For my final project for DSC 540 Data Preparation, I decided to pull flight data from the Amadeus Travel API. After registering and receiving my credentials, I looked through documentation and examples on GitHub. With such a vast amount of possible data pulls, one of the main challenges for me was choosing the right data and visualizing what it could look like as a data set. To narrow down my pull, I chose one departure airport, one destination airport, and a specific date.

Looking through the initial data pull, I had a difficult time trying to make sense of the structure. Through trial and error, I discovered that it was impossible to retrieve the data I wanted with only the lumped response data to go by. Finally, I remembered previously learning about JSON formatters, so I copied my response into one and was then able to follow the organization more clearly. Using the right keys and indices, I successfully accessed specific variables and stored them into a Pandas data frame. Previewing the data frame reminded me that the arrival and departure time columns included both a date and a time, so I decided to split them into multiple columns for better readability. In addition, I adjusted some of the header names to make more sense.

As an aspiring data scientist, the importance and relevance of this process comes through not only in the technical steps taken, but also the research and troubleshooting. Since every API and data source in general that we will encounter will vary in structure and format, this is just one example of wrangling the data required for an analysis. The general approach I took on this project will help me tackle any other project I take on, in terms of gaining access to data, studying the structure, looking at others' examples, forming the appropriate syntax through trial and error, and cleaning up the data to prepare for an analysis. Some sources will be much easier than others to pull from because of existing wrappers for instance, but the general wrangling and troubleshooting process will be similar overall.

References:

GitHub, Inc. (2020). Python library for the Amadeus self-service travel APIs. Retrieved from https://github.com/amadeus4dev/amadeus-python

Amadeus for Developers. (2020). Build the future of travel. Retrieved from https://developers.amadeus.com/