Report| Analyzing & Visualizing Insights For 'WeRateDogs' Data

Udacity-FWD Initiative| Professional Data Analysis Nanodegree

Prepared By: Muhammed Abdul-Fattah Salem

Project Overview

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

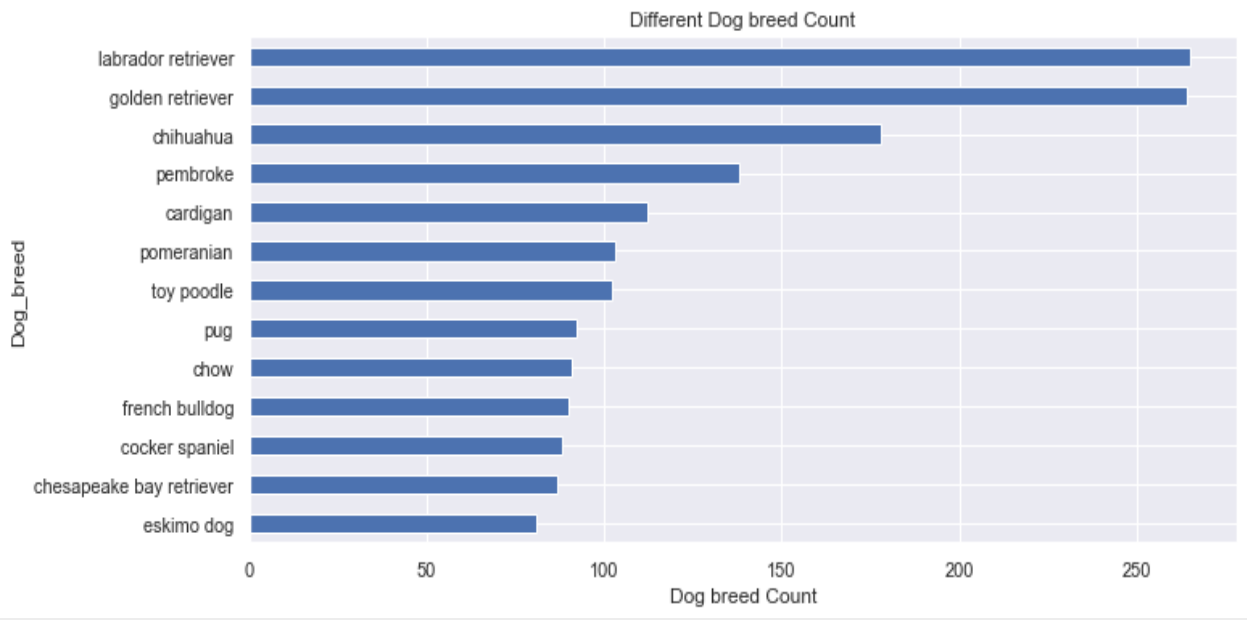
Wrangling 'WeRateDogs' Twitter data for insightful and trustworthy analysis and visualizations will be covered in this report.

Resulting Insights & Visualizations:

'WeRateDogs' datasets contain about 5000+ tweet data for their account on twitter, started from August 1, 2017. After analyzing datasets, I had cleaned versions of these datasets containing about 1900+ of tweet-data attached in csv file: 'Twitter_archive_master'. Let us focus on insightful and interesting facts out of their data.

