

|  |
| --- |
| Capstone Project  Final |
|  |
| Final Report  fnfloresr@outlook.com  Authored by: Fernando Noe Flores Ramirez |

# Assignment Capstone Project

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Introduction – Business Understanding I use an actual situation I need to define for my next trip to Tokyo, Japan but I think this idea may be converted into a business opportunity to build an application where anyone may indicate a target city or neighborhood of interest and what activity is desired and application may return options that may be refined time a time based on rankings people gave to other third party applications (also selectable) and based on location, so, people may decide to visit a certain area instead or other, this case only applies as a proof of concept. My personal argumentation of the case is as follows:  Therefore, let’s make a kind of story about a current need of any tourist may have before to decide where to stay in a city due a certain specific interest in, so, let’s say, I am planning to visit Tokyo during 2022, one of my main expectations regarding this trip is to define an accommodation site in Tokyo Metropolitan area with a plenty options of food variety and quality keep an eye in my favorites, sushi restaurants (this is my case but this will be other by choice, Chinese food, Mediterranean food etc.), my goal is to have full access by walk from the hotels I plan to stay to the best ranked restaurants near around, I mean, I need to define one neighborhood to stay most of the time; based on my research two districts are the most interesting to get a good and affordable hotel, attractions and of course, food, those are Asakusa and Shinjuku (neighborhood battle, right?), the first one is well known as a traditional site near Sensoji Temple and Tokyo SkyTree and the second one is most known by nightlife and shopping, in any case, I need to choose just one neighborhood to stay, therefore, neighborhood fight will be between those districts.  Well, the aim of this project is to identify which neighborhood is better in terms of high-quality sushi bars restaurant’s locations near the hotels and based on the previous customer’s ratings we may check it out in Foursquare, but it would be in any other site information site. Data  1. Data Description   As I stated before, we need to find out locations of sushi restaurants and ratings to define a “winner” neighborhood between those two sites in Tokyo, to get that information, I will use the following data:   * Actual addresses of two hotels I have defined previously, one in Asakusa, Taito City and the other one in Shinjuku, Shinjuku City, both are located in the Tokyo Metropolitan area. For business purposes, we may select any district or venue as starting points, for this case, we are using two, therefore, these addresses will be fixed in this project.   The addresses of those hotels are the following:   * + Richmond Hotel Premier Asakusa International, 2 Chome-6-7 Asakusa, Taito City, Tokyo 111-0032, Japan.   + APA Hotel Shinjuku Gyoemmae, 2 Chome-2-8 Shinjuku, Shinjuku City, Tokyo 160-0022, Japan.   Note: I have used the address notation in following format:  <Street Number> <City><Postal Code> <Country>   * I will use Geopy to convert those addresses into coordinates, I mean, to get latitude and longitude for those sites. * Using <latitude> and <longitude> data we are going to use Foursquare API to search locations near the spots of interest. The area to look around those sites for sushi bars restaurants will be 600 meters from the hotel’s addresses, a fixed value for our project, although it may be an input to extend or reduce the reaching area.   With the information received from Foursquare API -formatted in JSON format (Java Script Object Notation), we need to process it and change format to data frames to be manipulated later with Pandas Library.   * We are going to use Folium library to visualize the sites in a map, I mean the initial spots (hotels) and sushi restaurants around them to confirm locations are correct. * Regarding the ratings, only ratings more than 7.0 indicated in Foursquare sites venue information will be considered for the analysis between neighborhoods, this is a fixed value also, although for next versions this could be an input defined by the user. Rating information will be obtained by Foursquare API using its searching capabilities inside the JSON information per venue. * Criteria to declare a “winner” between two neighborhoods is the cluster of restaurants with be more restaurants with the higher ratings defined, i.e., ratings more than 7.0.  MethodologyGetting the data Using the two addresses from Asakusa and Shinjuku we get the coordinates <latitude> and <longitude>, therefore, let’s start with Shinjuku are:   |  |  |  |  | | --- | --- | --- | --- | | Site | Address | Latitude | Longitude | | Shinjuku | Chome-2-8 Shinjuku, Tokyo 160-0022, Japan | 35.6940077 | 139.7070511 |   To confirm site location, we use Google Maps, entering the coordinates, we get the map, please refer to Figure 1.    Figure 1 - Shinjuku site - Google Maps coordinates using Geopy library.  We identify an issue, the coordinates we got using Geopy library are not correct, actual site is far away from the location we got using Google Maps, please refer to Figure 1.  Note: I could not identify what the issue was, Geopy library works fine with other addresses - for example, US postal codes, perhaps Geopy is not optimize for all countries.  Therefore, we made a research about alternative libraries to get coordinates from addresses and we decided to use Google Maps API [1] that has a similar function like Geopy library.  