Insight Report

Description of dataset

- The dataset is a **merged table** of twitter_archive, image_prediction and tweet_extra_data
- It contain 2027 records and 17 columns (variables) about dog_rates from 2015 to 2017
- Analysis and Visualization is carried out on the combined dataset "**twitter_archive_master**"

Insight 1: Which year has the highest number of Dog tweets?

year	number_of_tweets
2016	962
2015	655
2017	354

- The result table shows number of dog_tweets per year sorted from highest to lowest. The year 2016, has the highest number of dog_tweets of about 962.

  This might have been as result of the online fight about dogs and the popular meme "They're good dogs, Brent," which first gained popularity in September, 2016.

Insight2: What is the most popular rating for dogs?

year	rating_numerator	total
2015	10	192
	11	104
	9	89
	12	75
	8	66

2016		
	11	253
	12	243
	10	222
	13	81
	9	61

2017		
	13	160
	12	128
	11	35
	14	24
	10	3

The table above shows the top 5 most popular rating for dogs for each year,

Dogs are rated from 0-14, with 14 been the highest rating

- In 2015, the popular dog rating is 10/10 as shown with a total count of 192 ratings
- In 2016, the popular dog rating is 11/10 as shown with a total count of 253 ratings
- In 2017, the popular dog rating is 13/10 as shown with a total count of 160 ratings
- In general, most dogs are rated above 10

Insight3: Which dog rating has the highest number of retweets?

year	rating_numerator	retweet_mean
2015	13	4880.666667
	420	4324.000000
	1	2512.200000
	12	1644.626667
	11	1483.846154

year	rating_numerator	retweet_mean
2016	14	9096.111111
	13	7560.148148
	75	7069.000000
	12	2844.930041
	1776	2772.000000

year	rating_numerator	retweet_mean
2017	5	10407.000000
	14	10081.791667
	13	7054.250000
	10	6576.000000
	12	4578.125000

The table above show the top 5 dog ratings which has the highest retweets for each year

- In 2015, the rating 13 has the highest retweets mean of 4880.67
- In 2016, the rating 14 has the highest retweets mean of 9096
- In 2017, the rating 5 has the highest retweets of 10,407.
- In 2017, the topmost rating was 5 irrespective it is a low rating score. This could be that the dogs rated 5 had a feature that could have prompted so much retweets

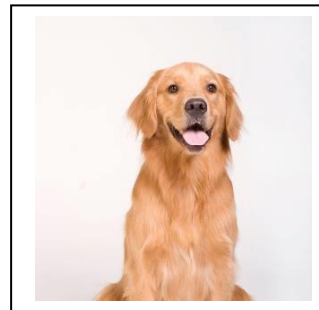**Insight4: Which dog rating has the highest number of likes**

year	rating_numerator	likes_mean
2015	13	10907.500000
	420	7989.000000
	1	5599.800000
	12	3824.893333
	11	3391.028846

year	rating_numerator	likes_mean
2016	14	22537.111111
	75	20296.000000
	13	17675.419753
	12	8731.316872
	27	7320.000000

year	rating_numerator	likes_mean
2017	5	46860.000000
	14	32731.125000
	13	26965.543750
	10	26304.666667
	0	24574.000000

The table above show the top 5 dog ratings which has the highest likes for each year

- In 2015, the rating 13 has the highest likes of 10,907
- In 2016, the rating 14 has the highest likes of 22,537
- In 2017, the rating 5 had the highest likes of 46,860
- In 2017, the ratings of 5 had the highest likes irrespective it is a low rating score. This tends to correspond with the high retweets which it garnered. This could be that the dogs rated 5 had a feature that prompted so much likes

Insight5: Which dog is most favorite (based on Prediction1_name)

predict1_name	total_count	total_likes
golden_retriever	137	1693475
Labrador_retriever	94	1116811
Pembroke	88	994755
Chihuahua	78	673526
Samoyed	40	542991
French_bulldog	26	465738
chow	41	404241
pug	54	320223
cocker_spaniel	27	315825
Pomeranian	38	299114

Golden retriever (1)

Labrador retriever (2)

The table above shows the top 10 dog which are most liked

From the table, the dog with the name ***Golden Retriever*** is the most liked dog with a total likes of 1,693,475. This could be due to the nature of the dog which includes gentleness, affectionate and having a striking golden coat.

The second most liked dog is ***Labrador Retriever*** with a total likes of **1,116811**. This dogs also possesses some of the features of golden retriever

Insight6: which dog name is most predicted and how accurate is prediction 1 algorithm
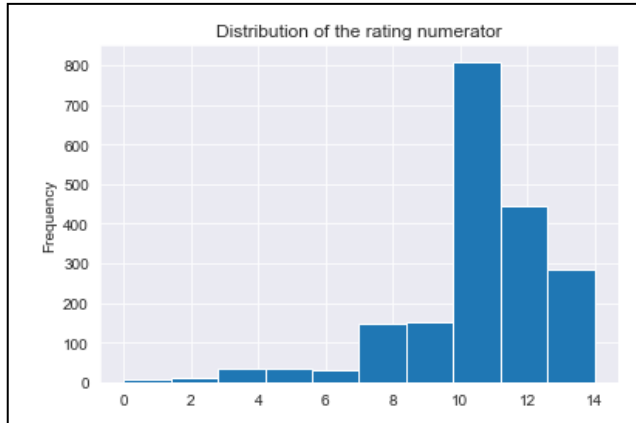
predict1_name	total_count	confidence_mean
golden_retriever	137	0.728116
Labrador_retriever	94	0.656606
Pembroke	88	0.717223
Chihuahua	78	0.587447
pug	54	0.759223
chow	41	0.615648
Samoyed	40	0.747088
Pomeranian	38	0.743606
toy_poodle	37	0.611089
malamute	29	0.589451

The table above shows the most predicted dog name based on algorithm one and it confidence mean.

