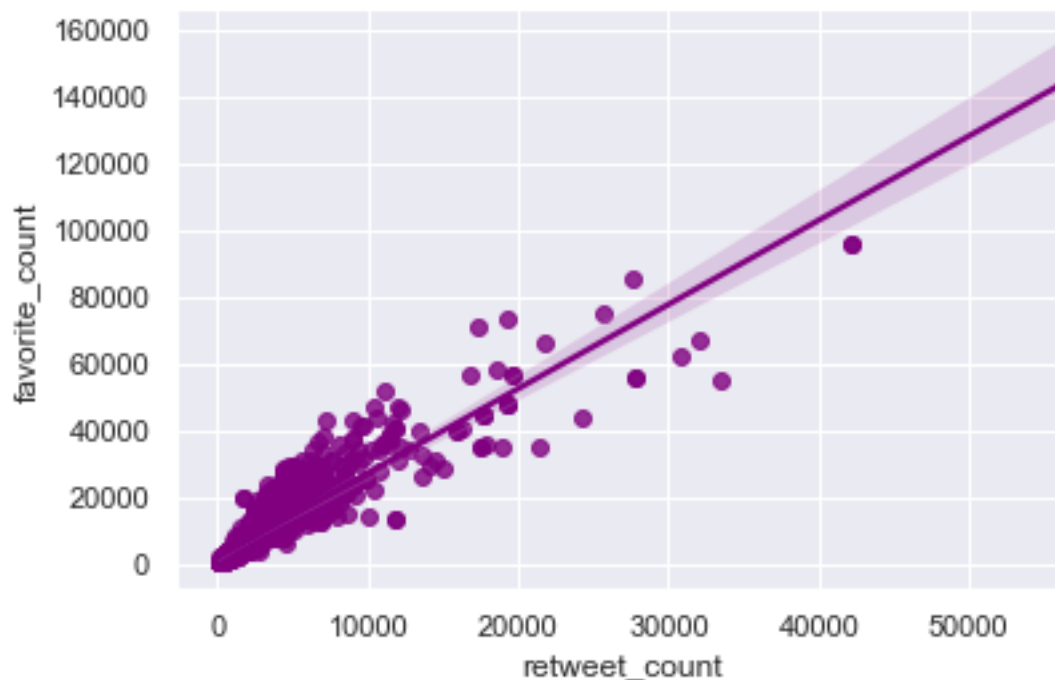


## Introduction

For the data wrangling project on Udacity Data Analyst Nanodegree, I had the chance to go through the whole data analysis process, from collecting the data to cleaning and analyzing it and finally visualizing trends from the data. The data collected was from the Twitter account 'WeRateDogs', a humorous account which gives most dogs a rating above 10.

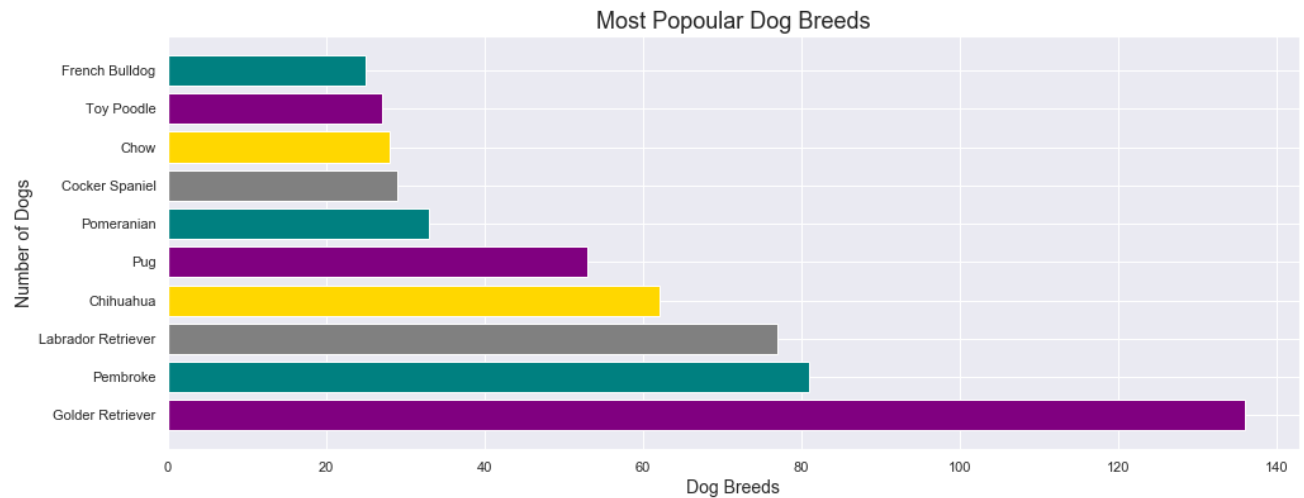
## Favourite Counts and Retweet Counts have a Strong Relationship

A strong relationship can be seen between a tweet's retweet counts and favourite counts since a tweet with a higher favourite count is also likely to have a high retweet count, this intuitively makes sense. Plotting these two values against each other produced the plot below, and it clearly shows a high correlation with a well-fitted, straight line passing through it.



