“WE RATE DOGS” TWEETS’ ANALYSIS

Paula Jasper, November 12, 2018

I analysed tweets from the **We Rate Dogs** twitter feed for my Data Wrangling project on the Udacity Data Analysis Nanodegree course.

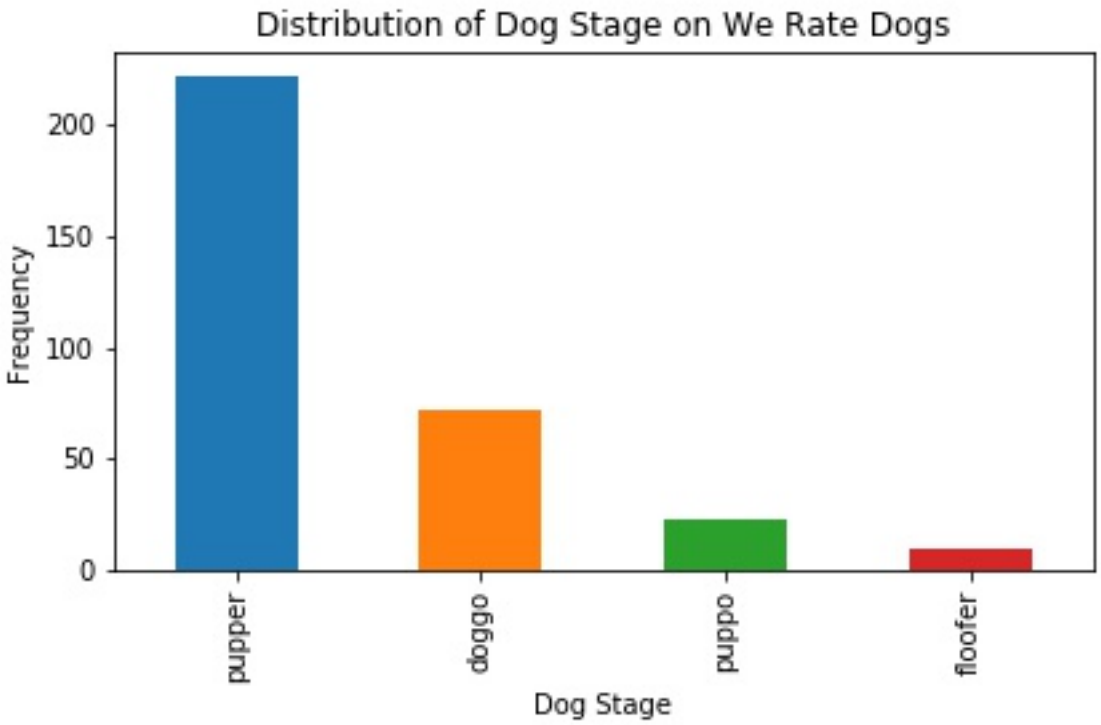
My analysis concentrated on **dog stage analysis**, **relationships and correlations between the ratings, favorite counts and retweet counts variables** and **dog breeds analysis**. Here are my findings.

Dog Stage Analysis.

Question: Which dog stage occurs most frequently on “We Rate Dogs”?

Dog stages are based on the definitions in the Dogtionary

- Doggo: A big pupper, usually older
- Pupper: A small doggo, usually younger
- Puppo: transitional phase between doggo and pupper
- Floofer: dog with excess fur.



Answer: In this sample of data, most of the dogs are Puppers. There are very few floofers.

Question: Which dog stages get the highest ratings, most favorites and most retweets?

