

1. Introduction

Tokyo is one of the busy metro cities in the world. The lunch and dinner hours are extremely busy and have a variety of cuisines available with reasonable cost per meal. We will go through the impact of opening a restaurant in places which have multiple office spaces which increases the chances of success of the place. In such places, the profit is okay about 15 - 20 % but can also increase to about 35 % as we discuss here.

Here we focus on 5 busiest business areas of Tokyo and they are Shibuya, Chou, Shinjuku, Shinagawa and Chiyoda.

2. Data Preparation

Data cleansing

Using Wikipedia for getting Names of Wards, Major Districts and the Population

Wiki page of Tokyo Wards contains the table of 23 wards of Tokyo, area, population, and major districts. Using BeautifulSoup4 and Pandas, we have created initial df and cleansed the data a bit. The first entry of the major districts column in the Wikipedia table is considered for analysis. It gives a detailed picture of the corresponding ward. In the later part of the analysis we have considered topmost venues within 1-kilometer radius of the major district. The next step is to obtain coordinates using Geopy library.

Getting coordinates of the Major Districts

The dataframe obtained looks like 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	35.684058	139.774501
3	Minato	12180	248071	Odaiba	35.619128	139.779403
4	Shinjuku	18620	339211	Shinjuku	35.693763	139.703632
5	Bunkyo	19790	223389	Hongo	35.708800	139.760100
6	Taito	19830	200486	Ueno	35.711788	139.776096
7	Sumida	18910	260358	Kinshicho	35.696752	139.814151
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	35.588400	139.727900
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	35.730103	139.711884
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	35.446369	139.430925
22	Katsushika	12850	447140	Tateishi	34.176335	132.226020
23	Edogawa	13750	685899	Kasai	35.663400	139.873100

Obtaining average Land Price Data

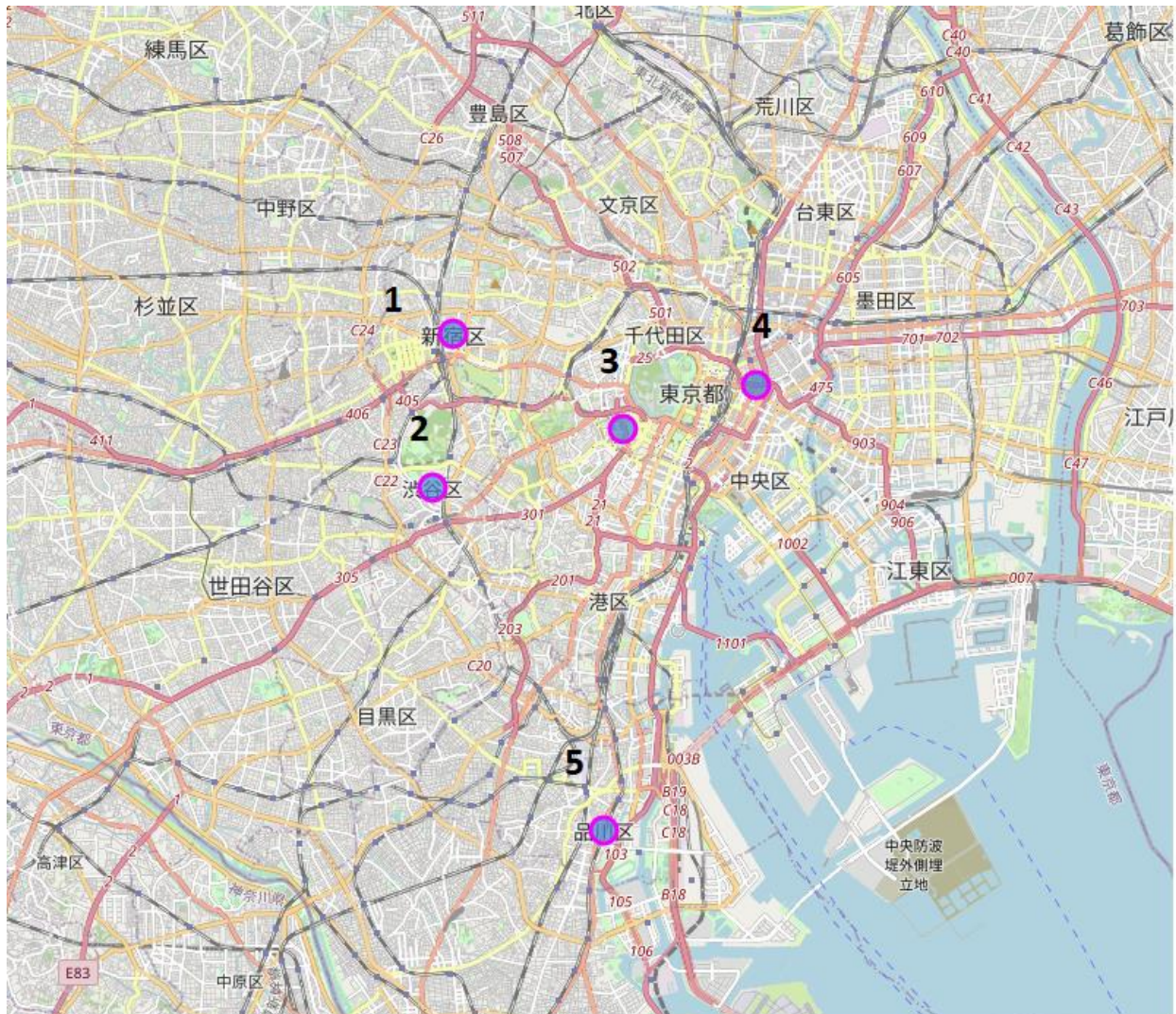
The average land-price data for each ward of Tokyo was obtained from Tokyo land market value page. Even though this data was not used for clustering, but it helps us to compare different districts of Tokyo for potentially opening a restaurant.

