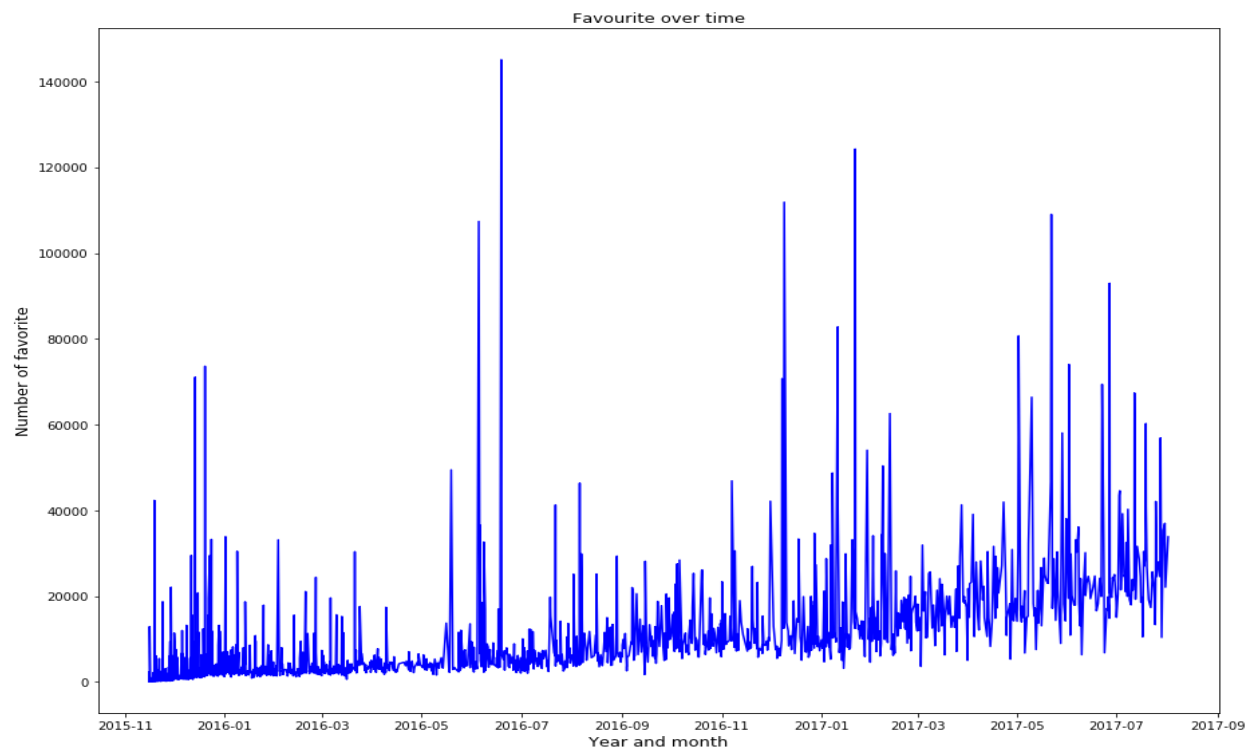
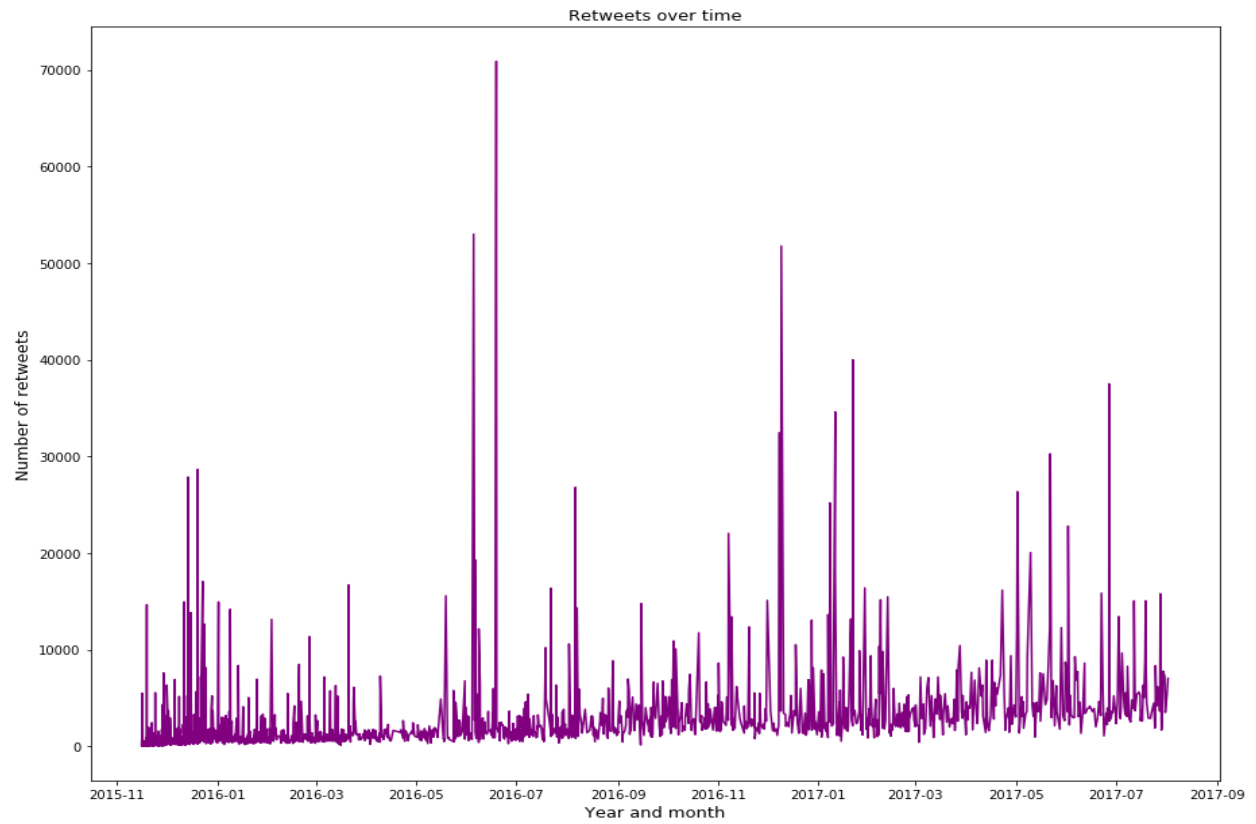
In the final section of my project analysis, I will be analyzing and also visualizing the data I just finished wrangling.

