Title: Olympics 2020 in Tokyo, Japan

Ayushi Shukla

24 December, 2019

1. Introduction:

The 'Good Vibes' is a company based in Japan that is responsible for organizing the event. It is working on a project for providing hospitality for visitors during the 2020 Summer Olympics which is an upcoming international multi-sport event, which will happen from 24 July to 9 August 2020 at New National Stadium in Tokyo, Japan. The purpose of the company is to provide the most excellent service which comprises information of the nearby place so that visitors have options to select based on their needs.

The idea of this study is to help people planning to stay and explore Japan during Olympics 2020 by providing data about neighborhood places.

WIth the help ofFourSquare, it will tell you all about places to go, things to see, restaurants to eat at, bars to drink in, nightclubs to part the night away in and then where to go in the morning to get breakfast and a strong coffee. This will help tourists to review their options and make choices about where to visit and eat up front before travel.

A high level approach is as follows:

1. The travellers decides on a city location (in this case Tokyo, Japan)

2. The ForeSquare website is scrapped for the top venues in the city

3. From this list of top venues the list is augmented with additional grographical data

4. Using this additional geographical data the top nearby restaurents are selects

5. Other places within a predetermined distance of all venues are obtained

6. A map is presented to the to the traveller showing the selected venues

Who is this solution targeted at: This solution is targeted at the traveller as well as people living in that area. The want to see all the main sites of a city that they have never visited before but at the same time, for whatever reaons unknown, they want to be able to do all that they can to make sure that all the information is easily available and they have more options leading to a delightful experience.

There are many data science aspect of this project including: Data Acquisition, Data Cleansing, Data Analysis, Machine Learning and Predictions.

2. Data:

Data Description: In this section, I will describe the data used to solve the problem as described

previously. As noted below in the Further Development Section, it is possible to attempt quite complex and sophisticated scenarios when approaching this problem. However, given the size of the project and for simplicity only the following scenario will be addressed:

1. Query the FourSqaure website for the top sites in Tokyo

2. Use the FourSquare API to get supplemental geographical data about the top sites

3. Use the FourSquare API to get top places recommendations closest to each of the top site

Top Sites from FourSquare Website Although FourSquare provides a comprehensive API, one of the things that API does not easily support is a mechanism to directly extract the top N sites / venues in a given city. This data, however, is easily available directly from the FourSquare Website. To do this, we will simply go to www.foursquare.com, enter the city of choise and select Top Picks from I'm Looking For selection field or we can define Foursquare credentials and the city to get it's latitute and longitude.

3. Methodology:

The methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.

Exploratory Data Analysis: The first round of eploratory analysis was to examine the Top Venues and dataframes to determine if there was any correlation between variables.Unforfunately the only data attributes that could be analysed were the Latitude and Longitude attributes and their relationship to the venuse score. Top Venues was examined First.

In this study for searching for places like Hospitals, hotels etc I have used the corresponding URL, by sending 'get' request I have find results. Then the relevant part of JSOn was assigned to venues and transformed venues into places. he data cleaning was done by keeping only columns that include venue name, and anything that is associated with location. Defined function that extracts the category of the venue anf filter the category for each row. Delete unnecessary columns, delete rows with none values, delete rows which its category is not Desired place or Event Space.

Generate maps to visualize venues and how they cluster together. For this we create dataframe by including all the places. Overall Map of all the Places . Map for Basic facilities like Hotels, Hospital and Restaurant and map to visualize neighbourhood including Cafeteria , Shopping Centers, Tourist Center, Spa and Park.

4. Result:

The idea for the Capstone Project is to show that when driven by venue and location data from FourSquare, backed up with open source data, that it is possible to present the required information about necessary placese to traveller with a list of attractions to visit supplemented with a graphics showing the occurance of places in the region of the venue.

A high level approach is as follows: The travellers decides on a city location (in this case Tokyo), The ForeSquare website is scrapped for the top venues in the city, from this list of top venues the list is augmented with additional grographical data, Using this additional geographical data the top nearby places are selected, The places within a predetermined distance of all venues are obtained, A map is presented to the to the traveller showing the selected venues and statistics of the area.

So goals have been achieved in this project. In this Results Section this goal is addressed. The purpose of this project was to explore various facilities for tourist including information they may need during emergency such as hospitals and tourist Help Centers to the places where they can go and relax such as hotls, parks and spa, also the information about places where they can have fun such as bars and clubs, is also provided. However, due to limited range   we have shown only best available options in that particular range but by increasing the distance range more options can be provided. The data set for few top hotels in the range of 2000 meters looked like this:

