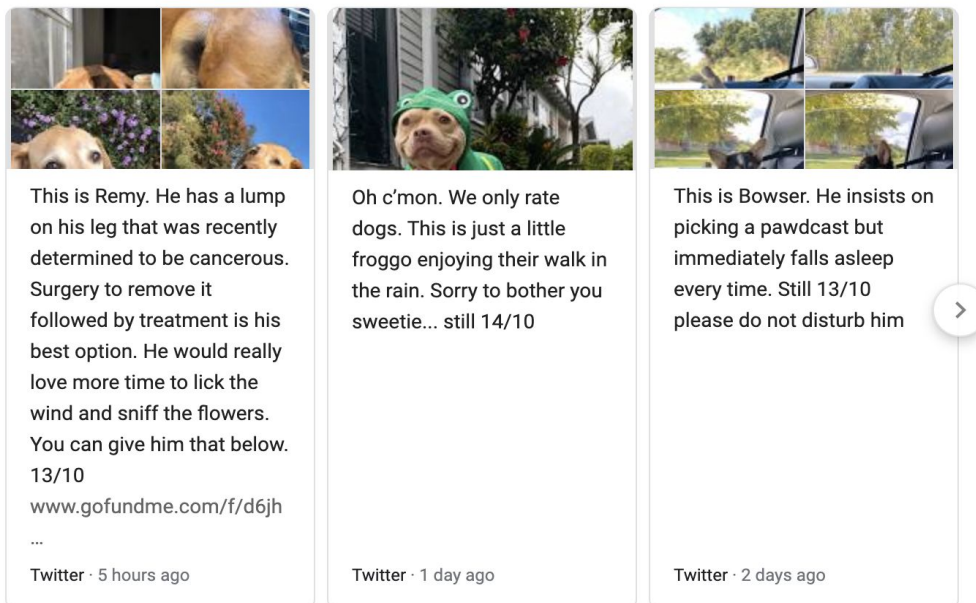# WeRateDogs Data Analysis

Synopsis: *This report is a brief analysis of WeRateDogs' Tweets. It provides the insights, analysis, visualizations that were produced from my wrangled data.*

WeRateDogs® (@dog_rates) is a Twitter account that has 8.9 million followers as of April 2020. Tweeters rate other people's dogs with humorous comments about the dog. Although ratings almost always have a denominator of 10, numerators however, can be greater than 10 because they are just "really good dogs".
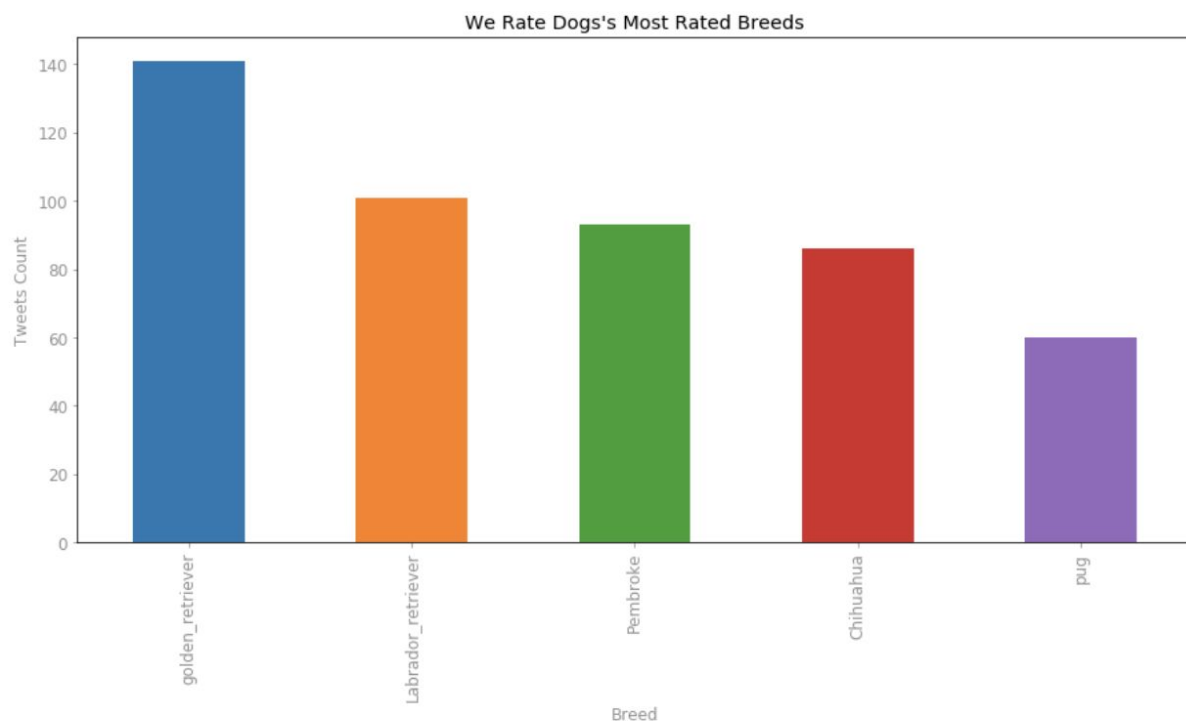

Data wrangling consists of assessing and cleaning three DataFrames (WeRateDogs Twitter Archive, Image Predictions File, and Twitter API), and they were then merged and stored as a single DataFrame. Having one enhanced DataFrame allowed me to analyze and visualize my wrangle data appropriately.