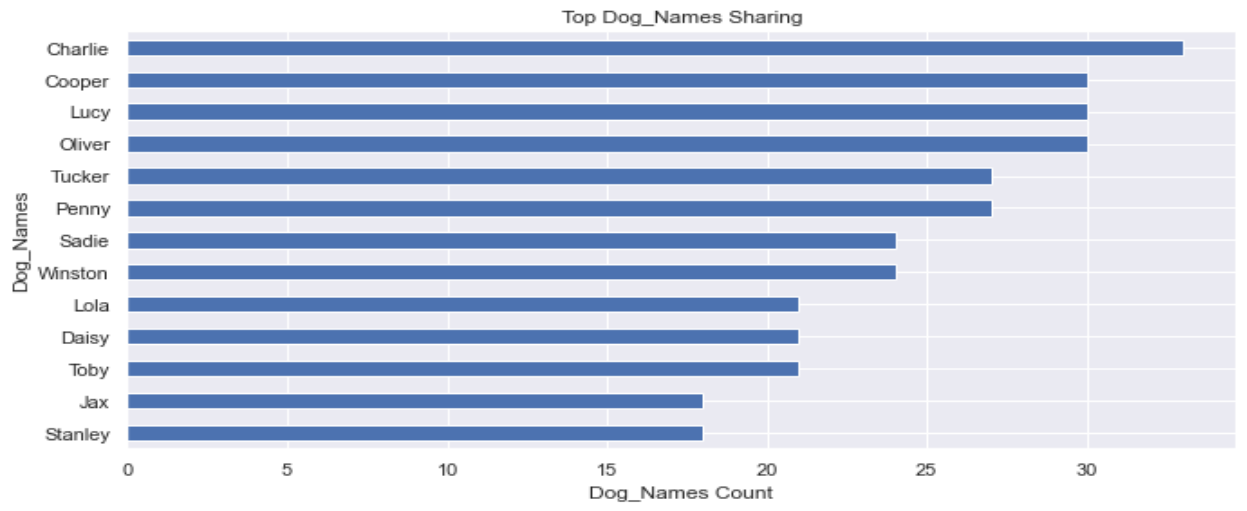
Most Common Dog Breeds

The figure as shown below demonstrates The highest 12 dog breeds. 'labrador retriever' and 'golden retriever' are the highest with 260 score, doubling the next 'chihuahua' score with about 130.

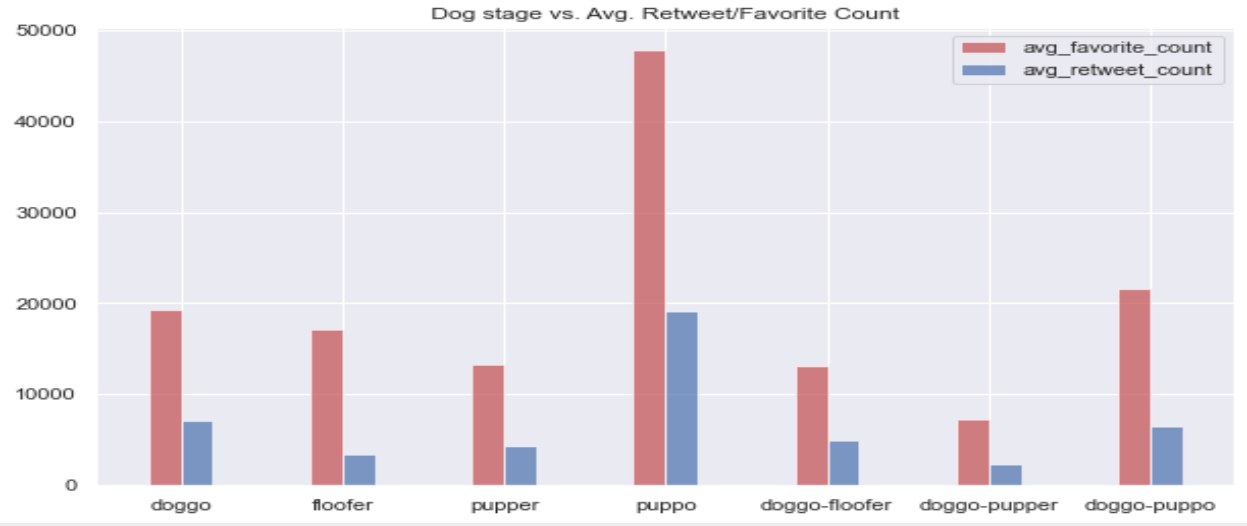


Most Common Dog Names

The figure as shown below demonstrates Top dog names. 'Charlie' is the most used with score of 33.



