Act Report

I analyzed the datasets I wrangled as shared in the wrangle_report.pdf file.

For the analysis, I set out to find 4 insights:

1. which dog has the highest tweet engagement on the WeRateDogs twitter account?
2. which dog has the highest rating on the WeRateDogs twitter account?
3. which dog has the lowest rating on the WeRateDogs twitter account?
4. what is the most common rating on the WeRateDogs twitter account?

Before I answered these questions, I read the stored twitter_archive_master.csv and image_predictions_master.csv into their individual dataframe.

#1

In order to answer question 1, I created an engagement column which is the sum of the retweet_count column and favorite_count column in the twitter archive master dataframe.

Then using the .max() function, I checked for the observation/row with the largest engagement. My finding is outlined below:

The dog with the highest engagement rating wasn't named, has an engagement of 210,590 (favorites and retweets). The dog was filmed while inside a pool which probably explains the reason for the highest engagement. Also, the dog is a doggo.

The link to the tweet and video: https://twitter.com/dog_rates/status/744234799360020481?lang=en

#2

To answer question 2, 3 & 4, I created a rating column which is the rating_numerator column divided by the rating_denominator column.

Then using the .max() function, I checked for the observation/row with the highest rating. I also checked for the jpg_url in the image predictions master dataframe using the tweet_id.

My finding is outlined below:

The dog with the highest rating is named Atticus and had a rating of 177.6 (1776/10). The dog was pictured donning the US flag inspired costume and the background has a US flag. This probably explain the reason for the highest rating.

jpg_url: https://pbs.twimg.com/media/CmgBZ7kWcAAlzFD.jpg

#3

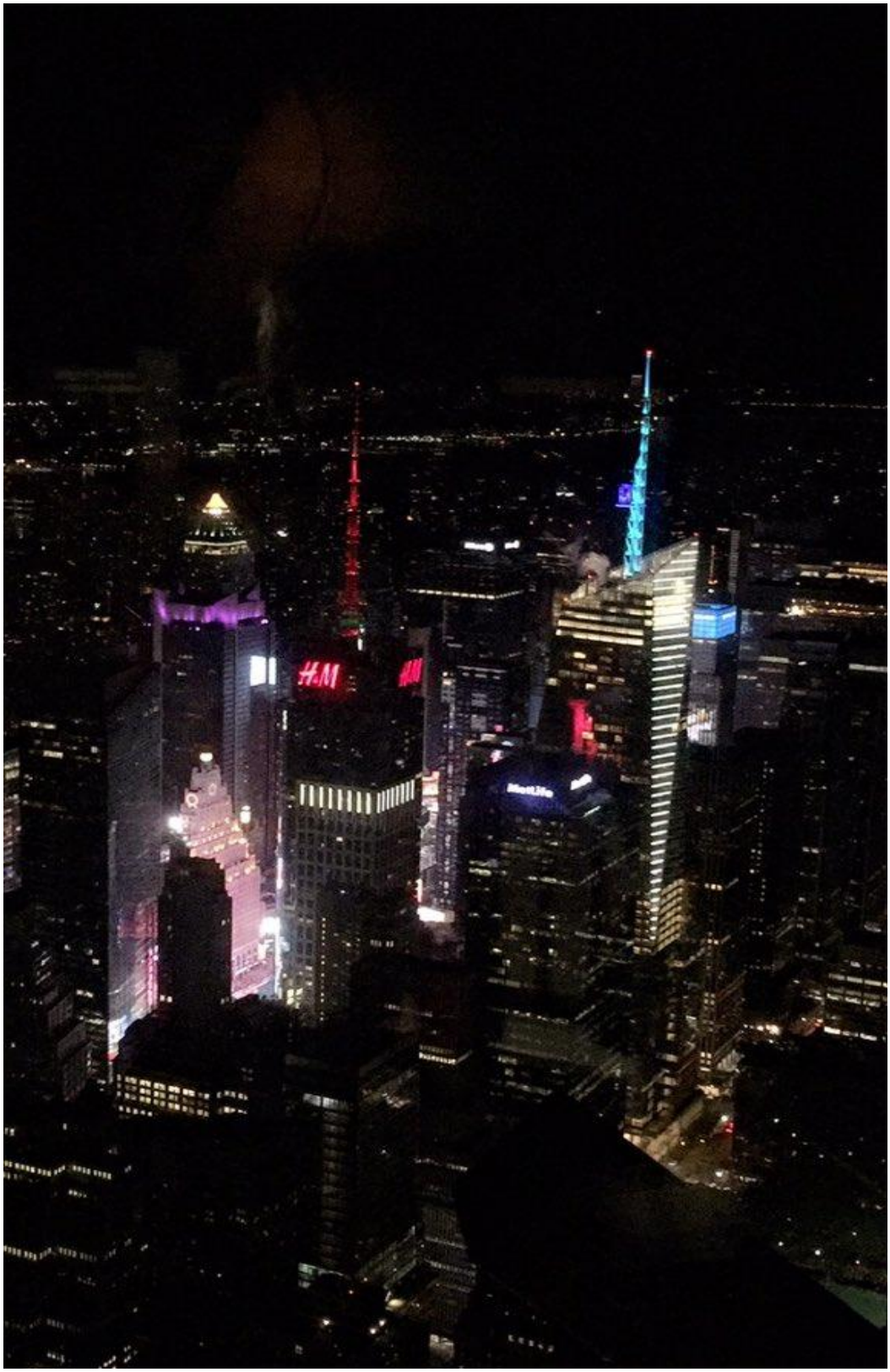
Using the rating column created in #2 above, I used .min() function to get the observation/row with the lowest rating. I also checked for the jpg_url in the image predictions master dataframe using the tweet_id.

