WeRateDogs Tweets Insights

During my study with Udacity, we were asked to analyse data gotten from a twitter user WeRateDogs. This account rates people's dogs with a funny comment about the dog. Collated data contains tweets between Nov 2015 to Aug 2017 as well as a prediction algorithm of some clear and distorted dog images passed through a neural network that was designed to be able to predict the type of dog breed. Several interesting insights were derived from this dataset which would delve into shortly. If you are a dog lover, it is imperative that you follow that twitter user for more information and ratings about dogs. I bet you wouldn't have any regrets.

Firstly, let's look at the frequently used words in the tweets. This was done using wordcount

The image above displays a word count with a dog image. As you can see the most frequently tweeted words are Pupper, dog, pup pet, meet amongst others. This indicates that the tweets mostly refer to introducing their dogs and its stages such as pupper and doggo amongst others.

The dataset provides us with many insights, but this report will focus on 5 listed below:

1) Most common dog names
2) Most common dog stage with high count of likes (favorite)
3) Relationship between retweet count and favorite counts over time
4) Top twitter source
5) Prediction algorithm accuracy.

Most Common Dog Names

In the dataset there were some inconsistent names which we disregarded during analysis. The bar chart below displays the top dog names tweeted.
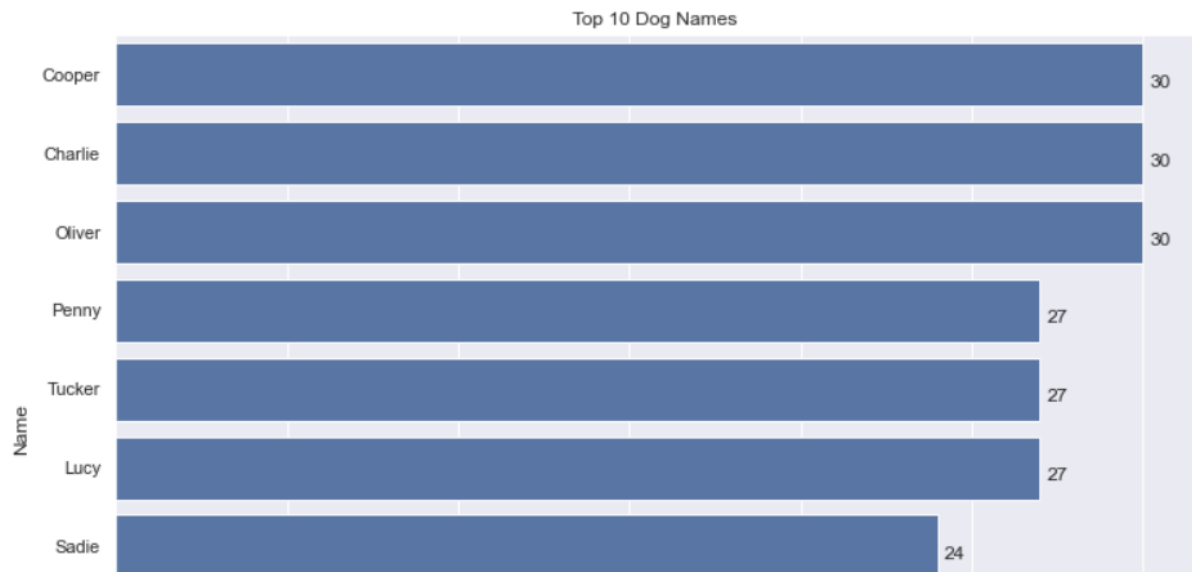
This shows that top names for dogs tweeted during this time frame was Cooper, Charlie and Oliver. I'm sure you are also wondering like I am right now regarding the famous dog's name BINGO 😊

Most Common Dog Stages with high count of likes(favorites)

You should know that there are various stages of dog which are doggo, floofer, pupper, and puppo (detailed information can be gotten from the book dogtionary). Let's look at the stage that has a higher count of like. This was plotted using a box plot. The chart below indicates that the puppo stage has more count of likes closely followed by doggo. My guess is that they are very cute dogs.