My data wrangling was done on Jupyter Notebook and I aimed to answer the following questions.

1. Which breeds have the most ratings on WeRateDogs Twitter account ?
2. Which dog stage is the most represented?
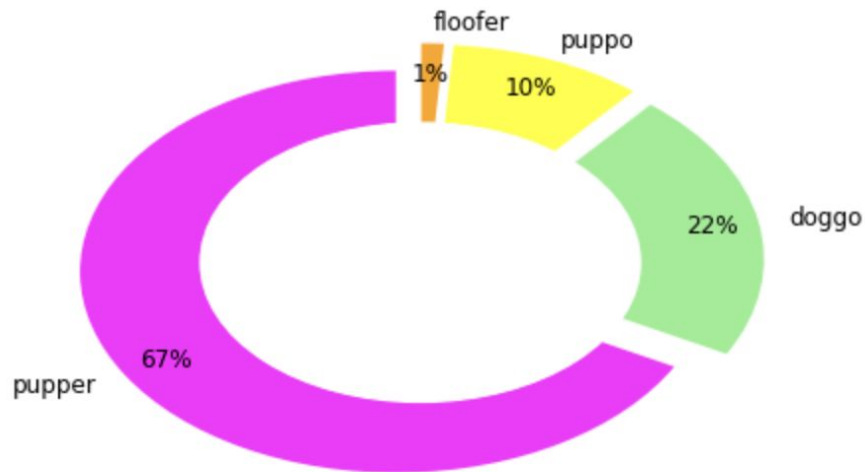3. Which breeds are most favorited by users?

Insights and Analysis

Insight 1: Which breeds have the most ratings on WeRateDogs Twitter account?
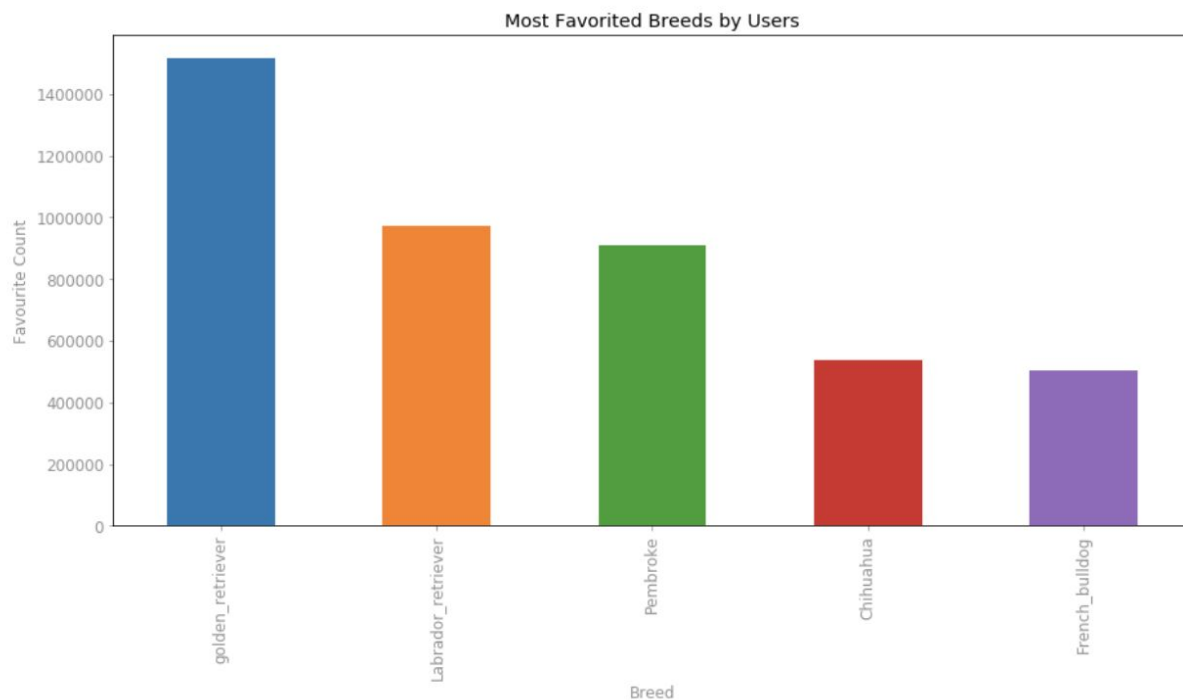

Looking at tweets count, the most rated breeds on WeLoveDogs Twitter account are Golden Retriever, Labrador Retriever, Pembroke, Chihuahua, and Pug. This was somewhat expected as they are the breeds that we see often or even have them as our four legged friends.

Insight 2: Which dog stage is the most represented?

Looking at the numbers of each breed represented via tweets, we see that "pupper" is the most represented with its frequency being 67%. In contrast, "floofer" frequency is 1% which is significantly low compared to the others. This shows that there may be data extraction issues.

Insight 3: Which breeds are most favorited by users?

As expected, the breed with the highest favorite counts is Golden Retriever then followed by Labrador Retriever, Pembroke, Chihuahua. I expected to see Pug at the end but French Bulldog was the 5th most favorite breed.

Conclusion

Data wrangling is a critical component and there's multiple steps that are required to help us analyze our data. It allows us to make trustworthy analyses and visualizations. Furthermore, having a well curated dataset is a necessity that helps us answer our questions so we can also be more creative with our inquisitive nature.