We Rate Dogs Twitter project Summary:

Data Gathering

Three sources of data

1. twitter-archive-enhanced.csv
2. image-predictions.tsv
3. tweet-json.txt

Tweet-json.txt was queried through twitter's developer portal.

a. sign in my twitter account and set up a developer account on twitters develop portal
b. submit my application for approval to query data from twitter
c. create an API object that gather twitter data. (see code section)

Data Access, Data Cleaning

Data Qulity Issue

1.NaN value in some column

2.some datatype is not collected, need to be convert to correct data type

3.some rating_denominator column is not equal '10'

4.some rating_numerator is way larger than '10'

5.Missing value in Name columns

Data Tideness Issue

1.date and time mixed in timestamp column

2.not all name in name column is first letter capitalized

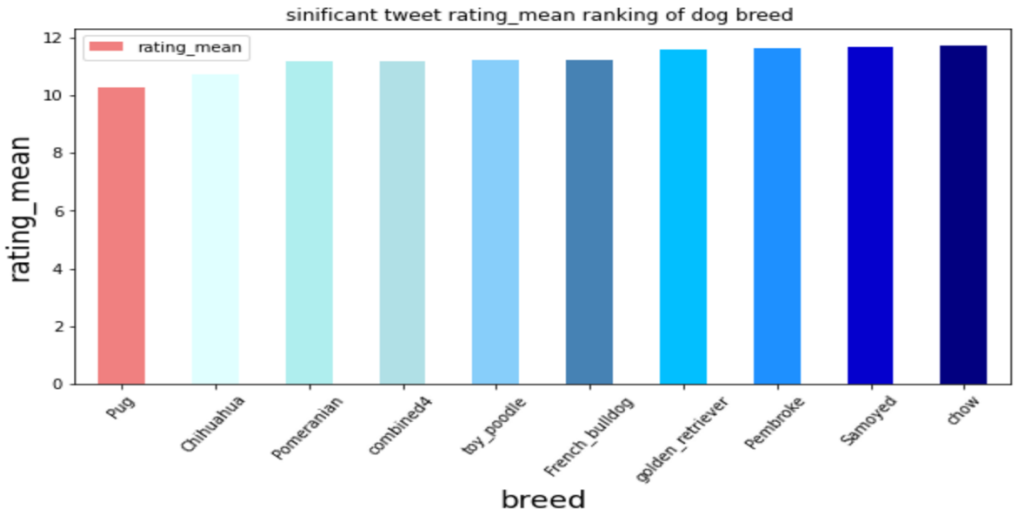
Data Analyzed

1. I combined all three achieve dataset into one
2. Selected the data only statistic significant for data analyze. (step include: a. Categorized p1_conf into four different level, only select the rows in ‘ very confidence ‘ and ‘ confidence level ‘, safe the new rows as ‘combined2’ . b. Only select the rows under combined2 I as ‘True’ in p1_dog. Save the new dataset into combined3. C. query those breed has sinificant tweet amount. Sor it and pick the top 10 . d calculate the mean rating _numerator and re-sort it.

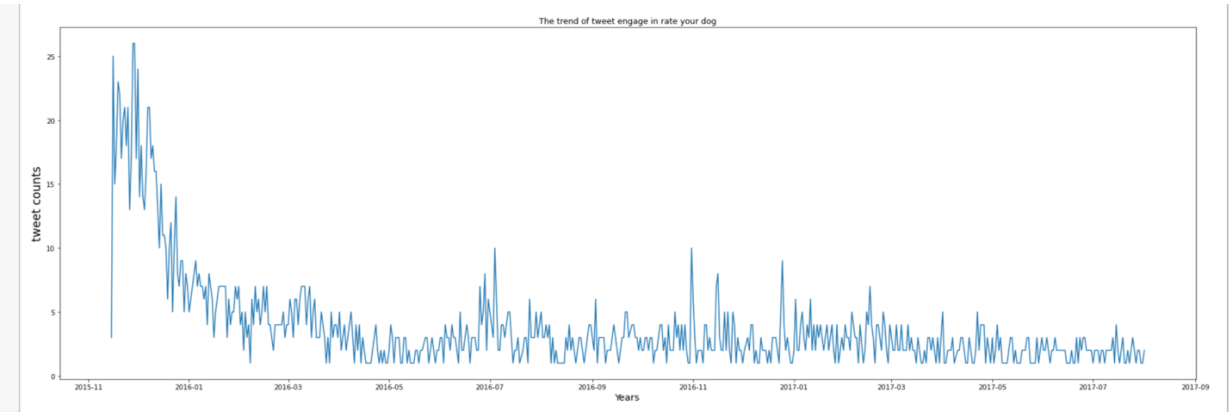
Data Visualization

2 visualizations

1. Visualized the sort result of the mean rating numerator of the breed. use different color from light to dark to indicating the sort result.



2. Visualized the trend of the tweet amount month by month base on timeline.



Data Storage

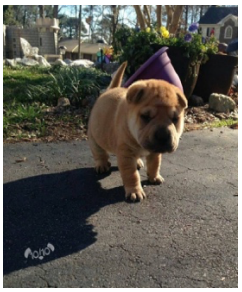
Store the data into ‘combined5’

Interesting insight

1. **Chow is the breed has highest mean rating amount the significant breed tweet.**

Here are the highest rate chow pictures (rating numerator: 13)

Lola



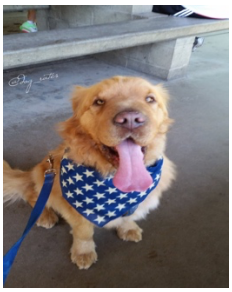
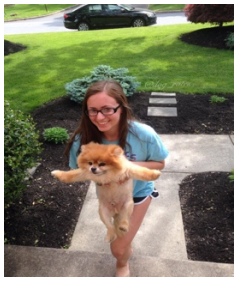
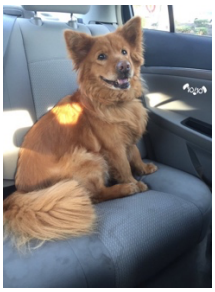
Grizzwald

No Name

Leo

Louis

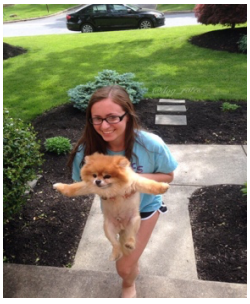
Daisy

2. for those breeds has unusual high rate, I would like to retrieve the pictures and find out how does it look like. / by looking at it, I can explains why only use the tweet with high confidence level (p1_conf_level), because some pictures are not dog picture or not clearly see the face.








3. The trend of tweet in ‘rate your dog’ was declining from Nov-2015 to Sep-2017.