Likelihood of New Restaurant in Tokyo Neigborhood

By: Khairul Azhar

1.0 Background Study

Tokyo is the capital city of Japan and it also known as Tokyo Metropolis. It is because the number of population in the city was estimated to be over 13.9 million. As the city is one of the busiest city in the world, so it suitable to do some analysis on the importance of restaurants near the office place at Tokyo neighborhood.

When regarding the Tokyo as the busiest city, the morning and lunch hour will be packed with people and this will create huge opportunities to build restaurant close to the office place.The core of Tokyo is made of 23 wards (municipalities) but this study will concentrate on 5 most busiest business wards of Tokyo — Chiyoda (千代田区), Chuo (中央区), Shinjuku (新宿区), Shibuya (渋谷区) and Shinagawa (品川区), to target daily office workers.

From this analysis, it will benefit the business personnel who wants to invest or open a restaurant at Tokyo neighborhood.

2.0 Data Preparation

2.1 Web Scraping and Cleaning

The Wikipedia page of Tokyo Wards contains the table of 23 wards of Tokyo, area, population and major districts. Beautifulsoup4 and pandas library were used to create the initial data-frame. For a clean and understandable data-frame some of the wards are renamed for example 'Chiyoda, Tokyo' to 'Chiyoda'. The first entry of the major districts column in the Wikipedia table to concentrate on. Even though not complete but it gives quite a detailed picture of the corresponding ward, as later on top most venues within 1 kilometer radius of the major district have been considered. After this inital preparation, obtain the coordinates using Geopy library.

Then, some of the coo-ordinates of the major districts returned by Geopy are wrong and need to figure out the reason for this is the name of the major districts in the data-frame are different from their actual names, for example Hongō to Hongo. In these cases (4 of them), had to google search and replace using pandas library. After little manipulation the obtained data-frame looks as Figure 1 below:

	Ward	Area_SqKm	Population	Major_District	Dist_Latitude	Dist_Longitude
1	Chiyoda	5100	59441	Nagatacho	35.675618	139.743469
2	Chuo	14460	147620	Nihonbashi	35.684068	139.774503
3	Minato	12180	248071	Odaiba	35.619050	139.779364
4	Shinjuku	18620	339211	Shinjuku	35.693763	139.703632
5	Bunkyō	19790	223389	Hongo	35.175376	137.013476
6	Taitō	19830	200486	Ueno	35.711759	139.777645
7	Sumida	18910	260358	Kinshicho	35.696312	139.815043
8	Koto	12510	502579	Kiba	35.672200	139.806100
9	Shinagawa	17180	392492	Shinagawa	35.599252	139.738910
10	Meguro	19110	280283	Meguro	35.621250	139.688014
11	Ota	11910	722608	Omori	35.588473	139.727933
12	Setagaya	15690	910868	Setagaya	35.646096	139.656270
13	Shibuya	15080	227850	Shibuya	35.664596	139.698711
14	Nakano	21350	332902	Nakano	35.718123	139.664468
15	Suginami	16750	570483	Koenji	35.704942	139.649909
16	Toshima	22650	294673	Ikebukuro	35.729799	139.710242
17	Kita	16740	345063	Akabane	35.778139	139.720800
18	Arakawa	21030	213648	Arakawa	35.737529	139.781310
19	Itabashi	17670	569225	Itabashi	35.774143	139.681209
20	Nerima	15120	726748	Nerima	35.748360	139.638735
21	Adachi	12660	674067	Ayase	35.446047	139.430823
22	Katsushika	12850	447140	Tateishi	33.481791	131.478154
23	Edogawa	13750	685899	Kasai	35.663400	139.873100

Figure 1: Dataframe

3.0 Data Exploratory Analysis

3.1 Foursquare

Foursquare API was used to obtain the 100 most common venues within 1 kilometre of each major district. As the focus is on 5 major business districts (Nagatacho, Nihonbashi, Shibuya. Shinjuku, and Shinagawa), from the data, there are 193 restaurants among the 500 top venues in these 5 districts. Folium library was used to plot a leaflet map of only these restaurants in these 5 major districts of Tokyo which is as shown in Figure 2, where the colors representations are the following:

Nihonbashi- Green, Nagatacho- Red, Shibuya- Orange, Shinjuku- Magenta, Shinagawa- Blue.

