

Data Analyst Nanodegree

ACT REPORT

TWITTER DATA WERATEDOG PROJECT



Prepared by : Lina Montrimaite

November 2, 2020

INTRODUCTION

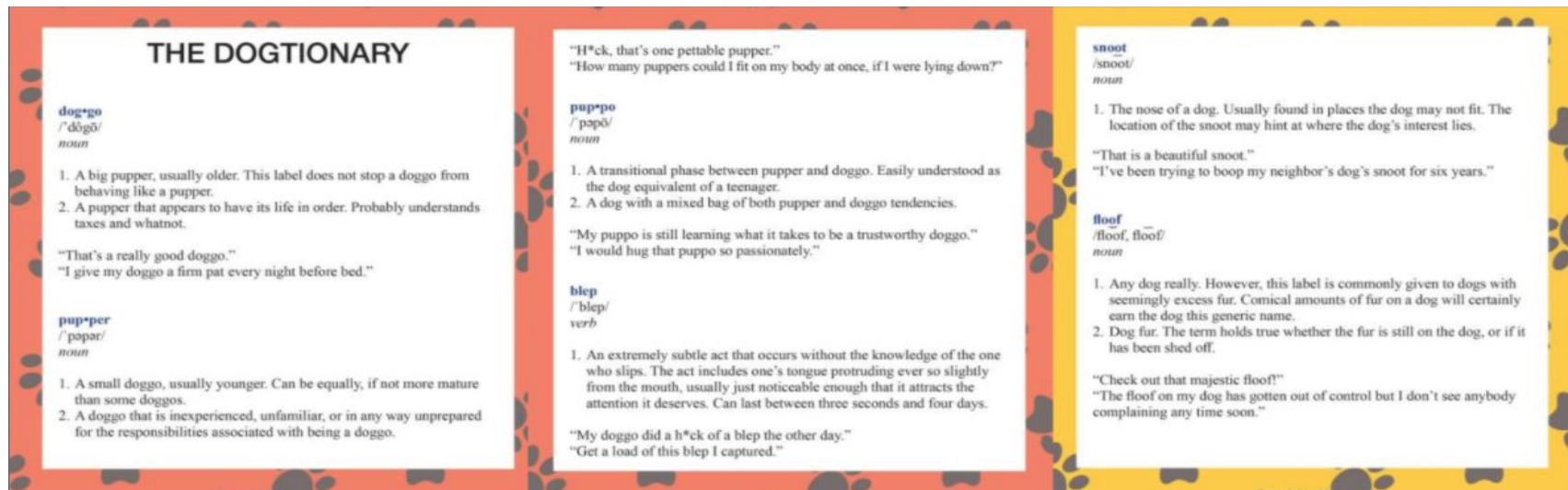
If you're the dog owner or love the dogs you probably heard about a Twitter account called WeRateDogs. On the account you can find people's tweets with their dogs' images, ratings and texts. The account is so popular with many tweets and has 8.8 Millions followers and 145.9 thousand likes.

I got familiarized more with the WeRateDogs account and data during my Udacity Data Analyst Nanodegree project. In this project I have wrangled and analyzed the tweet archive of Twitter user @dog_rates, also image prediction file and additional data via Twitter API. The goal was to wrangle Twitter data to create interesting and trustworthy analyses and visualizations.

I have choices analysis questions below:

1. Which are the most popular dog stages by count and most favorite?
2. Which are the most popular tweet sources by retweets?
3. Which are the most predicted dog breed of the tweets?
4. Which are the most likeable predicted dog breeds ?

One of the interesting things that WeRateDogs use is identifying the dog stages based on the Dogtionary. The Dogtionary explains the various stages of dog: doggo, pupper, puppo, and floof(er). See below for more information on these:



Which are the most popular dog stages by count and most favorite?

On this project I decided to look closer and analyze the dog stage popularity. After running pandas count function, I mention that the most of tweets don't have the dog state mentioned and decided to exclude the 'None' dog stage data from this analysis.