|	name	categories	address	lat	lng	postalCode	state
0	Palace Hotel Tokyo (パレスホテル東京)	Hotel	丸の内1-1-1	35.684644	139.761302	100-0005	東京都
2	Imperial Hotel Tokyo (帝国ホテル 東京)	Hotel	内幸町1-1-1	35.672468	139.758322	100-8558	東京都
4	Tokyo Dome Hotel (東京ドームホテル)	Hotel	後楽1-3-61	35.703617	139.753470	112-8562	東京都
5	Four Seasons Hotel Tokyo at Marunouchi (フォーシーズ...	Hotel	丸の内1-11-1	35.677912	139.766808	100-6277	東京都
6	Hotel Ryumeikan Tokyo (ホテル龍名館東京)	Hotel	八重洲1-3-22	35.682375	139.770832	101-0028	東京都
7	Hotel the Celestine Ginza (ホテル ザ セレスティン銀座)	Hotel	銀座8-4-22	35.669172	139.759801	104-0061	東京都
8	Hotel Main Kanda (ホテルメイン 神田)	Hotel	内神田3-1-6	35.689416	139.769730	101-0047	東京都
9	Hotel Villa Fontaine Otemachi (ヴィラフォンテーヌ 大手町)	Hotel	内神田1-8-8	35.690651	139.767015	101-0047	東京都
10	Hotel Mystays Premier Akasaka (ホテルマイステイズプレミア赤坂)	Hotel	赤坂2-17-54	35.670134	139.738332	107-0052	東京都
14	Solaria Nishitetsu Hotel Ginza (ソラリア西鉄ホテル銀座)	Hotel	銀座4-9-2	35.670863	139.767403	104-0061	東京都
17	Hotel Heimat (ホテル ハイマート)	Hotel	八重洲1-9-1	35.680521	139.769693	103-0028	東京都
18	GRANBELL HOTEL AKASAKA (赤坂グランベルホテル)	Hotel	赤坂3丁目10-9	35.675585	139.737296	107-0052	東京都
19	Hotel Villa Fontaine (ホテル ヴィラフォンテーヌ 神保町)	Hotel	神田神保町1-30	35.697120	139.758489	101-0051	東京都
20	Ginza Bellevue Hotel (銀座ベルビューホテル)	Hotel	銀座6-4-14	35.671160	139.761221	104-0061	東京都

I have merged those variables with related cluster informations in main master table. By joioning labels a cluster map of nearby places of Tokyo was created.