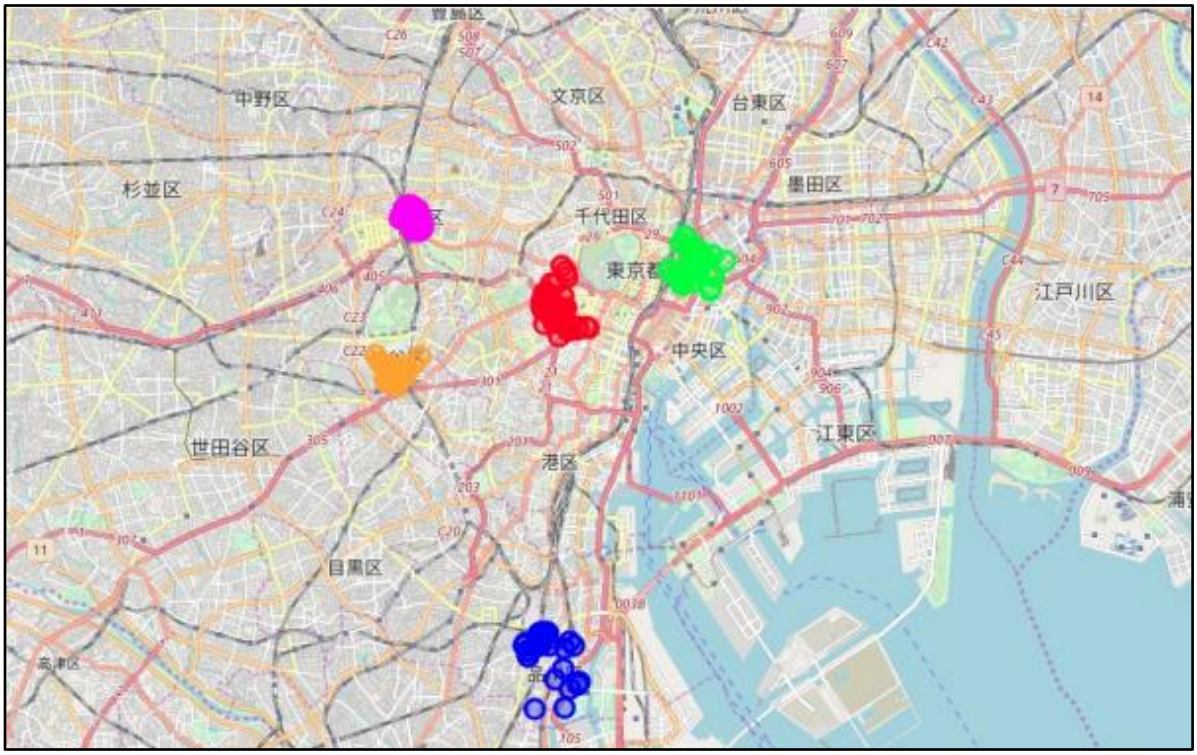

Figure 2: 5 Major Districts in Tokyo Neighborhood

From the obtained data, Japanese Restaurant and Ramen Restaurant both are top the chart of the common venues for 5 districts. It shows in the Figure 3 below:

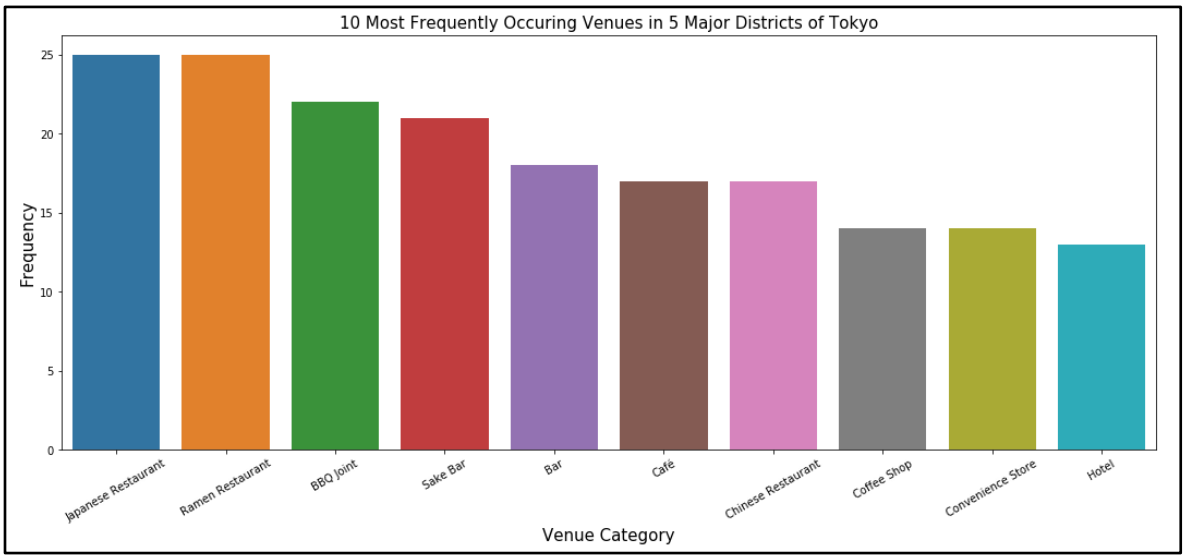

Figure 3: 10 Most Common Venues in Tokyo

Next step was to obtain information about the top 5 venues of each district. And to do that by proceeding with creating a data-frame with pandas one hot encoding for the venue categories. Then, use pandas groupby on District column and obtain the mean. Next, transpose the data-frame at step 2 and arrange in descending order. Implementing them in Pandas outputs the as follows:

```
%%%%%%%%%Nagatacho%%%%%%%%
                 Venue  Freq
0  Japanese Restaurant  0.09
1           BBQ Joint  0.07
2   Chinese Restaurant  0.06
3     Ramen Restaurant  0.06
4          Coffee Shop  0.06


%%%%%%%%%Nihonbashi%%%%%%%%
                 Venue  Freq
0  Japanese Restaurant  0.08
1                Café  0.05
2      Soba Restaurant  0.04
3           BBQ Joint  0.04
4    Department Store  0.04


%%%%%%%%%Shibuya%%%%%%%%
                Venue  Freq
0                Café  0.10
1         Record Shop  0.06
2         Coffee Shop  0.05
3    Ramen Restaurant  0.04
4  Seafood Restaurant  0.04


%%%%%%%%%Shinagawa%%%%%%%%
                 Venue  Freq
0  Convenience Store  0.14
1   Ramen Restaurant  0.08
2          Sake Bar  0.08
3         BBQ Joint  0.05
4     Grocery Store  0.05


%%%%%%%%%Shinjuku%%%%%%%%
                 Venue  Freq
0                  Bar  0.10
1            Sake Bar  0.09
2     Ramen Restaurant  0.06
3           BBQ Joint  0.05
4  Japanese Restaurant  0.05
```

In order to determine the district has the highest number of restaurants as the most common venue is shown in Figure 4:

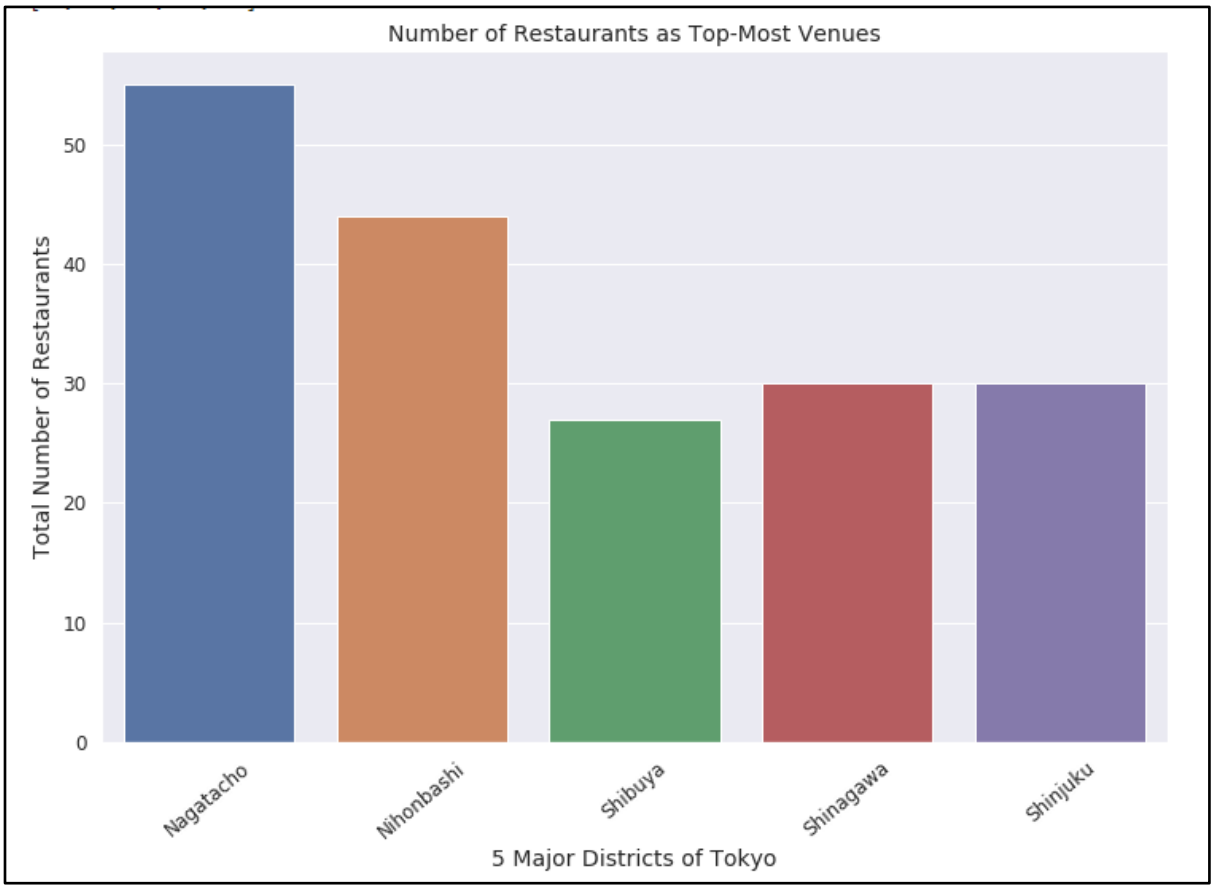

Figure 4: Number of Restaurant as Top Venues

3.2. Clustering the Major Districts of Tokyo

Finally, clustering these 5 districts based on the frequency of venue categories and, use K-Means clustering. So, the expectation would be based on the similarities of venue categories, these districts will be clustered. Using K-Means algorithm from Scikit-learn library and obtained 3 clusters as shown in Figure 5 below.