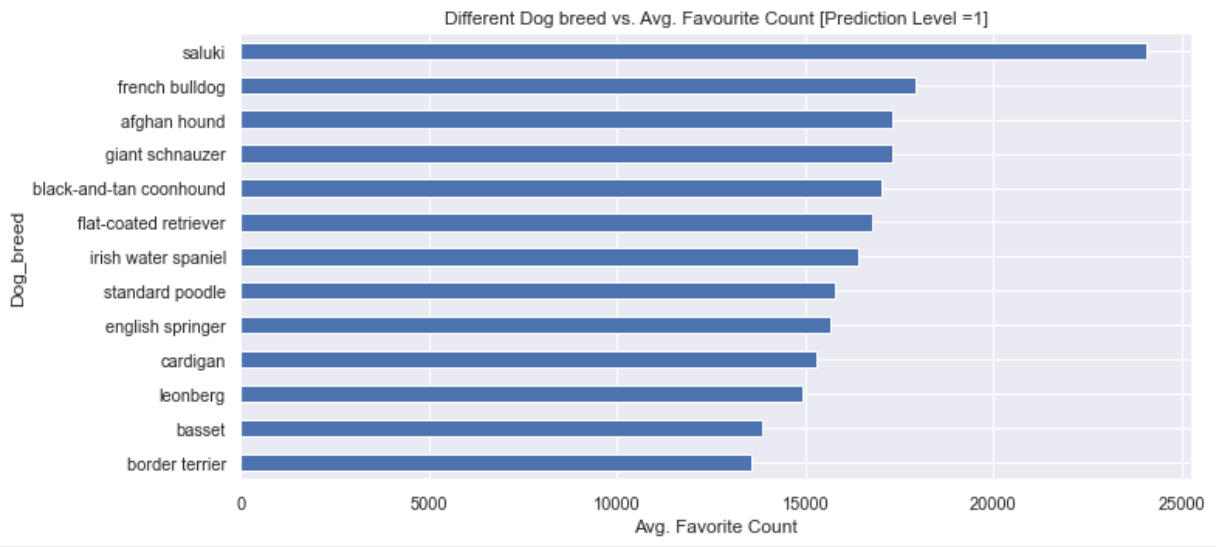
Dog Stages vs. Retweet- Favorite Count

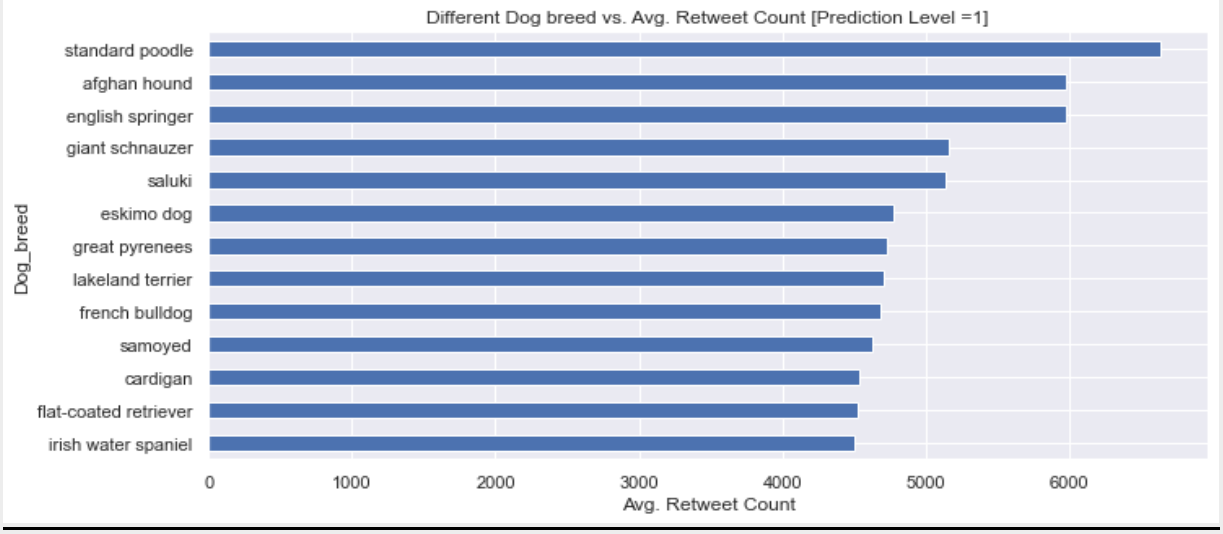


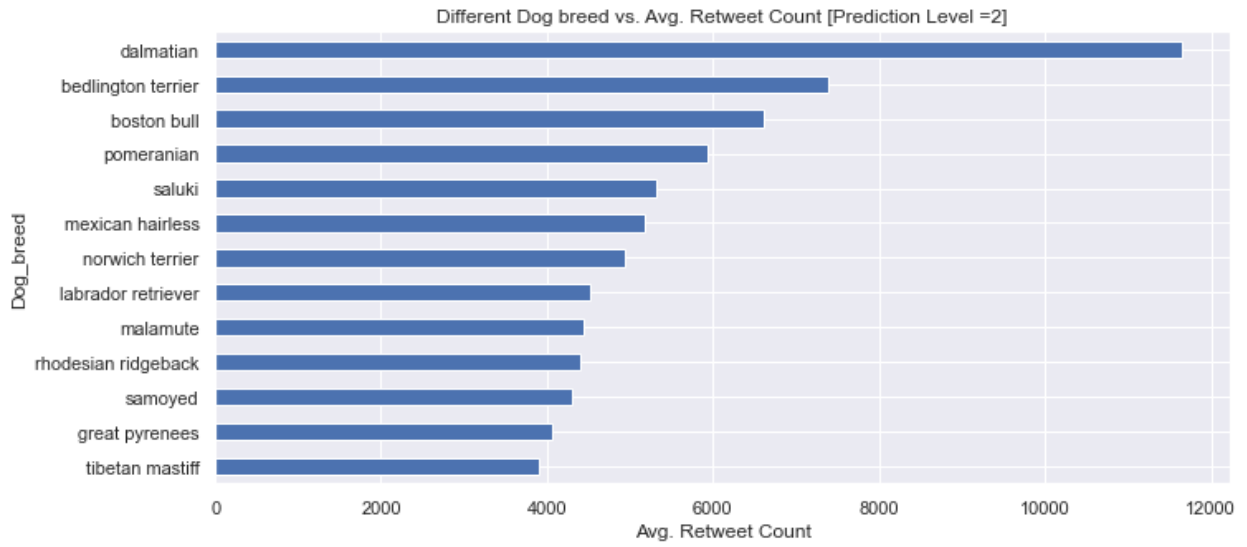
The figure as shown above tells us how much people interacting with favorites and retweets with 'puppo' dogs with the highest score between othe dog stages for around 140k+ avg. favorites score and 18k for avg. retweets. Another interesting insight is that people do love interacting with images for the same dog in different stages, make 'doggo-puppo' the next score in avg. favorites with around 22k+ score.

