

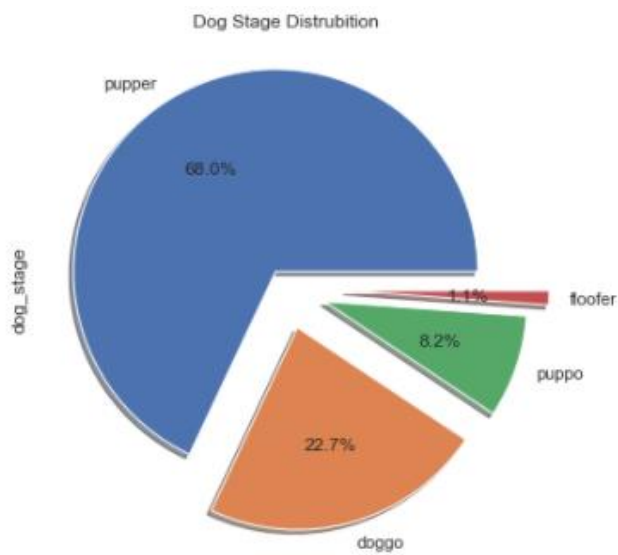
Data Wrangling Project

WeRateDogs Twitter data insights and visualization

Prepared by: Ahmed Senam Mostafa

March 2021

1- Tweets count distribution over dog stages:



From this chart, pupper has the most tweet counts while floofer is the least.

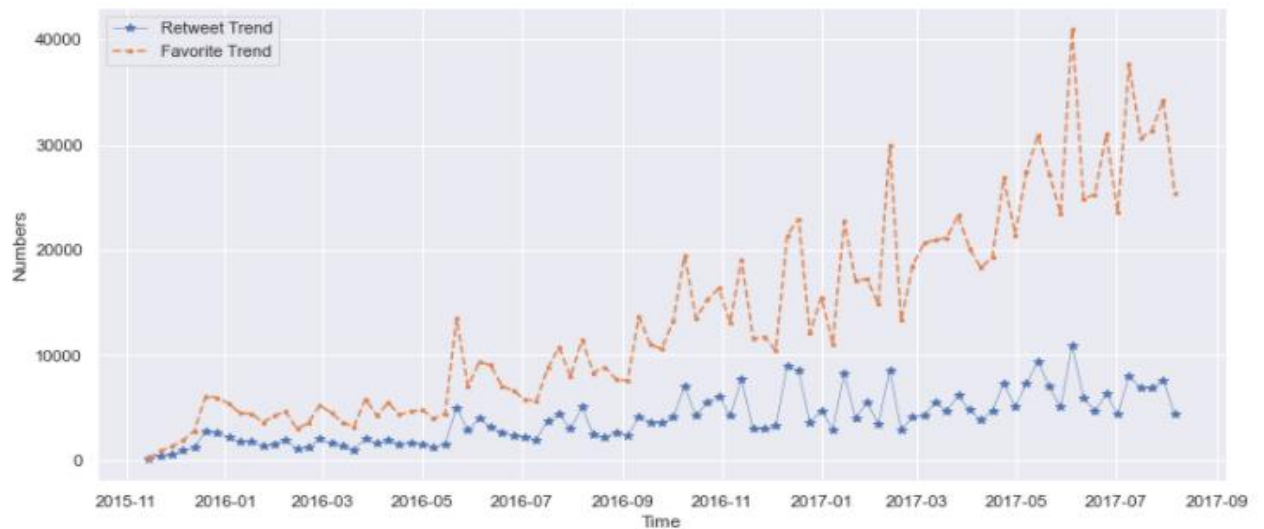
It seems that people like more small dogs that can live inside houses than large dogs