Ratings statistics

	count	mean	std	min	25%	50%	75%	max
dog_stage								
doggo	72.0	1.184722	0.145020	0.8	1.10	1.2	1.3	1.4
floofer	9.0	1.188889	0.105409	1.0	1.10	1.2	1.3	1.3
pupper	222.0	1.080631	0.203217	0.3	1.00	1.1	1.2	2.7
puppo	23.0	1.200000	0.127920	0.9	1.15	1.2	1.3	1.4

Favorite count statistics

	count	mean	std	min	25%	50%	75%	max
dog_stage								
doggo	72.0	18714.805556	25783.259184	2487.0	6865.25	11468.5	19047.25	165012.0
floofer	9.0	10665.777778	10314.789792	1550.0	3860.00	5811.0	12014.00	32140.0
pupper	222.0	7040.500000	11066.589658	257.0	2347.25	3213.0	7478.00	122542.0
puppo	23.0	21888.782609	29223.239264	3143.0	6800.50	14897.0	20392.50	141096.0

Retweet count statistics

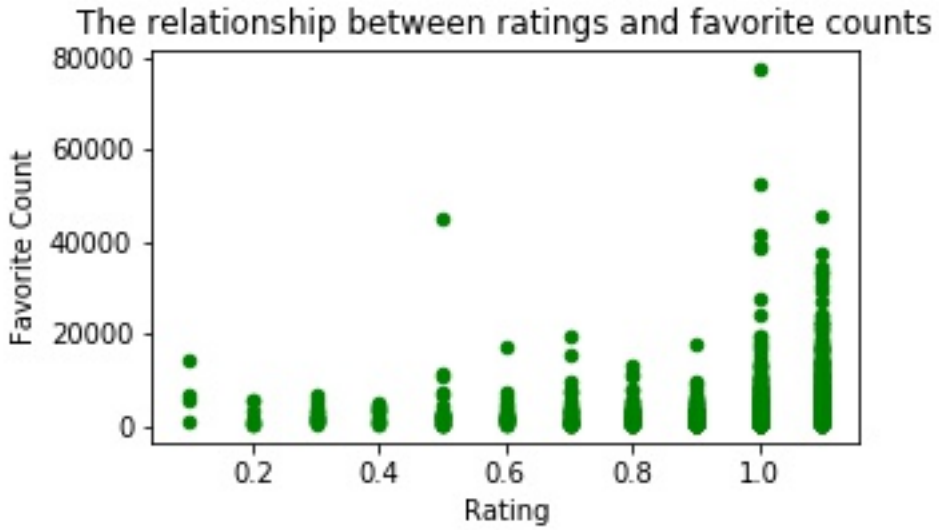
	count	mean	std	min	25%	50%	75%	max
dog_stage								
doggo	72.0	6577.486111	12968.733739	688.0	1891.25	2845.5	5115.0	84187.0
floofer	9.0	3944.000000	5293.277931	475.0	1080.00	2400.0	3628.0	17521.0
pupper	222.0	2362.986486	3757.539266	25.0	658.25	1188.5	2381.0	35679.0
puppo	23.0	6436.782609	10013.116888	671.0	1633.00	3059.0	7148.5	48018.0

Anwers:

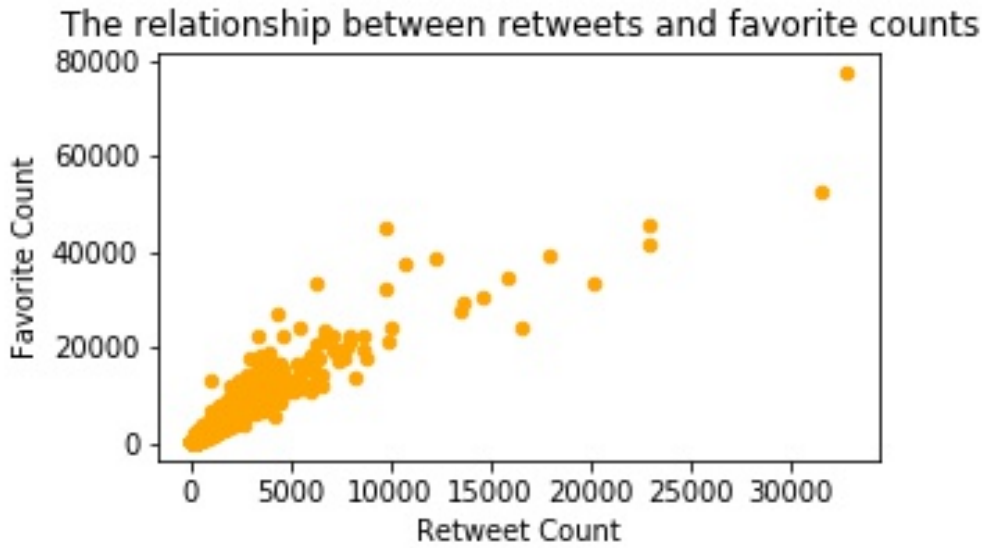
- **Puppers have the lowest ratings on average, but also the highest rating overall. They also account for the majority of the dogs.**
- **Puppos have the highest ratings on average but there are fewer of them.**
- **Puppers also have the lowest favorite count on average and puppos have the highest.**
- **Doggos have the highest retweet count followed closely by puppos**

Relationships & Correlations: Exploring the relationships between ratings and favorite counts, and retweets and favorite counts.

Question: Are these variables correlated and what is the strength and direction of the relationships?



Answer: There is a positive relationship between ratings and favorite counts – as ratings increase so do favorite counts. However the correlation – 0.19 – is not meaningful.

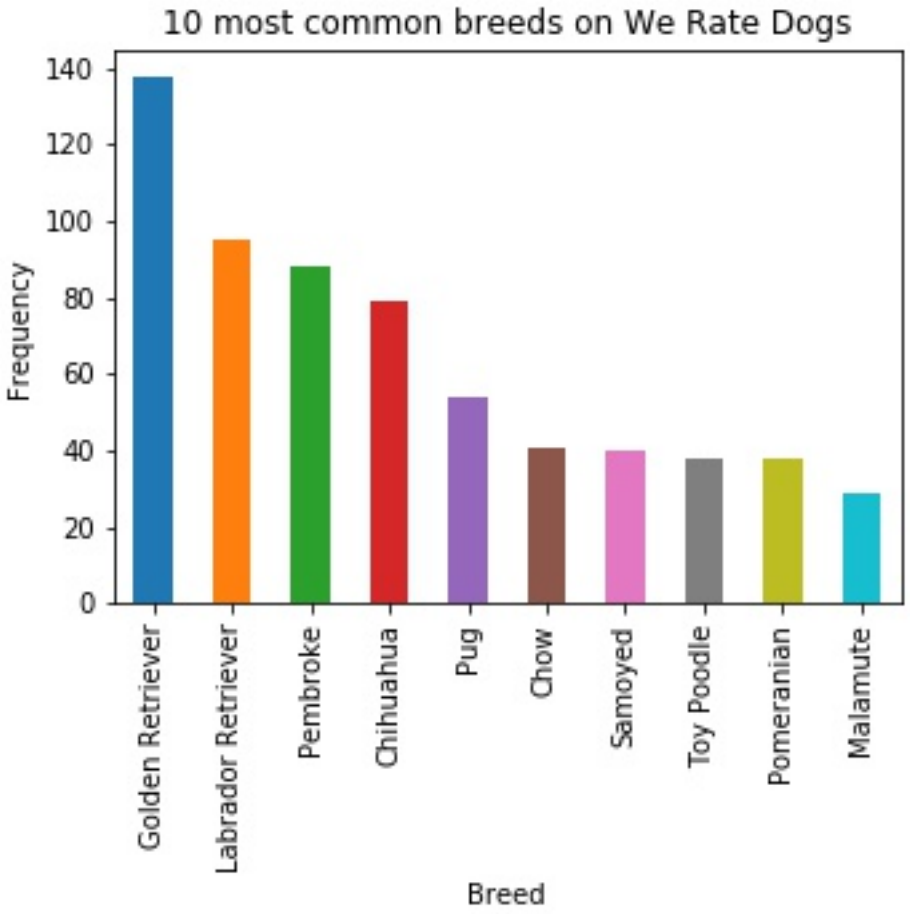


Answer: There is a positive relationship between favorite counts and retweet counts – as retweets increase so do favorite counts. This is as expected. The correlation is very strong at 0.93.

