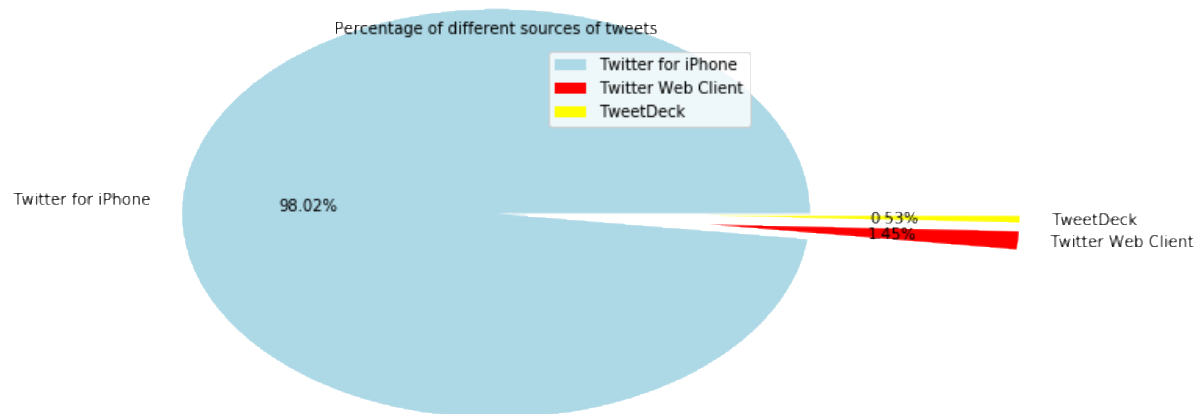


Insights and Visualization

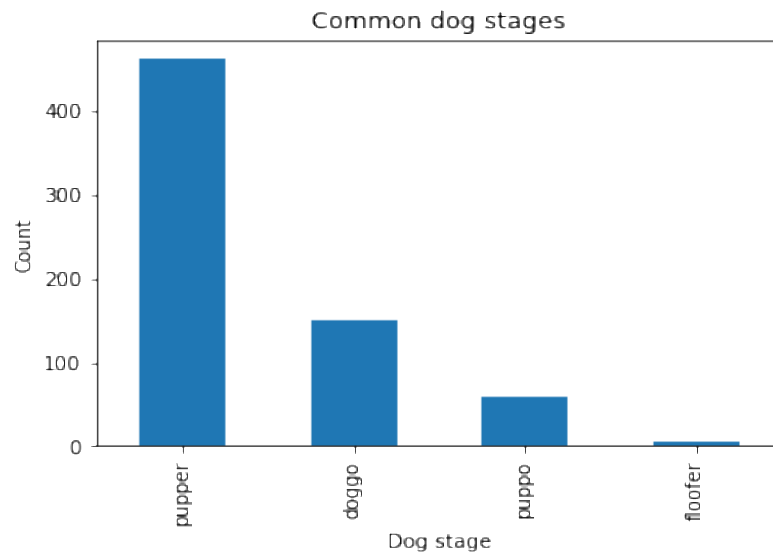
The Twitter account WeRateDogs is used for humorously reviewing pictures of dogs. Dogs are rated on a scale of generally one to ten, but in many instances they are given ratings in excess (e.g. 13/10). This initiative was started by college student Matt Nelson, and has received quite a lot of media coverage and popularity. WeRateDogs downloaded their Twitter archive and sent it to Udacity exclusively for Nanodegree courses. Additional analysis steps of Data Gathering, Assessing and Cleaning was required to present these insights and visualizations.

