This data wrangling and analyzing project consists of three datasets. The WeRateDogs Twitter archive is given in csv format, the tweet image prediction is in tsv format. We can easily import these two from python using pd.read\_csv(). The third dataset is tricky. We need to query the Twitter API to get each tweet’s retweet and favorite count. This is really new and challenge to me. Before this course, I have never done such similar projects. Query Twitter API needs a twitter developer account. It seems a high requirement for your coding skills. Fortunately, the project is well explained every step in details, it was not hard to register a developer account. Then the next step is querying the API and storing the tweet data in JSON format. At the beginning, I was very confused why we need to form the data in JSON format first, convert to a txt file, and then read the text file line by line to python data frame. It seems a redundant procedure. After a lot of searching on the project Q&A section, I understand that this is not a necessary procedure, but rather a good practice. Then next step for me is how to complete this task. I was very new to the JSON format, this is a lot of learning through website and Youtube videos, but I know the learning process is also a part of the project. Even an experienced programmer, he or she can’t know everything before work. Faster learning by self is an important skill for employer. After about five hours effort, finally, I gathered all three datasets into three data frames in python.

The second step is assessing. Firstly, I did visual assessment for all three datasets. Xxxxx

Secondly, I use programmatic assessment.