

Project 7 - Wrangle and Analyse Data

Analysis, Insights and Visualizations Report on “WeRateDogs” Twitter Dataset

by Dilip Rajkumar

Introduction

In this report we will describe the analysis, insights and visualisations of our cleaned [WeRateDogs™](#) twitter dataset. WeRateDogs is a famous twitter handle which rates dog in an unusual and humorous manner. The twitter account gained popularity after a [meme](#) (*They're Good Dogs Brent*) became viral.



For the analysis i looked at the most retweeted and most favorited tweets, relationship between retweets and favorites, gender of dogs and popular dog breeds.

Most Popular Tweets

The tweet with the highest retweet count of 77,675 belongs to [this tweet video](#) of this Labrador dog walking in the pool.

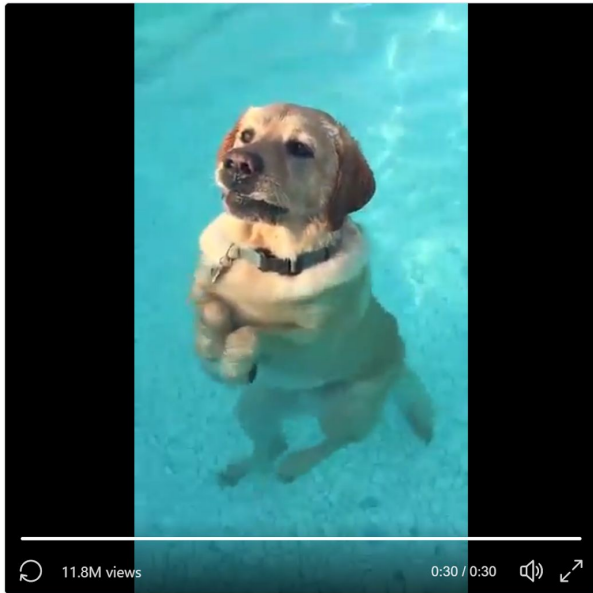
The tweet with the most favorites_count of 143,883 belongs to [this tweet](#) of a dog dressed up for a #WomensMarch event.



WeRateDogs™
@dog_rates

Follow

Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad)



11:26 AM - 18 Jun 2016

77,665 Retweets 128,660 Likes



1.1K 78K 129K



WeRateDogs™
@dog_rates

Follow

Here's a super supportive puppo participating in the Toronto #WomensMarch today. 13/10



10:26 AM - 21 Jan 2017

49,347 Retweets 143,869 Likes



Mary-Margaret Jones

720 49K 144K



WeRateDogs™ @dog_rates · 21 Jan 2017

Replying to @dog_rates

This dog's Instagram is @lennonsadventures if you'd like to give him a follow

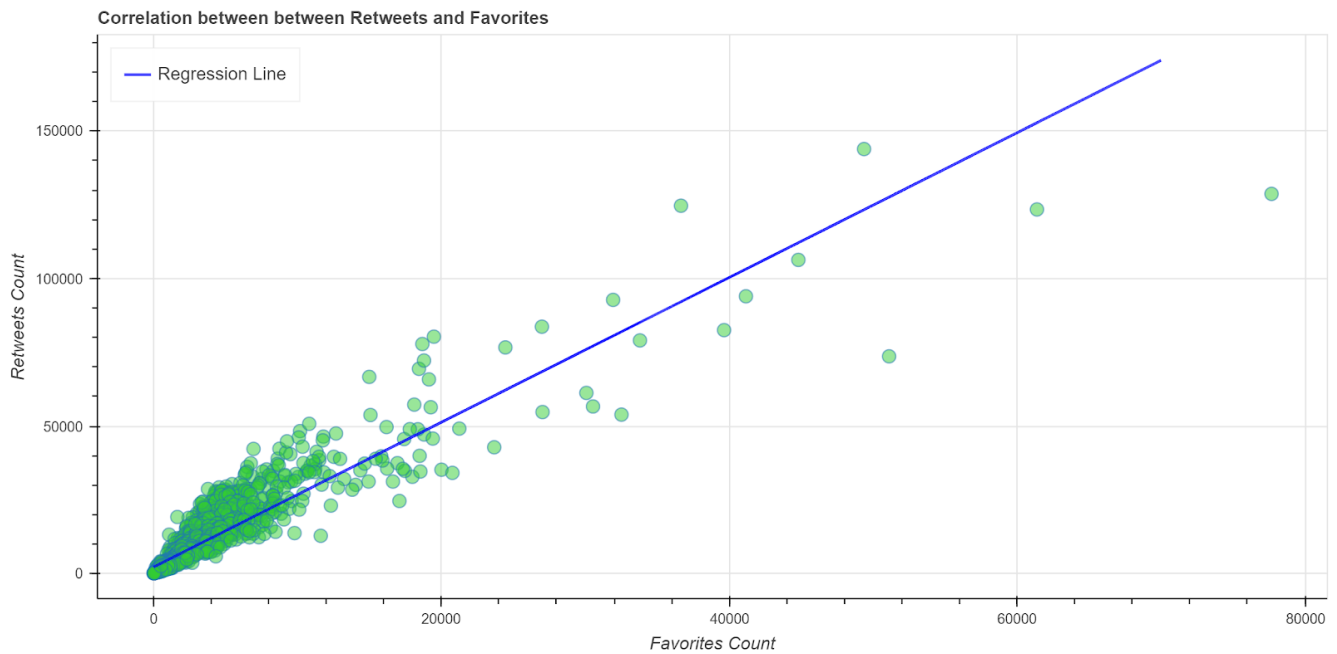
12 77 2.1K

Tweets by Day of Week:

- Quiet day of week with lowest tweets: **Saturday**
- Busiest day of week with most tweets: **Tuesday**

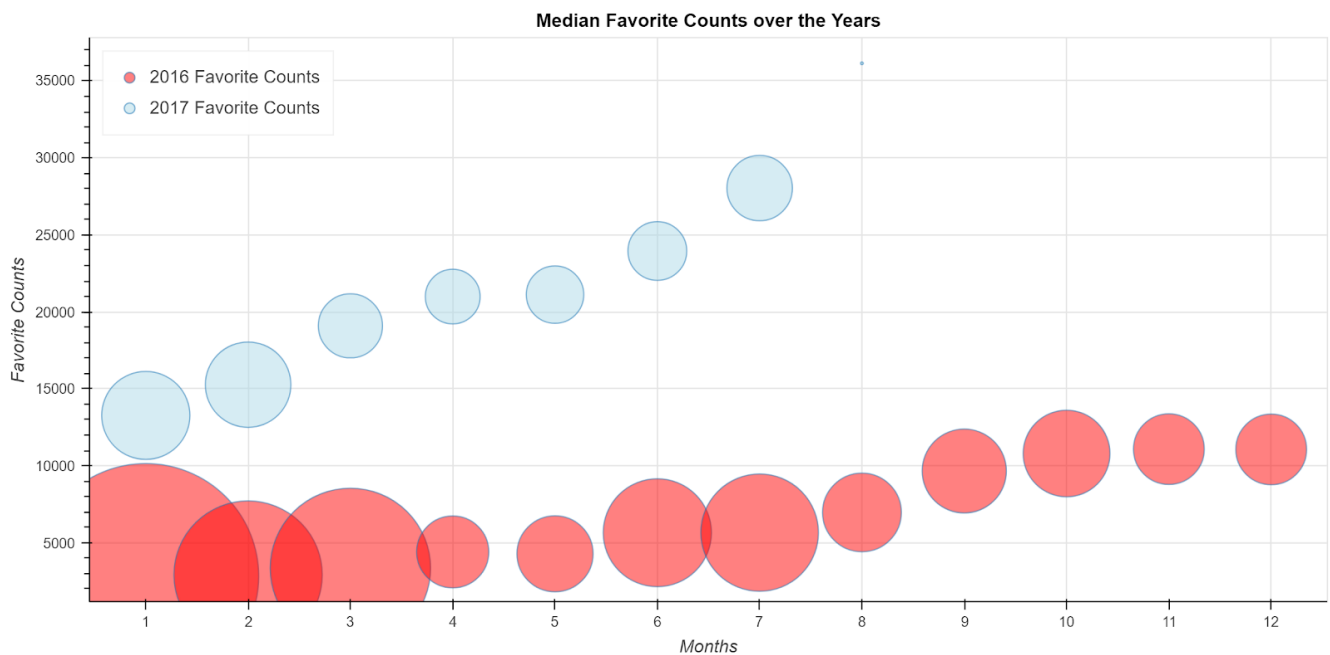
Relationship between Retweets and Favorites

The linear regression output between these two variables `retweet_count` and `favorites_count` gave an **r-value = 0.91**, which indicates a very strong, positive linear correlation between these two variables.



This positive correlation is confirmed in the above plot where we see that as the `retweet_count` (for any tweet) increases its `favorites_count` also increases.

Median Favorite Count over the years



The size of circle glyphs indicate the relative no. of tweets in a month. The plot shows year wise, how no. of tweets made decreased and the increase of favorite count over time. The popularity of WeRateDogs™ has certainly increased since its [inception](#) in late 2015.

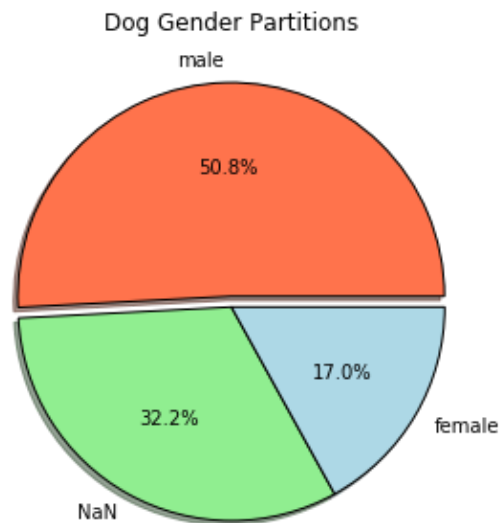
Dog Gender Ratio

The cleaned dataset had about 1993 tweets in which we could identify by text analysis the following counts of dog gender.

```
df['dog_gender'].value_counts()
```