Dog Breeds Analysis.

Question: What are the 10 most common breeds on "We rate dogs"?

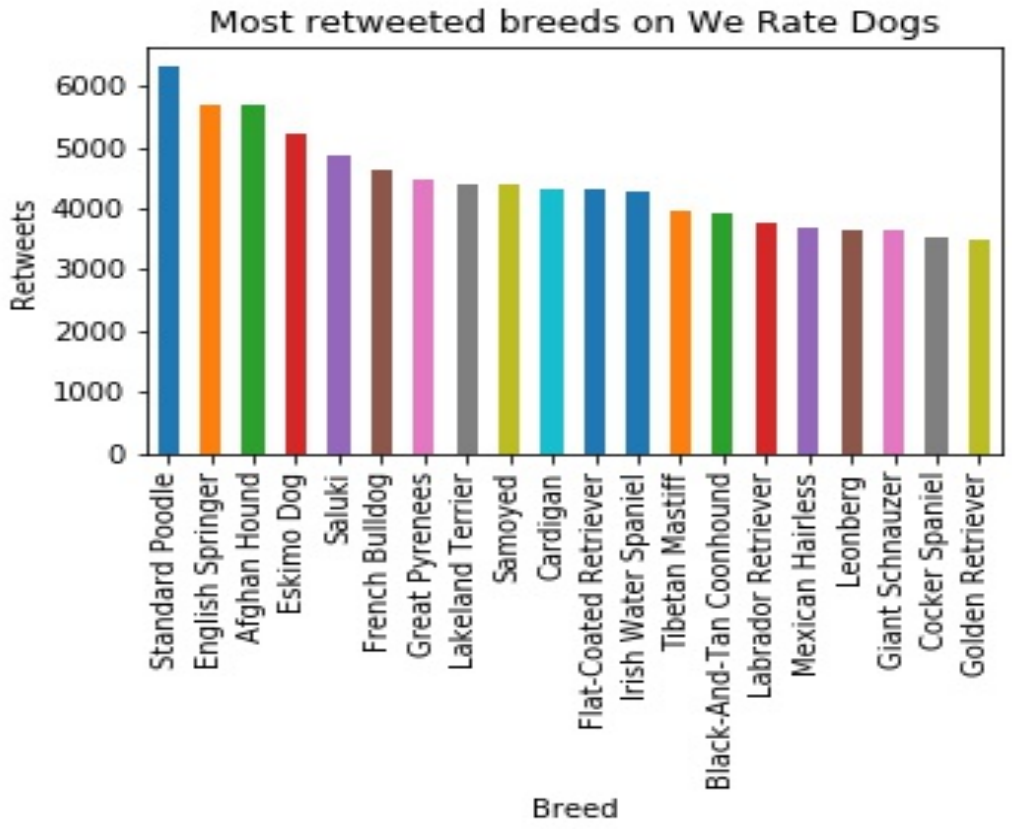
Answer: Based on prediction 1 - the neural network prediction with the highest confidence, these are the top 10 most common breeds on "We Rate Dogs."

