DATA ANALYSIS REPORT

This a brief report with some insights from twitter data WERATEDODS

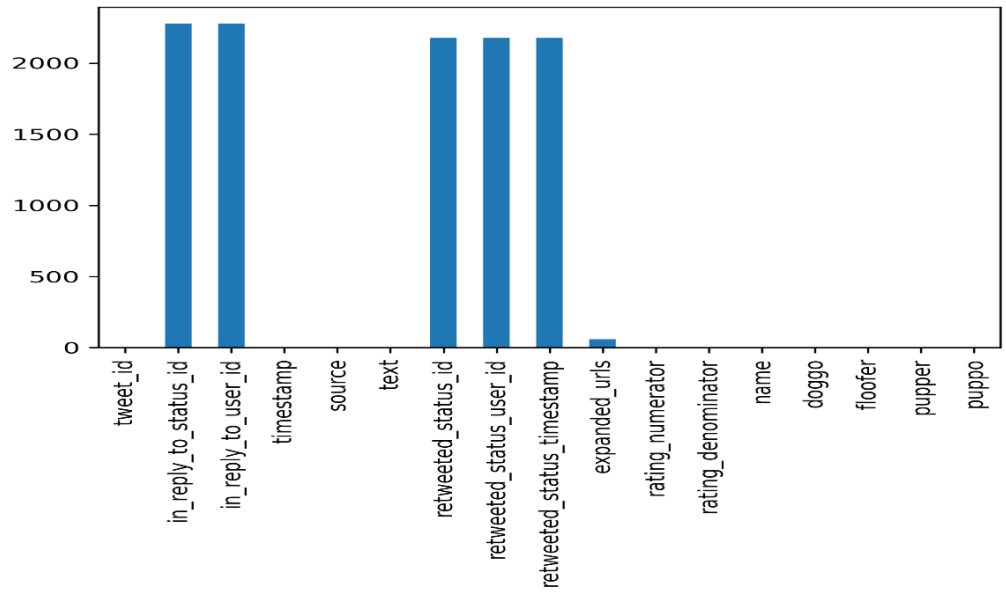
Nourhan Elaraby

Introduction:

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The account was started in 2015 by college student Matt Nelson, and has received international media attention both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter for breaking these aforementioned laws.

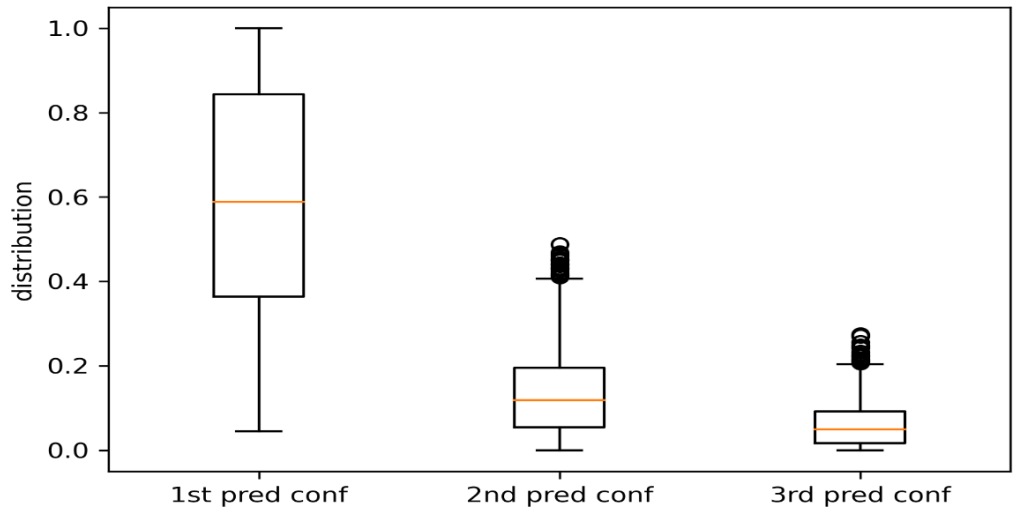
The dataset that I will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, 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.

