**Udacity: Wrangle and Analyze Data**

**Author: Yogesh Chaudhari**

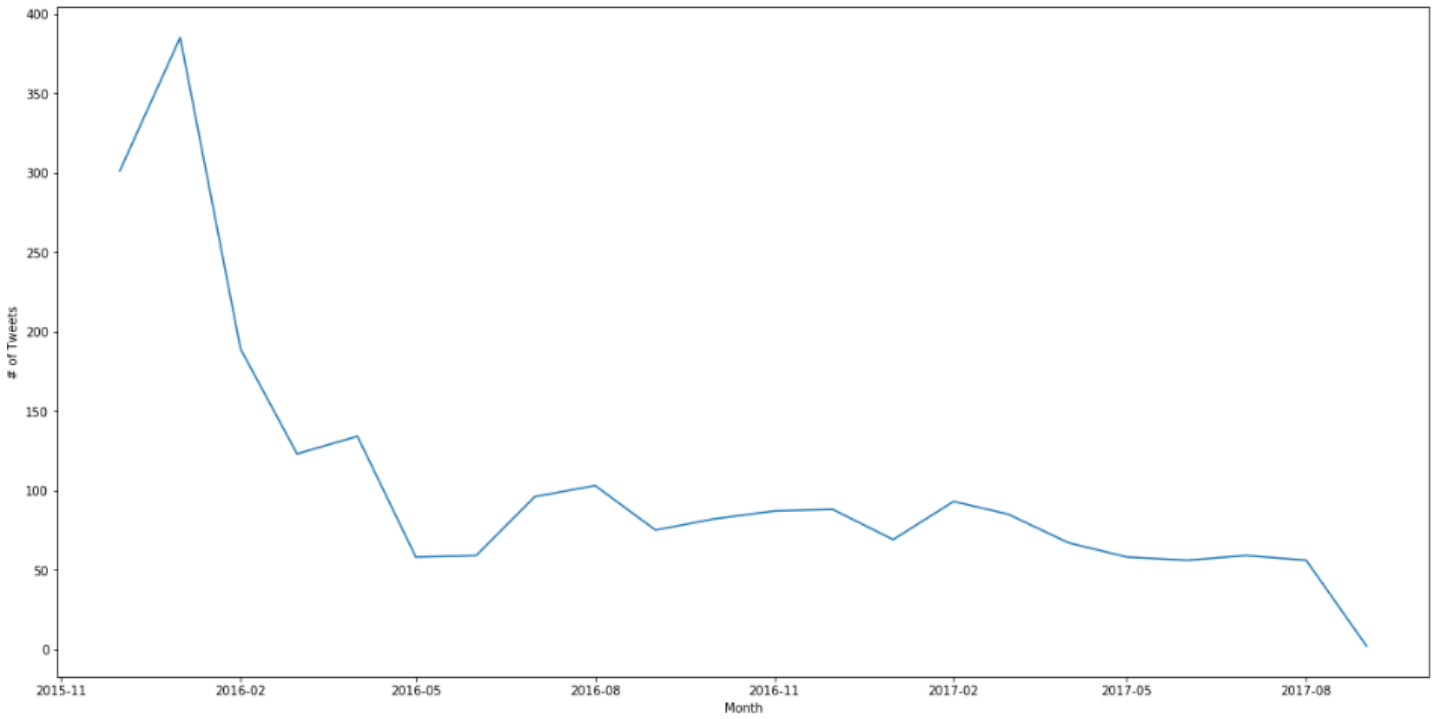
**Date: 1/16/2020**

## Analysis and Visualization

*[**Disclaimer:** This exercise leveraged matplotlib library to plot various charts. Alternatively, Tableau can be leveraged to create such charts quickly and with less code. However Tableau is not in scope for this course, hence matplotlib was leveraged to plot various charts.]*

