



# Who Are the Goodest Doggos on WeRateDogs?

## UDACITY Data Analysis Nanodegree

### Project:- Wrangle & Analyze Data from WeRateDogs Twitter, Analysis Report

*Grant Patience, 29th August 2019*

## Introduction

This project focused on wrangling data from the [WeRateDogs Twitter \(https://twitter.com/dog\\_rates\)](https://twitter.com/dog_rates) account using Python, documented in a Jupyter Notebook (wrangle\_act.ipynb and the subsequent analysis in act\_analysis\_notebook.ipynb). This Twitter account rates dogs with humorous commentary. The rating denominator is usually 10, however, the numerators are usually greater than 10. WeRateDogs has over 4 million followers and has received international media coverage. Each day you can see a good doggo, lots of floofers and many pupper images.

WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for us to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017. The data is enhanced by a second dataset with predictions of dog breeds for

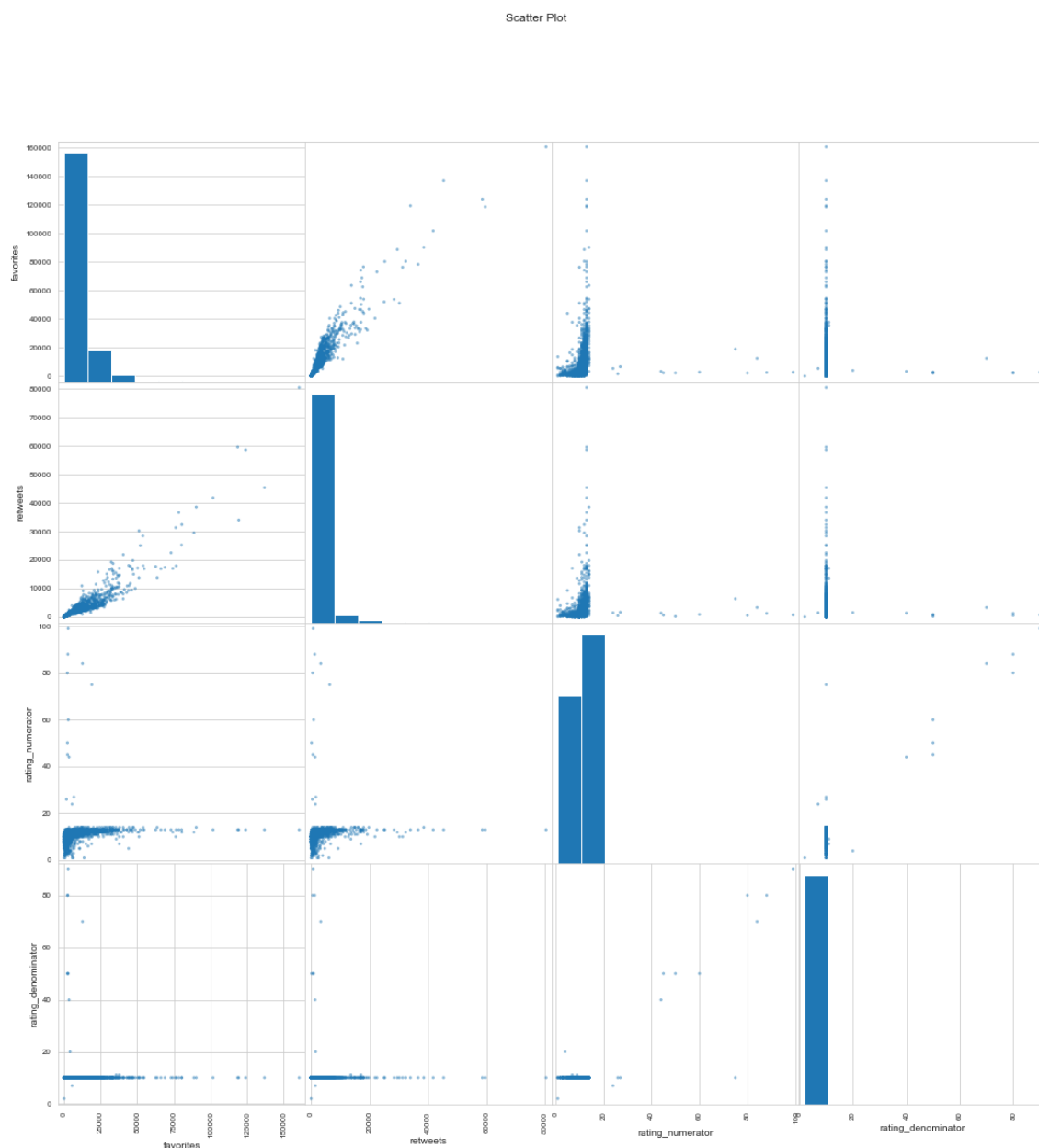
each of the Tweets. Finally, we used the Twitter API to glean further basic information from the Tweet such as favourites and retweets.