I tried to gain some insights in the following pages

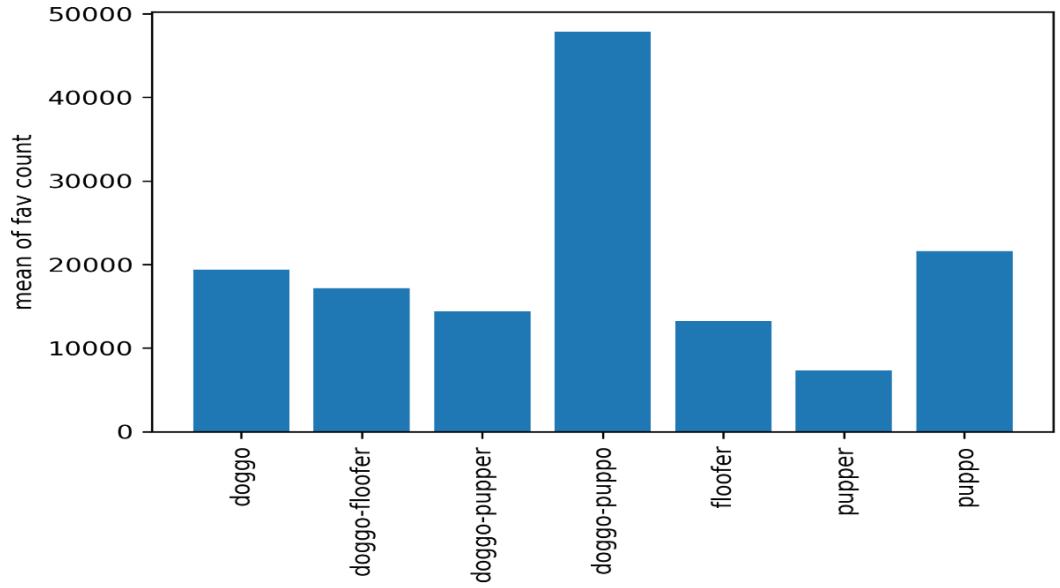


The high number of missing values(NaN) in

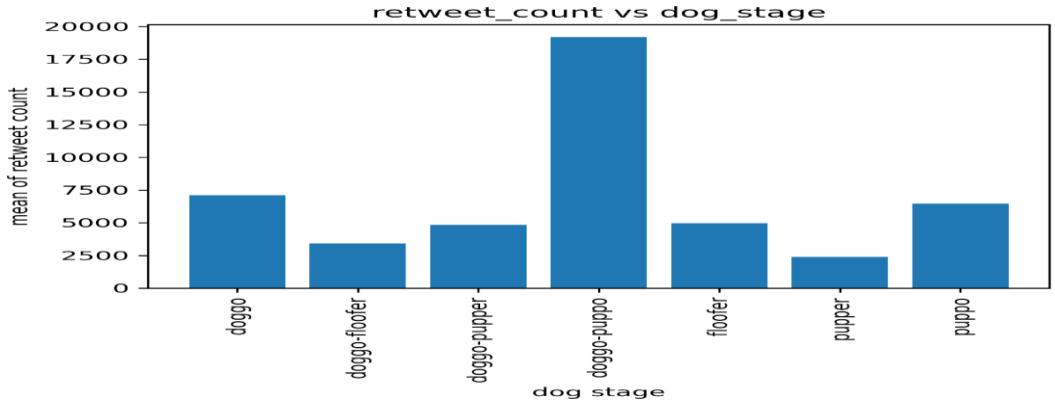
In_reply_status_id, In_reply_user_id,retweeted_status_id, retweeted_user_id, retweeted_status_timestamp indicates that most of our tweets are not retweets/replies



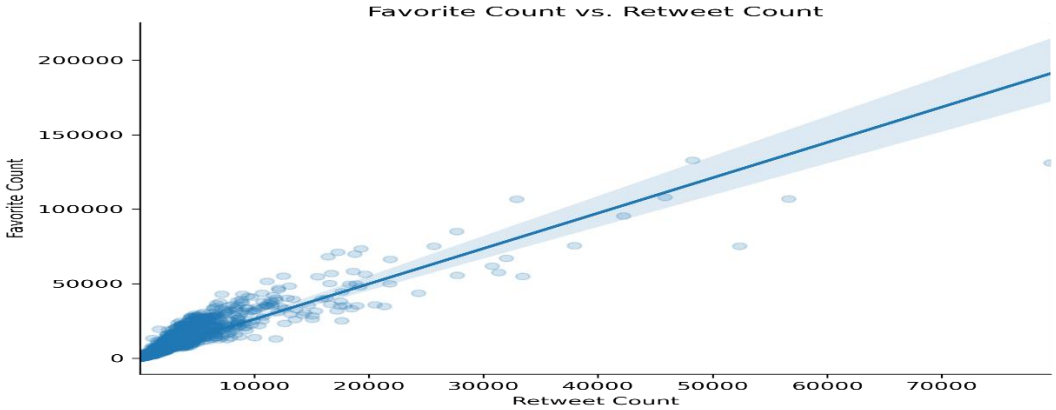
In image predictions data frame there are some outliers in second and third prediction confidence



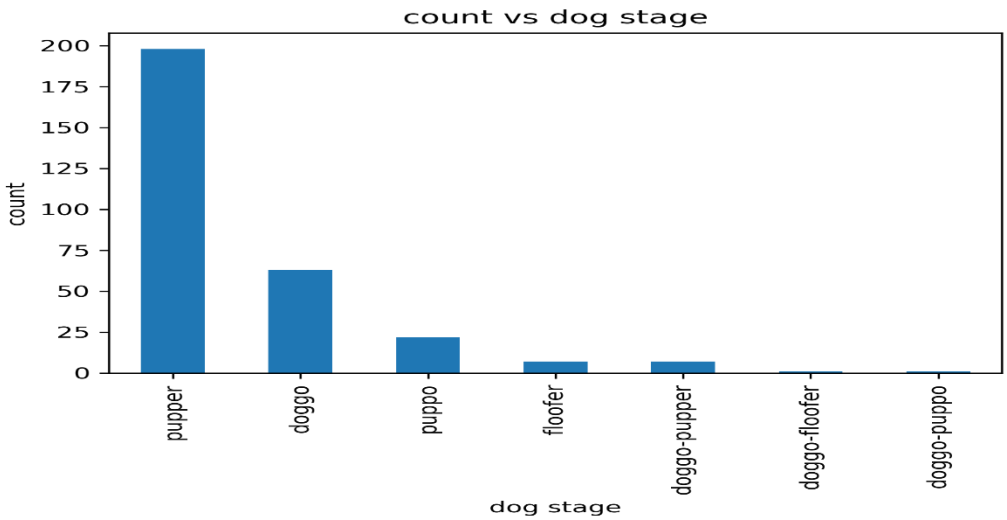
The images that are classified as doggo-puppo have the highest mean favourite count



The images that are classified as doggo-puppo have the highest mean retweet count And comes in second place doggo



By inspecting the scatter plot the favourite count is positively collated with retweet count which means by increase of favourite count retweet count increases.



The pupper has the highest count among all dogs