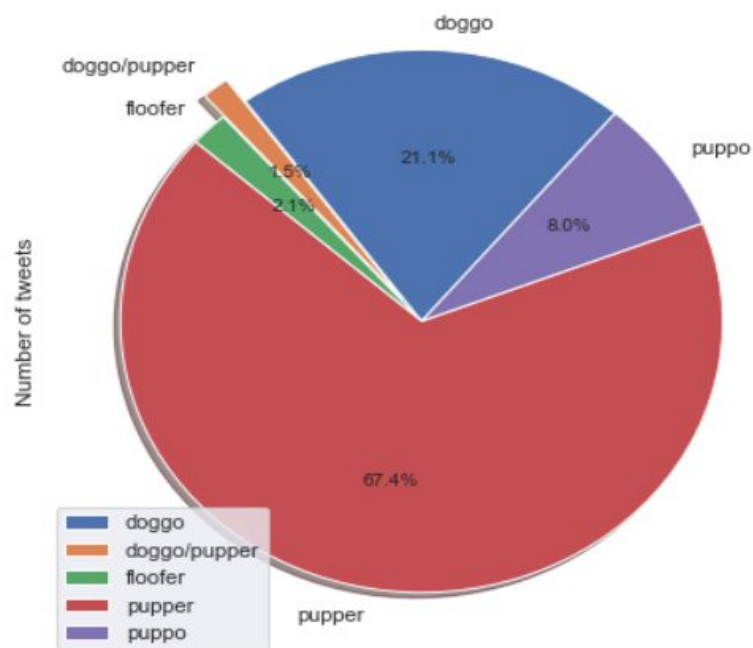
For this analysis I want to know which dog stage has the most tweets and which one receives more likes from the viewers.

So, I have prepared a table with total tweet counts, tweet likes counts and average likes per tweet value. You can check the table below:

| | Dog Stages | Number of tweets | Total likes received | Average likes |
|---|--------------|------------------|----------------------|---------------|
| 1 | pupper | 227 | 1445256 | 6366.766520 |
| 2 | doggo | 71 | 1218291 | 17159.028169 |
| 3 | puppo | 27 | 580764 | 21509.777778 |
| 4 | floofer | 7 | 82641 | 11805.857143 |
| 5 | doggo/pupper | 5 | 79647 | 15929.400000 |

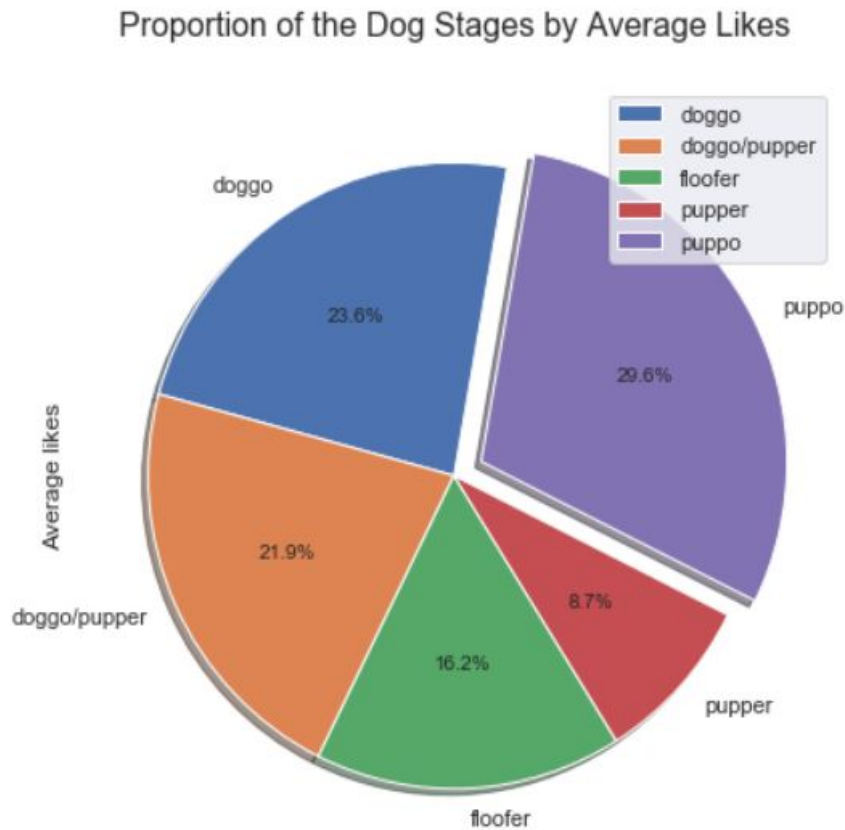
Let's see the proportion of the dog stages based on the count of tweets below:

