

Wrangle Report

August 2, 2021

I haven't executed the whole data analysis process before this project. I have tried a few analysis works only with the cleaned data without spending tremendous time gathering and cleaning the source data. At the beginning of the project, I supposed that I could finally complete and submit the assignment within a week; however, it was completely wrong because I passed almost two weeks as of today. Whenever I was stuck, I did self-researching and learned a lot from the process. This project made me grow pretty much as I believe.

The most challenging part of the project was gathering data using Tweepy API. Because the Tweepy administrator was skeptical about approving my developer account initially, it passed two days until I received my developer account. It would have me demotivated if it were not approved. Though I searched several posts from other websites and applied their advice to my coding note, it was not quickly done, especially when reading the downloaded JSON text by line and finally converting them into DataFrame. I almost spent half of the time getting the gathering stage done.

Accessing data would be fundamental to the entire process because I may find incorrect analysis from wrongly access source data. I reviewed related chapters in the course before I began working on accessing it. Accessing data looks superficial and straightforward; however, I realized it is the phase I should pay the most attention to. Especially when I tried to extract the name, rating, and stages by using the regular expression, I researched several sources and did test the Regular Expression a lot. Because regular expression is a practical and valuable tool, I would like to make it more useful by adopting it in my interest.

I figured visualization needs me to get intuition from the cleaned data. It was one of my favorite parts because I enjoy thinking to myself and draw a picture in my mind. Though I am not yet skillful and have limited knowledge about what kinds of charts would be available in Pandas and how to find the coding examples, I could get some meaningful results after reviewing the previous chapter of this course. I am already thrilled to learn the next chapter, Data Visualization.