Visualization



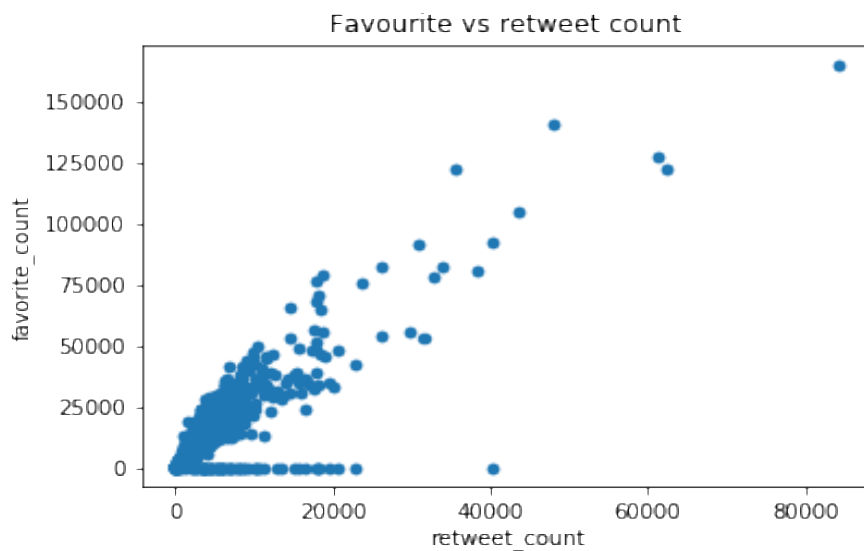
The pie chart shows the percentage source of tweets. It can be seen that majority of the tweets (~98%) were sourced from Twitter for iPhone app; followed by Twitter Web Client (1.45%), and TweetDeck (0.53%).

Insight 1:



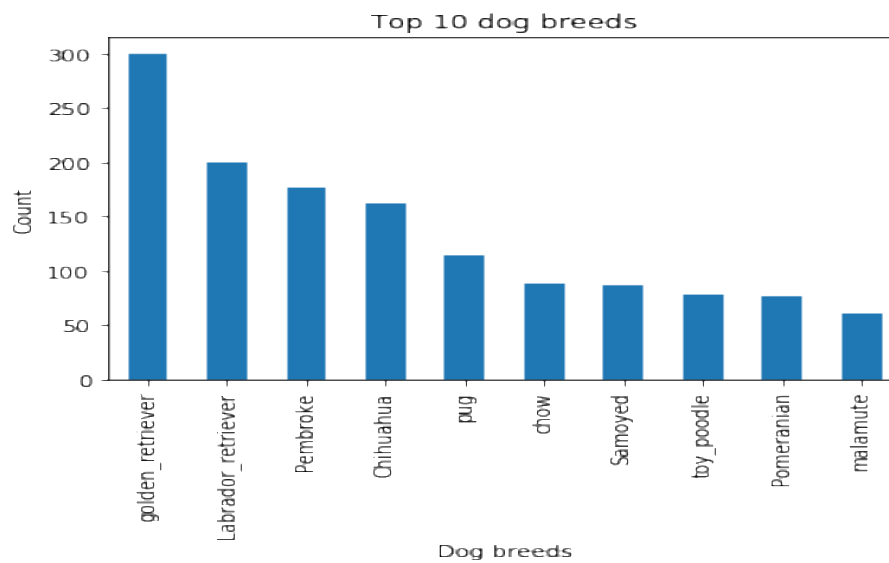
The bar graph above shows the value counts of the four different dog stages. It can be seen that pupper is the most popular, followed by doggo, puppo and finally floofer.

Insight 2:



The scatter plot above shows the relationship between favourite tweet count vs retweet count. It can be seen that there is a strong positive correlation between favourite count and retweet count, which implies that as there are more favourites for a particular tweet, it will have a higher chance of getting retweeted.

Insight 3:



The bar graph above shows the top 10 dog breeds with higher confident probability under p1. It can be seen that according to the prediction algorithm, the most famous breeds are Golden Retriever, Labrador Retriever and Pembroke, in the top 3, followed by Chihuahua, pug, chow, sampyed, toy_poodle, pomerian and malamute.