Proportion of Dog Stages by Counting Tweets



As you can see in this pie bar visualization on the left the 67.4 % of tweets containing a "pupper" stage. The "Doggo" stage was the next most frequent stage at 21.1%. "Floofer" , "puppo" and combined dog stage "doggo/pupper" all of them leading to a combined 11.6% of the total stages. So, when the stages are provided, the most popular dog stage is "pupper".

During review of dog stages it was also interesting to count the likes per dog stages and calculate average likes count per tweet. Let's see the proportion of average likes per tweet base on dog stage below:



As you can see in this pie bar visualization on the left, the most likeable dog stage is “puppo” with 29.6 % of average likes per tweet. The "Doggo" stage was the next most likeable stage with 23.6% and “Doggo/pupper” combined stages followed with 21.9%. "Floofer" and "puppo" together have a total 24.9% proportion.

Which are the most popular tweet sources by retweets?

Twitter archive dataset provided the source of the tweet which was “Twitter for iPhone”, “Twitter Web Client”, “TweetDeck” and “Vine - Make a Scene”. I wanted to know which sources have the most tweets and which one receives more retweets. So, I have prepared a table with total tweet counts, tweet retweet counts and average retweet per tweet. You can check the table below:

| | Tweet Source | Number of tweets | Total retweet received | Average retweet by tweet |
|----------|---------------------|-------------------------|-------------------------------|---------------------------------|
| 0 | Twitter for iPhone | 1948 | 4675066.0 | 2399.931211 |
| 1 | Twitter Web Client | 28 | 62338.0 | 2226.357143 |
| 2 | TweetDeck | 10 | 17721.0 | 1772.100000 |

The created table shows that Twitter for iPhone source is the most popular tweet source with the 1948 number of tweets and has the highest average retweet count with 2399.93. The "Twitter Web Client" source has 28 numbers of tweets and the “TweetDeck” source has 10 number of tweets.