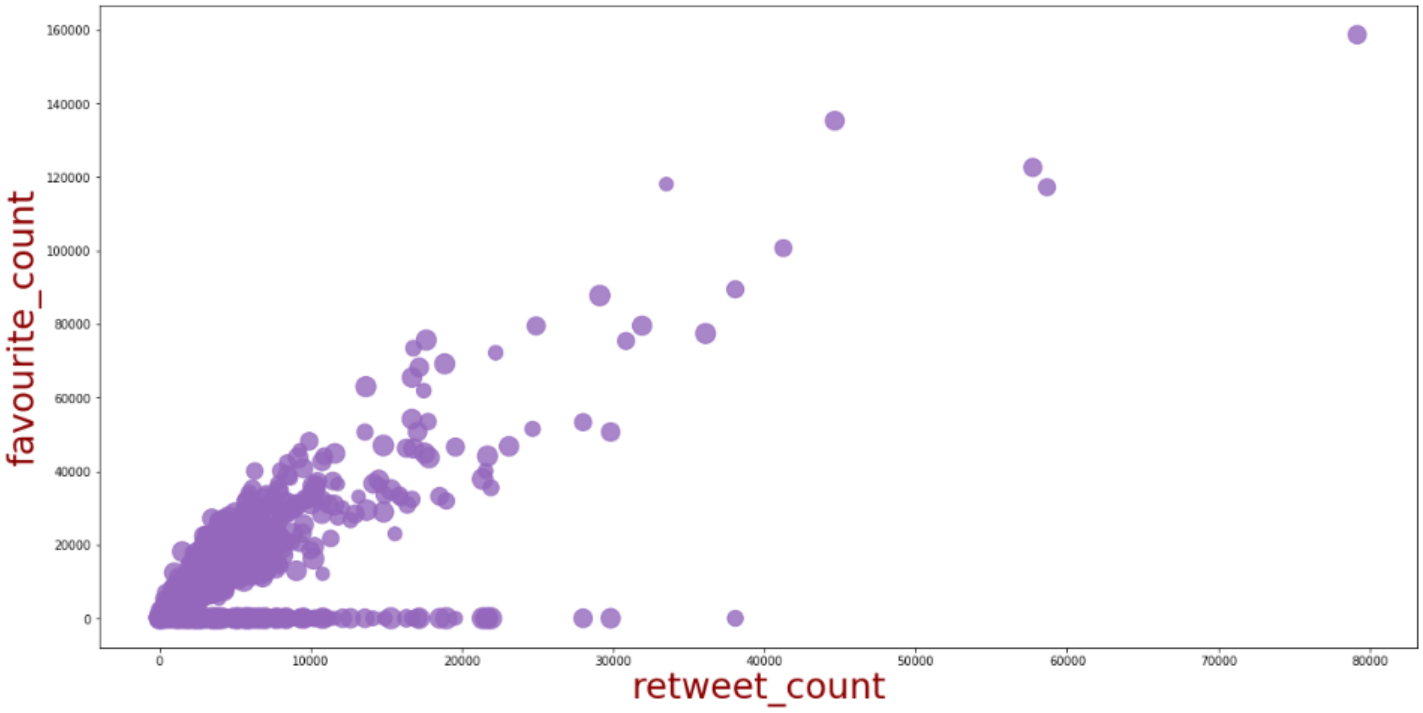
Using various data wrangling techniques, a final dataset was created and leveraged for plotting visualizations using seaborns and matplotlib library. Below are various visualizations and insights derived from collected data:

**a)** **# of Tweets over period:** To understand the activity of the twitter handle @dog_rates, a monthly trend for number of tweets over period was plotted. The plot is shown as below:



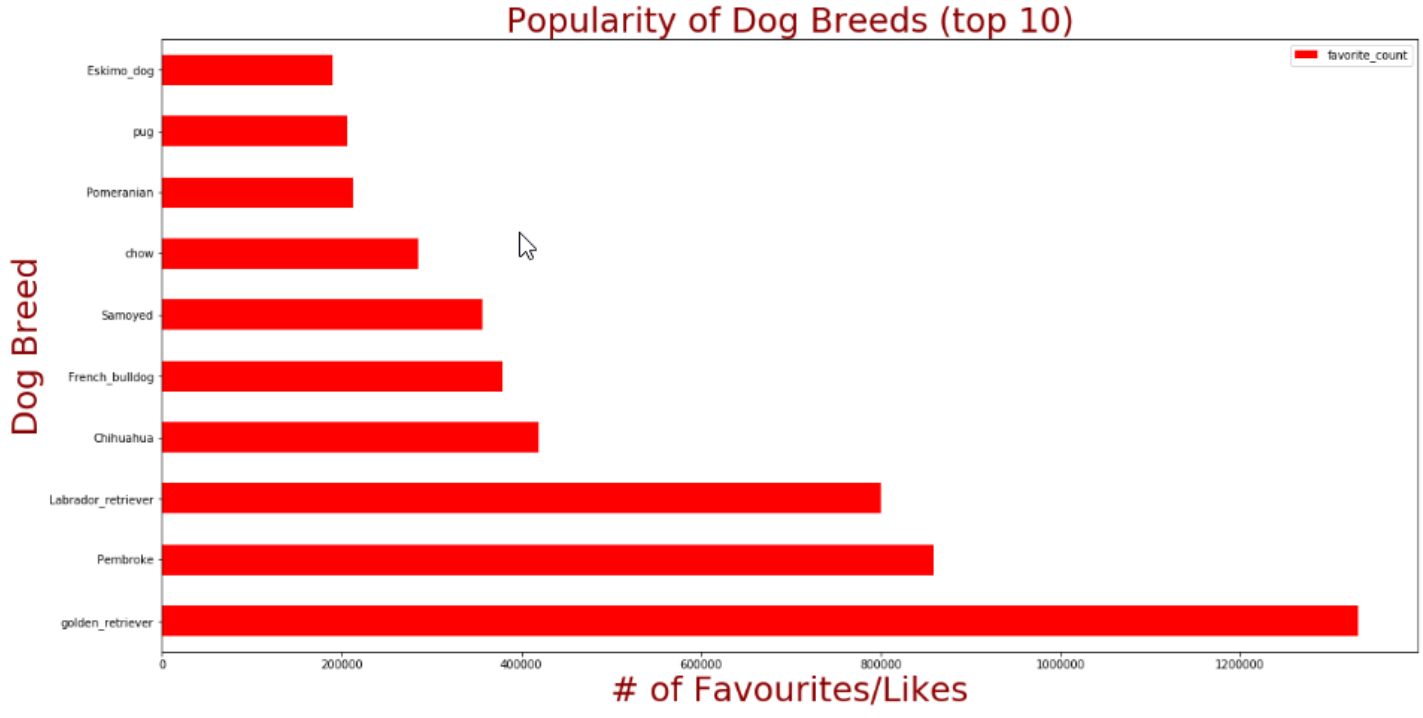

***Insight a:*** *The activity of twitter handle @dog_rates reduced drastically from January 2016 to February 2016. Looking at the trend, the overall activity declined over the period. The tail end of data shows that only 2 tweets were posted in the month of Sept 2017, however this behavior can be attributed to potential incomplete data for the month of Sept 2017.*

**b) <u>Correlation between retweet count and favourite count:</u>** A scatterplot was plotted to understand correlation between retweet_count and favourite_count. The scatter plot shown below indicates strong positive relationship between 2 variables.



