Wrangle Report

Been joining Udacity for 3 months, this is so far my most challenged project, especially the Twitter API part. Fortunately, with the help of mentor and every enthusiastic student from the student hub, I made it!

Wrangling data is fun, it actually contains three parts, gathering, accessing and cleaning. The data gathering process for the project was my greatest challenge. I had two sources provided with Udacity, the third one I need to query Twitter API. Though there' s guide about it in the project overview, I don' t think it really help some students, especially for student like me, who tried API scraping for the first time. I read all documents under the Twitter API. The Tweepy wasn' t difficult to understand. Then it' s about reading and writing JSON file. I went back to the Lesson 2 : Gathering Data, reviewing all the parts about JSON. And thanks to w3schools.com, I learned basic knowledge about JSON, helped me understand the documents provided by Udacity deeply and thoroughly.

Accessing was easy, while cleaning consumed most of the time. Once I gathered all I need, I started to define those issues, and I realized how much my data should have to adjust and add. The structure issues made analyzation more impossible, the unfixed quantity issues violated the reliability of the output. The erogenous datatypes, missing records, inaccurate or invalid data, all need to be fixed. Luckily, the quizzes about patient & treatment in the cleaning lesson provided lots of train of thoughts, as long as useful function that I could apply here. The most impressive one was regular expression, I learned a lot from regexone.com, applied it to extract rating with decimals in the text. Now I am still confused about the regex when deal with completed format, like email + phone number. But it' s fun, I will spend more time on it.

The project reveals me how the real data analytics looks like. The real-life data is always dirty and messy, for the most of time, we contribute efforts about gathering and cleaning. Overall, I think I did a good job and thanks again for everyone' s help.