I was able to produce 8 insights and 6 visualizations to represent my work and give a better explanation to my project. A similar method I used in my analysis part was introduced also; I asked a question I wanted to find an answer to with the use of visualization.

Below are my Insights and its Visualization:

How did the retweet count and favorite count improve over time?

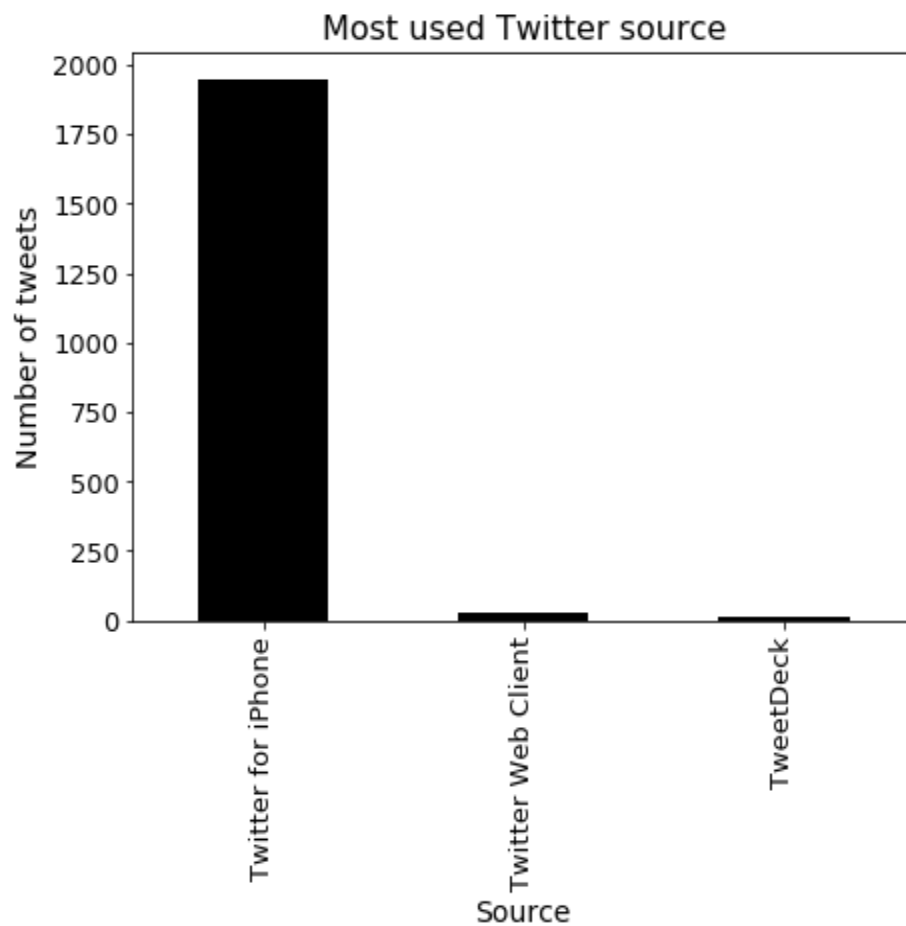
From the graphical visualization, the retweets had a really successful period in the month of June 2016 with retweets of little over 70000 which is also the highest ever they had btw 2015 down to 2017, while the favorite also had a really successful period during in the same month of June 2026 as well with retweets going over 140000 which is also the highest ever favorite count they had between 2015 to 2017.



