# Analysis and Visualization of WeRateDogs.

## Introduction

The dataset that used for analyzing and visualizing is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. It was started in 2015 by college student Matt Nelson.

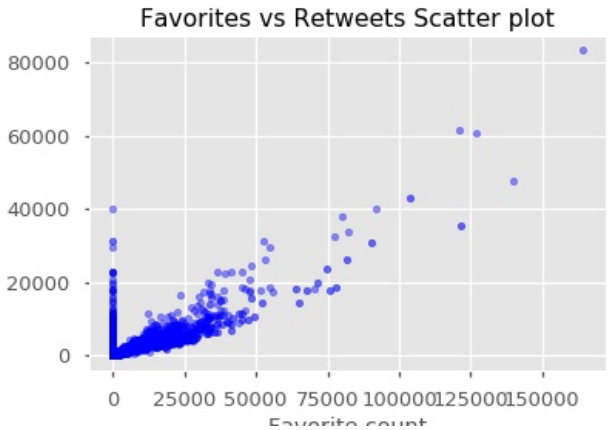
## Data Preparation

To analyse the dataset, we need to use three different files. The Twitter archive file,image prediction file and finally the JSON data file. To analyse the dataset we need to combine the three files into a single dataframe. The tweets from these files were from before August 2017.

Analysis of this dataset were mainly concentrated on the relationship between the favorites and retweets, finding the most popular dog among the dog stages,finding the most common breeds based on the prediction .

## Analysis and Visualization 1.

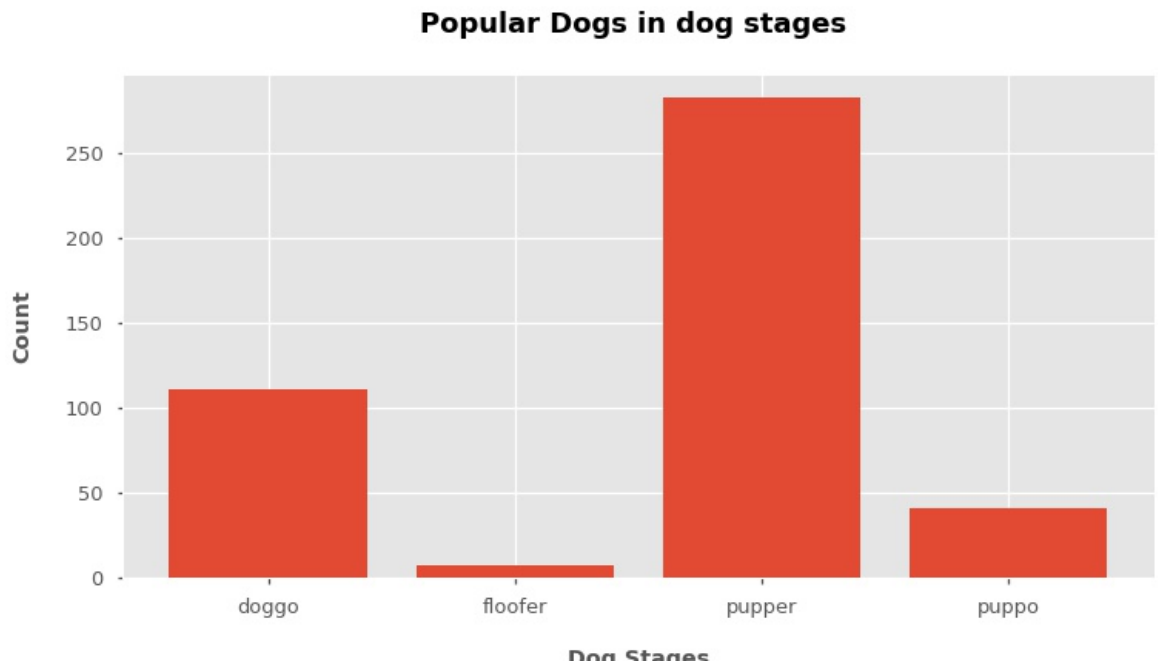
My first insight to the dataset is about the relationship between the retweet count and favorite count. A scatter plot is used ,which revealed a positive correlation between the two.




With refer to the above scatter plot, we can conclude that retweet count increases with favorite count,eventhough there are some outliers.

## Analysis and Visualization 2.

My second insight to the dataset is finding the popular dog among dog stages.Histogram plot revealed 'pupper' is the popular dog among the dog stages,followed by 'doggo' and 'puppo'.




## Analysis and Visualization 3.

My third analysis is to find the most common breeds based on the prediction algotithm.The histogram plot reveals Golden retriever is the topper, even though there are some 'not_dog' classfication is there.