DATA ANALYSIS AND VISUALIZATION REPORT ON WERATEDOG DATASET

INTRODUCTION

This report documents the data analysis and visualization of the WeRageDog dataset collected from Twitter using Twitter API.

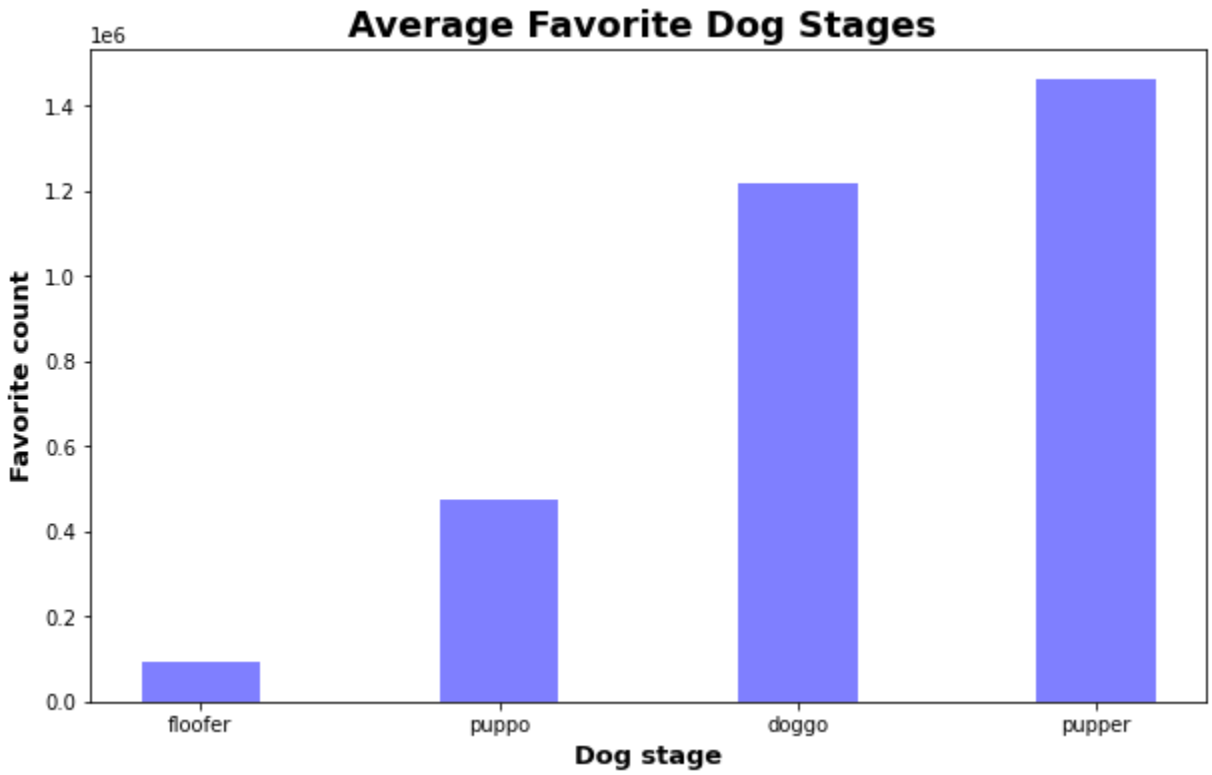
After performing the data wrangling, I organized the data to answer the following research questions:

1. Which type of dog got the highest favorite count?
2. What are the first 3 dogs with the highest rating numerator score?
3. Which type of dog got the highest retweet count?

1. Which type of dog got the highest favorite count?

To begin, I used the groupby function on the dog_stage and the favorite_count columns to get the sum of the dogs. Below is the result obtained from the analysis:

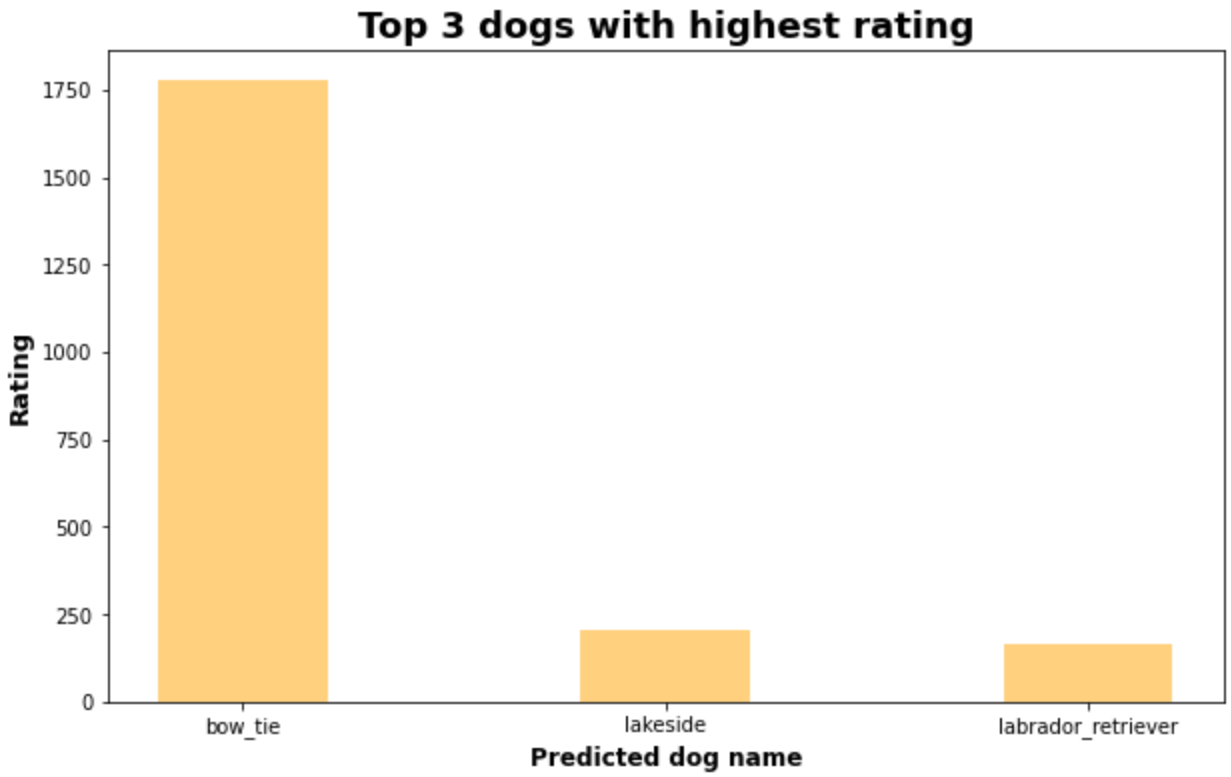
The data is better visualized and the result is shown below:



From the chart above, pupper has the highest average favorite stage, followed by doggo, puppo and floofer.

2. What are the first 3 dogs with the highest rating numerator score?

In order to get the highest rating numerator score, I sorted the dataset in ascending order and visualized it to obtain the first three dogs. Below is the visualization I obtained from working on top three dogs with highest rating.



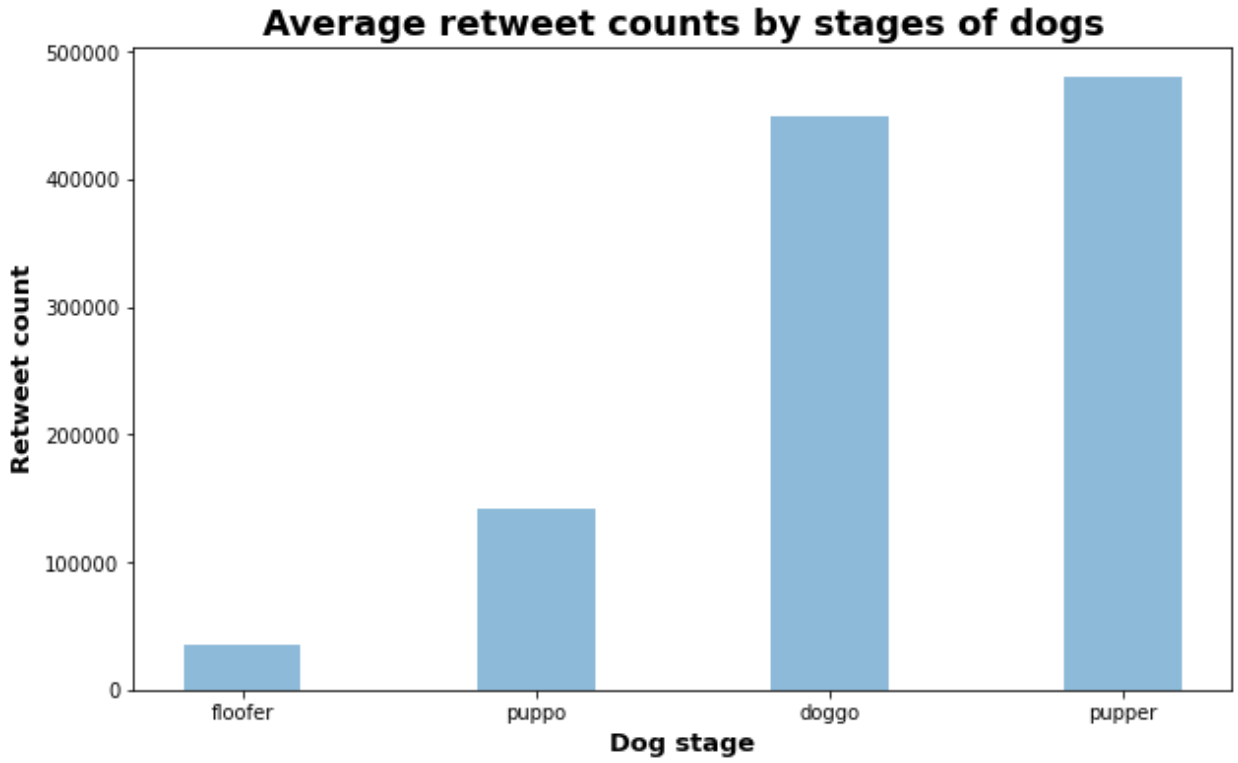
From the chart above, we can see that the image with a predicted name of bow_tie has the highest numerator rating followed by lakeside and finally labrador_retriever.

3. Which type of dog got the highest retweet count?

Last, I looked at the dog_stage and retweet_count columns in the cleaned dataset, and I used the groupby function in pandas on the dog_stag column and also called the sum function on it.

```
dog_stage
floofer    34781
puppo      142427
doggo      448919
pupper     479807
Name: retweet_count, dtype: int64
```

The data is better visualized and the result is shown below:



From the chart above, pupper has the highest retweet count, followed by doggo, puppo, and floofer.