from the graph above the favorite counts and the retweet counts have high correlation and are at a similar level. From 2015-11 to 2016-05 the favorite counts and retweet counts were lower compared to 2016-07 down to 2017, then the number of the favorite counts and retweet counts became higher and higher. As time goes on, more people are liking tweets than retweeting them.

Most used Twitter Source?

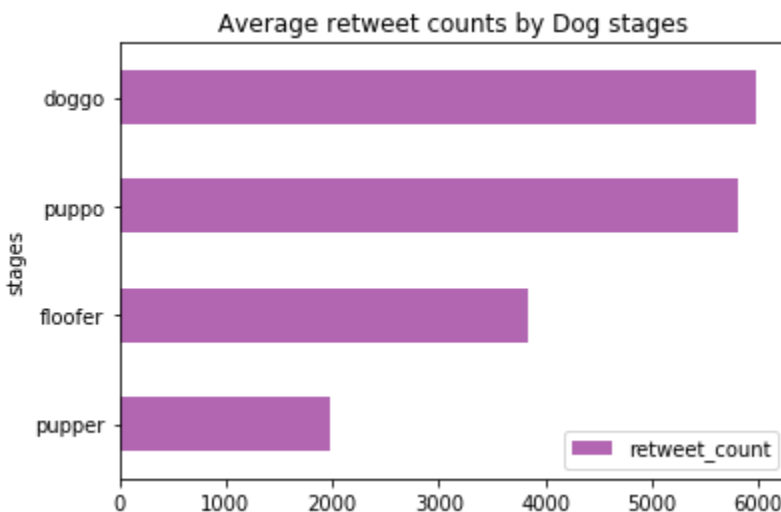
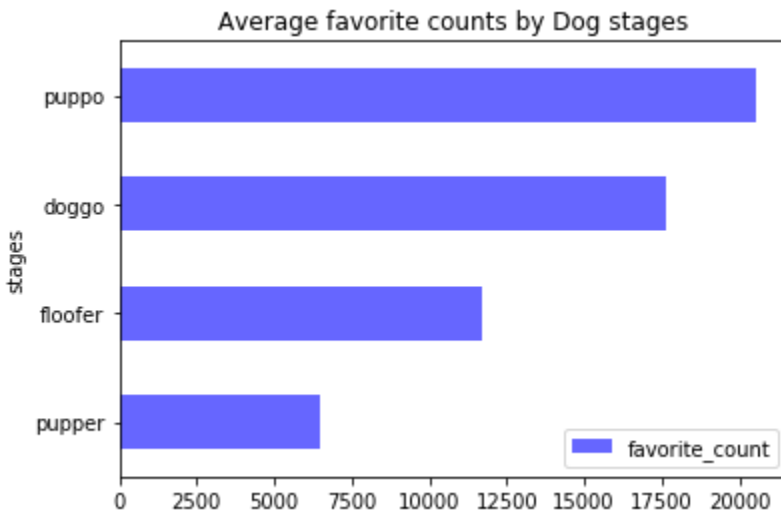
From my visualization we can see that The most use source of this twitter account dataset will be the Iphone user.



the iphone is the most used source

Which stage of dogs got the highest retweet counts and favorite counts?