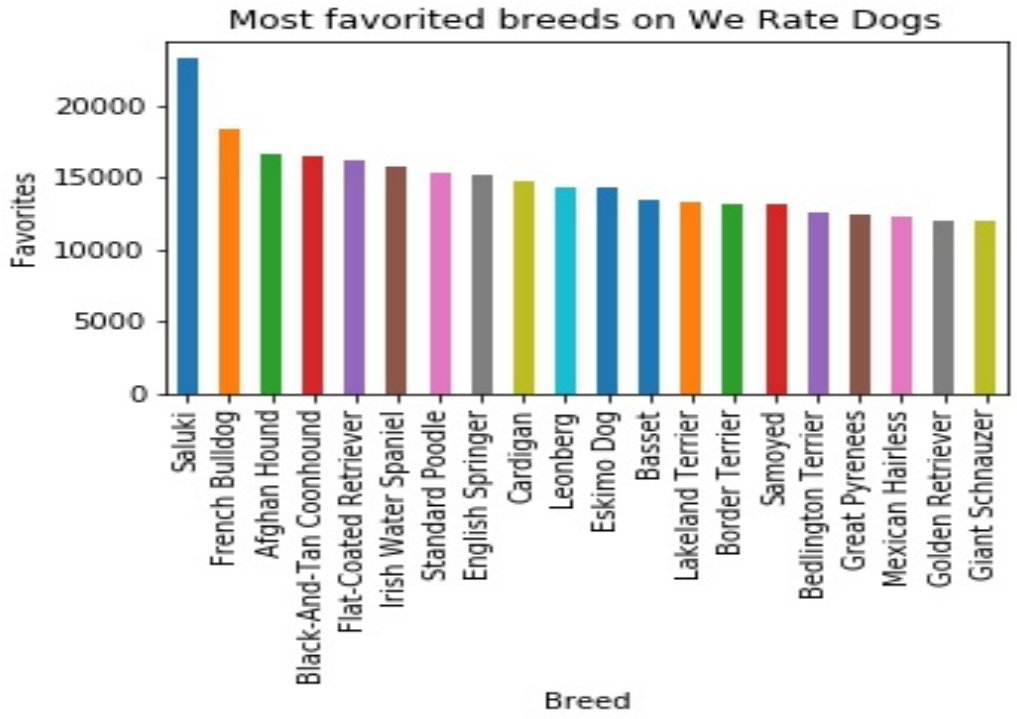


Breed	Count
Golden Retriever	138
Labrador Retriever	95
Pembroke	88
Chihuahua	79
Pug	54
Chow	41
Samoyed	40
Toy Poodle	38
Pomeranian	38
Malamute	29

So, retrievers are most common breed. They are indeed a common breed – I wonder if they are also the easiest to classify? I will address this at the end.

