

Project Data Wrangling of WeRateDogs Twitter Archives: Actual Report (External Document)

1. **Goal**. To wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations.
2. **Scope**: WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. I performed data wrangling (i.e. gathering, assessing and cleaning) on the WeRateDogs Twitter archive which is a summary of about 5,000 tweets from 2015 to 2017.
3. **Steps**. The steps involved in data wrangling are:
 - a. **Data Gathering**. This was achieved through three methods: (1) manually downloading the WeRateDogs Twitter data from Udacity's website, (2) programmatically downloading additional image prediction data via the Python Requests library from Udacity's servers, and (3) obtaining additional Twitter data by web-scraping the WeRateDogs Twitter website through querying the Twitter API.
 - b. **Data Assessing**. After gathering each of the above pieces of data, I assessed them visually and programmatically for quality and tidiness issues using the Pandas library.
 - c. **Data Cleaning**. The issues documented in the data assessing stage are worked on, to produce a high quality and tidy master Pandas DataFrame.
4. **Insights**. Data visualisation on the cleaned data produced the following insights. Please see the following pages for the summary of the visualisations produced using the Matplotlib and Seaborn libraries.
 - a. **Top Well-Rated Dog Breeds**. Black and Tan Coonhound is the dog breed with the highest ratings, followed by Saluki and Briard.

Black and Tan Coonhound

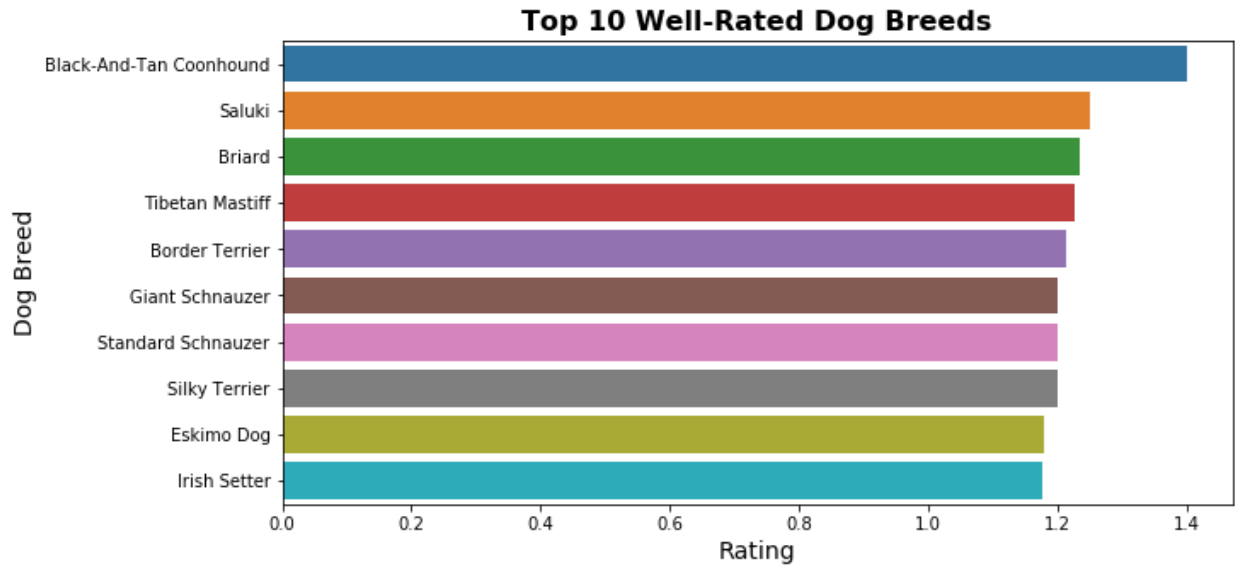


Saluki



Briard





- b. Top Favourite Dog Breeds. Black and Tan CoonHound is the dog breed with the most number of likes (i.e. favorites), followed by Saluki and French Bulldog.