Using Google Maps API, we got the following coordinates:   |  |  |  |  | | --- | --- | --- | --- | | Site | Address | Latitude | Longitude | | Asakusa | Chome 6-7 Asakusa, Tokyo 111-0032, Japan | 35.7141245 | 139.7936921 | | Shinjuku | Chome-2-8 Shinjuku, Tokyo 160-0022, Japan | 35.6888067 | 139.7090027 |   To confirm site’s locations, we get the following maps:    Figure 2 - Asakusa site - Google Maps coordinates using Google Maps API library.    Figure 3 - Shinjuku site - Google Maps coordinates using Google Maps API library.  Data is correct, sites are located correctly.  Next step is to get information about venues (sushi restaurants) around the sites already located, for that we are going to use Foursquare API [2], this service provides several information about venues through a GET method using a http request, this is a call to a Foursquare which as a response, we receive a JSON (Java Script Object Notation) file, this file includes a different set of data depending of the parameters we use in the GET method.  For our purposes, we need to get basically the following information:   * Venue ID, a unique identifier per venue. * Coordinates (latitude and longitude), per each venue. * Distance, from each venue to the original point for Asakusa or Shinjuku -to verify distance is in accordance with the specification.   Once, we got these two files in JSON format, we need to convert those into dataframe format -using a Python,this is useful to manipulate and clean up or filter the whole information contained in the JSON format and thus using Pandas library we may process the data easily.    Figure 4 – Asakusa site – dataframe\_filtered\_asak.  Interfaz de usuario gráfica, Texto  Descripción generada automáticamente  Figure 5 – Shinjuku site – dataframe\_filtered\_shin.  Another key advantage of using dataframe format is that we may inspect the data directly and visualize the data column by column and verify information in case we get missing data or inconsistence data. After the checking, it seems data is OK.  Also, important to include in our analysis is the site visualization using Folium library [3].    Figure 6 – Asakusa site – venues in blue.  Mapa  Descripción generada automáticamente  Figure 7 – Shinjuku site – venues in blue.  In Figure 6 and Figure 7 we may visualize the center points -in red and venues obtained -in blue, respectively.  For the next step, we need to isolate the ID information per each venue to be the input for the next step, therefore ID lists for both sites are the following:  Texto  Descripción generada automáticamente  Figure 8 – Asakusa site – ID per venue.  Interfaz de usuario gráfica, Texto  Descripción generada automáticamente  Figure 9 – Shinjuku site – ID per venue.  Based on these two objects, we are going to use Foursquare API again, in the case, ID per venue will be the input and rating will be the output, our results are as follows:    Texto  Descripción generada automáticamente  Figure 8 – Asakusa site – ratings.  Texto  Descripción generada automáticamente  Figure 9 – Shinjuku site – ratings. ResultsPresentation By comparing information in Figure 8 and Figure 9, we deduce the sites with higher ratings are:  Shinjuku – Sushizanmai – 8.3  Shinjuku - Sushi Rosan – 7.6  Asakusa – Sushi Ken - 7.4  Shinjuku – Numazuko – 7.3  Observing the outputs of both sites, we may infer the following facts:   * No all the venues have rating information available; this may be a handicap. * Shinjuku has more sites with ratings that Asakusa. * Shinjuku by far has higher ratings that Asakusa.   Winner neighborhood is Shinjuku. Conclusions We did apply the concepts learned during all the courses and practice extensible the code skills we developed course by course for Python, specially, functions and data structures.  This project uses Python and key libraries like Pandas, Folium and Geopy, furthermore, we found out difficulties using Geopy library, therefore, we did investigate how to use alternative libraries finding out Google Maps API -which was not part of the course and it was included replacing Geopy.  As key issues I found out I would like to mention how to handle Foursquare API, I did need to fine tune the operations using my own credentials, finally, FS API was successfully managed, other fundamental take away was the use of GitHub repository and Jupyter Notebooks and finally, I could code the latest version using Visual Studio Code linked with my GitHub account, synchronization between both platforms was an important take away I took as a must learning in any developer career.  I would like to thank Pedro Pessoa who provides me valuable feedback, guidelines and excellent support in coding to improve my capstone project, I do guess his interesting and advances jobs help me a lot to understand specific topics and definitely may help others to build better applications regarding geo reference topic.  In the business perspective, this small project maybe a foundation of an application as I described in the introduction if a web application includes part of this job.  I do guess a web application may help people to looking for any business based on previous ratings where such information is provided. References: [1] Google Maps API — https://developers.google.com/mapsWikipedia  [2] Foursquare API - https://foursquare.com/developers/login?continue=%2Fdevelopers%2Fapps | | |
| Let’s go for it and learn in the process! | | |
|  | | |
|  | | |
|  | |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |
|  | |  |  |