Foursquare Data

Using Foursquare API, we obtain the 100 most common venues within 1 kilometer of each major district.

3. Exploring the Data and Major Districts of Tokyo

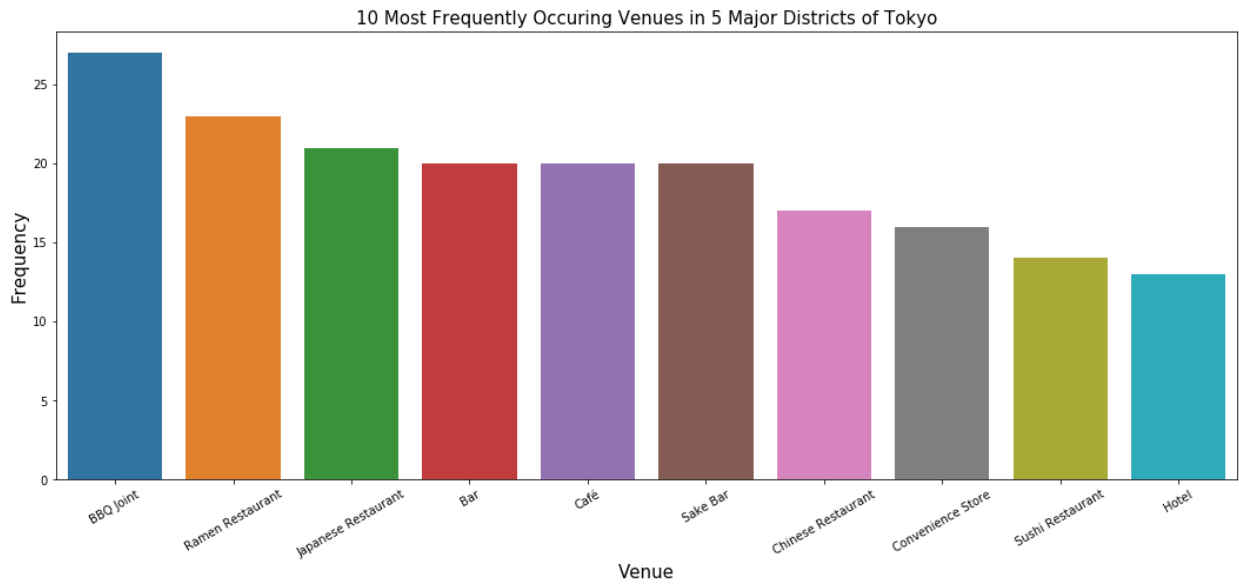
From the Foursquare data, we could see that there are \$134\$ unique categories, but for the current analysis lets concentrate only on restaurant options. Focusing on 5 major business districts as mentioned above, we find that there are 193 restaurants among the 500 top venues. Using folium library to plot a map of these restaurant we see as below 1) Shinjuku, 2) Shibuya, 3) Nagatacho, 4) Nihombashi, 5) Shinagawa.