My finding is outlined below:

2 dogs have the lowest rating and their rating was 0. Both weren't named and their dog stage is also unknown. The first dog seems okay so no reason as to the rating of 0 while the second dog image is not a dog image but rather a night picture of high-rise buildings which is probably a mistake in the dataset.

jpg_url: https://pbs.twimg.com/media/C5cOtWVWMAEjO5p.jpg

jpg_url: https://pbs.twimg.com/media/Cl2LdofXEAATl7x.jpg



#4

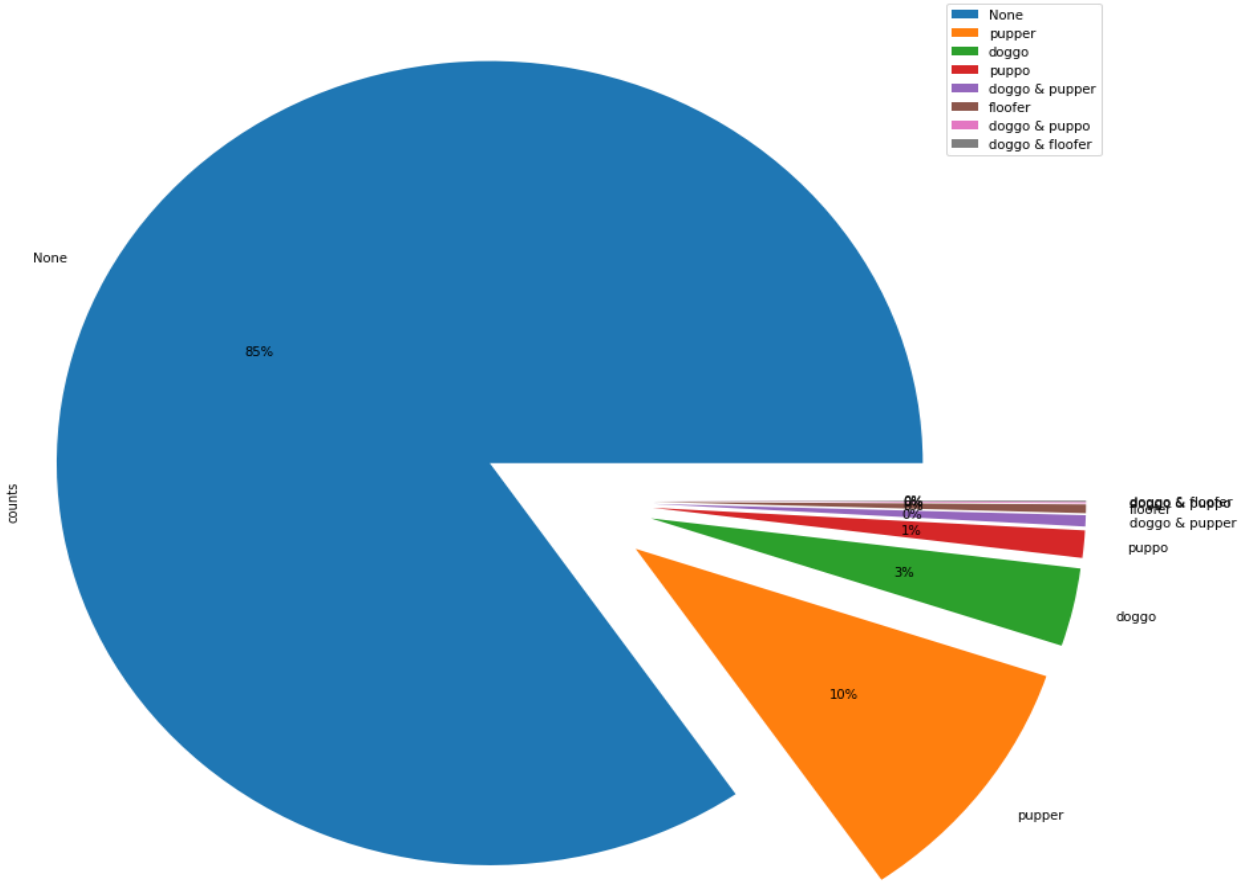
To answer question 4, I used the value_counts() function on the rating column and my finding is outlined below:

The most common rating of dogs on the WeRateDogs twitter account is 1.2 (12/10)

Visualization

For the visualization, I set out to check the proportion of each unique value in the dog_stage in the twitter archive master dataframe in a pie chart.

The pie chart is as below:



As seen in the visualization, 85% of the dogs are classified as None which probably means been unsure of which stage those dogs.

10% of the dogs are pupper, 3% are doggo while the remaining 2% are puppo, floofer, doggo & pupper, doggo & puppo and doggo & floofer. Interesting that some dogs share two stages.