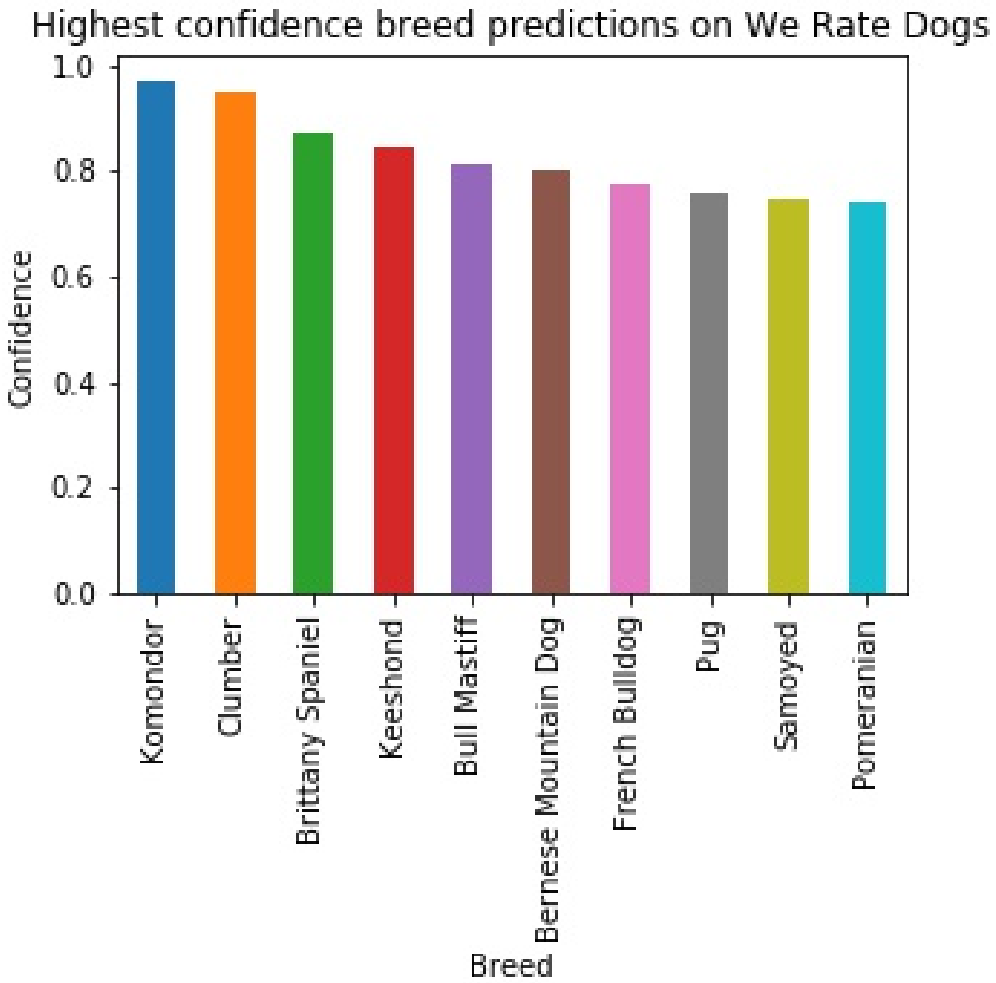
Question: Which breeds are the most retweeted and most favorited ones?





Answer: The Standard Poodle and English Springer are the most retweeted, while the Saluki and French bulldog are the most favorited. Many of the breeds appear in both.

Question: Which breeds have the highest and lowest confidence levels according to the Machine Learning Neural Network?



Answer: The komondor has the highest confidence rate, followed by the clumber. Surprisingly, the retriever (the most common breed) does not appear in the highest confidence list.

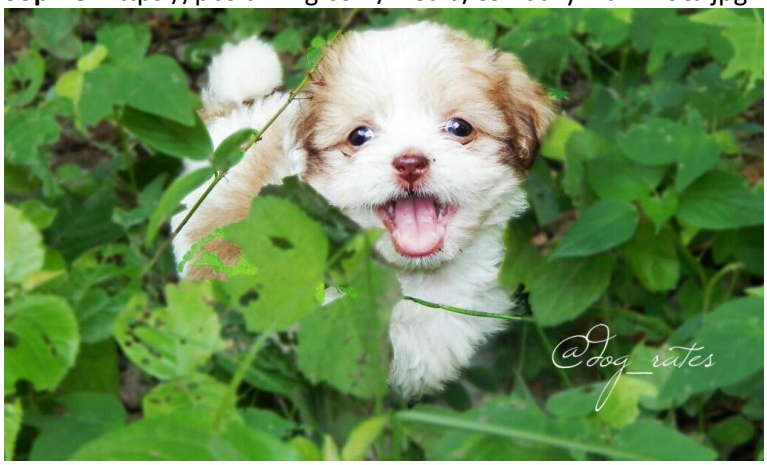
Below are the photos from the dataset of the three dogs predicted to be a komondor and the one clumber. These look to be accurate. The komondor looks like a very distinctive dog.