Dog categories via favorite count
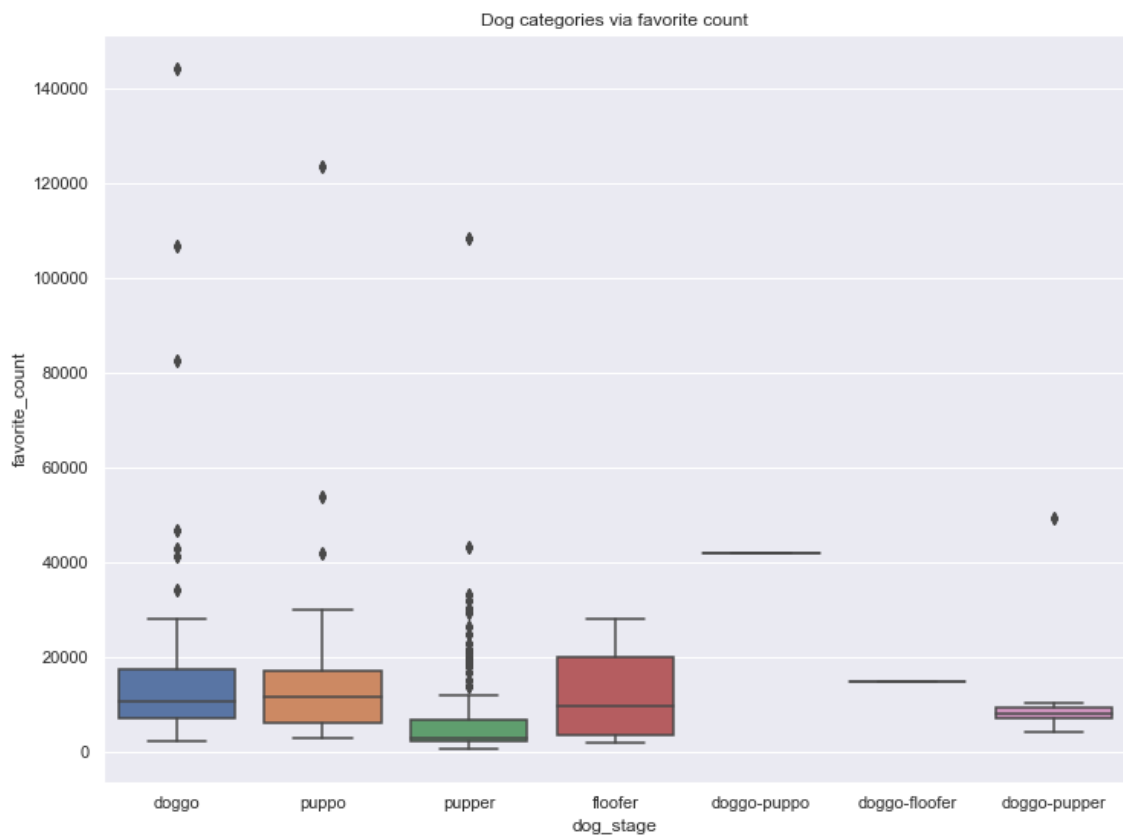
Relationship between retweet counts and favourite count over time.