2- Inspecting data of “twitter archive master.csv”

| | rating_numerator | rating_denominator | retweet_count | favorite_count | img_num | p1_conf | p2_conf | p3_conf |
|-------|------------------|--------------------|---------------|----------------|---------|---------|---------|---------|
| count | 2097 | 2097 | 1371 | 1371 | 1971 | 1971 | 1971 | 1971 |
| mean | 12.189 | 10.449 | 2677.555 | 8489.507 | 1.202 | 0.595 | 0.135 | 0.06 |
| std | 40.365 | 6.645 | 4099.537 | 11272.485 | 0.559 | 0.272 | 0.101 | 0.051 |
| min | 0 | 2 | 16 | 81 | 1 | 0.044 | 0 | 0 |
| 25% | 10 | 10 | 614 | 1860.5 | 1 | 0.363 | 0.053 | 0.016 |
| 50% | 11 | 10 | 1344 | 4026 | 1 | 0.588 | 0.117 | 0.049 |
| 75% | 12 | 10 | 3175.5 | 10529.5 | 1 | 0.848 | 0.196 | 0.092 |
| max | 1776 | 170 | 56625 | 107015 | 4 | 1 | 0.488 | 0.273 |

- Rating_denominator → should be fixed at 10 but it is not the case, so will not use it in analysis
- Rating_numerator → should be less or equal to 10 but this is not the case, so we will take the absolute value of rating_numerator as rate value
- $P1_conf > p2_conf > p3_conf$ → which is logic and indicates that the neural network developed to identify the breed of dog works effectively. Also shows that the first prediction (p1_conf) has a higher confidence of being correct than the following predictions.

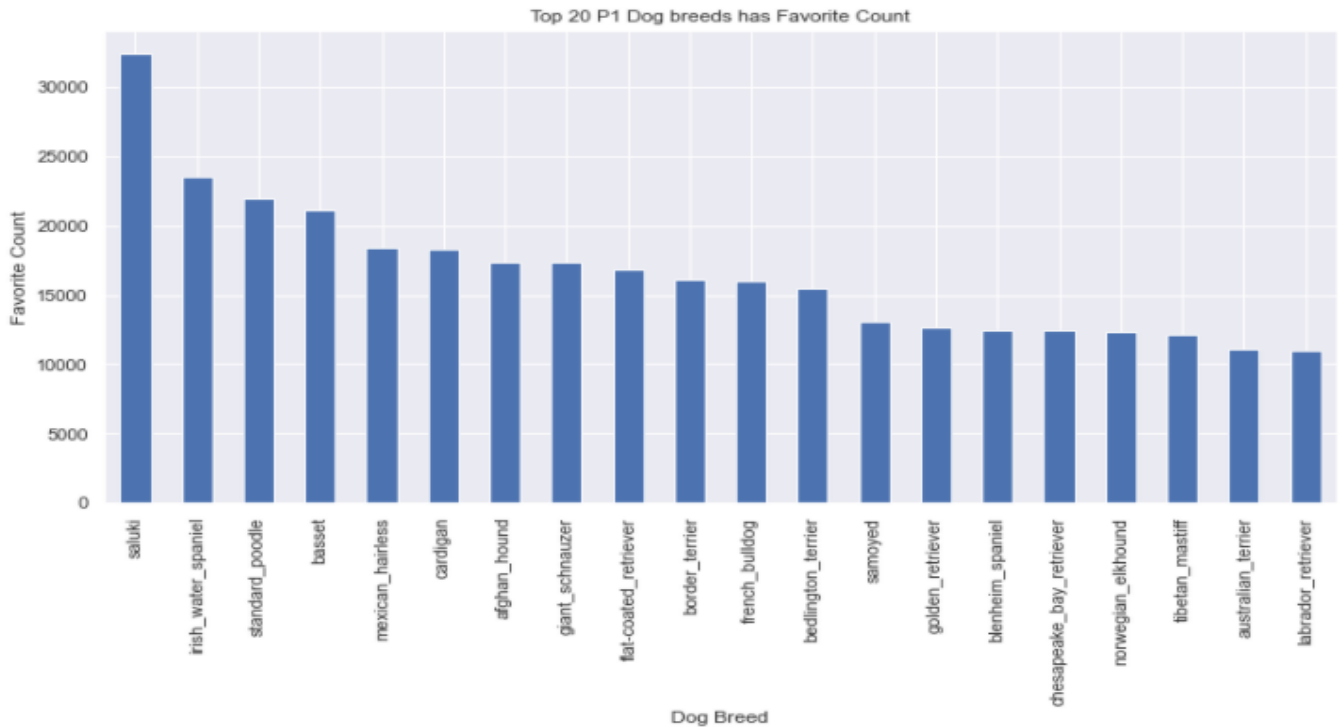
3- Retweets and Favorite counts trend:



