# Analysis and Insights Report

Description:

WeRateDogs is a twitter handle that posts about pictures and tweets about dogs along with a rating out of ten. This handle currently (June 2020) has around 8.7M followers and has a global base of followers.

After loading, assessing, and cleaning the individual datasets obtained from different sources, they are merged into a single dataset and are used for the following analysis.

Insights and Visualizations:

1) *What is the year-wise distribution of tweets posted by the twitter account?*

To find out this insight, first, I applied the group by function on the year extracted from the timestamp column and based on that, calculated the count for each year. For the year 2015, 587 tweets were posted; in 2016 the maximum of 884 tweets was posted and in 2017, 338 tweets were posted. It is represented using a pie chart as follows.
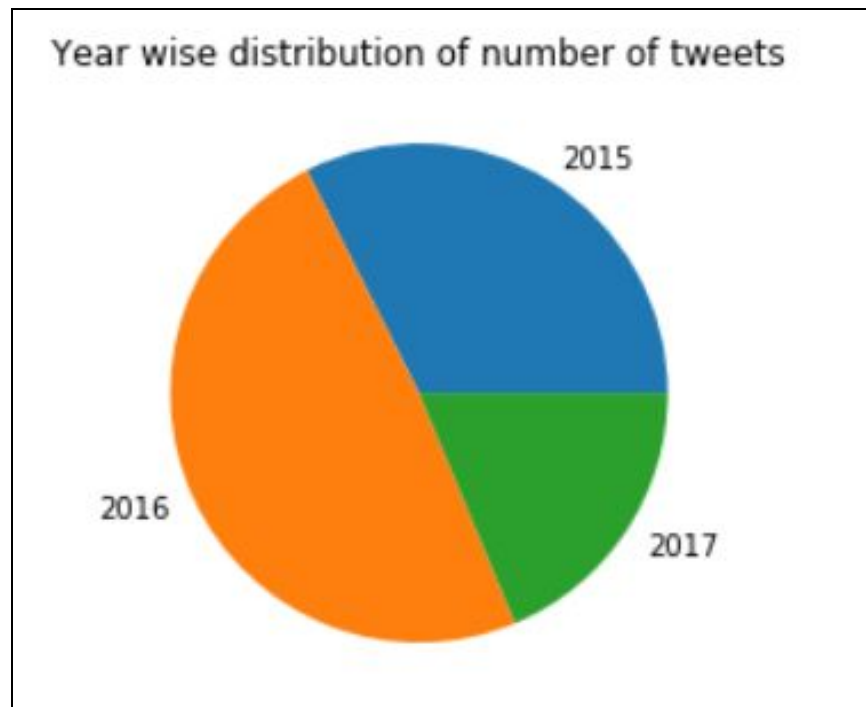
2) How the dogs in the dataset can be categorized based on the dog stage?

In the dataset, the dog stages are divided into four categories namely, Pupper, Doggo, Puppo, and Floofer. After applying the value counts function on the dog stage attribute, we get the following results; there are 188 'pupper' dogs, 57 'doggo' dogs, 21 'puppo' dogs, and 7 'floofer' dogs, 6 'doggo, pupper' dogs and 1 each in 'doggo, puppo' and 'doggo, floofer' stages. It can be substantiated using visualization as follows:
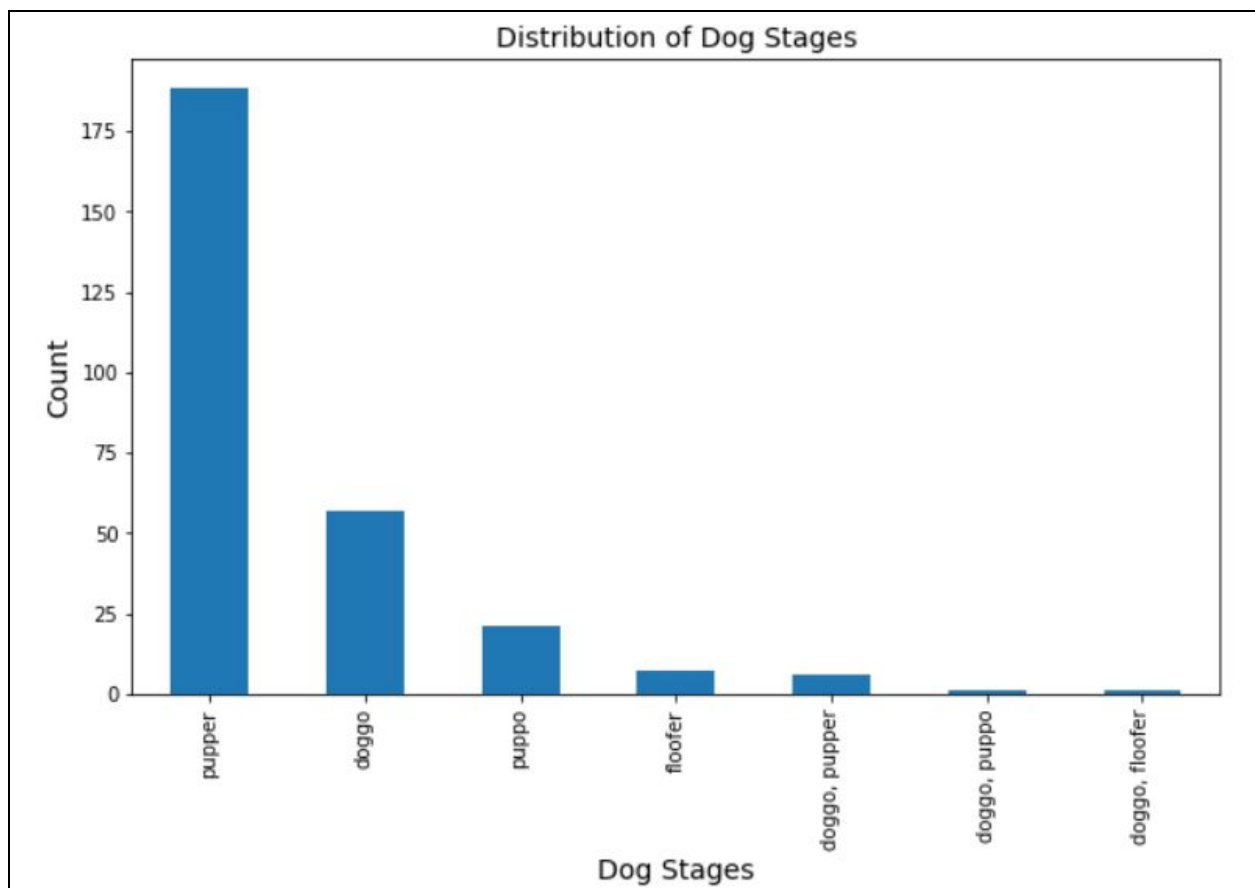
However, this insight could be further improved with the dog_stage data is available for all tuples of the dataset.

3) What is the comparison between mean favorite tweets with respect to correct and incorrect predictions in '1st_prediction_isdog'?

In the dataset, if the first dog prediction is True, then the average of the tweets being marked as favorites is 8959 whereas if the prediction is False, then the average is 7105. It indicates a trend that if the prediction was correct, then the tweet being marked as a favorite has a higher chance.