Project WeRateDogs Analysis and Visualization Report


This is Cassi of breed doggo rated 14/10

Introduction

WeRateDogs is a Twitter account that rates people's dogs with humorous comments 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.

Project Steps Overview

The main task for this project was to:

1. Gather the data
2. Assess the data
3. Clean the data
4. Store the data either as a .csv or SQLite database
5. Analyzing, and visualizing data
6. Reporting

The focus for this document will be on step five (5) only, Analyzing, and visualizing data.

In this section, at least three (3) insights and one (1) visualization was needed to display our wrangle effort. First, the twitter_archive_master.csv was loaded into the Jupyter notebook.

For the insight, my interest was to check which breed of dog is the most liked, favorited and has the most retweet count compared to other breeds of dogs, also, to confirm the correlation and how confident the dog image algorithm is, hence the following insight:

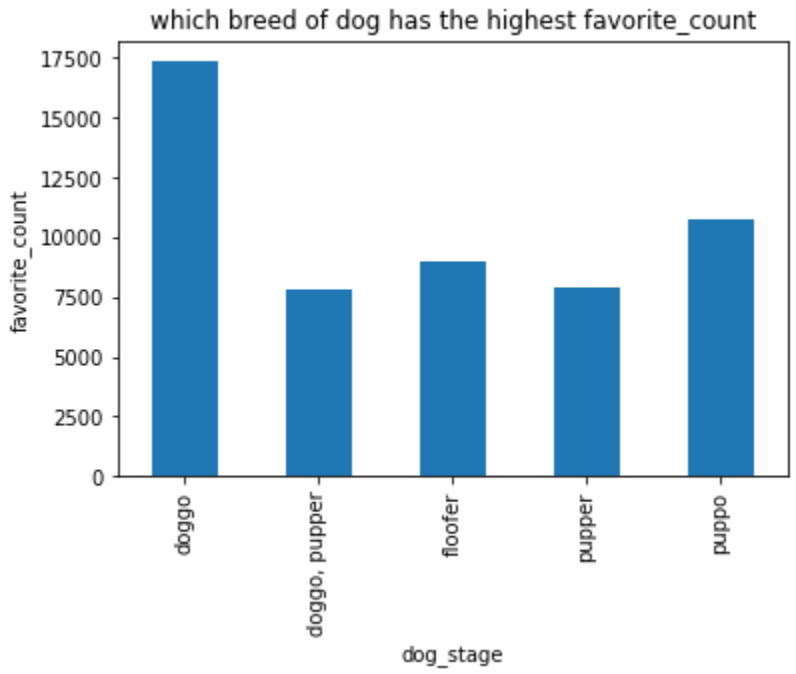
Insights:

1. Check which breed of dog has the highest favorite count
2. Check correlation between p1,p2 and p3
3. Check which breed of dog has the highest rating
4. Check which breed of dog has the highest retweet count

The Discussion

Check which breed of dog has the highest favorite_count

To confirm this, dog_stage and favorite_count columns were considered.



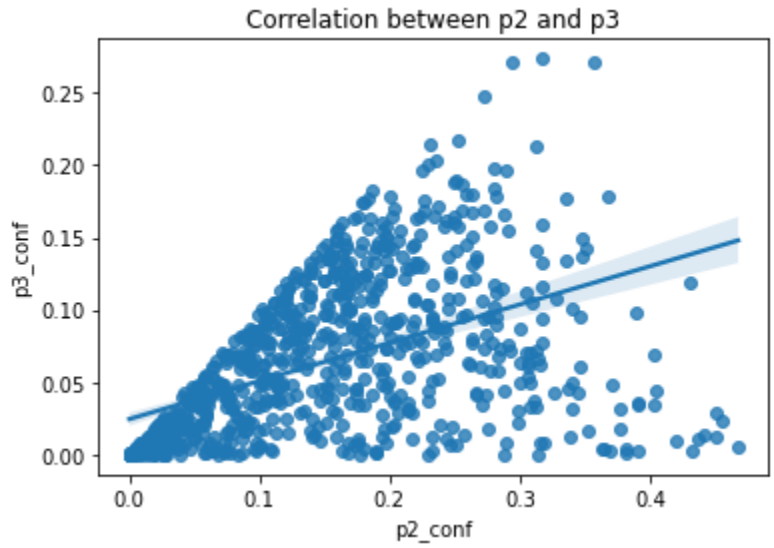
From the plot above, it is seen that dog breed doggo has the highest favorite count compared to other breeds of dog.