Here, we could see that

The most common venues visited are 1) Ramen restaurant, 2) Japanese restaurants 3) BBQ joint.

Below is the chart for top 10 most visited venues in these 5 districts



Now, we obtain information of the top 5 venues using the below steps

- 1) Creating a df with pandas for venue category
- 2) Using pandas groupby on District column and obtain their mean
- 3) Transposing the df at step 2 and arrange it in descending order

This give the result as below

-----Nagatacho-----

	Venue	Freq
0	BBQ Joint	0.09
1	Japanese Restaurant	0.08
2	Coffee Shop	0.07
3	Ramen Restaurant	0.05
4	Chinese Restaurant	0.05

-----Nihonbashi-----

	Venue	Freq
0	Japanese Restaurant	0.07
1	Café	0.07
2	BBQ Joint	0.05
3	Soba Restaurant	0.05
4	Hotel	0.04

-----Shibuya-----

	Venue	Freq
0	Café	0.09
1	Record Shop	0.07
2	Bar	0.05
3	French Restaurant	0.04
4	Coffee Shop	0.03

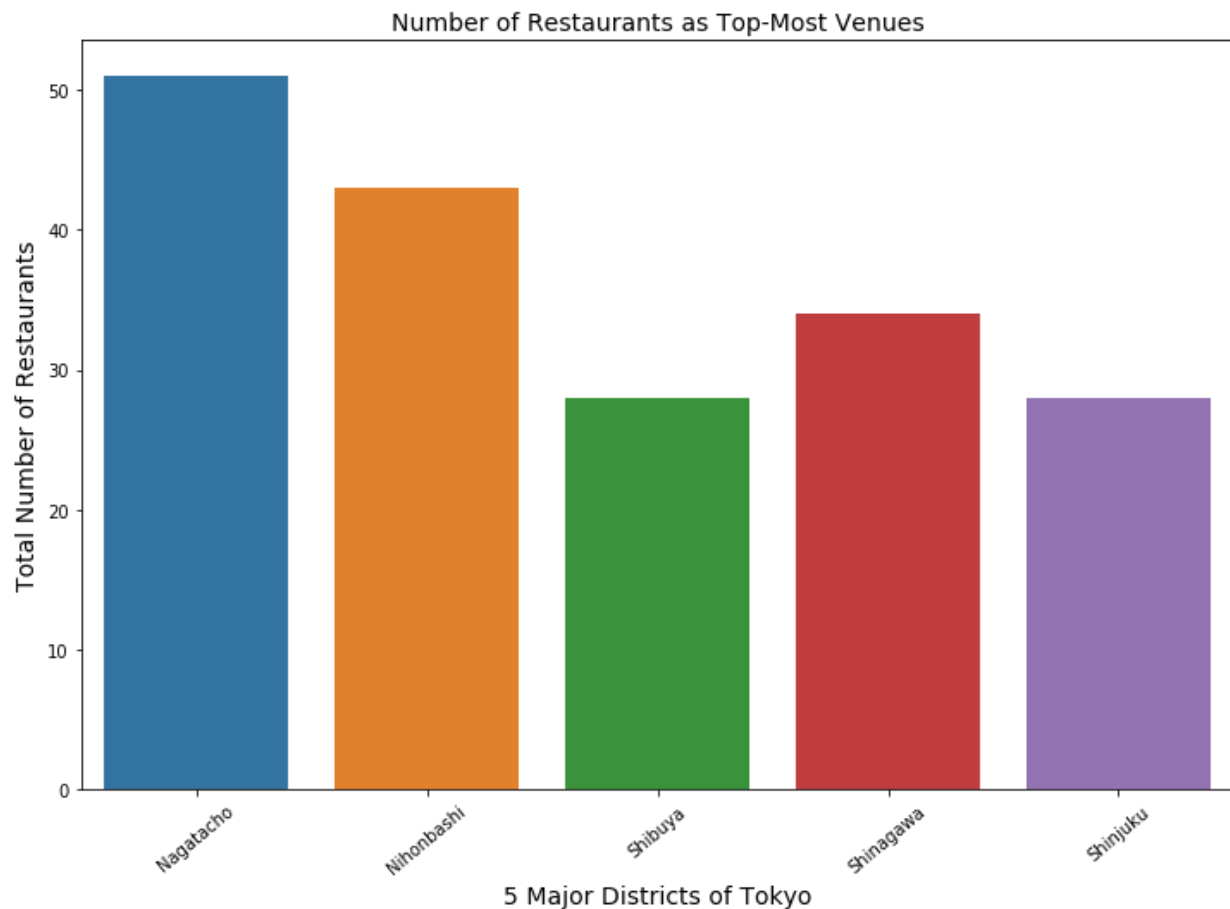
-----Shinagawa-----

	Venue	Freq
0	Convenience Store	0.16
1	Ramen Restaurant	0.08
2	Sake Bar	0.06
3	Grocery Store	0.05
4	BBQ Joint	0.05

-----Shinjuku-----

	Venue	Freq
0	Bar	0.09
1	Sake Bar	0.09
2	Ramen Restaurant	0.06
3	BBQ Joint	0.06
4	Pub	0.04

Looking at which district has highest number of restaurants as the most common venue, we get below plot



Clustering the Major Districts of Tokyo

Now, using k-means clustering, we cluster the 5 districts based on frequency of venue categories. So, our result would be based on the similarities of venue categories, these districts will be clustered. Using K-Means algorithm, we obtain clusters as shown below



From the most common venues this clustering makes a complete sense as Shibuya, Shinjuku are dominated by pubs, bars and cafe falls under purple cluster, whereas Nagatacho, Nihombashi dominated by Japanese and Chinese restaurants falls under red cluster and Shinagawa stands under green cluster.