Using this freshly cleaned WeRateDogs Twitter data , interesting and trustworthy analyses and visualizations can be created to communicate back our findings.

## What Correlations can we find in the data that make a Good Doggo?

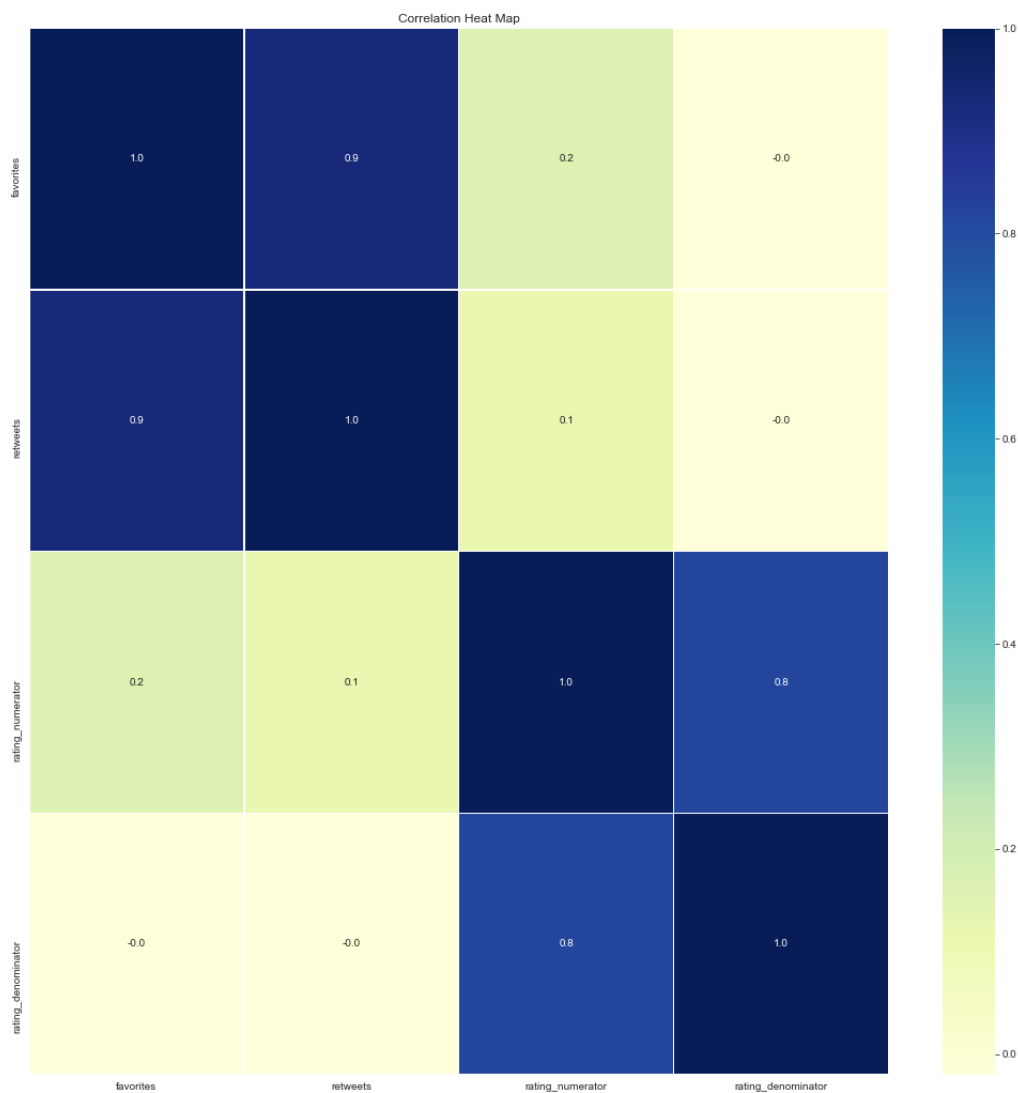
First, we wanted to determine if there were any correlations in the data to find any interesting relationships. To do this we performed some correlation analysis and produced visuals to support that. Prior to the analysis, we assumed that Favourite & Retweet would be correlated since these are both ways to show your appreciation for a tweet on Twitter.