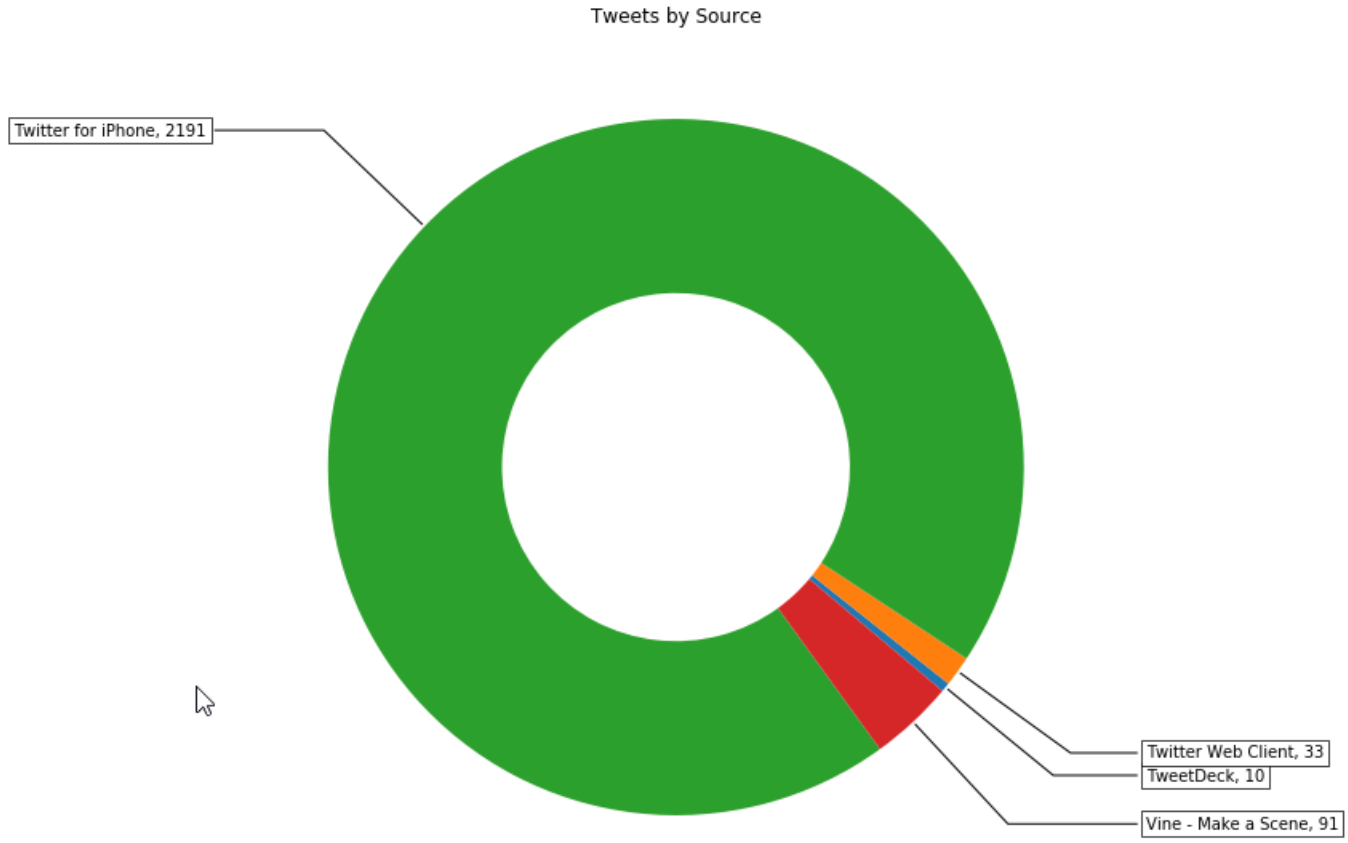

***<u>Insight b:</u>*** *There is positive correlation between retweets and favourites. Based on this initial assessment we can conclude tweet with higher retweets usually receives higher number of likes/favourites. However, additional analysis (p-value, R-squared) is required to conclude the strength of this relationship.*

**c) <u>Popularity among Dog Breeds:</u>** In order to understand which dog-breed is popular among dogs, a bar chart was plotted for top 10 breeds by favourites/likes in the dataset.



