WeRateDogs Analysis

WeRateDogs is a Twitter account that rates people's dogs with 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."

We'll be analyzing is the twitter archive data of **WeRateDogs**.

The data sent over doesn't contain important metrics such as ***Favorite Count*** & ***Retweet Count***. We downloaded this missing data using Twitter API. We also ran the images of these tweets through a neural network to get the predictions of the dog breeds of the tweets. After gathering all the data, we assessed each data and found couple of Quality and Tidiness issues. Next, we cleaned these identified issues and stored this data into a master data file.

We used the master data file for our analysis.

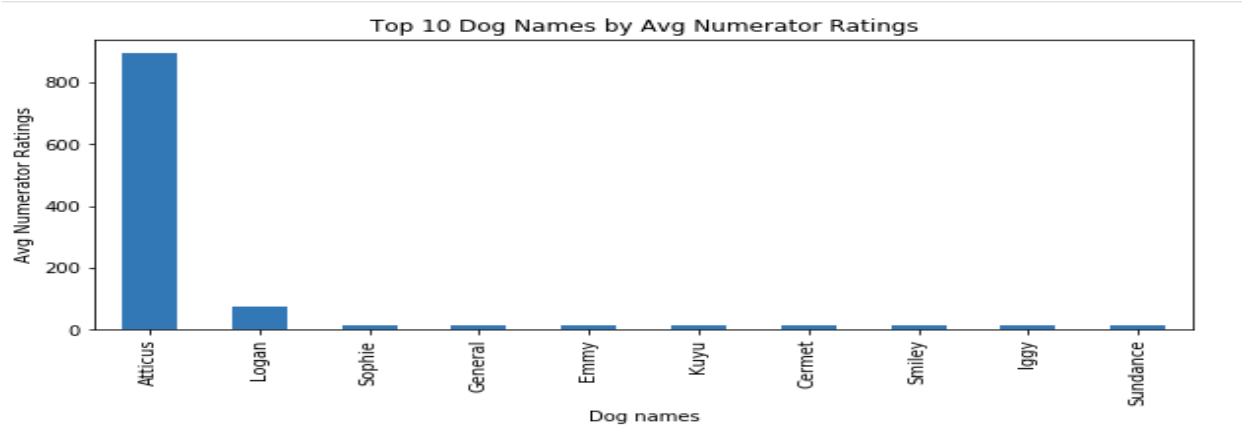
Insight 1

Charlie	11
Lucy	10
Cooper	10
Oliver	10
Penny	9
Tucker	9
Winston	8
Sadie	8
Lola	7
Daisy	7

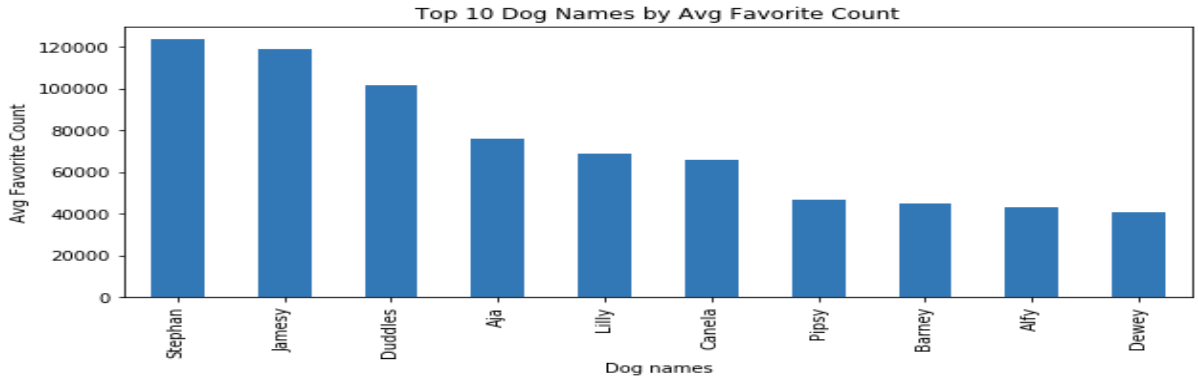
Common 10 Names of dogs.