The output of our analysis is as follows



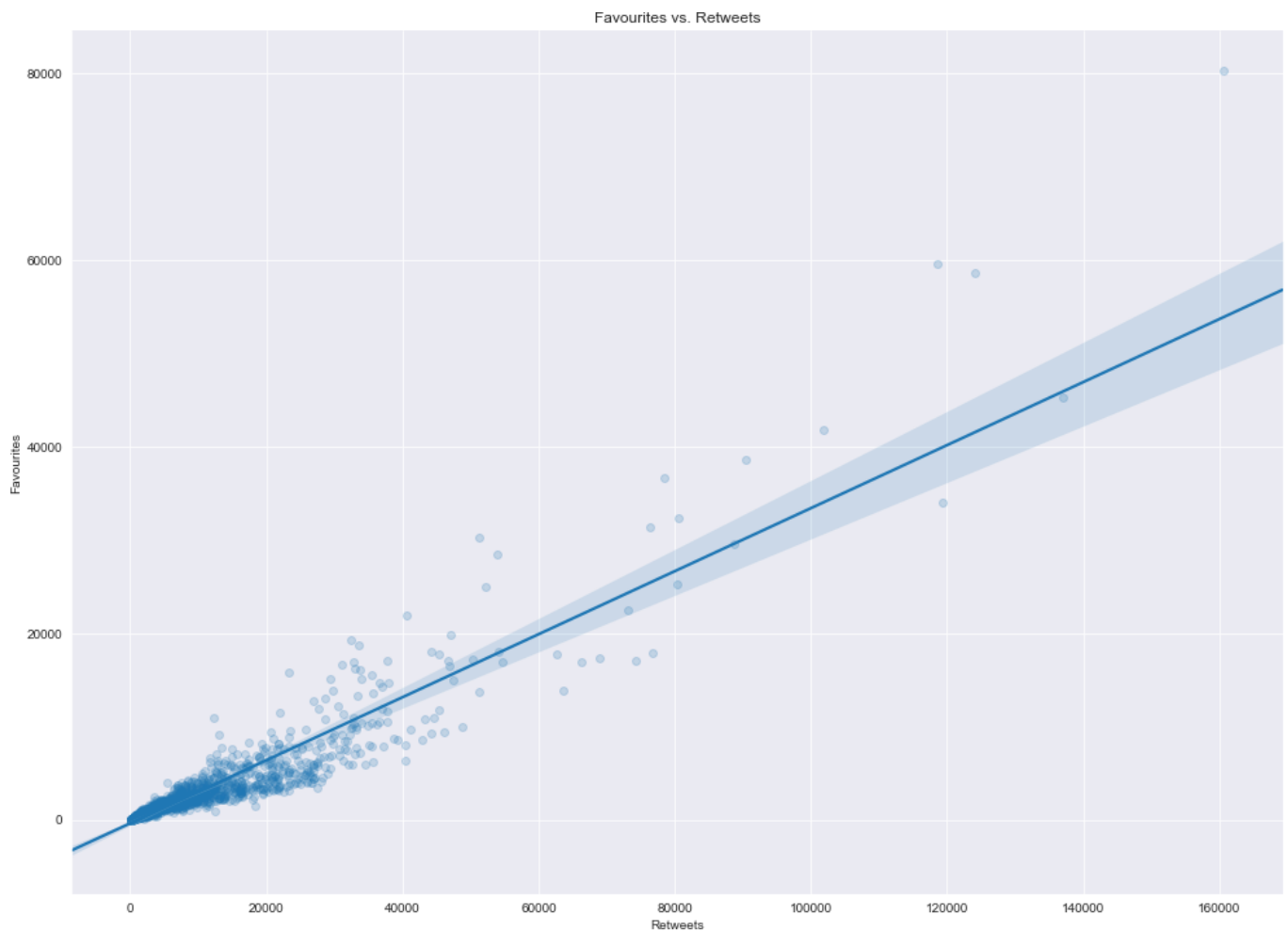
This scatter plot matrix shows the relationships between each of the variables. It shows that while there looks to be a strong linear relationship between Favourites and Retweets, there no othe relationships were highlighted.