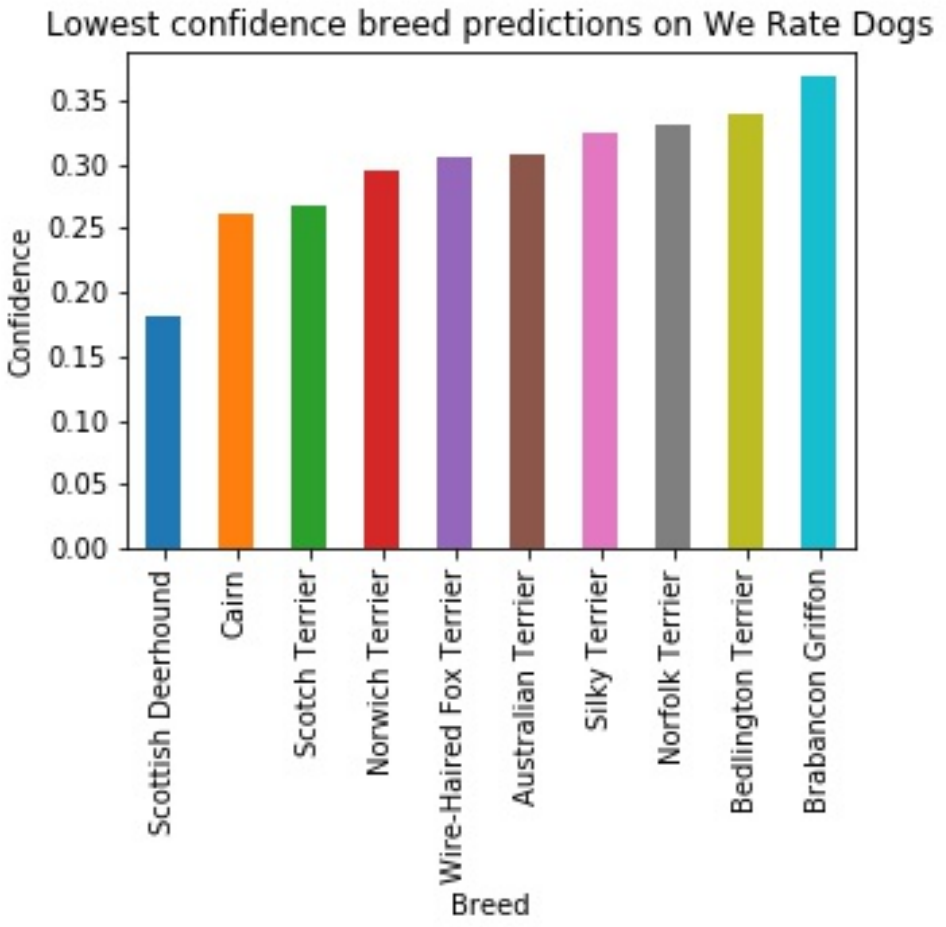
Napoleon: https://pbs.twimg.com/media/DBg_HT9WAAEeIMM.jpg

Remus: https://pbs.twimg.com/media/Cv3tU38WcAASFas.jpg

No name: https://pbs.twimg.com/media/CUd5gBGWwAA0IVA.jpg

Sophie: https://pbs.twimg.com/media/Cswbc2yWcAAVsCJ.jpg





Answer: The Scottish Deerhound has the lowest confidence, followed by the Cairn.

Below are the photos from the dataset of the three dogs predicted to be Scottish Deerhounds and the three Cairns. I can see that these photos would be difficult to classify as the dogs are either just a small part of the photo – with people, other dogs or other objects included, or the dogs wearing disguises. Therefore this shows that the confidence level is not related to the breed itself but the photo.

Freddery: https://pbs.twimg.com/media/CaBP7i9W0AAJrIs.jpg

Louis: https://pbs.twimg.com/media/CWzDWOkXAAAP0k7.jpg

No name: https://pbs.twimg.com/media/CV_r3v4VAAALvwg.jpg

Winston: https://pbs.twimg.com/media/C25d3nkXEAAFBUN.jpg

Calbert: https://pbs.twimg.com/media/CbSqE0rVIAEOPE4.jpg

Butters: https://pbs.twimg.com/media/CU2FsRnVAAA3TEg.jpg