Insight 2

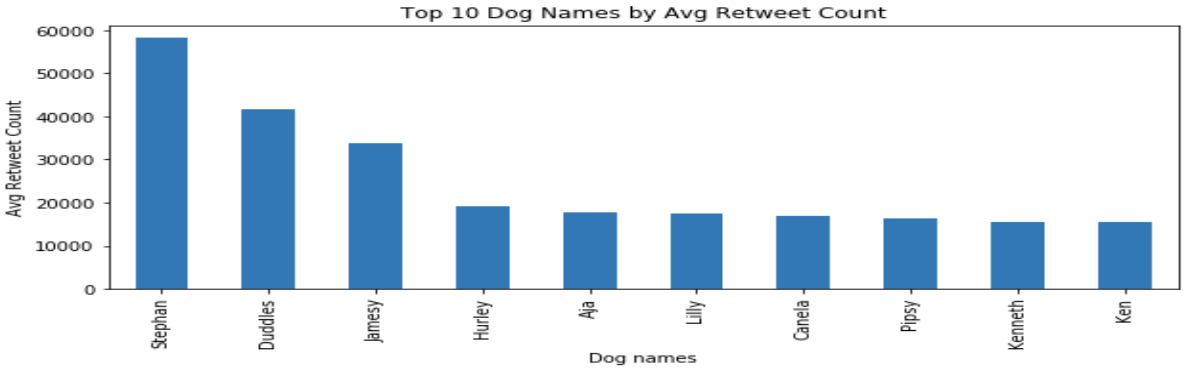
Top 10 dog names by Avg Ratings: -



Top 10 dog names by Avg Favorite Count: -



Top 10 dog names by Avg Retweet Count: -



Numerator Ratings aren't influencing **Favorite Count** & **Retweet Count**. But there is some relationship between **Favorite Count** & **Retweet Count** as *Top 5 Names* by Avg Favorite Count also are in *Top 6 Names* by Avg Retweet Count.

Insight 3

Count of dogs at each stage: -

Pupper 194
Doggo 70
Puppo 23

% of dogs at each stage: -

Pupper 67.59%
Doggo 24.39%
Puppo 8.01%

Avg Numerator Ratings of dogs at each stage: -

Pupper 11.75
Doggo 10.69
Puppo 12.04

Avg Favorite Count of dogs at each stage: -

Pupper 18807
Doggo 6766
Puppo 22257

Avg Retweet Count of dogs at each stage: -

Pupper 6427
Doggo 2129
Puppo 6485

Most common dog stage classifications/records are for **Pupper(67.5%)** i.e. *"A small doggo."* **Puppo** i.e. dogs that are in a *"A transitional phase between **Pupper** and **doggo."*** have **least** numbers of classifications/records **8%** among the tweets but have the **highest** *Numerator Ratings, Avg Favorite Count & Avg Retweet Count.*