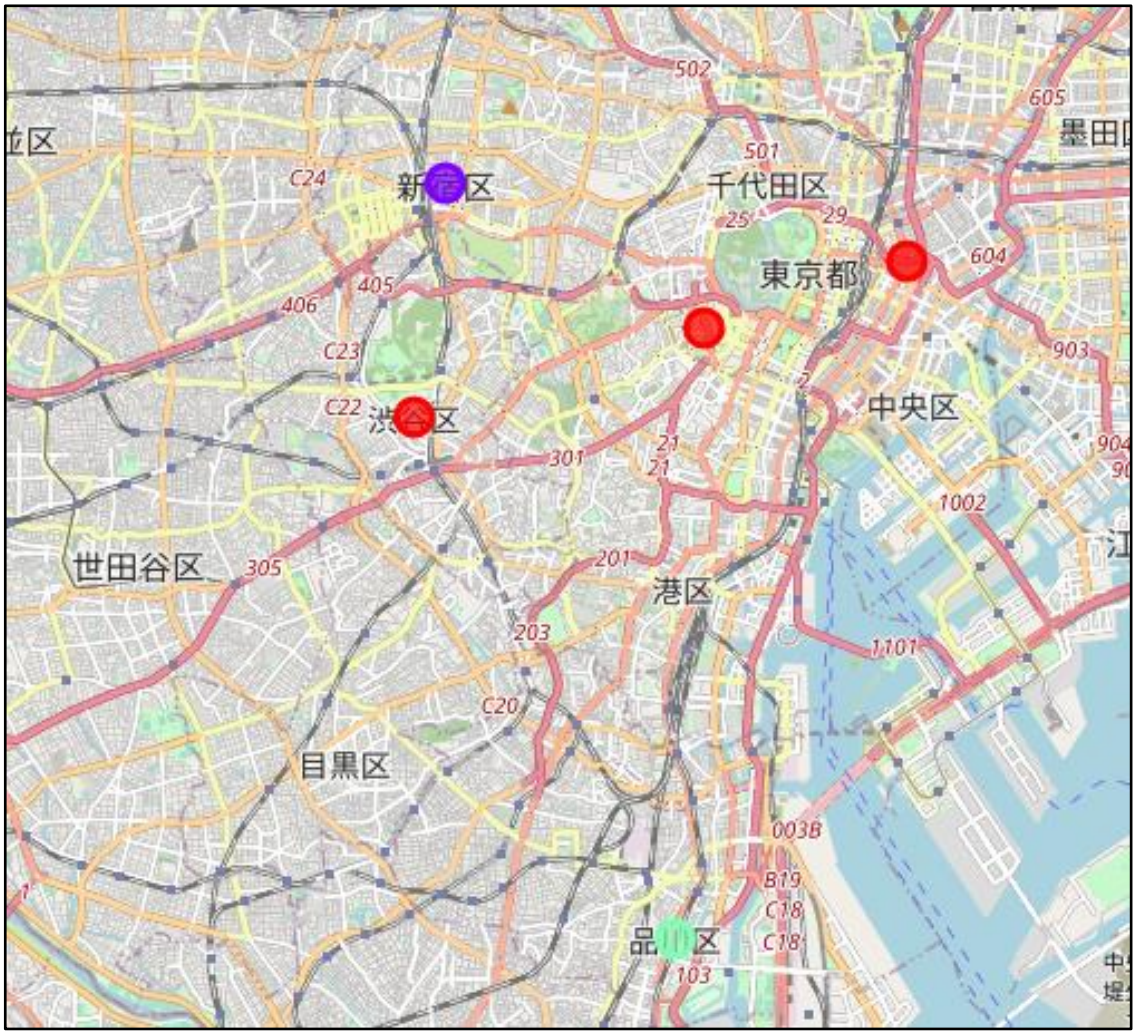

Figure 5: Cluster Venues

Here the radius of the circles represents the number of restaurants as most common venue for the corresponding district. From the most common venues this clustering makes a complete sense as Shinjuku are dominated by pubs, bars, and ramen restaurant falls under the purple cluster, whereas Nagatacho, Nihonbashi and Shibuya dominated by Japanese restaurant and cafe falls under red cluster and Shinagawa stands alone in green cluster.

4.0 Result and Discussion

According to this analysis, Shinagawa area will provide least competition for an upcoming lunch restaurant as convenience store is the most common venue in this area and the frequency of restaurants as common venue are very low compared to the remaining districts. So, definitely this region could potentially be a target for starting quality restaurants.

Some drawbacks of this analysis are the clustering is completely based on the most common venues obtained from Foursquare data. Since distance of the venues from closest stations, number of potential customers, benefits and drawbacks of Shinagawa being a port region, could all play a major role and thus, this analysis is definitely far from being conclusory.

However, it definitely give some very important preliminary information on possibilities of opening restaurants around the major districts of Tokyo. Also, another pitfall of this analysis could be consideration of only one major district of each ward of Tokyo, taking into account of all the areas under the 5 major wards would give an even more realistic picture.

5.0 Conclusion

As conclusion, Tokyo is known as one of the busiest city in the world. People who are working most likely want to have a good lunch break and business need to be done properly. The place like Shinagawa has least competition compared to others. So, Shinagawa is the best place to build a restaurant for a start-up.