5 October 2020

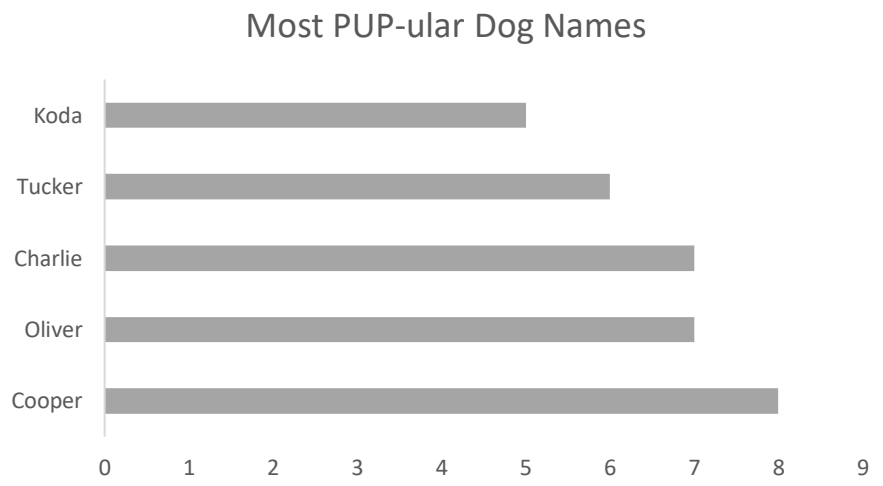
Dogs are man's best friend, that's a fact. But…how great are they? In this project, I analyzed the WeRateDogs Twitter data to gather three key insights into how great dogs are. Additionally, I created a dedicated visual to further emphasize how great dogs are.

Insight 1

Questions
What is the most PUP-ular dog name? What names make up the top 5?

Method
To do this, I created a series using the Pandas 'value_counts()' function on the 'name' column of the dataset. This returned a series with the name as one column and the count of that name in another column. Using Excel, I graphed the top 5.

Insight
The most PUP-ular dog name in the dataset is "Cooper"! And the top five, with their counts in parentheses, is as follows: Cooper (8), Oliver (7), Charlie (7), Tucker (6), and Koda (6). For reference, there were roughly 643 unique dog names!



Insight 2

Questions
What is the average rating for the dogs? What is the maximum rating? What about the minimum rating?

Method
To do this, I used the Pandas 'describe()' function on the 'rating_numerator' column. This showed me descriptive numerical statistics on that column.

Insight
The average rating is 10.97/10. The maximum dog rating in the dataset is 14/10. The minimum rating is 2/10 ☹.

Insight 3

Questions

What dog or dogs had the most total interactions as measured by retweets + favorites? What about just favorites? What about just retweets?

Method
To do this, I first created a column titled 'total_interactions' that was the sum of 'retweet_count' and 'favorite_count'. Then, I used the Pandas 'max()' functions to filter the dataset where the respective count equaled the max of that count. I then printed these counts, the dog's/dogs' name, and the rating for the dog/dogs.

Insight
The most PUP-ular dog for ALL three metrics is…Stephan! Stephan had the most total interactions, the most retweets, and the most favorites. Stephan was rated a 13/10!

Visualization 1

Goal
The goal of this visualization is to plot the frequency of ratings in the clean dataset. Here we can see what the most common rating was, as well as spread of the frequency.

Method
To do this, I used Pandas and Matplotlib functionality. Specifically, I used the 'df.plot.hist()' function.

Result
The most common dog rating in the clean dataset is 12/10. There are some ratings below 10 ☹, however, the ratings are largely skewed towards 10 and above. This falls in line with **Insight 2** as the average rating for the clean dataset was above 10.