Report: act_report

- A report that communicates my insights and displays the visualizations produced from my wrangled data.

Introduction

This report is based on the data from twitter handle 'WeRateDogs'. WeRateDogs takes various dog's images with keen interest on dogs. They frown at any image other than dogs. Their rating is special because they often have higher numerator rating than the rating denominator. There are four dog stages puppo,floofer,pupper,doggo.

After cleaning, I proceeded to analyzing my data. From the analysis, I had five insights and three visualizations to explain some of my insights which are:

1. ***Text column containing 'We only rate dogs' suggest from the data that the tweets are either non-dogs or contains humans posing with dog***: From the text column of twitter archive, I discovered that texts containing 'We only rate dogs' are tweets that weren't accepted by the dog rating platform

2. ***There are tweets that are not dog's images. Some people made post of other images other than dogs*** : The reason why some of the tweets weren't accepted is because they are either tweets with non-dog images or people pose with their dogs. There could be other reason but these are some I discovered.

3. ***There are dogs that are flagged as non-dogs because their breed makes them look like other animals***: The dog rating twitter page didn't accept some tweets not because they are not dogs. Dog breeds like Samoyed look like bears, because of this some of them are unaccepted assuming them to be bears.

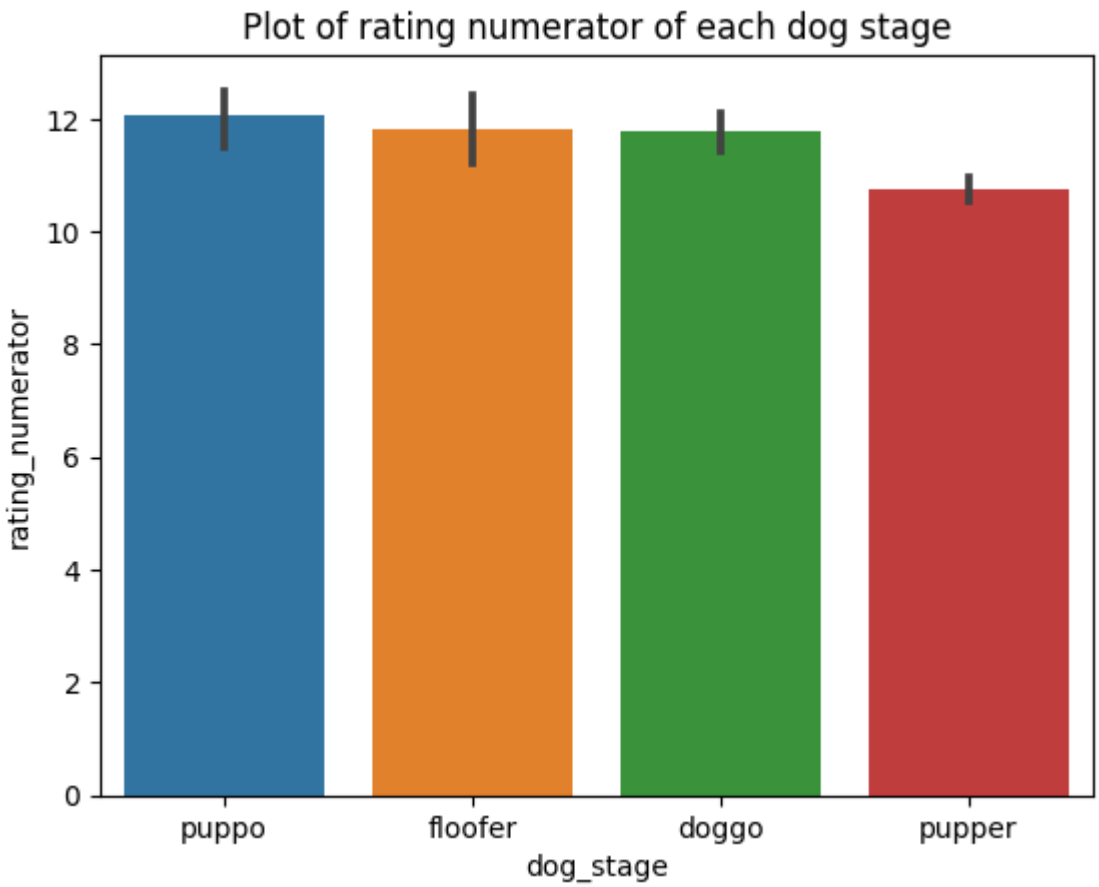
4. ***Dog stage (puppo) have the highest rating numerator**: Pupo has the highest rating among other dog stages, doggo is next and fluffer is last. The rating was measured using their mean (See Visuals 1)**

5. ***Dog breeds (Rotweiler) at dog stage (puppo) has the higest rating (rating_numerator)***: Using p1(dog breed prediction) and p1_conf (The confidence level of dog breed prediction). It was observed that dog breeds 'Rotweiler' at dog stage 'puppo' had the highest rating numerator.