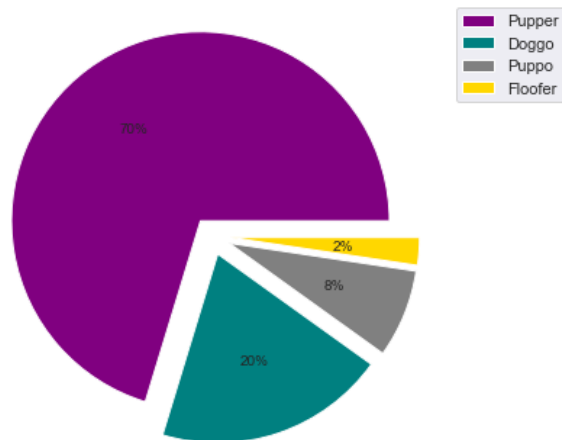
## Popular Dog Breeds

Grouping all the tweets by the dog breed showed that the top 3 breeds of dogs with the most number of tweets were golden retrievers, pembrokes and labrador retrievers.



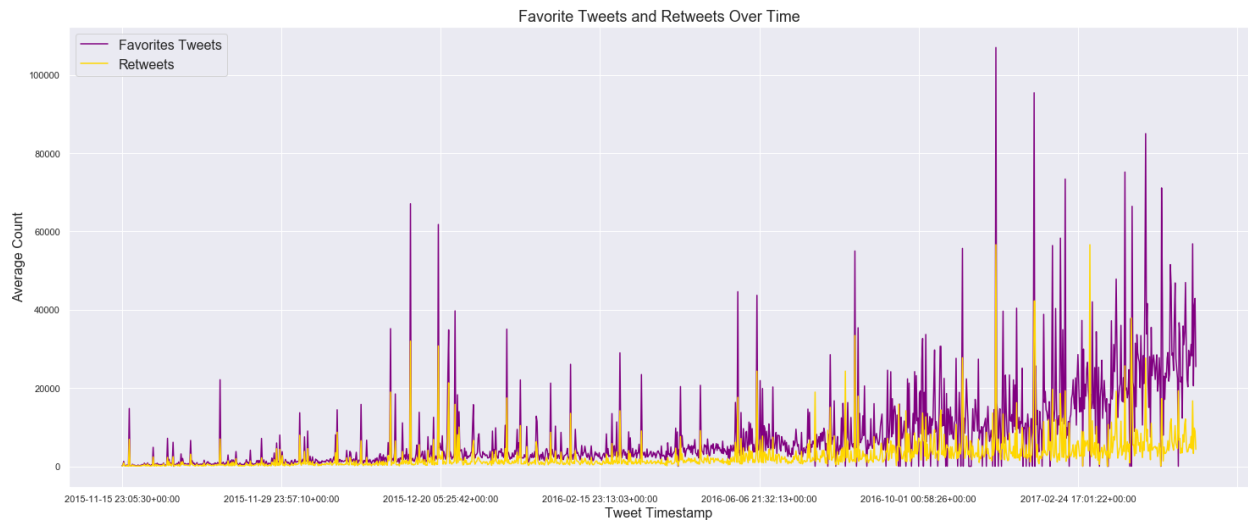
## Dog Life Stages

There are four main dog life stages that some of the dogs in the dataset were divided into. From the tweets it can be analyzed that puppies were the most common type, occurring in 70% of the dataset. This was followed by doggo, puppo and floofer.



## Tweet Trends Over Time

In the beginning, the favourite counts and the retweet counts are relatively lower and at a similar level. After 2016 and during 2017, the number of retweets per time seems to increase at a much higher level than the number of favourites. However, both generally follow an increasing trend overtime.



## Most Retweeted and Favorited Dog

The dataset was used to identify the most retweeted about and favorited dog, by analyzing the highest number of retweets and favorites. This turned out to be Stephan who was retweeted over 56000 times and favorited over 107000 times!