From the above trends shows that:

- Both retweet and favorite counts are positively correlated
- People most likely to like the tweet and their number are increasing, while retweet trend is increasing slowly
- There are spikes in the trends, it seems that these intervals people become more active due to events or certain days of year

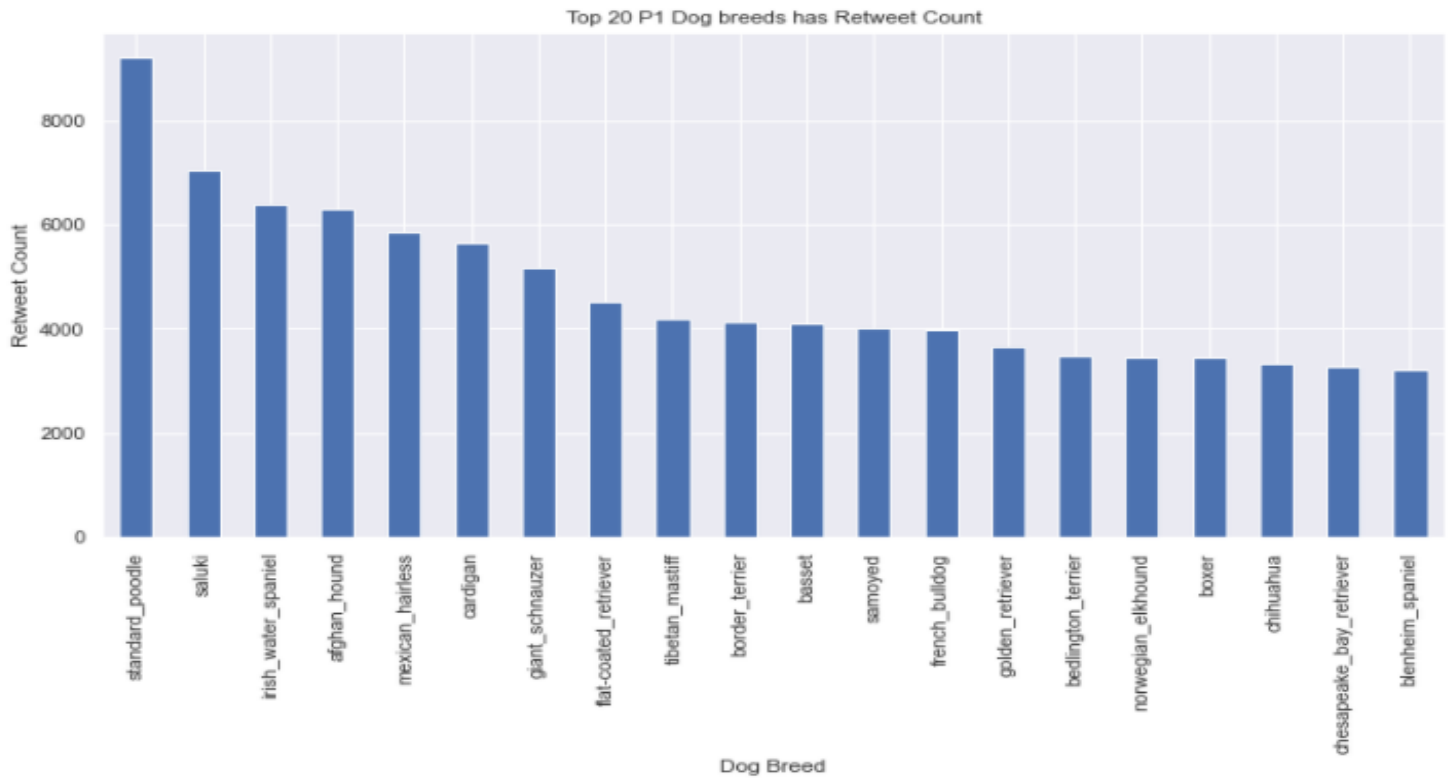
4- Top 20 dog breeds with favorite counts



