Term Project

Summary

For this project I have used <https://openweathermap.org/current> API to extract current weather data for various locations in North Carolina using ZIP codes and do some data transformations to display weather data in a readable format. Here instead of directly using the ZIP codes list I am scraping a web page [www.zip-codes.com](http://www.zip-codes.com/) to collect ZIP codes of North Carolina.

I have used BeautifulSoup and requests libraries for web scraping. Created a ZIP codes list for further use. I have used this list to pull weather data from API. After collecting data from the API when I checked the data length, I got it as 1080. But my initial length of ZIP codes was 1083. And my account was blocked due to call volume of 540/min which was exceeding the permissible call volume of 60/min.

Here two things bothered me. First thing was if at all my account was blocked for higher data volume, why only three records left out from the list and what are they? Second thing was I didn’t check for duplicates in my ZIP codes list before data collection from API. To correct this I have decided to follow these steps:

* First find out the duplicates from the ZIP codes list and remove them: Found three duplicates from the ZIP codes list. Removed them from the list.
* Need to create another account and use it sensibly. For limiting the call volume I have used time.sleep() to set waiting time before moving to the next step in the program.
* Created used list to cross check the list of ZIP codes for weather data collection and compare the results. This helped me reassure to eliminate the problem I have faced during the starting phases.
* Saved the weather data collected into a json file for further analysis without having to access the API every time which can create a problem again for call volume. Checked the json file for data. Always it’s a good idea to save data in some format.

After collecting weather data successfully I have done some data formatting to display the data in a readable format. For this

* I have converted time stamp to Date and Time.
* I have extracted some key elements from the data to display. Because I feel some elements like latitude and longitude are not major concern for the user while checking the weather information.
* Finally I have used this data to display in readable format.

It is not always the case for any Data Scientist to get data in csv or json format directly. Sometimes it is necessary to scrape a web page or collect data from API. For this we have to be always prepared for collecting data from various sources. This project helped me to recollect all the concepts I have learnt in this course and use them successfully for collecting current weather data from API.