This project was the most challenging and exciting I did as a python beginner. It started with gathering data and this was the first time I wrote code which took so long to run. The results of that code were very overwhelming at first, presence of 31 columns made me confused for two days. https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object this site helped me a lot in understanding the significance of every variable. As I got basic understanding of every column I referred back to udacity project page and quality solution notebook in order to understand what is being asked and how to start. Every part(Gather, Assess, Clean) of this project had a lot of inertia, they were tough when starting but after that initial push it became easier. Many stackoverflow answers and pandas documentation came extremely handy, few of them are:

1. http://pandas.pydata.org/pandas-docs/version/0.20.3/advanced.html#advanced-mi-slicers
2. https://pandas.pydata.org/pandas-docs/stable/visualization.html
3. https://stackoverflow.com/questions/29919306/find-the-column-name-which-has-the-maximum- value-for-each-row

the tweepy page http://docs.tweepy.org/en/v3.2.0/api.html#API
and of course https://twitter.com/dog_rates/

Some of the quality issues that I spotted are addressed in groups. The language quality issues were spotted by chance when I was looking for full forms of the code in lang variable in order to replace them.

I dropped the columns which had almost same or null values across all observations because as I understood from twitter development website the data in those columns was associated with either my interaction with those tweets or was about the publisher of those tweets and I also dropped those which were not used in cleaning or analysis process.

There were mainly two tidiness issues, first the data present in three tables and the second was dog stages in four column. Both these issue looked straight forward at first but they had their hidden complications like some of the observations had more than one dog stage which created multiple observation while using melt function, I was able to spot this because I was continuously monitoring the overall size of data frame

Future work on this project include improving the name variable, dog stage. While working I spotted many tweets which didn't had real dogs in it and because currently the end results(visualisation) was more focused on tweets rather than dogs if in future the focus is shifted to dog and dog rating then somehow making better use of extended entities to ensure that

the contents in tweets are about dog will be a challenging and exciting task .