Analysis, Insight and Visualization Report.

To give insight on what this dataset entails, some analysis were done as detailed below.

Process: Using .describe() on this master_dataset, the minimum, mean and max of important variables can be seen.

	tweet_id	img_num	p1_conf	p2_conf	p3_conf	retweet_id	favorite_count	rating_numerator	rating_denominator	time
count	8.400000e+01	84.000000	84.000000	84.000000	84.000000	84.000000	84.000000	84.000000	84.0	84
mean	8.259344e+17	1.440476	0.653526	0.148063	0.051630	5261.550524	18981.238095	12.547619	10.0	3 days 17:53:31.428571
std	3.411578e+16	0.811828	0.248607	0.115545	0.049669	8901.146448	19757.385008	1.839361	0.0	1 days 23:36:51.981762
min	7.680839e+17	1.000000	0.113992	0.000068	0.000034	855.000000	3563.000000	11.000000	10.0	0 days 04:08:40
25%	8.019583e+17	1.000000	0.470472	0.054627	0.011980	1991.500000	8559.250000	12.000000	10.0	1 days 19:25:35
50%	8.228879e+17	1.000000	0.699911	0.121579	0.034451	2895.000000	13559.500000	12.000000	10.0	3 days 16:53:50
75%	8.510174e+17	2.000000	0.862381	0.214879	0.087457	4010.500000	20099.250000	13.000000	10.0	5 days 07:21:20
max	8.902403e+17	4.000000	0.999828	0.487678	0.196399	39752.000000	123688.000000	27.000000	10.0	7 days 18:01:20

1. p1_conf has the highest prediction while p3_conf have the lowest prediction value.
2. The highest favorite_count is 123688 while the mean is 16152.
3. The highest rating is 27/10 while the mean rating is 12.54 /10 and minimum rating is 11/10
4. Maximum of 4 images and mean of 1.44

To see the row(s) that have the highest favorite_count,

Process: Let's use the .loc function to view the properties of the row(s).

	tweet_id	jpg_url	img_num	p1	p1_conf	p1_dog	p2	p2_conf	p2_dog	p3
1744	822872901745569793	https://pbs.twimg.com/media/C2tugXLXgAArJO4.jpg	1	lakeland terrier	0.196015	True	labrador retriever	0.160329	True	irish terrier

p3	...	favorite_count	source	text	expanded_urls	rating_numerator	rating_denominator	name	date	time	stage_name
sh er	...	123688	Twitter for iPhone	Here's a super supportive puppo participating	https://twitter.com/dog_rates/status/822872901...	13	10	None	2017-01-21	23:59:20	puppo

We can see that the tweet id is 22872901745569793 was from an iPhone user, the first, second, and third predictions predicted the dog to be a Lakeland terrier, Labrador retriever, and Irish terrier respectively and stage_name puppo present, rating of 13, tweet contains one image and the tweet on 2017-01-21 at 23:59:20.

Using the group to see the name that has more favorite count, we see it is Barney. Using the index to see its features, the following can be seen;

tweet_id	846514051647705089
jpg_url	https://pbs.twimg.com/media/C79sB4xXwAEvwKY.jpg
img_num	2
p1	golden retriever
p1_conf	0.650003
p1_dog	True
p2	leonberg
p2_conf	0.0651992
p2_dog	True
p3	norfolk terrier
p3_conf	0.0529553
p3_dog	True
retweet_id	10346
favorite_count	41088
source	Twitter for iPhone
text	This is Barney. He's an elder doggo. Hitches a...
expanded_urls	https://twitter.com/dog_rates/status/846514051...
rating_numerator	13
rating_denominator	10
name	Barney
date	2017-03-28 00:00:00
time	3 days 17:00:20
stage_name	doggo

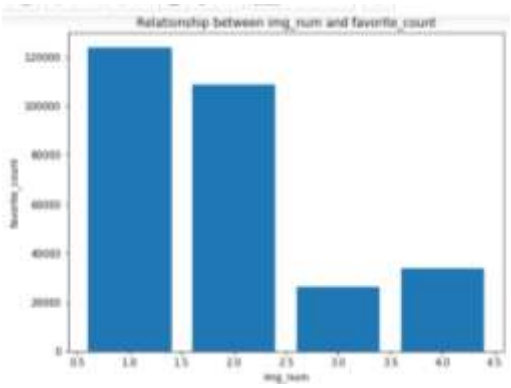
Using the .loc function to see the features of the row(s) with the highest rating, we see that;

* It was tweeted from an iPhone with id 1778027034220126208 on the 16th of September, 2016 contains one image, is a female dog of stage pupper, name Sophie and received 6175 favorite counts.

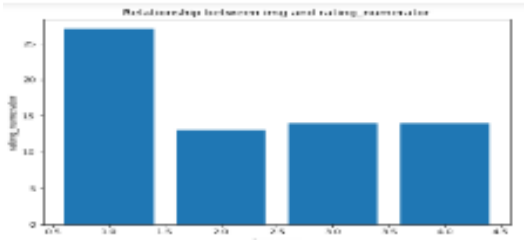
Using bar charts and scatter plots to examine the correlation in this dataset, it can be seen that there is;

1. Tweets with one image received more favorite count compared to the rest, as number of images increases, favorite count decreases (inversely related). This can be as a result of an image being definite and specific rather than having multiple different pictures. Most

social media users will likely view one picture assigned to a name, then react to it rather than view multiples pictures.