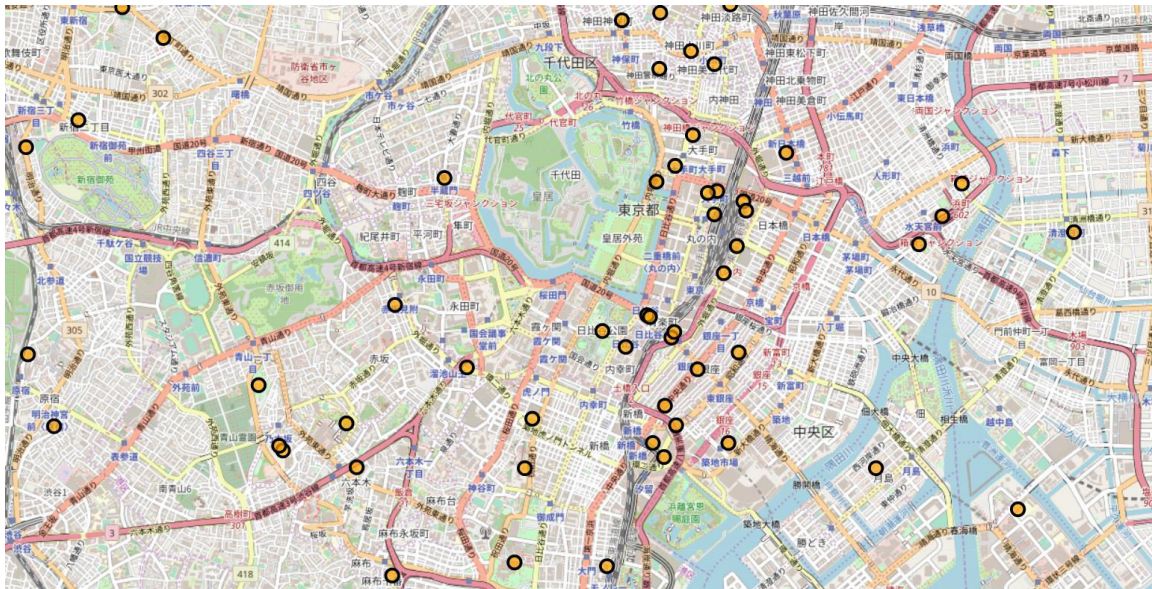Overall Map of all the Places identified.
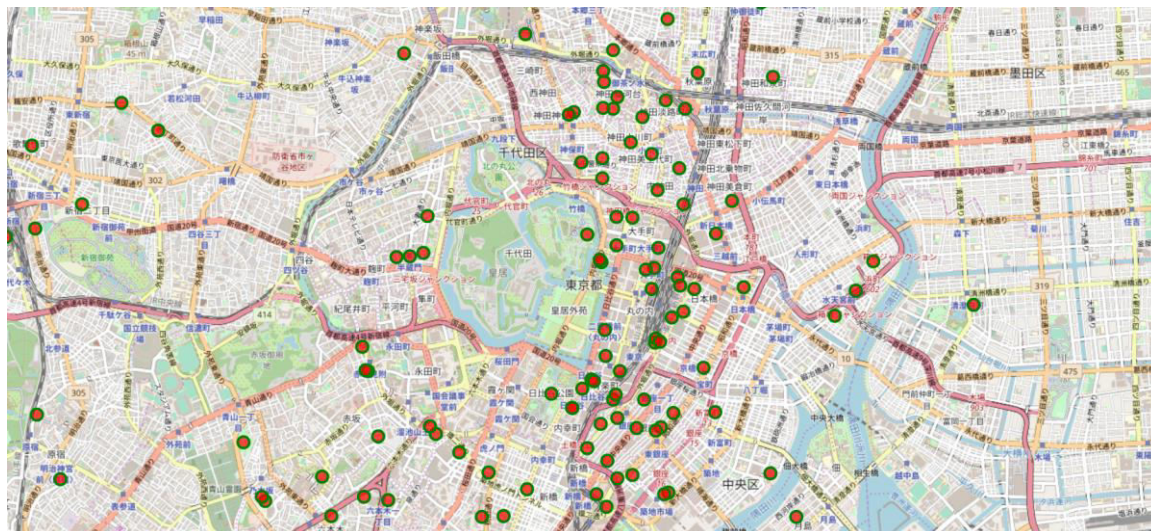
Map for Basic facilities like Hotels, Hospital and Restaurant
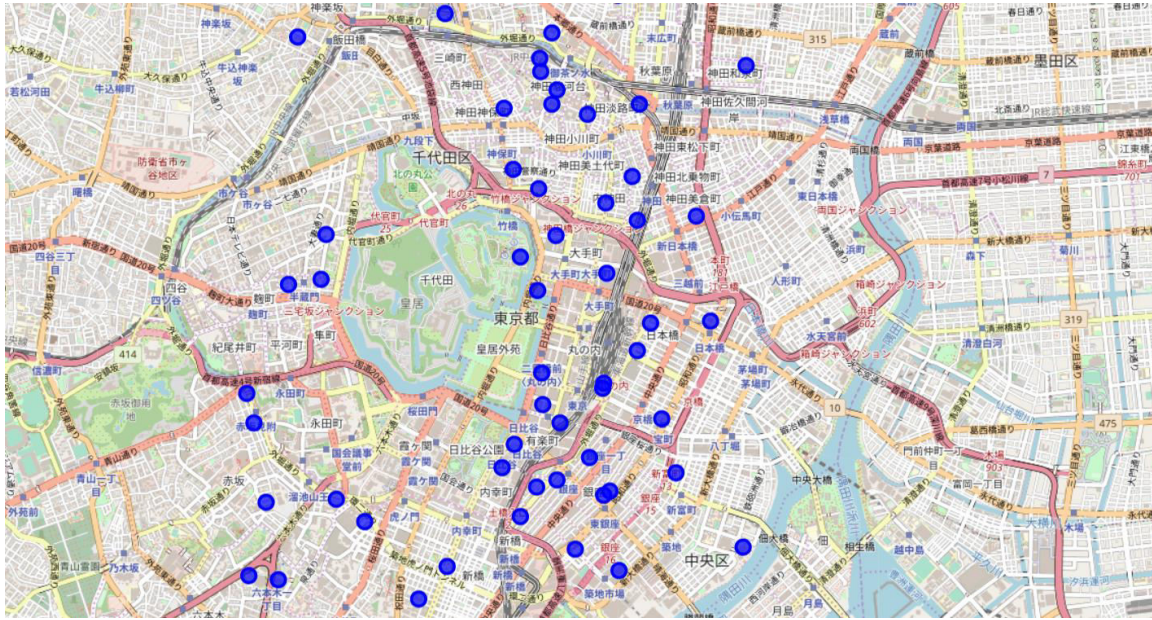
Map to visualize neighbourhood including Cafeteria , Shopping Centers, Tourist Center, Spa and Park

5. Discussion:

As I mentioned before, Tokyo is a big city with a high population density and have various options to explore. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results. Although the goals of this project were but met there is definitely room for further improvement and development as noted below. However, the goals of the project were met and, with some more work, could easily be devleoped into a fully phledged application that could support the traveller in an unknown location. Of the contributing data the Tokyo data is the one where more data would be good to have. FourSquare proved to be a good source of data.

6. Conclusion:

In this study, I analyzed the various places in Tokyo, Japan   where Olymipcs 2020 will happen. I identified various options for tourist to explore such as Hospitals in case of emergency, hotels to stay, shopping places, cafeteria, bars and clubs, Restaurants, tourist cenetrs for any kind of help, spa to relax and parks. To enjoy any trip it is very important to be aware of that place so that one can have maximum benefits from it. These informations can be very useful in helping tourists.