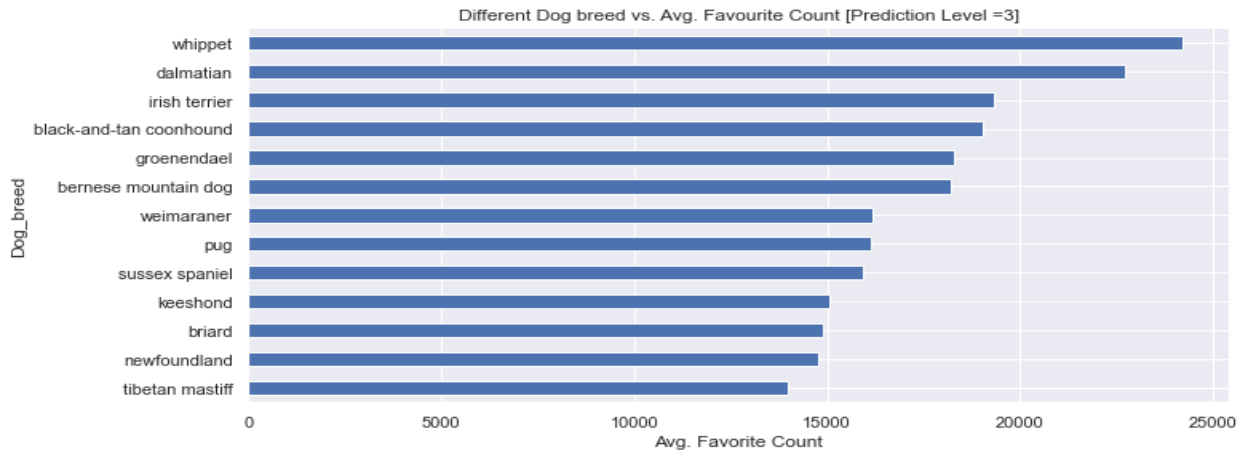
Dog Breed vs. Retweet- Favorite Count

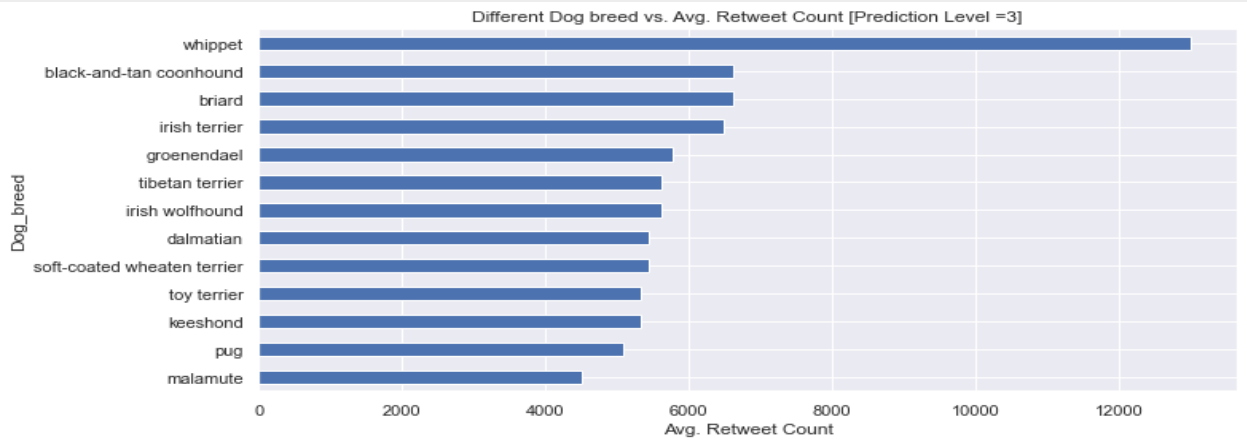
For this insight, I am trying to investigate how much people interacting with different dog breeds referenced to prediction_level; this results contribute my investigation for orignal images showing how it is important to make analysis confined to the highest prediction_level: 3.


Different Dog breed vs. Avg. Favourite Count [Prediction Level =1]
saluki
french bulldog
afghan hound
giant schnauzer
black-and-tan coonhound
flat-coated retriever
irish water spaniel
standard poodle
english springer
cardigan
leonberg
basset
border terrier
Dog_breed
0
5000
10000
15000
20000
25000
Avg. Favorite Count


Different Dog breed vs. Avg. Retweet Count [Prediction Level =1]
standard poodle
afghan hound
english springer
giant schnauzer
saluki
eskimo dog
great pyrenees
lakeland terrier
french bulldog
samoyed
cardigan
flat-coated retriever
irish water spaniel
Dog_breed
0
1000
2000
3000
4000
5000
6000
Avg. Retweet Count


Different Dog breed vs. Avg. Retweet Count [Prediction Level =2]
dalmatian
bedlington terrier
boston bull
pomeranian
saluki
mexican hairless
norwich terrier
labrador retriever
malamute
rhodesian ridgeback
samoyed
great pyrenees
tibetan mastiff
Dog_breed
0
2000
4000
6000
8000
10000
12000
Avg. Retweet Count





Statistics' Visualizations

Visualizations as shown below tell us intersting findings such: avg. retweet/favorite count for each of dog breeds, most source of intaracts, most used name for dogs.

```
<Doggo>
Mean Retweet Count:      7126
Mean Favorite Count:     19356
<Doggo-Floofer>