Dog name **Golden Retriever** was the most predicted dog name and has a high confidence mean score of **0.728116** which according to Microsoft confidence score, a score of above 0.7 is a high confidence for the predicted image

Labrador retriever has a confidence mean score of **0.6566** which is below 0.7 and does not entail a high confidence for the predicted image
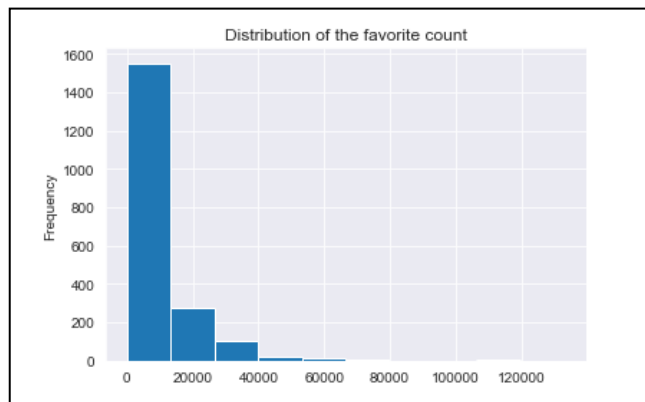
Distribution of numerator ratings

The chart above shows the distribution of the numerator ratings and it show that it is not normally distributed.

The distribution is skewed to the left which implies that mass of the distribution of ratings are in the range of 10 to 14
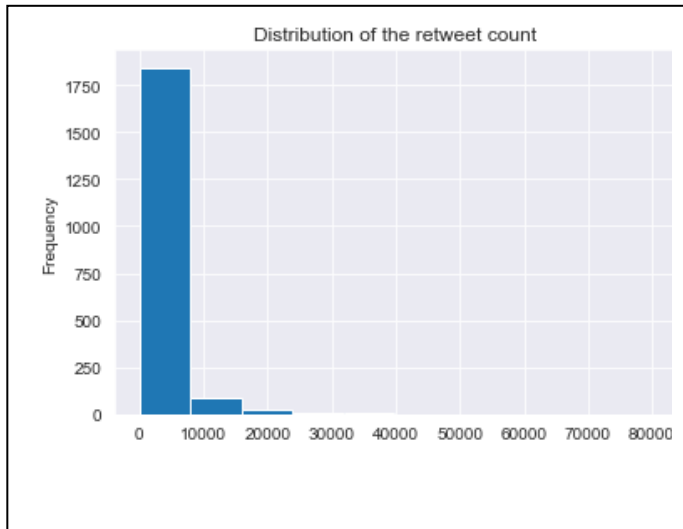
Distribution of favorites count

The chart above shows the distribution of the favorite count and it show that it is not normally distributed.

The distribution is skewed to the right which implies that mass of the distribution of likes are in the range of 0 to 20,000 likes
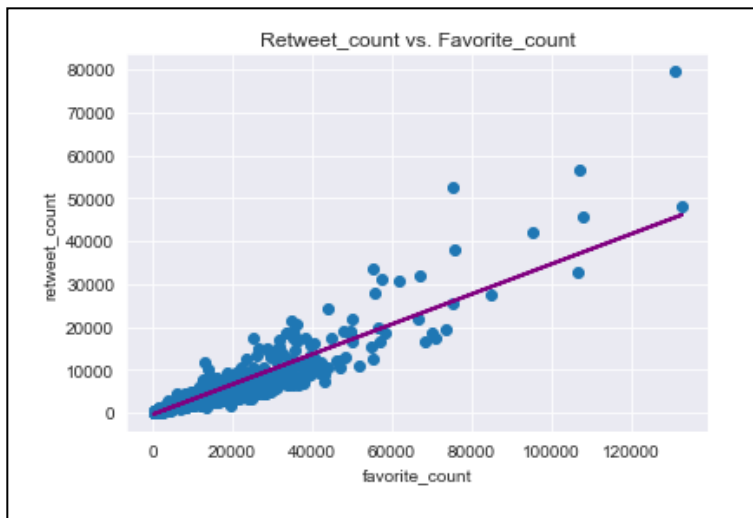
Distribution of retweets count

Distribution of the retweet count

The chart above shows the distribution of the retweets count and it show that it is not normally distributed.

The distribution is skewed to the right which implies that mass of the distribution of retweets are in the range of zero (0) to 10,000 retweets

Insight8: Is Likes associated with retweets

Retweet_count vs. Favorite_count

The Chart above shows that there is a *positive correlation* between favorite_count and retweet_counts and a correlation coefficient of **0.9130** indicates a **very strong positive correlation**. This implies that retweets are associated to likes(favorites)