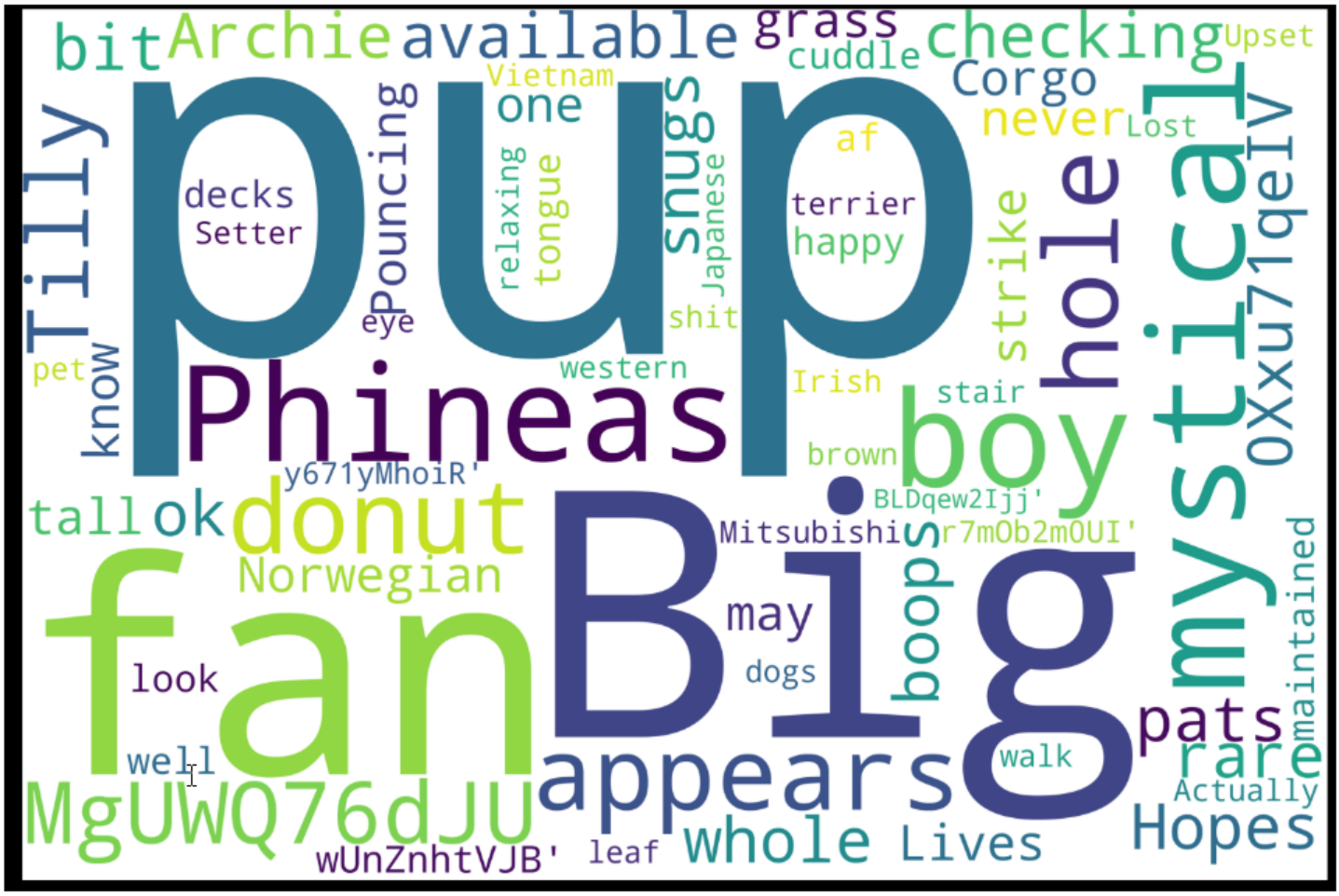
***Insight c:*** *Based on the available data, we can conclude that Golden Retriever is most liked breed among dogs.*

**d) Source of Tweets:** In order to understand which platform used for tweeting a donut chart was plotted to understand the distribution.




***Insight d:*** *Based on the distribution above and available data, we can conclude that iPhone twitter app is most common source of tweets.*

**e)** **<u>Word Cloud of Tweets:</u>** A word cloud was created to understand what word(s) are commonly used in a tweets related to dogs. Below is the word cloud of the words used in the tweet text.




***<u>Insight e:</u>*** *Based on the word cloud, it appears that people commonly use words such as pup, big, fan most commonly in tweets.*