

## Insights and Visualization from WeRateDog Twitter Archive Dataset

I used python and its libraries to wrangle (gather, assess, and clean) and perform exploratory data analysis EDA on a dataset from the tweet archive of Twitter user @dog\_rates with an account that rate people's dogs. The user also known as WeRateDogs which have over 4 million followers and have received international media coverage is a good option for learning wrangling and visualization of data. The dataset are scattered around, have quality issues and likewise tidiness issues. I gathered the data by downloading dataset programmatically and querying twitter API, so you can never go wrong practicing your wrangling skills on such dataset.

After I used json, requests, and tweepy libraries to gather data from different sources. Pandas and Matplotlib libraries were used extensively in assessing, cleaning, and exploring the dataset. I derived the following insights from the master table I created:

1. The ten most common dog breed on the said twitter space are:

S/N	Dog breeds	Number of dogs
1	Golden_Retriever	173
2	Labrador_Retriever	113
3	Pembroke	96
4	Chihuahua	95
5	Pug	65
6	Toy_Poodle	52
7	Chow	51
8	Samoyed	46
9	Pomeranian	42
10	Cocker_Spaniel	34

2. The five dog breeds with the highest average ratings are:

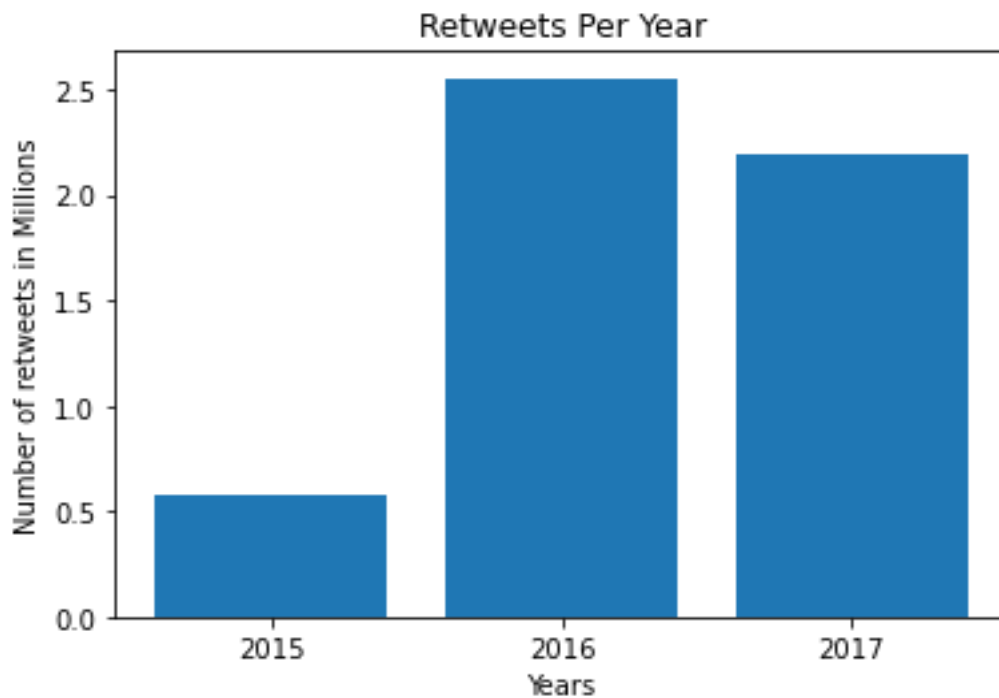
1. Clumber 27.000000
2. Soft-Coated\_Wheaten\_Terrier 20.800000
3. West\_Highland\_White\_Terrier 14.687500
4. Great\_Pyrenees 14.666667
5. Chow 14.039216

3. The dog breed with the single most likes is Lakeland\_terrier

4. The top 3 most liked dog breeds as a whole are:

1. Golden\_Retriever 1941007.0
2. Labrador\_Retriever 1274526.0
3. Pembroke 1039484.0

For Visualization, I asked the question, which year did the most retweet occur? I presented the information on a bar graph below:



In conclusion, the dataset is very rich and many insights could be derived from it. Also, note that the following insights are from year 2015 to 2017 and are only provisional as the dog breeds are derived from image prediction using Artificial Neural Networks. Thank you.

Victor Oladoja  
Author