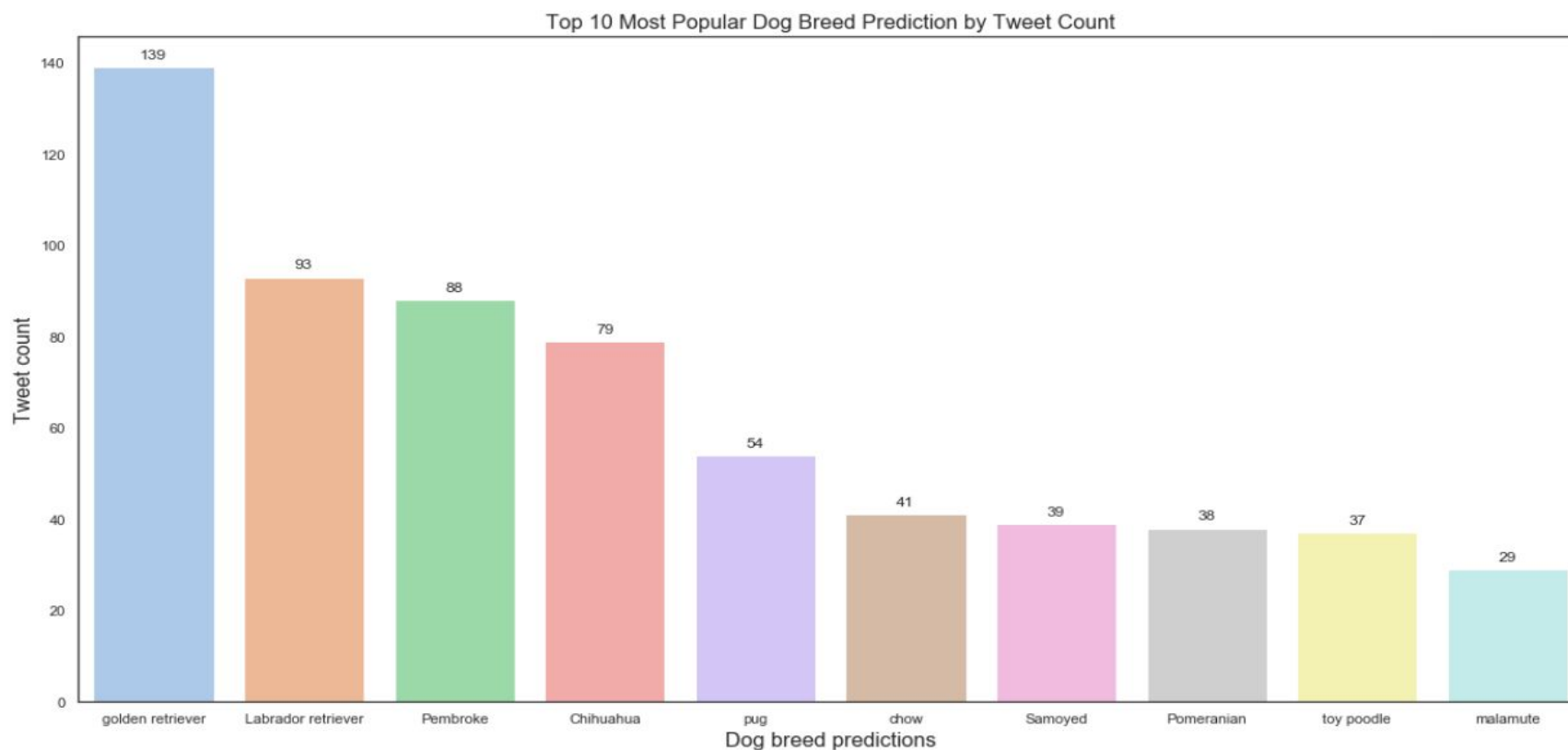
As the project dataset of Tweet image predictions was provided, I wanted to analyze which dog breed had most predicted and was most likeable. After combining three image predictions in one column the count value, the total likes count and the average likes per tweet value was calculated.

| | Prediction | Number of tweets | Total likes received | Average likes per tweet |
|-----------|--------------------|-------------------------|-----------------------------|--------------------------------|
| 1 | golden retriever | 139 | 1547912 | 11136.057554 |
| 2 | Labrador retriever | 93 | 972809 | 10460.311828 |
| 3 | Pembroke | 88 | 911241 | 10355.011364 |
| 4 | Chihuahua | 79 | 630369 | 7979.354430 |
| 5 | pug | 54 | 285838 | 5293.296296 |
| 6 | chow | 41 | 371767 | 9067.487805 |
| 7 | Samoyed | 39 | 470287 | 12058.641026 |
| 8 | Pomeranian | 38 | 278413 | 7326.657895 |
| 9 | toy poodle | 37 | 213214 | 5762.540541 |
| 10 | malamute | 29 | 228119 | 7866.172414 |

So we can answer the questions followed.

Which are the most predicted dog breeds of the tweets?