Check correlation between p1,p2 and p3

	p1_conf	p2_conf	p3_conf
p1_conf	1.000000	-0.609723	-0.756300
p2_conf	-0.609723	1.000000	0.504258
p3_conf	-0.756300	0.504258	1.000000

From the insight above, we can see that P2 shows more confident in the algorithm

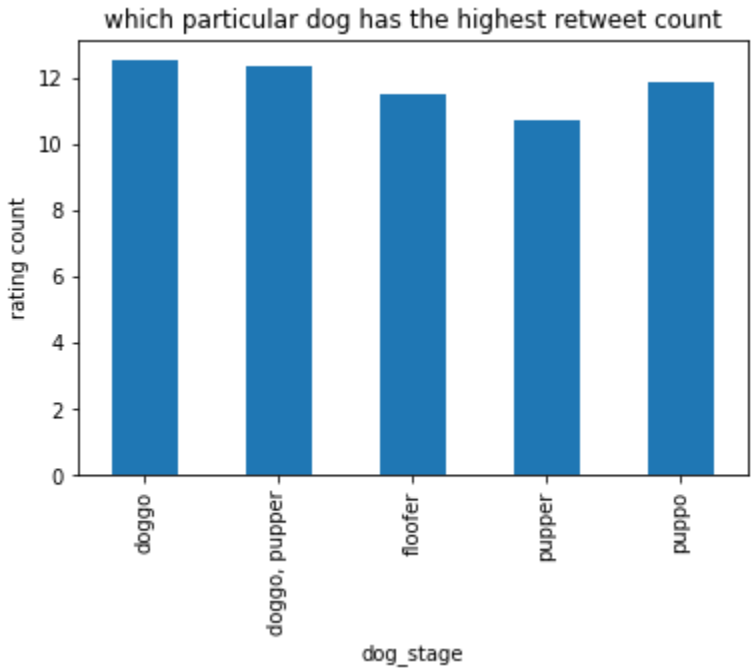
Visualization for insight 2, shows the correlation between p2 and p3



From the plot above, it is seen that there is a positive correlation between p2 algorithm and p3 algorithm

Check which breed of dog has the highest rating

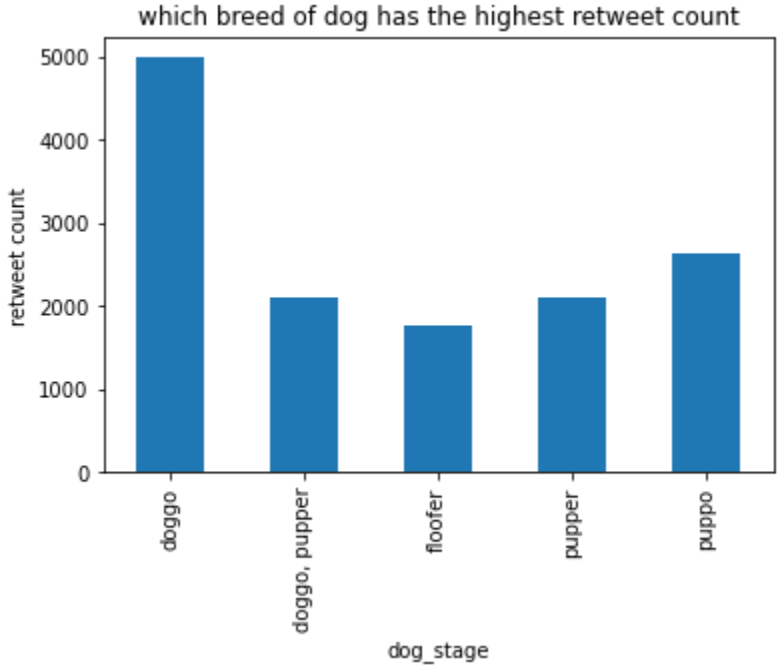
To confirm this, we will look into dog_stage and rating_numerator columns



The plot above shows that, doggo has the highest rating count compared to puppo with a slight difference of about +2 points

Check which breed of dog has the highest retweet count

To confirm this, we will look into dog_stage and retweet_count columns



The plot above shows that doggo has the highest retweet count compared to other breeds.

In Conclusion, it is seen that breed doggo is the most liked, favorited and has the most retweet count compared to other breed breeds of dogs