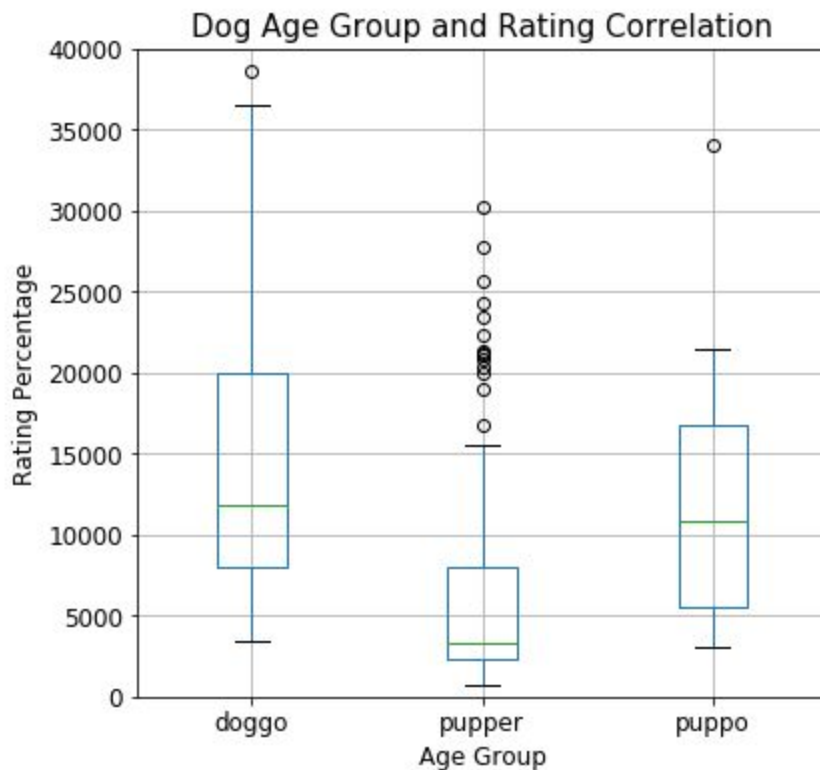


Act Report

By Steven Coleman

Udacity Data Analyst Nano Degree

Visual Insight:



The above visualization was created using pandas built-in plotting functionality and matplotlib to examine the correlation between a dog's age group and the rating percentage given to it by the Twitter community. A boxplot was chosen because it provided a better visual understanding than a scatter plot, and it does a good job showing outliers. As we can see, dogs tend to get higher rating percentages with older age.

Insight 1:

Insight one tackles the topic of what breed of dog tends to get the most favorites from the Twitter community based on p1 breed predictions from the predictions_clean dataframe. By joining information from the archive_clean and predictions_clean dataframes, it appears that golden retrievers claim that honor with 1,210,855 favorites.

Insight 2:

Insight two answers the question of whether the most popular dog names have higher numbers of retweets. Retweet by name information was stored in a python dictionary where a dog's name served as the key, and the sum total of a dog's retweets acted as the value. Afterwards this information was showcased in two print statements. Based on what can be gathered, it appears that posts for dogs with the most common name did not amass the largest number of retweets. That honor belongs to dog's name Bo (six dogs total) with a combined 69,914 retweets.

Insight 3:

Insight three pertains to media sources used to make posts to Twitter and the average rating score given by people using these sources. Based on the information, iPhone is the most popular method to make these posts. Additionally, those using the TweetDeck source give the highest average score to dogs which is 2,071.