The time these tweets were made between Nov 2015 to Aug 2017. The retweet count and favorite count showed a very strong correlation at 0.93, this goes to show that most people that like a tweet often tend to share and retweet this for others to view as well. The figure 4 below is a time series with the weekly average count of both retweets and favorite
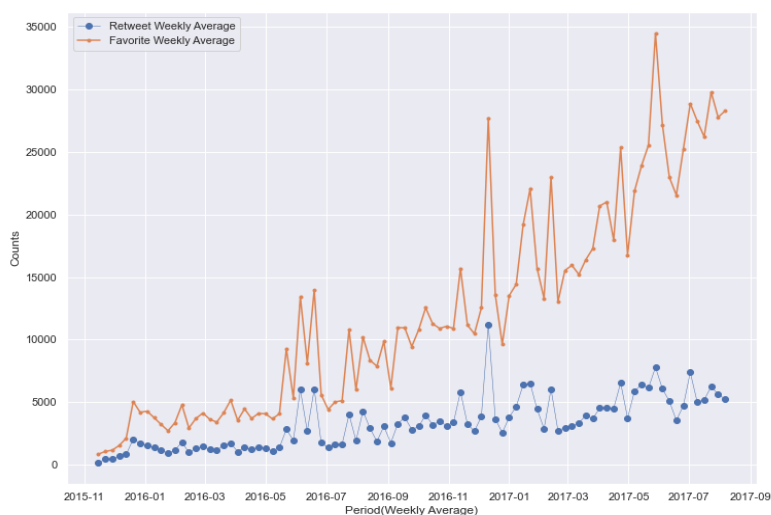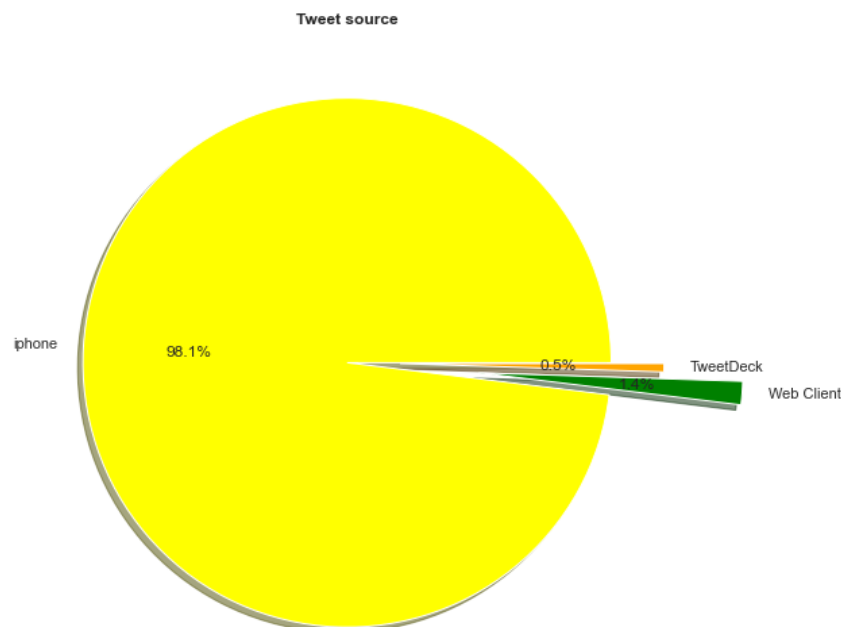
From the above graph, we can that there are 4 spikes/increase in retweets and favorite counts which deviates from the normal continuous increase between the below months. Taking a closer look, we would observe that these spikes occur during summer and winter holidays respectively. This indicates that most twitter users have more time to tweet and take their dogs out.

Another insight derived from this chart is the continuous growth in counts over the years showing the popularity of the twitter page increasing and I bet having more twitter followers as well

Top Twitter Source

Here a pie-chart was used to know the percentage of source of tweet from the users. The chart shows that users with iphones represents 98% of tweet source. This could mean most iphone users have at least a dog.

Tweet source

iphone   98.1%
                                              0.5%   TweetDeck
                                                   1.4%   Web Client

Prediction algorithm accuracy

Images gotten from this tweet was ran through a neural network in order to predict the dog breed as well as the prediction confident. Analysis was carried out to confirm why some dogs had 100% prediction confident while others had about 10%. The result was that dogs with higher percentage level had a clear view of the image with the dog while does with low prediction confident either does not have a dog image or the dog image is distorted with an obstruction. This means that the confident level is true but the neural network can be improved to be able to detect dogs even within distorted images.The figure below show images of a high and low prediction.

Low prediction confident

High prediction confident