With that in mind, we wanted to quantify these relationships to solidify our understanding



Again, the above heatchart shows our correlation relationships, showing a strong relationship between favourites and retweets with a correlation coefficcient of aprox ( $r = 0.8$ )

Let's narrow in on just that relationship.



With this chart we can see Favourites verses Retweets and it's strong linearly positive relationship

### Observations

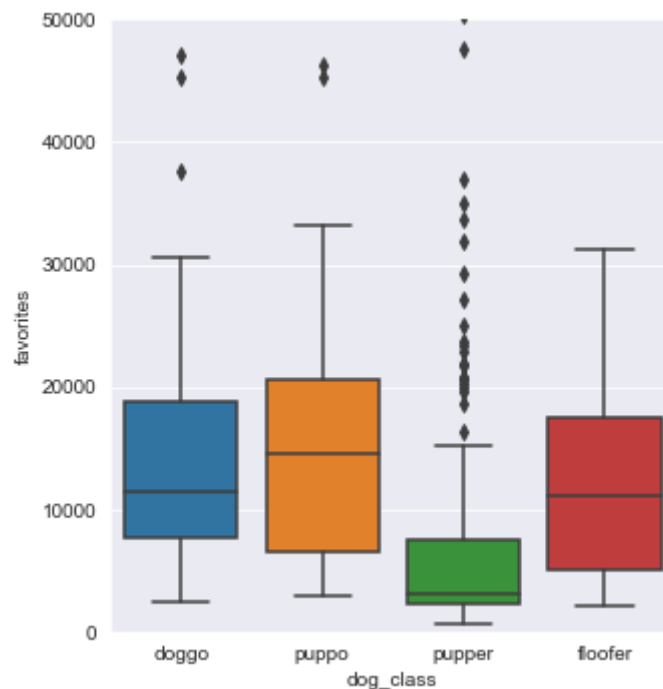
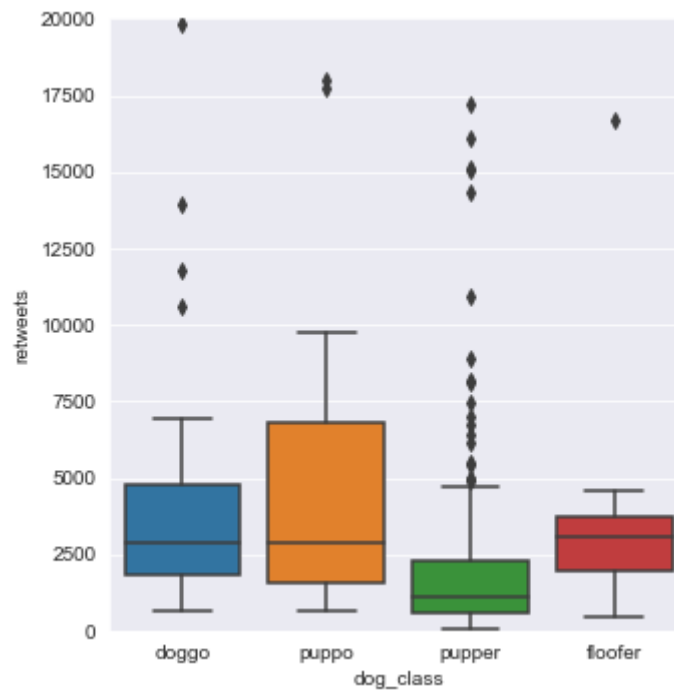
- As we assumed, There is a **strong linear relationship between Favourites and Retweets**.
- The regression coefficient for this relationship is ( $r = 0.797$ )
- From the points we plotted, we cannot find any other correlations.
- In future, we could try and categorize the source and dog\_stage to investigate correlations there with popularity of the Tweet.

### Which are the more popular; doggos, puppies, fluffers or poppos?

We performed some data wrangling on the tweet\_archive dataset to integrate 4 different "Class" of doggos down into one column which would be easier to analyse.

These classes are fun terminologies used by WeRateDogs so it would be really cool to see the popularity of these different types ( `dog_class = [doggo, pupper, fluffer, puppo]` )

Can we ascertain which category of dog is more popular?