- The Saluki is the top one of favorite counts, far from other breeds. It seems this is due to:
One of the oldest of dog breeds, Salukis were once considered a gift from Allah. They're fast as the wind, skinny as a supermodel, and quietly devoted to their people. Make great family pets for people who have enough time to dedicate to a demanding canine companion. Salukis form very strong bonds with their families and owners which means they like to be around them rather than to be left on their own



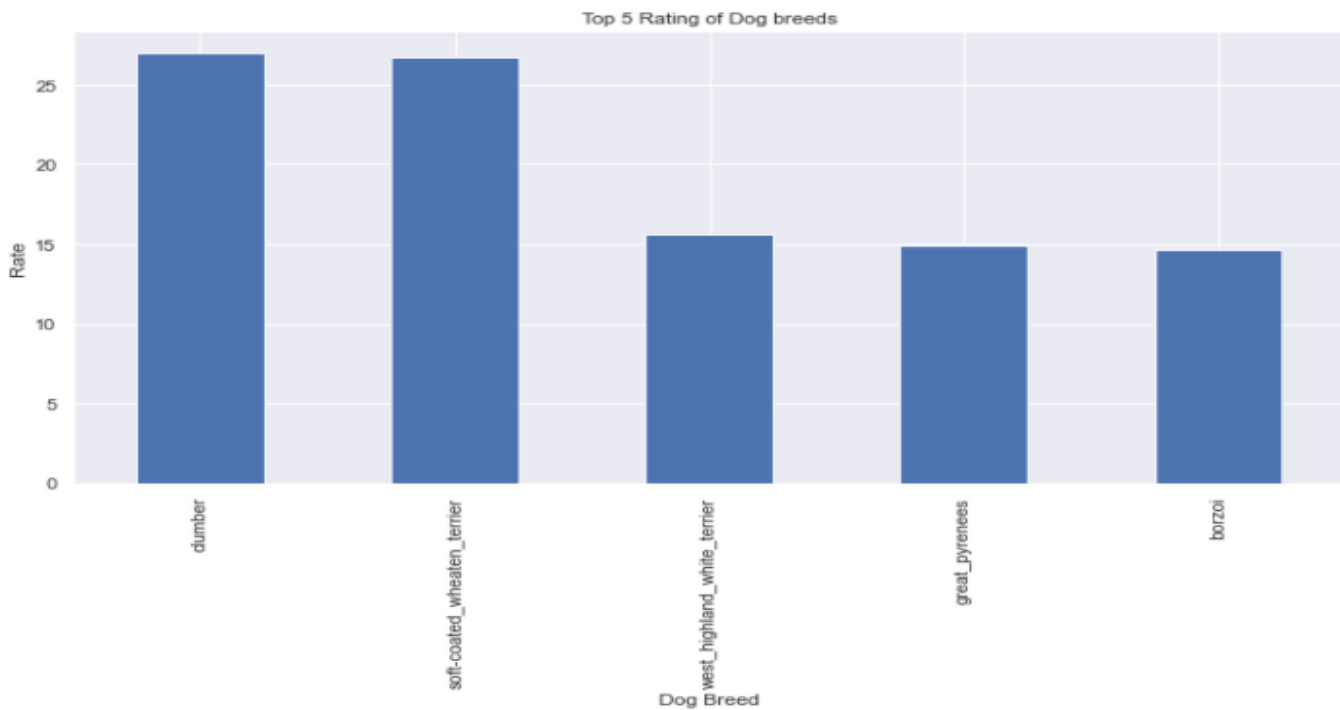
5- Top 20 dog breeds with Retweet counts



In retweet counts Standard poodle come in first place as poodles are good family dogs — fun, energetic, smart and easy to train. They do best with plenty of exercise for both mind and body and prefer to be with people most of the time. And Saluki in second place which is not far.