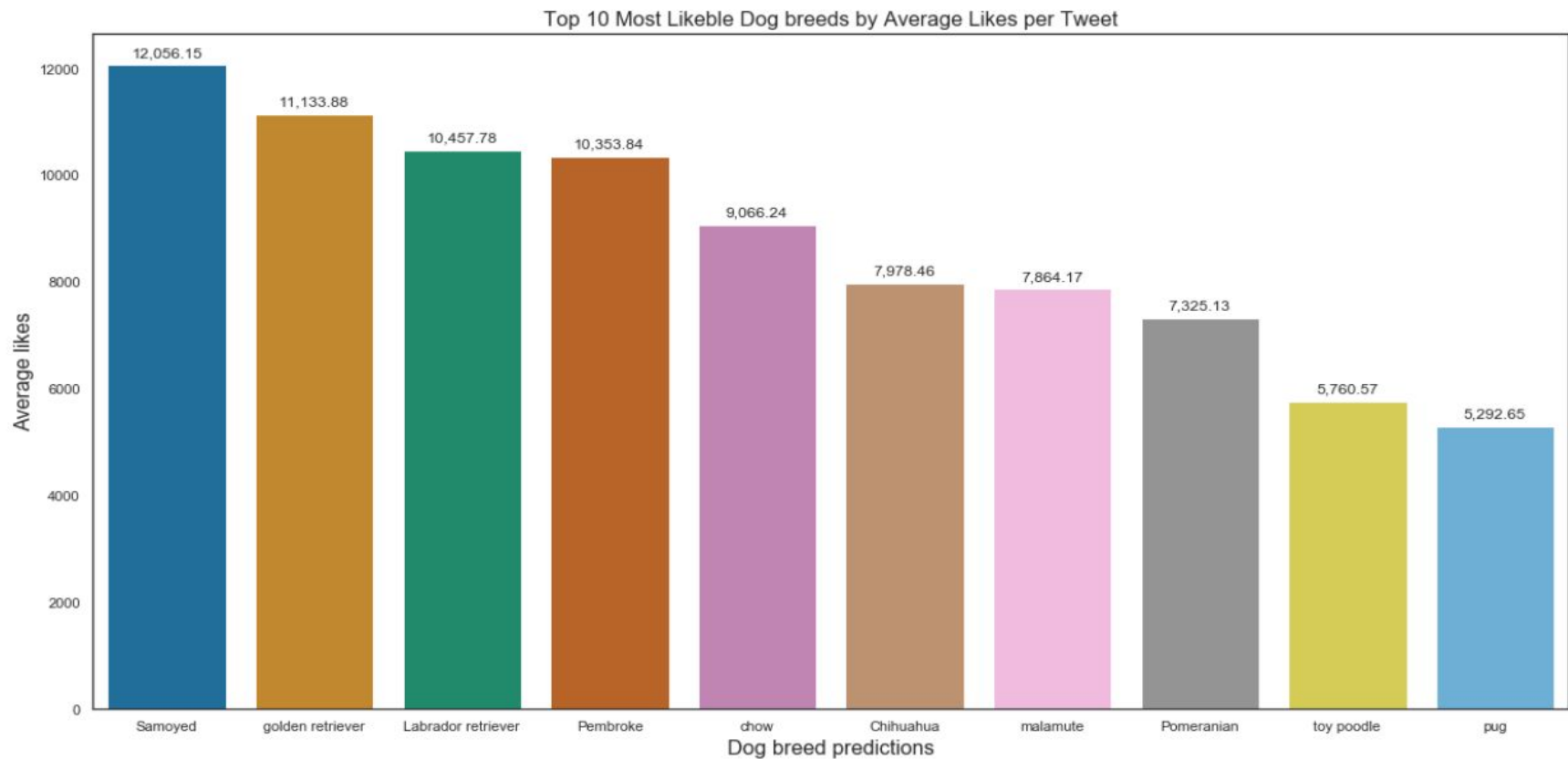
My analyst was focused on the top ten most popular dog breeds detected, so just the first top breed was displayed on the bar plot below.



From the bar plot we can see that Golden retriever dog breed prediction is most popular in the dataset and has appeared 139 times. Second place taken by Labrador retriever appeared 93 times. After two retrievers breeds follow Pembroke (88), Chihuahua (79), Pug (54) and others.

Which are the most likeable predicted dog breeds ?

As my analyst was focused on the top ten most popular dog breeds I wanted to show which dog breed received more likes. I have calculated total likes count by breed predicted and multiplied by count of tweets and received average likes per tweet value. The first ten top breeds were displayed on the bar plot below.



The bar plot above shows that the Samoyed dogs images received more likes (12,058.64 avg.

likes) than the rest of the dog breeds based on average likes by count of tweets. Second and third places taken by Golden and Labrador retriever with 11,136 and 10,460 average likes. After them follow Pembroke, Chow, Chihuahua, Malamute and others.

I hope that you have enjoyed looking through this analysis and I wish you a good day.