Wrangling Efforts

Po-Ching Yang

2018/7/15

1. Introduction:

This project is a challenging one for me because I almost never use twitter and not a really dog lover. Everything vocabulary is very unfamiliar for me at the beginning, and it took some a fair amount of time for me to just get started.

First, to get familiar with “tweets”, “retweets”, and “tweet id”, I actually reactivate my twitter account (never login for nearly 3 years) and go inspect the URL structures of a tweet. Then I went to WeRateDog’s page, follow it, and start to study the tweets for this project.

After I got a good understanding of what this project wants me to do, I was finally able to go through the gather-assess-clean process of data wrangling.

2. Gathering:

The gathering process involves reading data from three kinds of data source: a csv file, a tsv file on a remote server, and twitter API. Reading csv is simple, one line of code is enough. For reading data from a url, I needed to use requests library and python’s file library, which is also not very hard. But the last one from twitter API is a little bit challenging. Fortunately, my former nanodegree experience (Full Stack Web Developer) has save me a lot of time in the process.

3. Assessing:

Assessing is the most challenging section for me. It was hard to find 8 quality issues and 2 tidy issues at the beginning. I’ve gone through the whole process several times, and frequently come back from later cleaning or visualization sections. I think iteration is very vital for data assessment.

4. Cleaning:

The cleaning section is all about the application of python and pandas’s dataframe and series. With proper assessment in the former section, I was able to concentrate on the cleaning tasks in this part. During the work, I found myself frequently refer to the documentation of pandas, and after frequent consultation of the documentation, I finally got familiar with some techniques and were able to perform some tasks all by myself.

5. Conclusions:

This was an awesome project! Before I started this project, “data wrangling” sounded a little scary and stressful for me. But after I put into some real work and the continuous study of the pandas library, I feel more confident in performing data assessment and cleaning. I’m sure the techniques I learned through the project will help me in the future and in my current job as well.