6- Top 5 Rated dog breeds

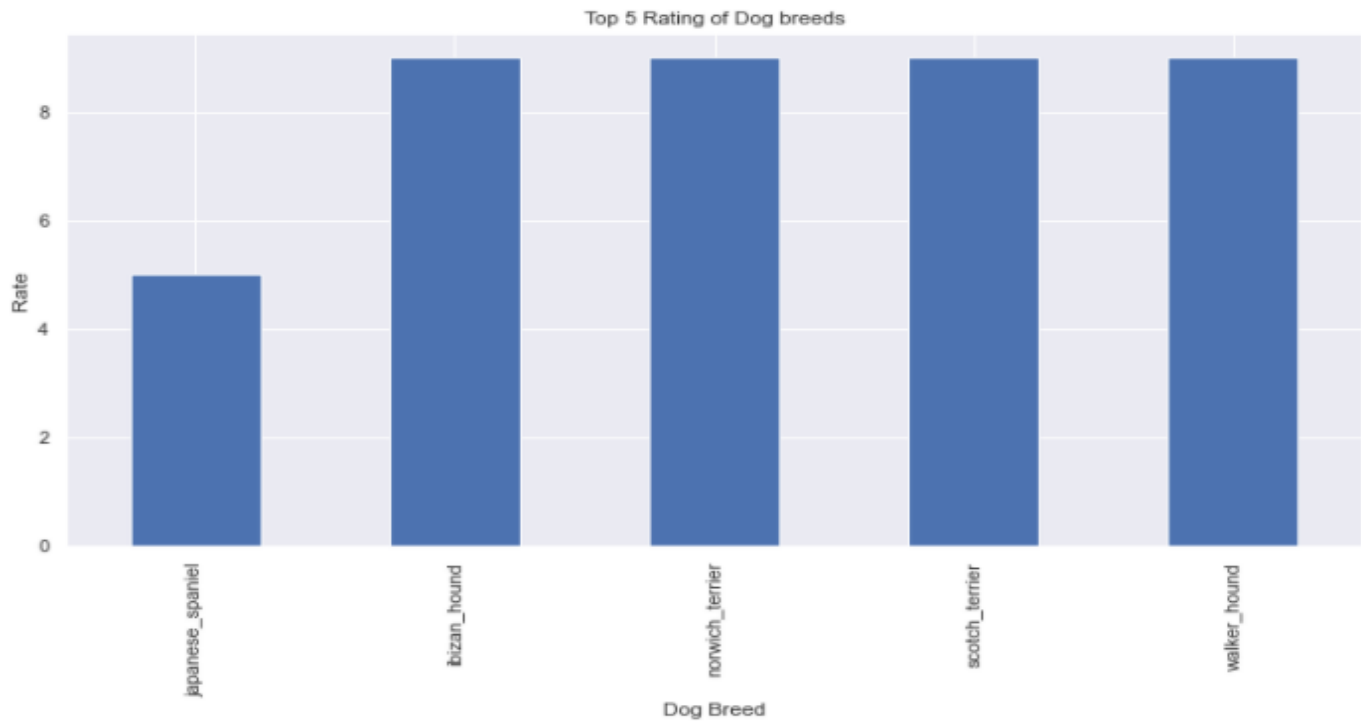


The top rated dog breed is Dumber.

The Soft Coated Wheaten Terrier comes next, as is full of energy, loves people, and makes a great family pet. They are native to Ireland. They were a "Poor Man's Dog". They performed every farm job. They are born in a variety of colors.



7- Least 5 Rated dog breeds



Japanese spaniel is the least ranked breed, also it is not included in the top favorites or retweets. Which makes it not favorable, lovable or owned by most of dog lovers and owners.

This could be due to common health issues in the Japanese Chin include luxating patellas (slipping kneecaps), cataracts, and early-onset heart murmurs. The Chin, as with smallest breed dogs, can also have a risk of hypoglycemia when aged under six months or weighing four to five lbs or less.

