{‘Team_8’ : [‘Riya Prajapati’, ‘Vivek Chowdhury’, ‘Jason Riffaterre’]}

Group Project Question 02

The final python script written to complete the requirements for CIS 9650’s Group Project Question 2 is called “uni_climate_data.py”. This code is designed to scrape the names of the top 30 colleges and universities in the United States, as well as their location. This data was obtained from tables on the “QS Top Universities” website, which is a service dedicated to ranking schools based on a large number of data points, which are then “distilled” into user-friendly metrics for applicants to find the right college fit. Our team took this ranked data and added 5 additional data points that applicants may want to consider when making such a momentous commitment. For each school we provide table data that includes the average temperature in each of three academic trimesters, as well as total precipitation and total snowfall for the school year.

The code scrapes the first 30 universities on the top 100 list. We limit it to 30 records to safely stay within the API rate limits. The API request to www.weatherbit.io retrieves historical climate data for unique locations. The data being queried here is based on average temperatures, precipitation, and snowfall for a given month based on data collected between 1991 and 2020. Once the data is retrieved, this script then calculates the average temperatures for each trimester of the college year (Fall = September - December, Winter = January - March, Spring = April - June). The JSON data requested lists the historical averages for temperature by the month of the year, so the code had to perform additional calculations on this data to return values for the average temperature in a given trimester. The script also returns the total precipitation and snowfall for the entire school year (Sept - June). These totals were, for convenience and efficiency, calculated for each trimester, and then were summed as new column data right in the final data frame itself.

The getInfo() function performs the call that returns the historical climate data for a particular location. And since the API will only return historical climate data based on the latitude and longitude of one of our cities in the list, the latlong() function was designed to translate our list of cities into usable geolocation points. A separate script (API_edge_case_testing.py) was written to make sure that the geolocation sub-routine actually returned the correct values. Since there may be several cities in the US that share a name, this code uses the same list of college cities as our main solution to question #2, as well as the same function to determine lat and long for these as in our main solution. This script then takes the lat and long values and reverse locates them to return city, state, and country. The state data revealed that the correct city was being geolocated by the latlong() function. During the development of this code a static, made up set of JSON data similar to what the API would return was employed. This was done so that debugging would not be impacted by network issues or rate limits. Once everything worked, the API call itself replaced the stand-in JSON data used during testing. A short python script (API_usage_stats.py) was also used to monitor API request limits, and informed our use of actual calls when performing the final functionality tests for this code. The final output of this code is a single data frame with the universities and their average climate data. Rate limit issues continued to plague us to the very end, and so we could return only a maximum of 24 rows in the final table in a single run. It appears that weatherbit.io prefers to sell its historical data through bulk orders of data sets rather than through free API calls. The data was exported to both csv and database files.

Given that the goal of visitors to the QS Top Universities website was to use various metrics to decide on what college would be a good fit, it seemed logical that the climate at the site of each school would be an important data point to possess. Today’s students take into consideration self care and mental health more than ever before. And the climate in a school’s location could play a big part in that equation. New York may seem like a dream, but perhaps not a great fit once someone takes into consideration how cold it gets and how much it snows. A location that sees a tremendous amount of rain, as with the Pacific NorthWest, may not be ideal for someone who needs more sunshine to thrive. A student has a stark choice of climates when comparing schools in Ann Arbor to options in New Orleans. This data set provides greater clarity about the potential impact of local weather on student performance and morale in many places other than those overtly famous for their weather like Los Angeles.