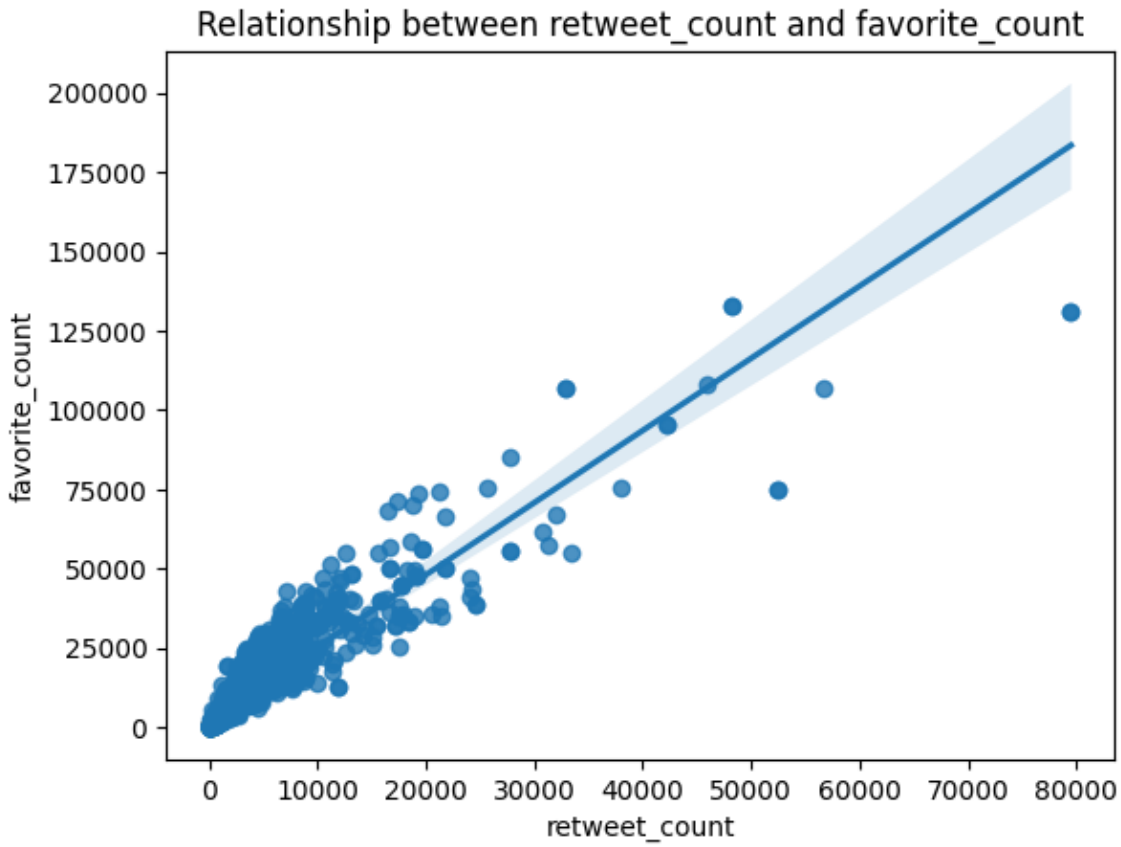
6. ***Favorite count increases as retweet count increases**: Both variables are positively correlated. Tweets with higher favorite counts have higher retweet count (See Visual 2)**

7. *__Visual 3 shows the correlation heatmap of all numerical variables.__* Favorite count and retweet count shows correlation of 0.9

Visual 1



Visuals 2


Relationship between retweet_count and favorite_count
favorite_count
200000
175000
150000
125000
100000
75000
50000
25000
0
0
10000
20000
30000
40000
50000
60000
70000
80000
retweet_count

Visual 3


Correlation heatmap of the dataset
tweet_id
rating_numerator
rating_denominator
id
retweet_count
favorite_count
img_num
p1_conf
p2_conf
p3_conf
1.0 0.0 -0.0 1.0 0.4 0.6 0.2 0.1 0.0 -0.1
0.0 1.0 0.2 0.0 0.0 0.0 0.0 -0.0 -0.0 -0.0
-0.0 0.2 1.0 -0.0 -0.0 -0.0 -0.0 -0.0 -0.0 0.0
1.0 0.0 -0.0 1.0 0.4 0.6 0.2 0.1 0.0 -0.1
0.4 0.0 -0.0 0.4 1.0 0.9 0.1 0.0 -0.0 -0.0
0.6 0.0 -0.0 0.6 0.9 1.0 0.1 0.1 -0.0 -0.1
0.2 0.0 -0.0 0.2 0.1 0.1 1.0 0.2 -0.2 -0.1
0.1 -0.0 -0.0 0.1 0.0 0.1 0.2 1.0 -0.5 -0.7
0.0 -0.0 -0.0 0.0 -0.0 -0.0 -0.2 -0.5 1.0 0.5
-0.1 -0.0 0.0 -0.1 -0.0 -0.1 -0.1 -0.7 0.5 1.0
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6