Insight 4

Dog breed predictions by neural network: -

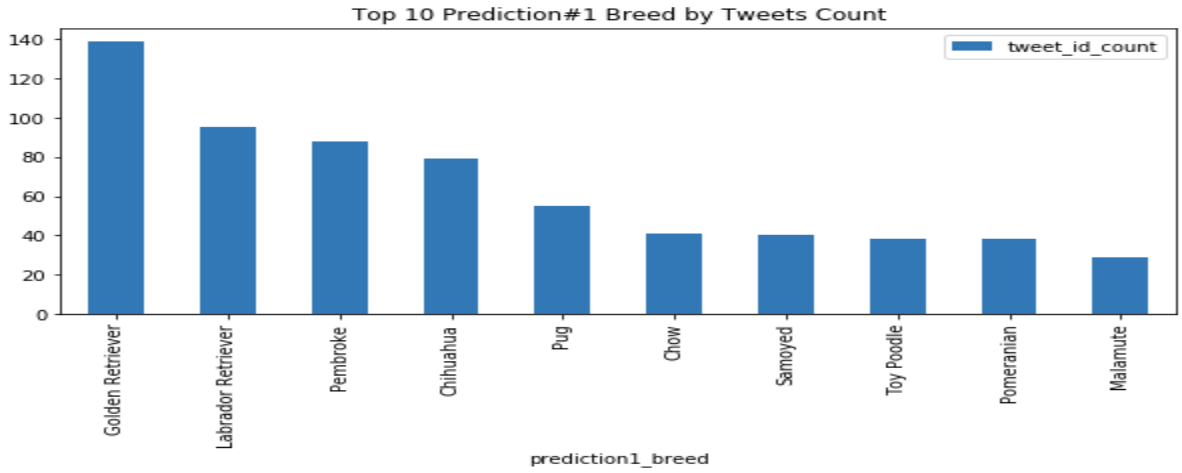
p1_dog	p2_dog	p3_dog	
False	False	False	15.828771
		True	3.086112
	True	False	2.638128
		True	4.778497
True	False	False	2.140368
		True	4.280737
	True	False	7.217521
		True	60.029866

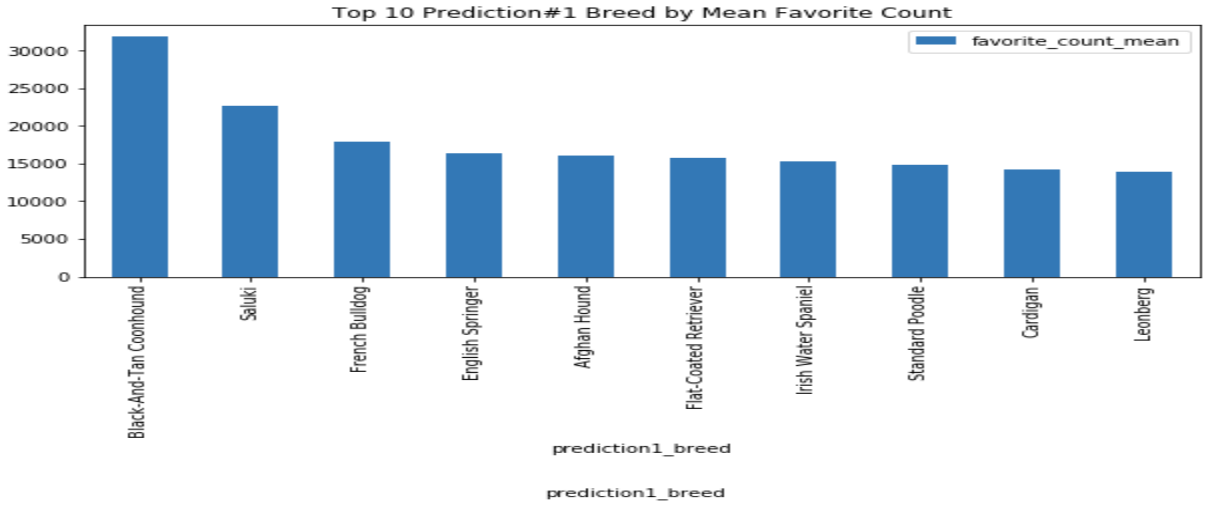
60% of the total predictions of each tweet image resulted in prediction of a dog breed & **15%** of all the predictions were not dog breeds.

