Analysis of tweets by We Rate Dogs

by Souvik Das

Motivation

The idea of this project is to apply the data leaning techniques I learned as a part of Udacity Nanodegree Program. The coding part of the project is included in wrangle_act.ipynb and the theory part is present in wrangle_report.pdf. This file provides the insights uncovered after analysing the cleaned data.

The Data

A typical tweet looks like this.

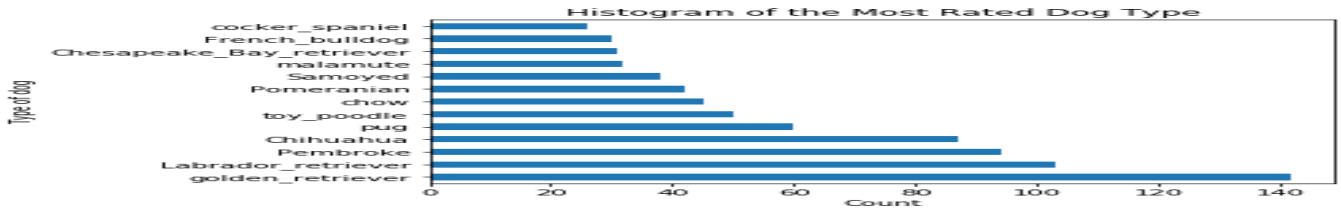
We use the tweepy api provided by twitter to retrieve information about the tweets and along with the data already gathered we build our pandas dataframes for analaysis. An important point to consider is that we have tweets from the year 2015 to 2017 present in our dataset and consequently when we run the API we find 2000+ tweets (seems great to learn about data wrangling techniques)

Next Steps

The data wrangling techniques included both manual and programmatic inspection. The programmatic inspection was broadly done in Python Programming Language with the help of pandas library. After cleaning and assessing the data we remove the columns from our dataframe which we would not be using. Finally we have in our hand clean and tidy data ready for some analysis!

Insight 1:

We simply use pandas .groupby() function to find the most common breed which has been rated by the twitter handle. Our answer comes out to be the well-known breed ‘Golden Retriever’!



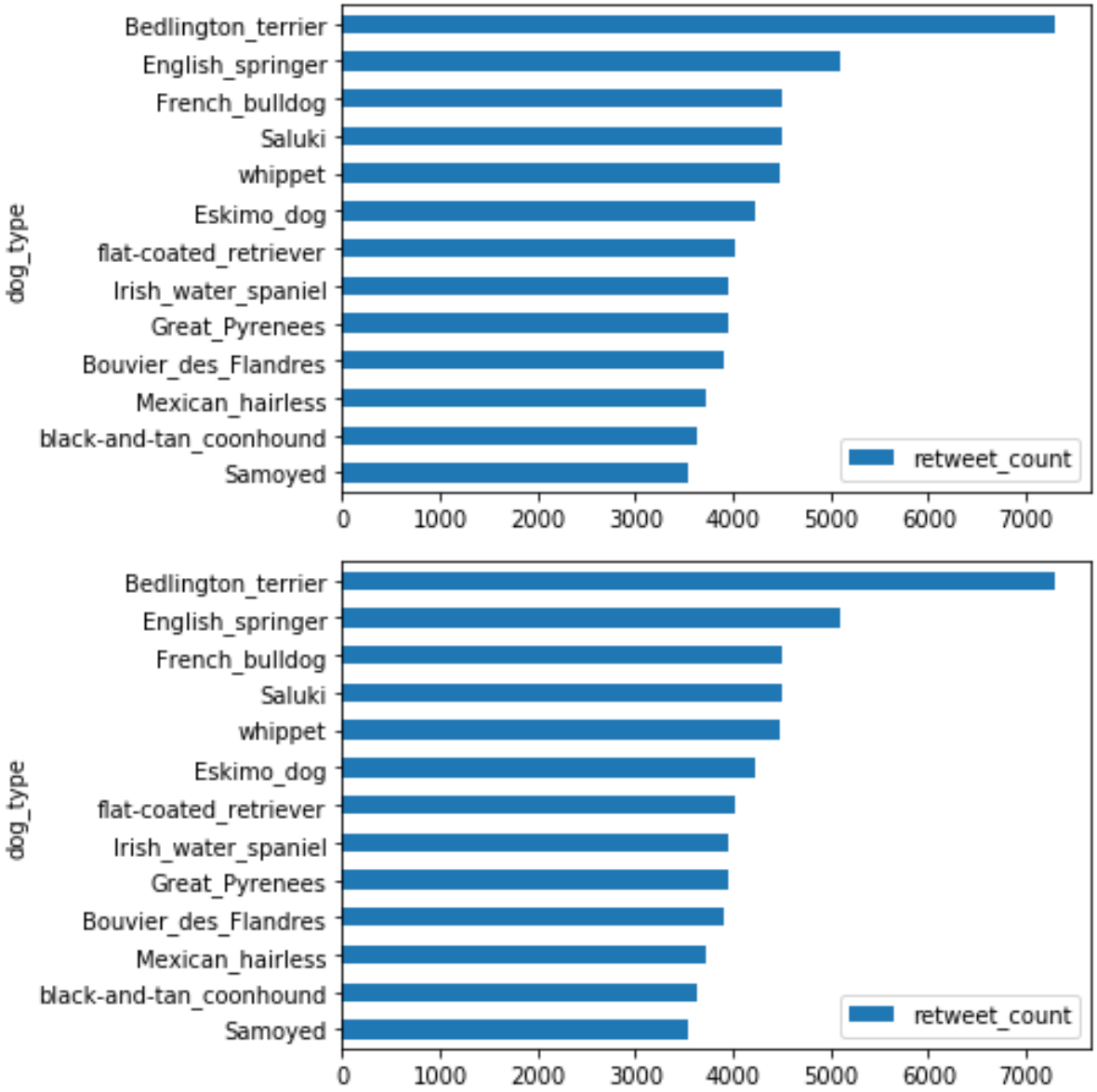
Insight 2:

Sorting the types of dogs by the average rating received we find the Japanese Spaniel has received lowest ratings and Bouvier Des Flandres has received highest rating.

Here is a tweet showing Japanese Spaniel and another of Bouvier Des Flandres

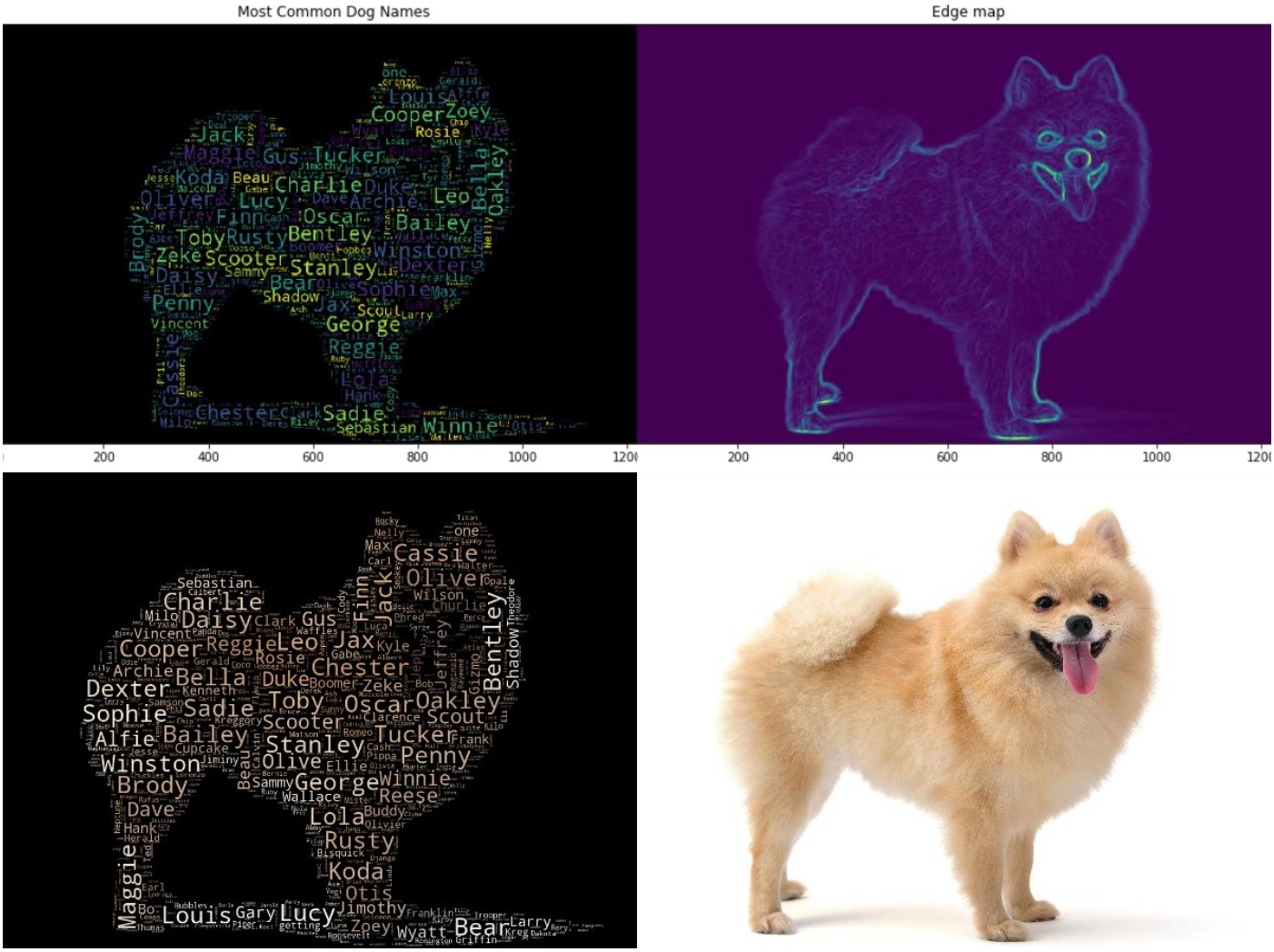
Insight 3:

We find out the dog breeds who have received highest interactions in terms of likes and retweet counts. We observer that the most common dog breeds are not present in the list so we can deduce the rare dog breeds get more interactions per tweet, which seems plausible.



Visualization :

Here is a WordCloud made by using the most common dog names tweeted out by @dog_rates



Closing remark:

The project provided unique insights towards the ratings of dogs as adjudged by the twitter handle @dog_rates. I personally had a lot of fun in working on a dataset made for 'man's best friend'. Hope you enjoyed the read!