

In the interest of consistency, American spelling has been used throughout this document.

## Introduction

In this project, I have taken data from the WeRateDogs Twitter page and wrangled through the data in order to extract meaningful information and create visualisations.

The insights explored includes a top 10 list of Twitter's favorite dogs, as well as which dogs most often make an appearance on Twitter.

## Wrangling

A number of cleanliness issues were identified and addressed over the course of this project. First the data was tidied, to do this it was necessary to combine the data I had extracted from various sources into one dataframe. As well as melting columns in order to consolidate data relating to the same variable.

Quality issues were also identified, some of the issues addressed in this analysis include:

- Duplicated images
- Inappropriate datatypes for series' in the dataframe
- Unnecessary columns
- Tweets with pictures that were not dogs
- Unoriginal Tweets

## Analysis

We can see from this visual that golden retrievers are tweeted about most regularly.

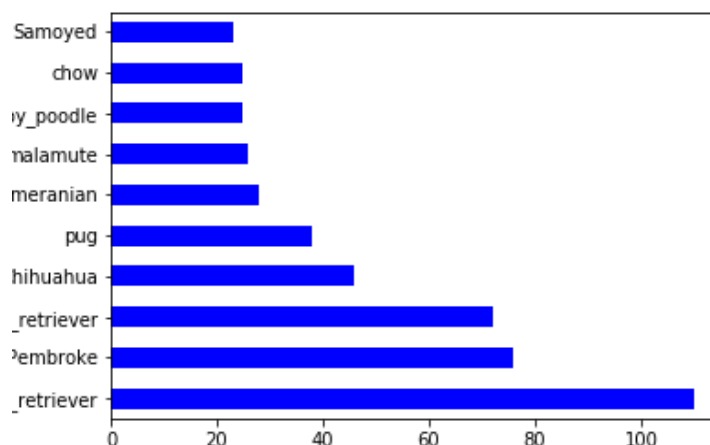


Figure 1. Most common dog by number of tweets

Unsurprisingly, we can see that there is a strong positive relationship between the number of favorites and the number of retweets of any particular post.

We can also see that the majority of tweets get under 30,000 favorites and under 10,000 retweets (Figure 2). However, there is clearly the potential for tweets to go viral (Figure 3).

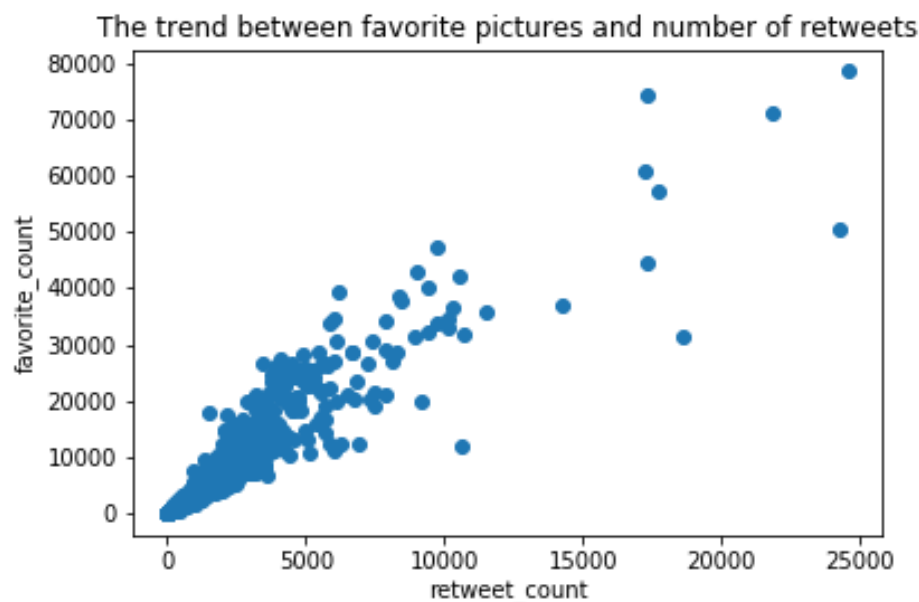


Figure 2. A graph to show the correlation between favorites and retweets



Figure 3. The tweet with the highest number of retweets and favorites!

## Scope for Further Cleaning and Exploration

Another area to clean in future iterations include identifying incorrect names of dogs and using the name mentioned in the tweet to correct these. This could then be used to identify the most common names for various breeds of dogs.