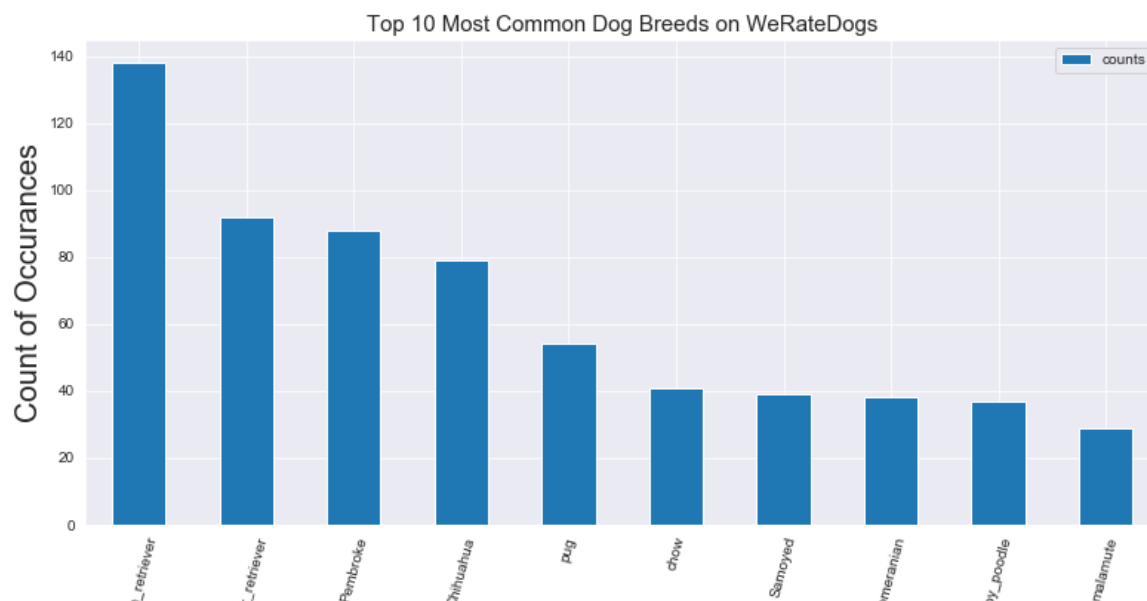
## Observations

- Interesting, as we look at Retweets and Favourites, **Puppops** are by far the more popular on average containing the higher number of favourites and retweets
- From the points we plotted, we can see that **Puppers** have the lower nubmers on average, there are a lot of outliers.

## Which are the more popular doggo breeds and why is it Spaniels?

Everyone loves doggos, but we all have a different favourite kind. With so many to choose from, **which breed really is the goodest doggo and why is it Spaniels?**

By integrating the image\_prediction data into our dataset, we have three columns denoting the probability chance of the image being of a particular breed. This is some really interesting data to use, lets use it to see if we can determine the popularity of certain breeds of doggos.



## Observations

- We can see the most common types of dog here are **Golden Retrievers and Labrador Retrievers**, this seems sensible since these dog types are very common. Other dog breeds rounding out the top 5 are **Chihuahuas, Pugs and Pembroke**.
- We could also limit the probability to ensure it meets a minimum probability level
- Some incorrect values like Seat-Belt, hamster, bath towel still exist in the data which we could clean given more time in future
- We only used the 1st prediction column, we may have been able to use all 3 to determine the overall probability or popularity of dog breeds
- There must be some mistake, **Spaniels** were not even in the top 10!?

## 4. Observations and Conclusion

- During our analysis, we found that there is a **strong linear relationship between the number of Favourites and the number of Retweets** of a given Tweet. The regression coefficient for this relationship is ( $r = 0.797$ )
- We did anticipate this relationship already since there is a fair chance that if a user enjoys a tweet they have the choice option to Favourite or Retweet it - both are a measure of the users enjoyment of the tweet.
- We have also found through visualisation and data wrangling that the **pupper** is the most popular doggo, with on average, more Retweets and more Favourites per tweet than the other 3 categories Doggo, Fluffer and Puppo.
- **Golden retriever** are the goodest doggos, **Labrador Retriever, Pembroke, Chihuahua and Pugs** complete the top 5 common dog breeds in the data. -

## Goodest Doggos





---

## References

- Title Image (<https://pixabay.com/photos/rottweiler-puppy-dog-dogs-cute-1785760/>  
(<https://pixabay.com/photos/rottweiler-puppy-dog-dogs-cute-1785760/>))
- Conclusion Image (<https://pixabay.com/photos/puppy-cocker-spaniel-cute-dog-318161/>  
(<https://pixabay.com/photos/puppy-cocker-spaniel-cute-dog-318161/>))

In [ ]: