WeRateDogs wonderful journey

Introduction.

The dataset I was working with is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." For now, WeRateDogs has over 9 million followers and has received international media coverage.

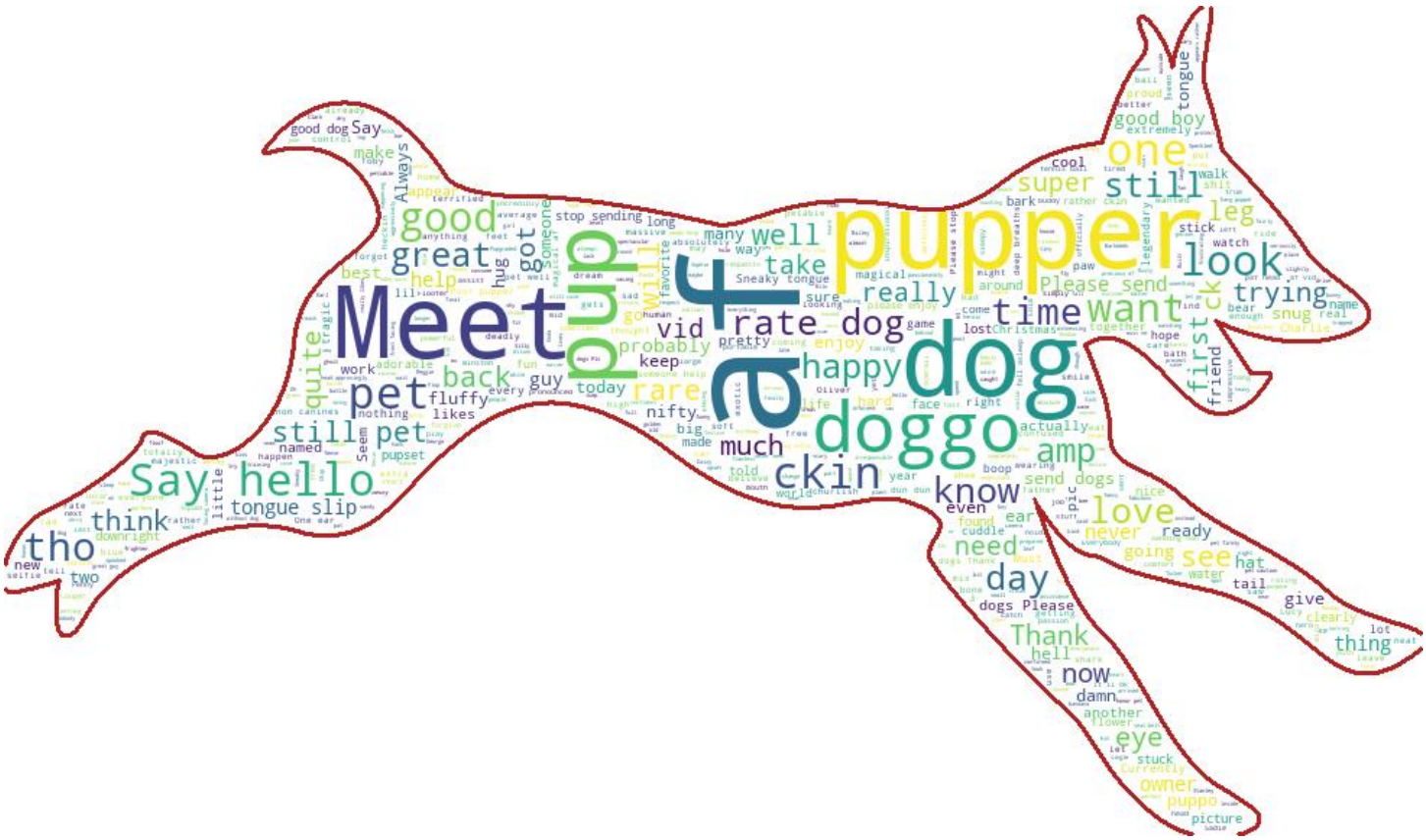
Data gathering and cleaning.

I had three parts of data in different formats and in different places.

- The WeRateDogs Twitter archive, which was available for downloading.
- The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file (image_predictions.tsv) was hosted on Udacity's servers.
- Using python library and Twitter API gathered additional data about tweets, 'likes', followers etc. in JSON format and then extracted only what I needed.

Analyzing and visualizing data.

After everything was done, I had lots of information. The most interesting part for me was to analyze text data to reveal most common words, hence catch the main, say 'vector' of the entire twitter account. So, the results are below:



I managed to draw these words using a cover of a dog image . Looking at these words, we can conclude that in general people who own a dog are very kind and cheerful; the most prominent words are: good,great,meet,pupper,dog,say,hello,love, pet and etc. I myself couldn't keep from smiling, reading tweets and laughing at images.

Intersting, that according to the data, the most retweeted picture of the dog was at the stage of "doggo".