Mean Retweet Count:      3433
Mean Favorite Count:     17169
<Doggo-Pupper>
Mean Retweet Count:      4397
Mean Favorite Count:     13220
<Doggo-Puppo>
Mean Retweet Count:      19196
Mean Favorite Count:     47844
<Floofer>
Mean Retweet Count:      4969
Mean Favorite Count:     13206
<Pupper>
Mean Retweet Count:      2383
Mean Favorite Count:     7251
<Puppo>
Mean Retweet Count:      6474
Mean Favorite Count:     21582
```

```
# Visualizing Most Sources iteracting with WeRateDogs Twitter Account

master_df['source'].value_counts()
```

```
Twitter for iPhone    5796
Twitter Web Client      84
TweetDeck               33
Name: source, dtype: int64
```

```
# visualizing Most Rated Dog_names
master_df.name.value_counts()
```

```
Charlie    33
Cooper     30
Oliver     30
Lucy       30
Tucker     27
           ..
Obie        3
Brat        3
Emma        3
Strider     3
Fwed        3
Name: name, Length: 931, dtype: int64
```

Conclusion

There are a lot of insights for further working on these data, we are just picking few interesting ones.

References

- https://nfpdiscussions.udacity.com/t/weratedogs-data-wrangling-gathering-data-part/67534/2
- https://nfpdiscussions.udacity.com/t/cleaning-quality-issues/67893
- https://www.kite.com/python/answers/how-to-check-if-a-character-is-uppercase-in-python
- https://nfpdiscussions.udacity.com/t/concerning-name-column/42212/5
- https://stackoverflow.com/questions/7353968/checking-if-first-letter-of-string-is-in-uppercase/7354011#7354011
- https://nfpdiscussions.udacity.com/t/project-2-assessing-data/44649/6
- https://nfpdiscussions.udacity.com/t/insights-and-visualization-weratedogs/40725
