

DATA ANALYSIS AND VISUALIZATION (EXTERNAL DOCUMENT)

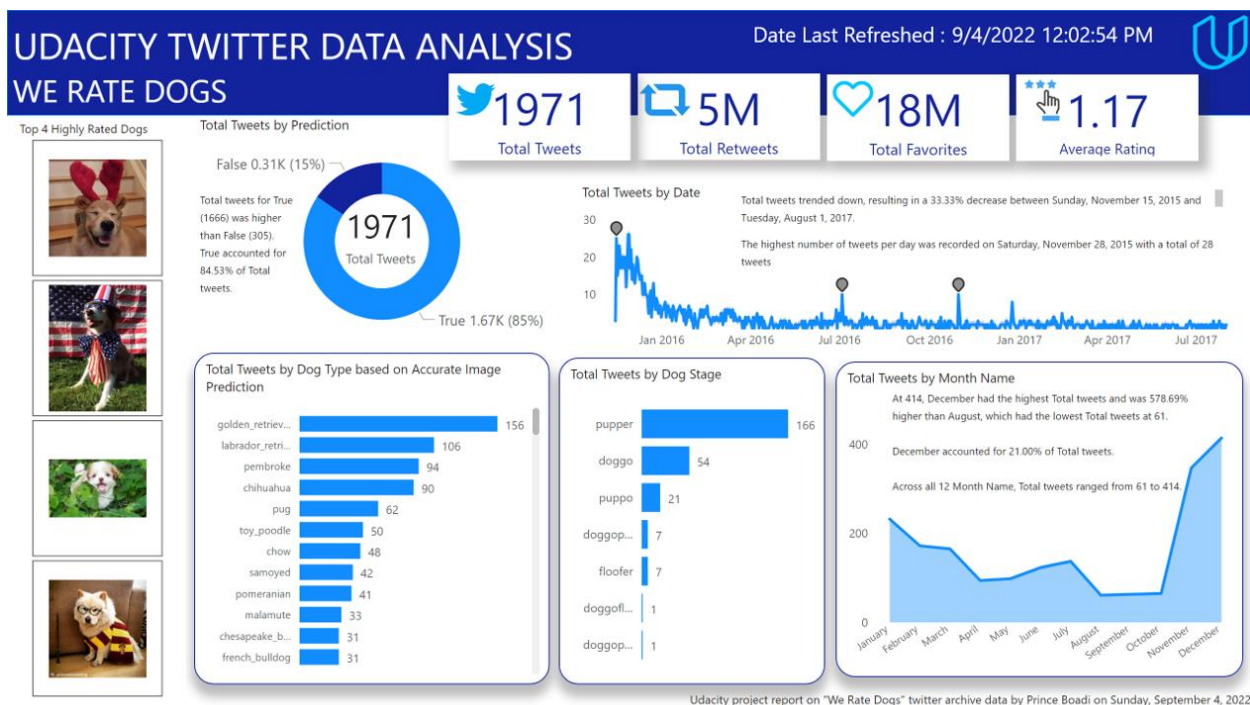
INSIGHTS AND VISUALIZATIONS FROM “WE RATE DOG” TWITTER DATA

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

After the wrangle process, in order to understand the dataset and reveal its hidden insights through visualizations, I asked the following questions:

- Is there a correlation among rating, favorite count and retweet count?
- Which dog image received the most favorite count and which dog type is it?
- Which dog image received the most retweet count and which dog type is it?
- What are the top 4 highly rated dogs?
- Which dog types [top 3] were correctly predicted the most?
- What is the proportion of correctly predicted images?
- What do our raters say in their tweets?
- When do our raters tweet?

To begin with, the image below has its interactive version embedded in the jupyter notebook.

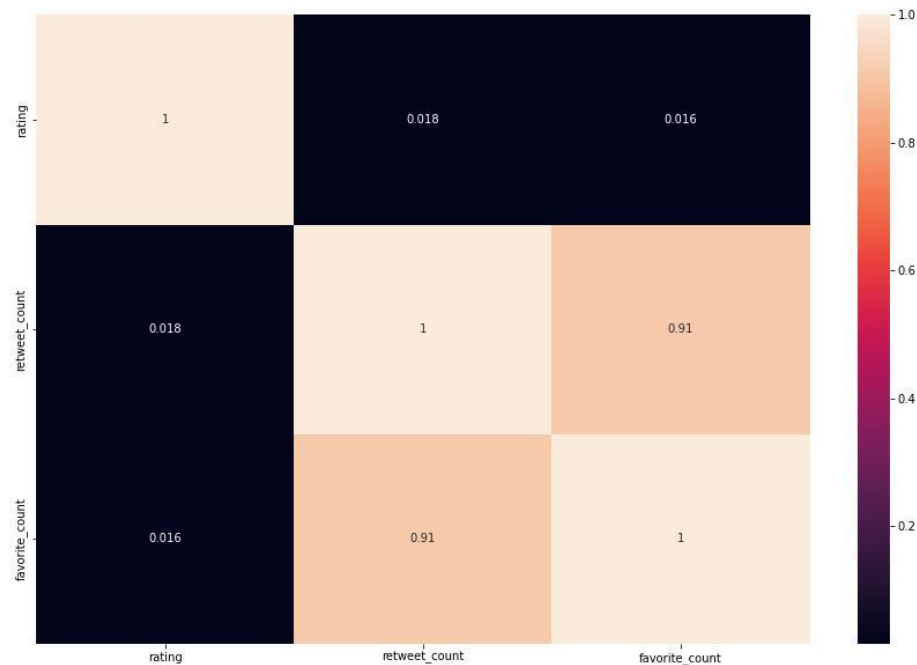


I created this dashboard with the merged dataset of all the cleaned datasets to give a complete overview of the data and summarize its insights. Key insights among these are:

- ✓ Golden Retriever, Labrador Retriever and Pembroke are the top 3 most tweeted dogs

- ✓ About 85% of the image predictions were accurate
- ✓ Most tweets are done in the months of November and December
- ✓ Average rating is 1.17 (rating is normalized by dividing rating numerators by their rating denominators)
- ✓ Among the dog stages, pupper is the most tweeted.

A heatmap on correlation between the variables



- ✓ There is a very strong positive correlation between retweet count and favorite count with a correlation coefficient of 0.91
- ✓ There is a very weak or no correlation between rating and favorite count with a correlation coefficient of 0.016
- ✓ There is a very weak or no correlation between rating and retweet count with a correlation coefficient of 0.018

A wordcloud on tweet contents



Dog image with the most retweet count is Labrado Retriever (doggo)



Dog image with the most retweet count is Labrador Retriever (doggo)

Dog image with the most favorite count is Lakeland Terrier (puppo)



Golden Retriever, the most correctly predicted dog image, perhaps the most popular



Dog image with the highest rating