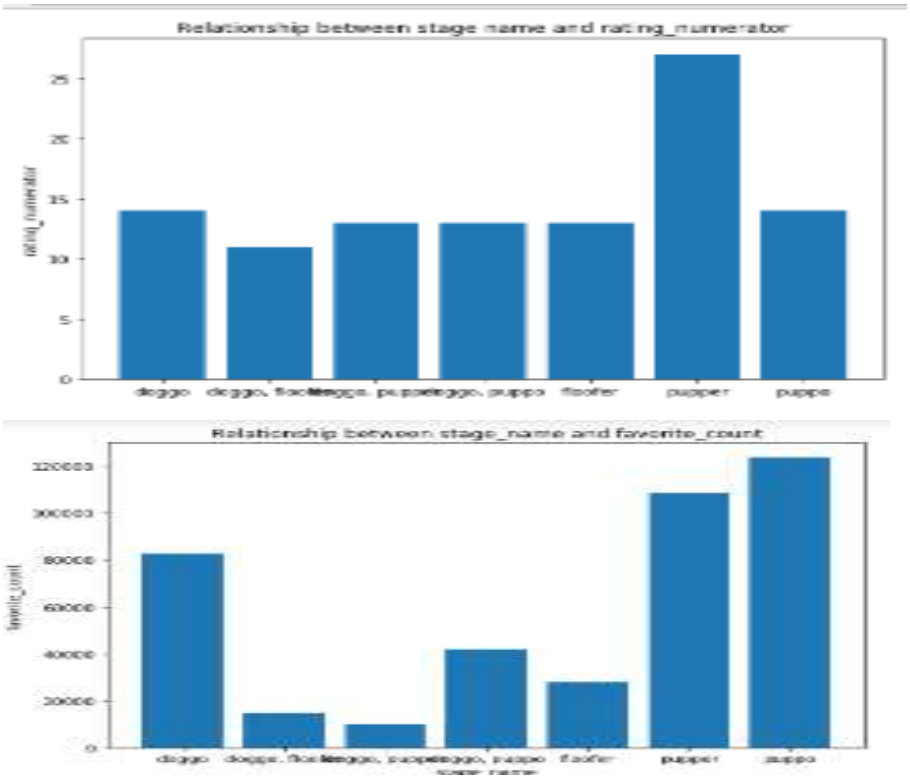


2. Tweets with one(1) image received the highest rating of 27. Tweets with 3 and 4 images received same amount of rating. As earlier stated,viewers will view and react more to post having few images than those with more images.

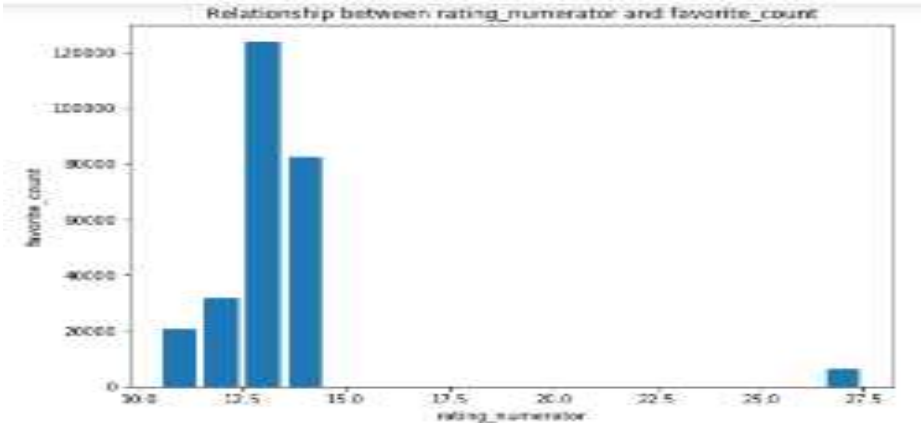


3. As can be seen from the charts below, pupper received highest ratings followed by puppo and doggo while doggo-floofer got the lowest rating.

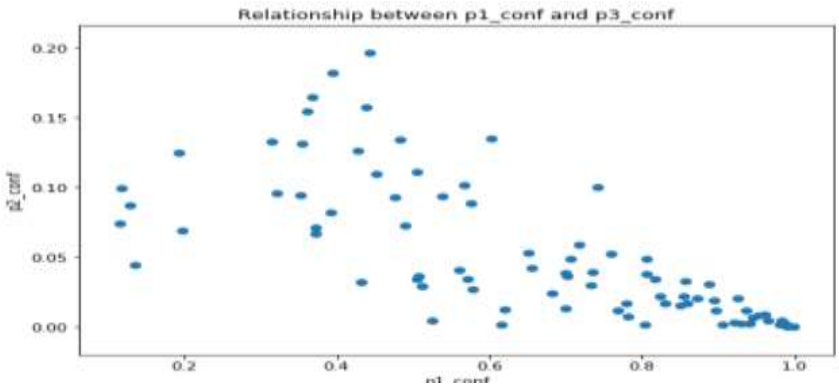
   Puppo received the highest favorite count followed by pupper while pupper-doggo received the lowest favorite count.



4. Tweets with ratings in the range of 12.5-15 received the highest favorite count compared to those between 11-12 and 15-27 as seen in the chart below.



5. The scatter plot of p1_conf and p3_conf which is the first and last predictions shows a negative correlation as the points tend to slope downwards from left to right as shown below.



NOTE: The insights and analysis in this dataset are numerous. More can be done depending on what you want to achieve with the analysis.