male	1013
NaN	642
female	338

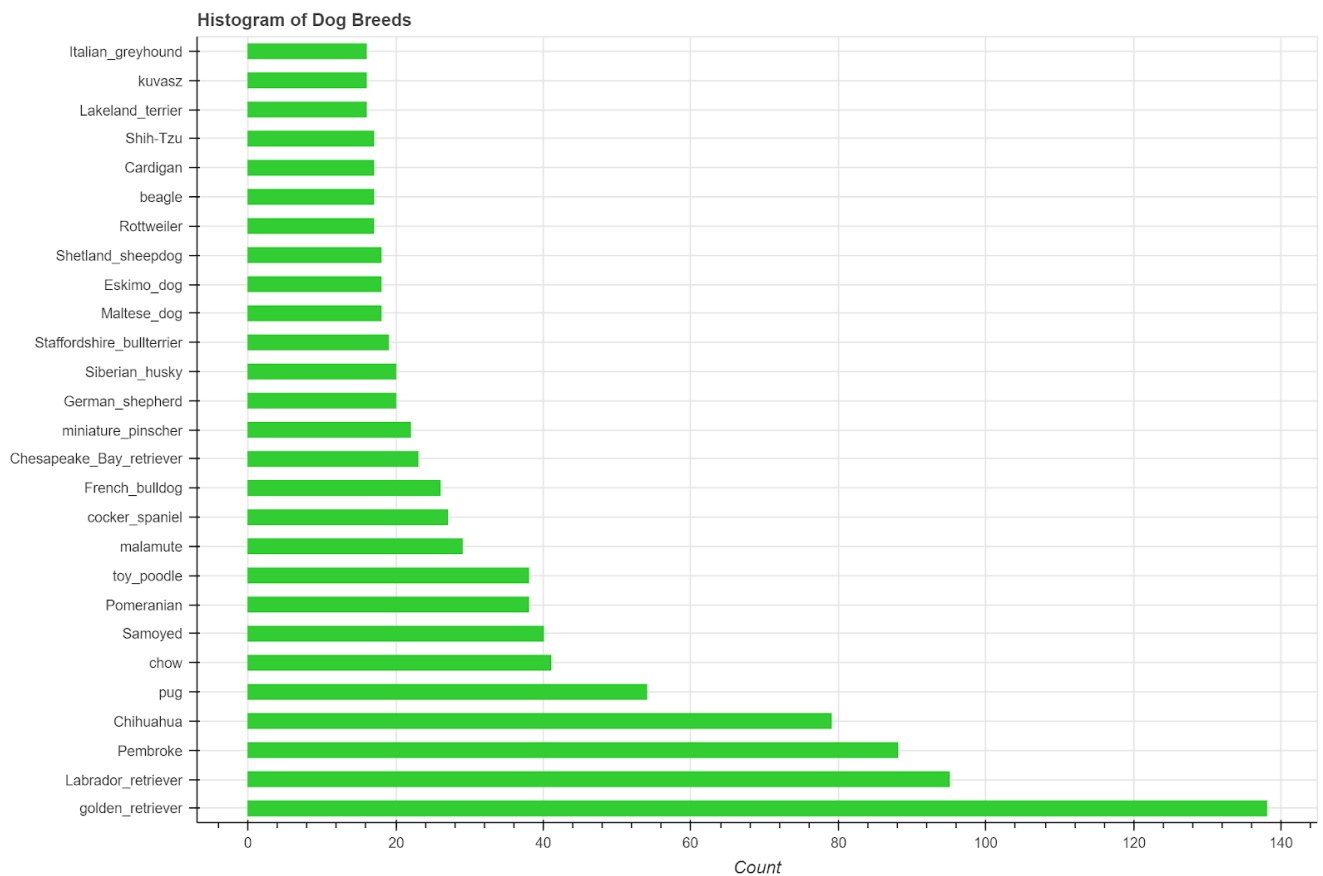
- Male Dogs appeared to be the most popular gender amongst all the tweets, as 50% of the tweets are about male dogs.
- About 32% of the tweets did not contain any mention or tags indicating dog gender (NaN).



Popular Dog Breeds

For the final visualization, a horizontal bar plot of the various dog breeds with their count was made.

- The dataset contained predictions of dog breeds from a neural network. Using this column of predictions a custom dataframe containing the various dog breeds and their respective count was made.
- To make the plot and tables easier to read we ignored dog breeds with count < 15.



The most popular dog breeds appear to *Golden Retriever* and *Labrador Retriever*.

Notes

- [Bokeh](#) was used to make all the visualizations, except for the pie chart which used [matplotlib](#)
- This dataset contained tweets upto August 2017, so it was possible that the following insights and visualizations could be different if the dataset was updated till the present.