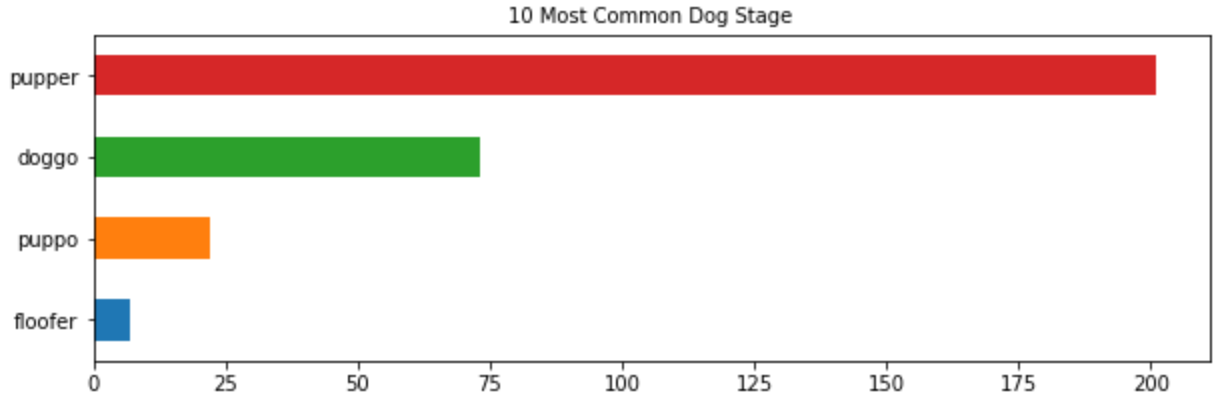
Introduction:

This is the act report for the ‘we rate dogs’ project in udacity’s data analysis professional course, the projects use data from a twitter account that rates dogs based on their pictures, the act report aims to document insights found through visualisation and Here are the 3 insights that i found after gathering, assessing, cleaning and visualsing the data

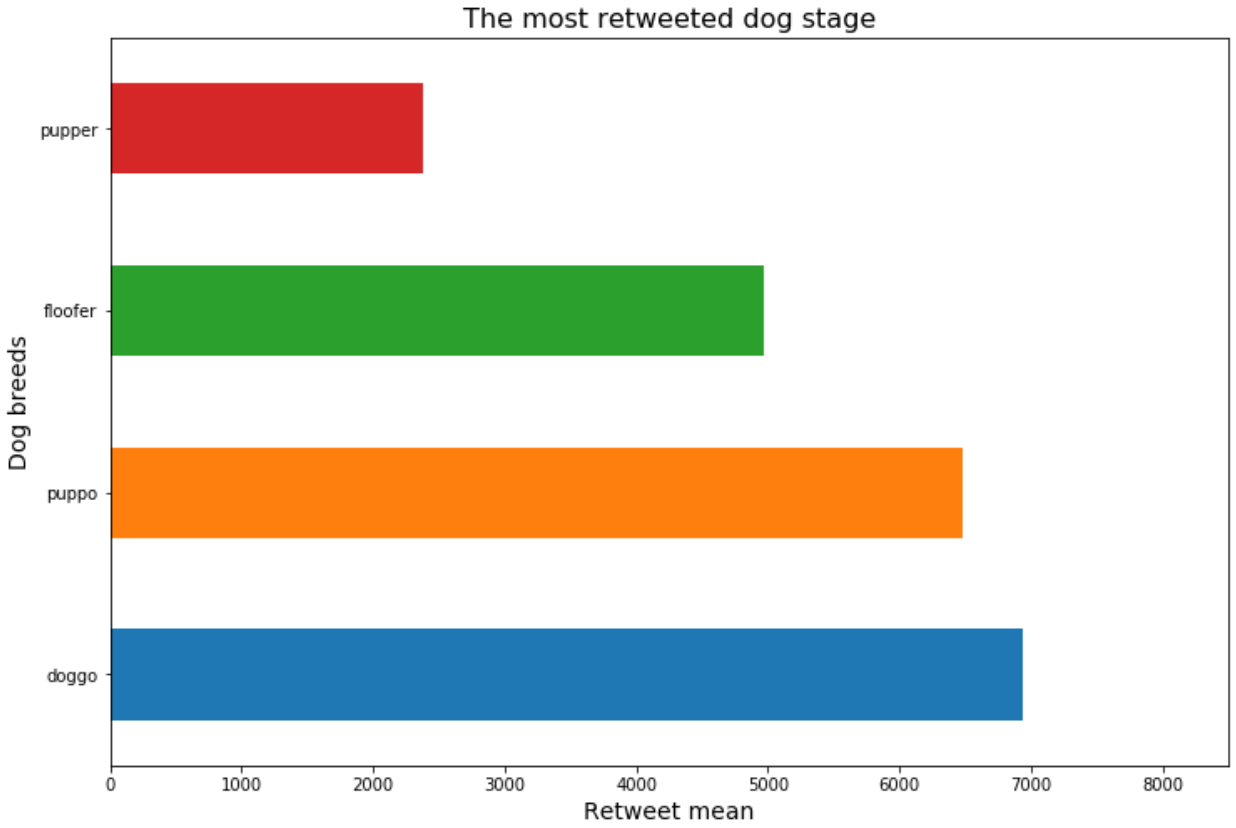
1. The Most Common dog stage:



After analyzing 1971 observations got from the wrangled data process and plotting it, it was found that the most common posted dog stage was ‘Pupper’ and the least common was ‘floofer’

2. The most retweeted dog stage:

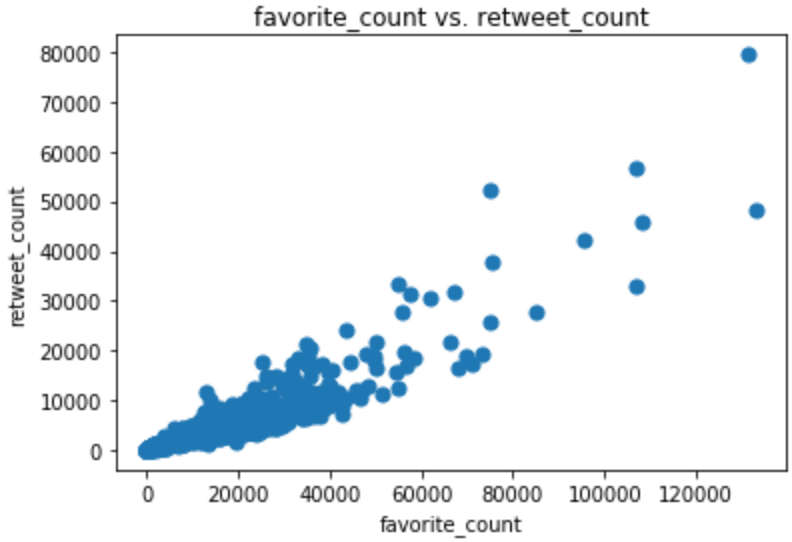
Twitter retweets indicates how many users loved the dog so it make great sense to use it to find out which dog stage people loved more



After plotting the data on a horizontal bar graph, it was found the the ‘doggo’ stage was the most retweeted also even though it was the most posted dog stage ‘pupper’ was the least retweeted stage

3. Retweets Vs Favorites Count:

There are two ways to express in your interest in a tweet, you either retweet or favorite(like) it So are the two correlated? If they're not we can't depend on the Retweet numbers only to decide which breed is the most popular so i decided to make a scatter plot to find the relation between them



Turns out they are Strongly Correlated with a value 0.91 which validates the 2nd graph

Conclusion:

Visualsing the data give the analyst a chance to discover interesting new features or patterns and discover relationships between the data and help him predict the data behaviour and makes it easier to explain to other people.