4. Results

- Ramen restaurants lead as the most common venues in the 5 districts.
- Nagatacho district in Chiyoda ward and Nihombashi in Chuo ward have Japanese and Chinese restaurants as the most common venues.
- Shibuya and Shinjuku areas have bars, pubs, and cafe as the most common venues.
- Nagatacho has maximum number of restaurants as the most common venue whereas has Shibuya area has the least. But Cafe and BBQ joints are common.
- Since the clustering was based only on the most common venues of each district, Shinjuku, Shibuya fall under the same cluster and, Nagatacho, Nihombashi fall under another cluster. Shinagawa is separated from both clusters as, convenient stores stand out as the most common venue

5. Discussion

Shinagawa district is common for convenience stores and is not common for lunch restaurant and hence will provide very low competition for new restaurants

Some drawbacks of this analysis are-- the clustering is completely based on the most common venues obtained from Foursquare data. Since land price, 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 not much conclusory. However, it definitely gives some very important initial information on possibilities of opening restaurants around the major districts of Tokyo.

6. Conclusion

Finally, here we have got some insight of how real life we can use data-science in real life projects. Using Foursquare API and different types of maps and plotting graphs, we can easily analyze the data visually. This analysis was focused on seeing which district is better for opening a new restaurant and I hope it has helped in the same.