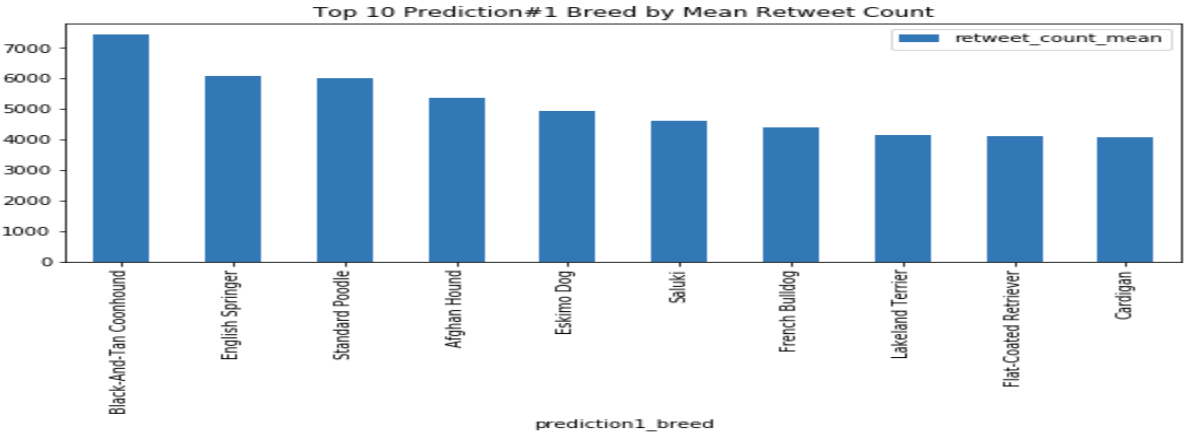
Insight 5

Top 30 Prediction1 dog breed names by count: -

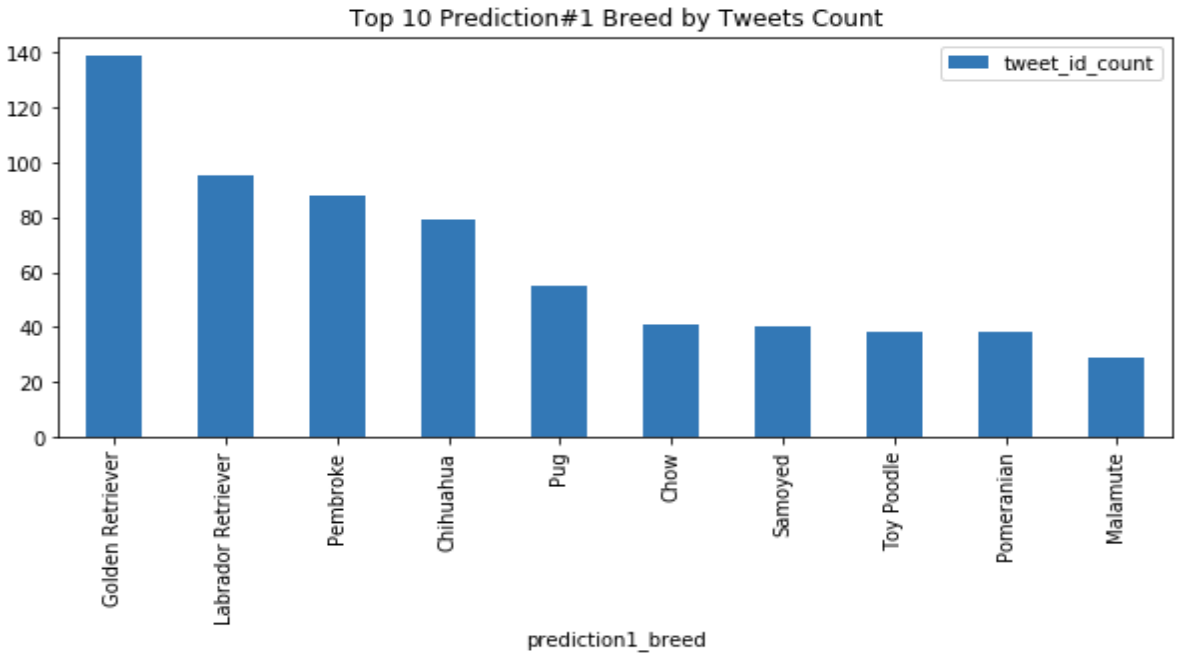
Golden Retriever	139
Labrador Retriever	95
Pembroke	88
Chihuahua	79
Pug	55
Chow	41
Samoyed	40
Toy Poodle	38
Pomeranian	38
Malamute	29
Cocker Spaniel	27
French Bulldog	26
Chesapeake Bay Retriever	23
Miniature Pinscher	22
Siberian Husky	20
German Shepherd	20
Staffordshire Bullterrier	19
Eskimo Dog	18
Maltese Dog	18
Shetland Sheepdog	18
Rottweiler	17
Cardigan	17
Shih-Tzu	17
Beagle	17
Kuvasz	16
Lakeland Terrier	16
Italian Greyhound	16
West Highland White Terrier	14
Great Pyrenees	14
Basset	13

