Connor Zarkowski

Wrangle Report

For me, this was the most challenging project in the nanodegree program. I had not used an API to gather data before, so that was interesting to learn about. Most of the wrangling work was straightforward, looking for unusual names and ratings, and keeping the rules of tidy data in mind. I did take some time to learn more deeply about Pandas's datatypes and when they are appropriate. I found it helpful to be able to open the data with a program like Excel, it helped me to realize that just about all of the mistaken names were lowercase words, which made it easy to write code that corrected for that. The most challenging code to write was to manually replace the ratings where another fraction had appeared first. It was difficult because in past examples the error had been the only entry of that value, but in this dataset, for example, two numerator entries could be 14, but one of them would be an error if you compared it to the text. It wasn't something you could mathematically figure out. I liked that I got practice using pandas.melt(), when I saw it in the example video I thought it was a very helpful tool. I did end up iterating over the cleaning process again while I was making visualizations, I had not yet become familiar with the bool and category datatypes, and when my code wasn't working properly, I realized I had to go back and convert a few columns. I found it helpful that the rubric required 'define, code, test' sections during the cleaning phase, it made me catch times when I thought my code had worked, but upon testing it I realized that I had made a mistake. Overall this project was challenging and I feel like I stretched my abilities a lot further than they were before.