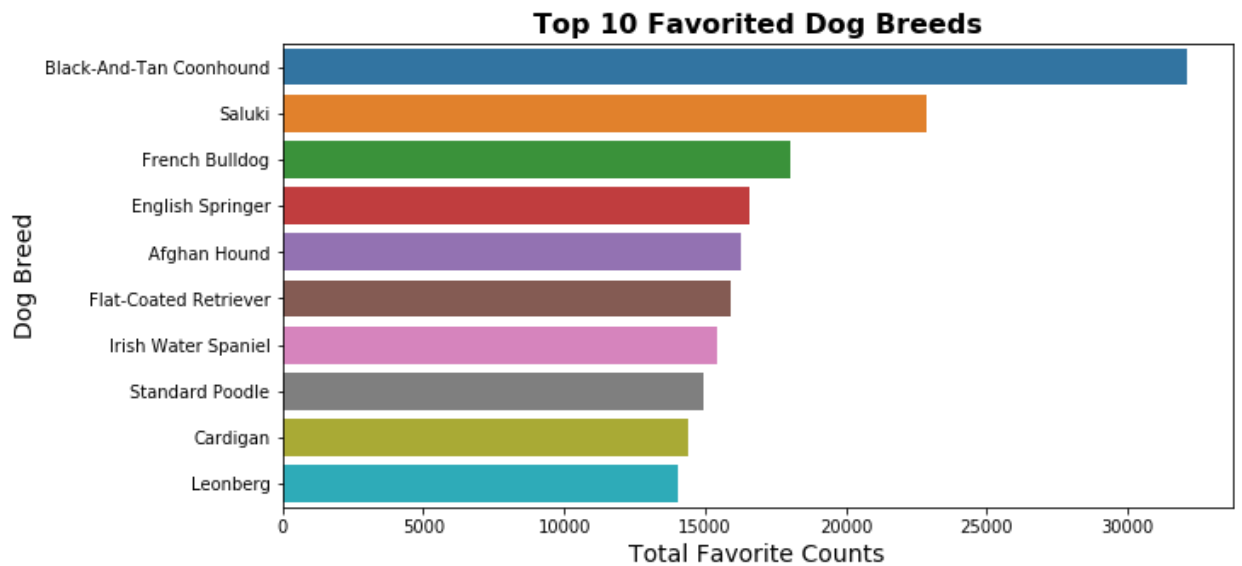
Black and Tan Coonhound



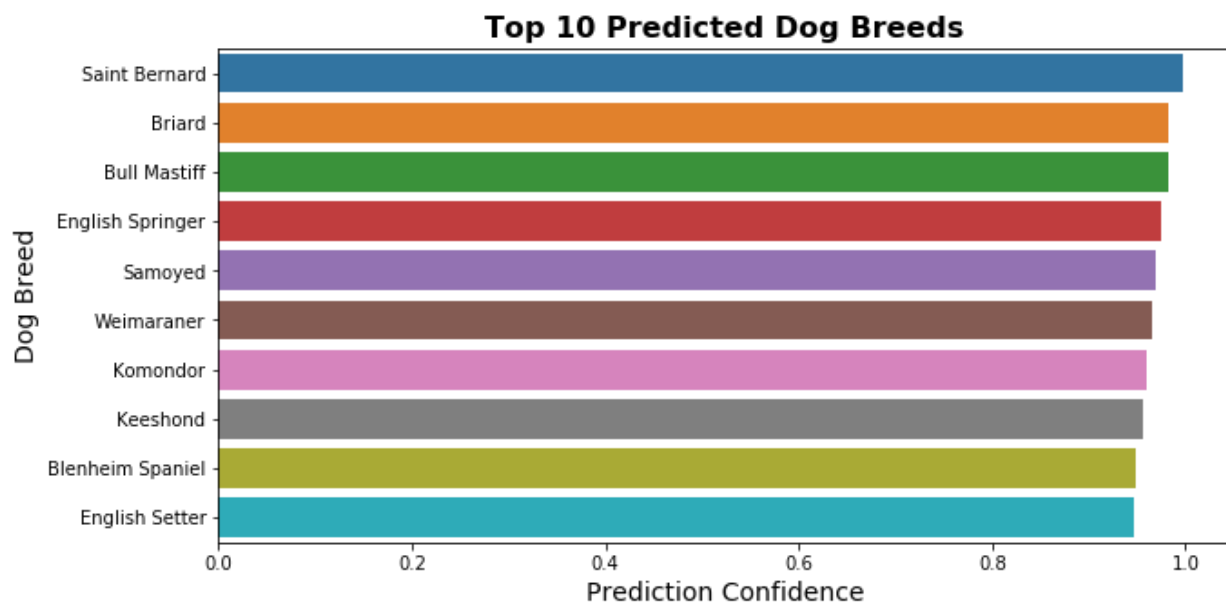
Saluki



French Bulldog

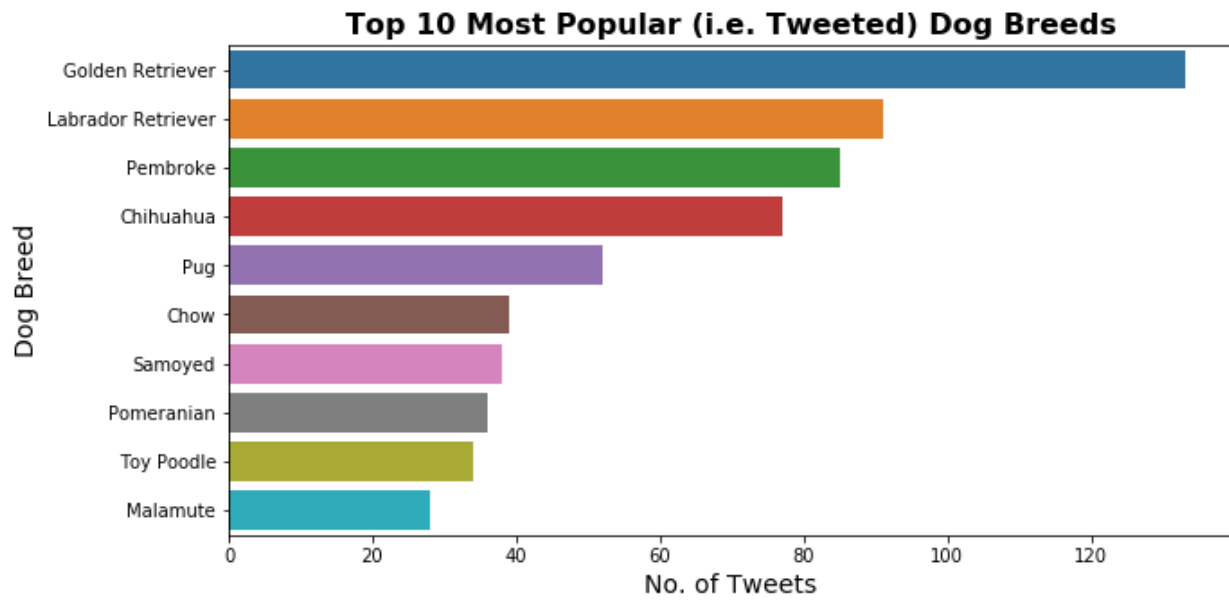


- c. Top Predicted Dog Breeds. In terms of prediction, Saint Bernard is the dog breed with the best prediction performance, followed by Briard and Bull Mastiff.



- d. Top Tweeted Dog Breeds. The most-tweeted dog breed is Golden Retriever, followed by Labrador Retriever and Pembroke.





- e. The top predicted items (that was the result of a mistaken classification of the image of the dog for a non-dog item) is ping-pong ball, peacock and snail.

Ping-Pong Ball



Peacock



Snail