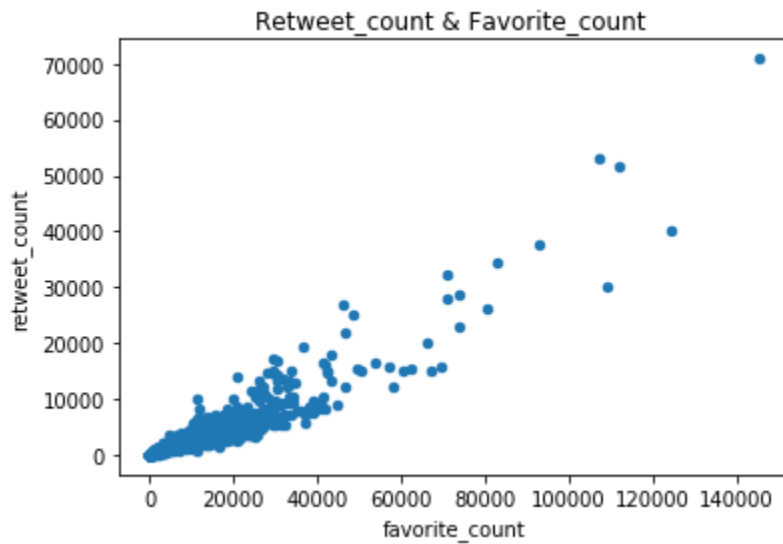
The stage of Puppo got the highest favorite counts and the Doggo got the highest retweet counts and both stages are almost 3 times higher than the stage of Pupper.



The stage of Puppo got the highest favorite counts and the Doggo stage got the highest retweet counts

Is there a correlation between Favorite and Retweet ?

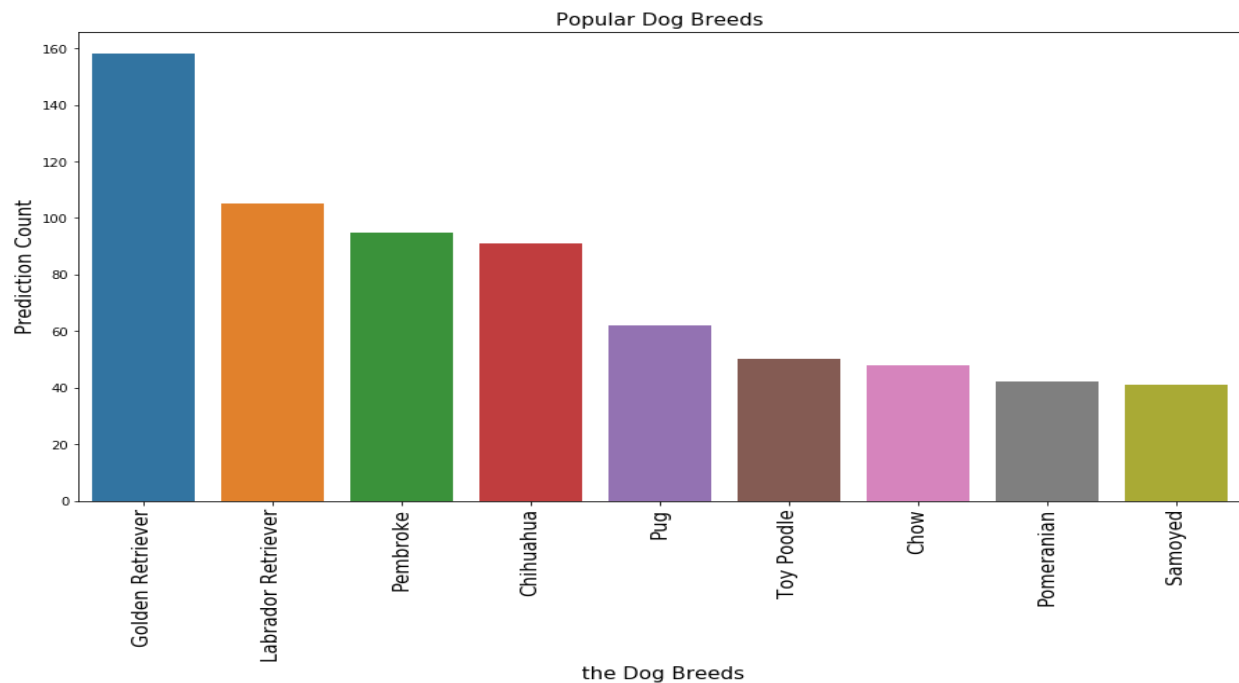
It is safe to say there is a linear relationship between the retweet counts and the favorite counts. They both have a strong positive correlation between them which can be confirmed by the below scatterplot.



As expected after looking at the scatter plot, the correlation is 0.93 which is very strong.

The most popular Dog Breeds?

Based on my data analysis on the breeds I picked out the *top 10 most popular dog breeds*.



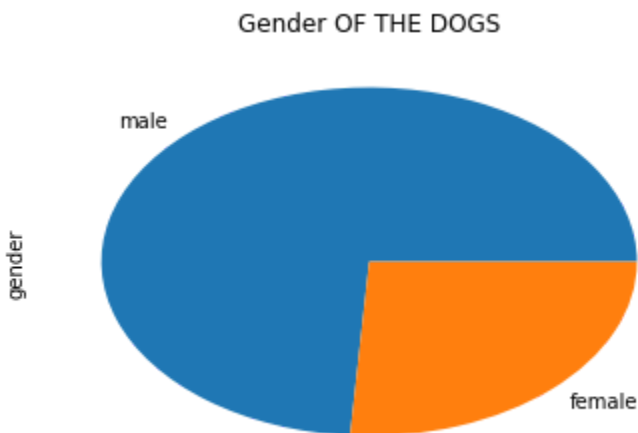
The golden retriever is the highest dog breed in this data set, followed by the Labrador and then the Pembroke, Chihuahua, Pug, Toy Poodle and Chow, every other breed gradually decreases in count.



The Golden Retriever is the most popular dog. These dogs have the most tweets hence that made it the most popular dog to be tweeted about on the list of dog breeds on this data set.

What is the gender distribution like?

From the visualization the male dogs are far more than the female dogs in this dataset.



There are more male dogs identified than female dogs from my graph.

Conclusion

Data Wrangling Process carried out and the below analyses are based on the wrangled data,

- The most popular tweet is also the same as the most retweeted tweet; it has been retweeted over 70,892 times and favorited over 145,088 times, which has the breed name Labrador Retriever and stage doggo.
- The iPhone is the most used source in this data set.
- The stage of Puppo got the highest favorite counts and the Doggo stage got the highest retweet counts.
- The scatter plot shows there is a positive correlation(0.93) between favorite and Retweet.
- from the graph above the favorite counts and the retweet counts have high correlation and are at a similar level. From 2015-11 to 2016-05 the favorite counts and retweet counts were lower compared to 2016-07 down to 2017, then the number of the favorite counts and retweet counts became higher and higher. As time goes on, more people are liking tweets than retweeting them.

- The golden retriever is the highest dog breed in this data set, followed by the Labrador and then the Pembroke,Chihuahua, Pug, Toy Poodle and Chow,every other breed gradually decreases in count.
- There are more male dogs identified than female dogs from my graph.

Limitations

Based on my previous biology Knowledge,the gender column is likely not representative of the dog population at large. this might be due to more male dogs having their sex announced by either by the @dog_rates twitter account or their owner or maybe both cases.it is possible that some of the dogs were misgendered either by the @dog_rates twitter account or my function.that is it's possible that @dog_rates is misgendering some females by using male pronouns as gender-neutral,which is a possibility.i cant draw a concrete conclusion since only a small portion of the rated dogs where not gendered.