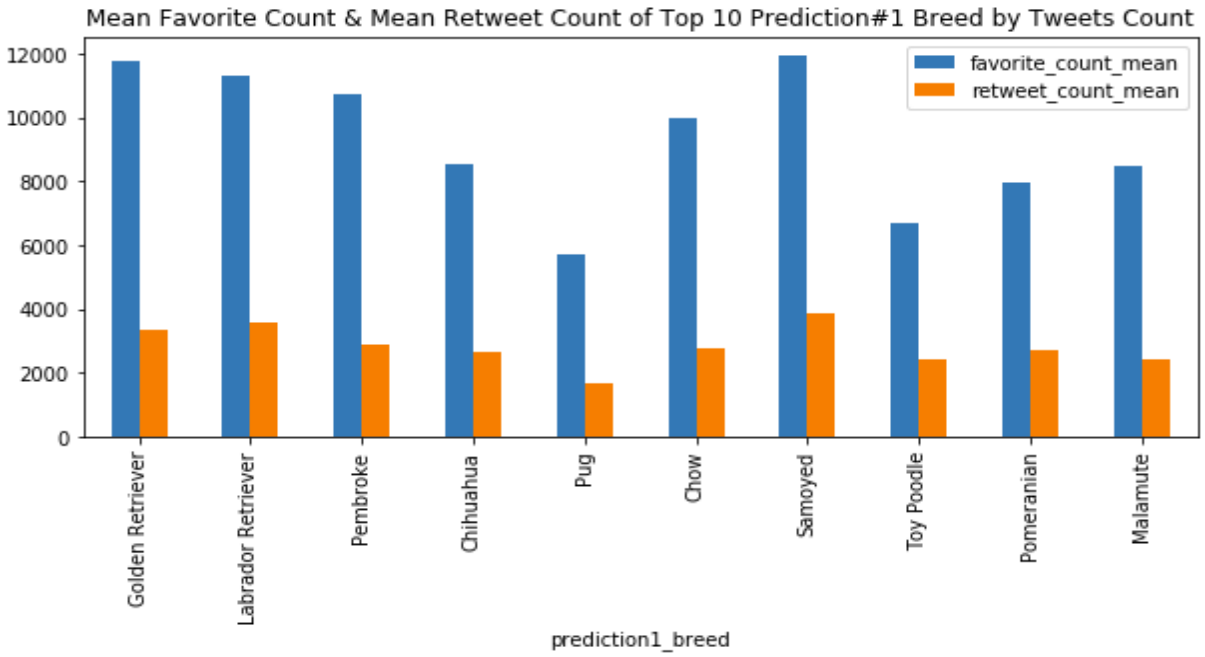






From above images we can visually analyze there is no correlation between Top 10 Prediction#1 Breeds by value count and Top 10 Prediction#1 Breeds by Favorite Count and Retweet count but the **breeds** with **Top 10 Prediction#1 Breeds by Favorite Count and Retweet Count** overlap with each other.





From above image we can analyze that it's **not certain** that all Top 10 Prediction#1 Breeds by count may also have same hierarchy with respect to *Top 10 Prediction#1 Breeds by Favorite Count* & *Retweet Count.*