

Act Report

After all the datasets were combined into one dataframe and cleaned, it was saved as a csv file.

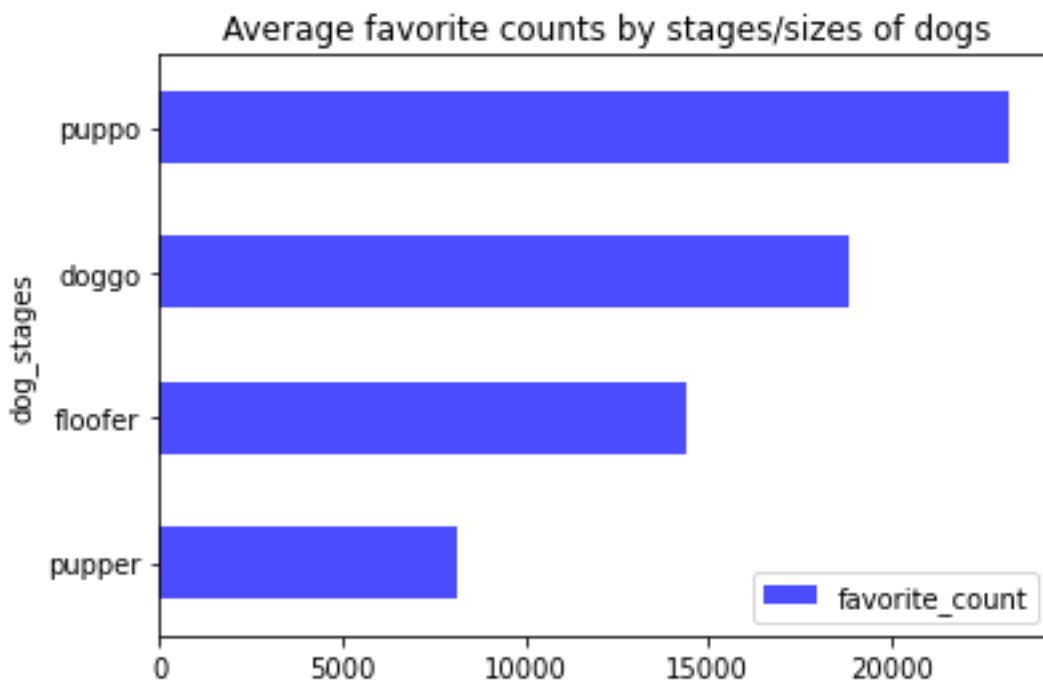
I read in the file into the notebook using the pandas read_csv function in order to perform some analysis and some visualization of the data that can give me some more insights about the data.

Some questions were asked.

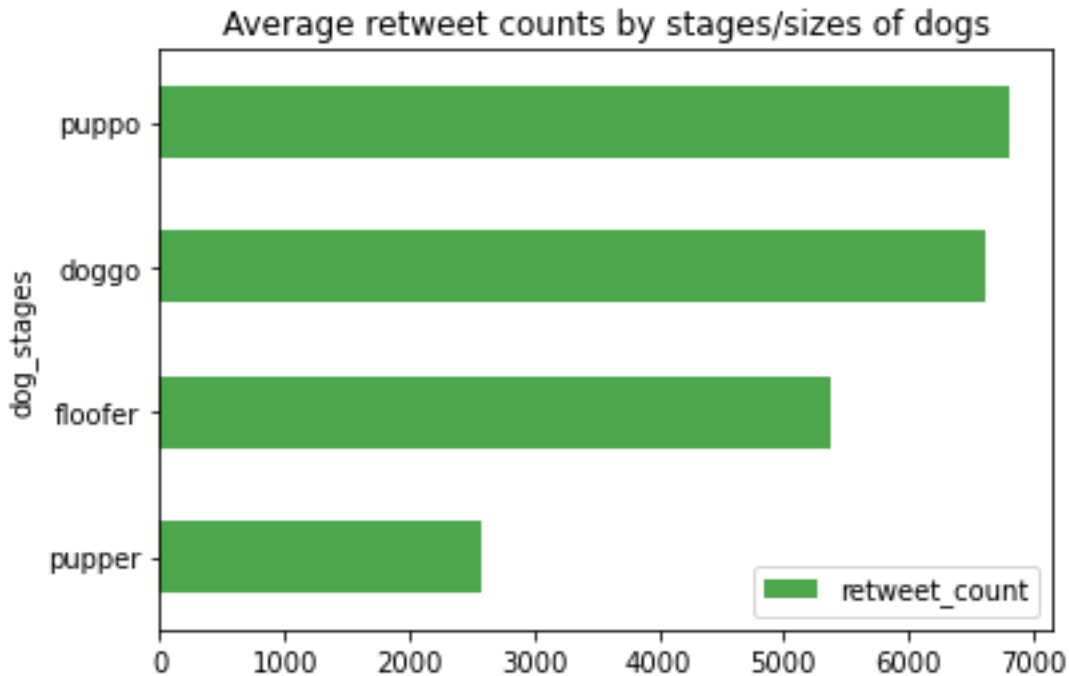
Question 1: Which stage or size of dogs got the highest retweet counts and favorite counts?

To answer this question, rows with invalid dog stages/sizes (i.e. null) were excluded.

Below is a visualization of a barh plot that shows the average favorite counts by stages/sizes of dogs:



The next visualization is a plot of the average retweet counts by stages/sizes of dogs:



The two plots show that the stage/size “puppo” got the highest favorite and retweet counts.

Question 2: The next question asked was “What are the three top dog names?”

Using, the pandas value_counts function, the names Coper, Charlie, and Oliver top the list i.e, they are the three top dog names according to my analysis.

Question 3: I was also interested in the breed of dog that is the most popular.

Also using the pandas value_counts function, the breed of dog most used in tweets is golden_retriever. The sum of the fvorite count for each breed also shows golden_retriever as topping the list, followed by Pembroke, and Labrador_retriever. However, the mean of the favorite_count shows otherwise. It suggests the three most popular breeds among the followers are Irish_water_spaniel, Saluki, and standard_poodle.