

Battle of Neighborhoods in Japan

Description of the Problem and Discussion of the Background

(Introduction Section)

Exploring the neighborhoods of Tokyo to obtain the 10 most frequently occurring venues in the top 5 districts of Tokyo.

Tokyo, the most populous metropolitan area in the world. Currently ranked 3rd in the global economic power index, Tokyo is definitely one of the best places to start up a new business.

During the daytime, especially in the morning and lunch hours, office areas provide huge opportunities for restaurants. Reasonably priced (one lunch meal 8\$) shops are usually always full during the lunch hours

(11 am -- 2 pm) and, given this scenario, we will go through the benefits and pitfalls of opening a breakfast cum lunch restaurants in highly dense office places. Usually the profit margin for a decent restaurant lie within 15 - 20% range but, it can even go high enough to 35%, as discussed. The core of Tokyo is made of 23 wards (municipalities) but, we concentrate on 5 busiest business wards of Tokyo – Chiyoda, Chuo, Shinjuku, Shibuya and Shinagawa, to target daily office workers.



We will go through each step of this project and address them separately.

Target Audience

1. Business personnel who wants to invest or open a restaurant. This analysis will be a comprehensive guide to start or expand restaurants targeting the large pool of office workers in Tokyo during lunch hours.
2. Freelancer who loves to have their own restaurant as a side business. This analysis will give an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.
3. New graduates, to find reasonable lunch/breakfast place close to office.
4. Budding Data Scientists, who wants to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it and, finally be able to tell a story out of it.

2. Initial Data Preparation:

2.1. Web-Scrapping and Cleaning (Week 1)

2.1.1. Get the Names of Wards, Major Districts and Population from Wikipedia

The Wikipedia page of [Tokyo Wards](#) contains the table of 23 wards of Tokyo, area, population and major districts. I have used [Beautifulsoup4](#) and pandas library 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'. Here I have taken the first entry of the major districts column in the Wikipedia table to concentrate on. Even though not complete but it gives us quite a detailed picture of the corresponding ward, as later on I have considered top most venues within 1 kilometer radius of the major district. After this initial preparation, I moved on to the next step to obtain coordinates using [Geopy](#) library.

2.1.2. Get the Coordinates of the Major Districts

Some of the coo-ordinates of the major districts returned by Geopy are wrong and I have figured 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), I had to google search and replace using pandas library. After little manipulation the obtained data-frame looks as below

Tokyo_df

	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

2.1.3. Obtain the Average Land Price Data from Web-Scrapping

The average land-price data for each ward of Tokyo was obtained from [Tokyo land market value](#) page. Even though this data wasn't used for clustering but it definitely helps us to compare different districts of Tokyo for potentially opening a restaurant.

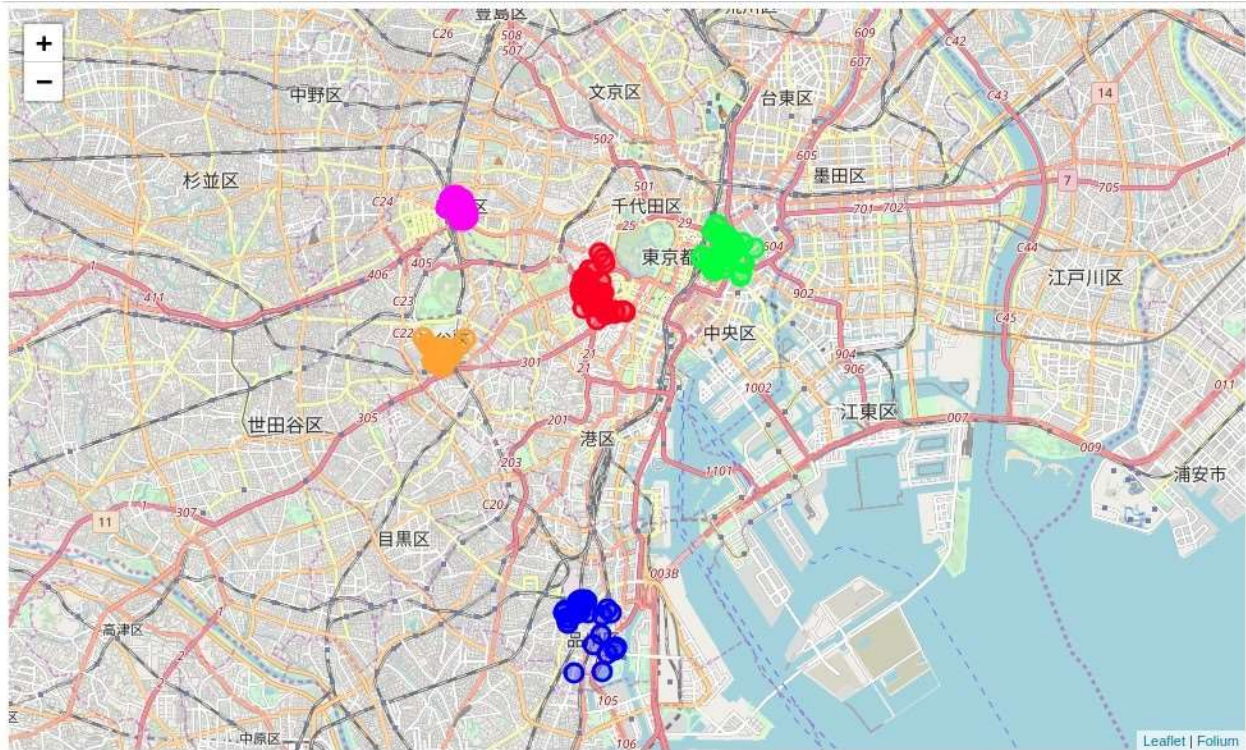
2.2. Foursquare Data (Week 2)

Finally, I make use of [Foursquare API](#) to obtain the 100 most common venues within 1 kilometer of each major district.

3.1. Exploring the Data and Major Districts of Tokyo

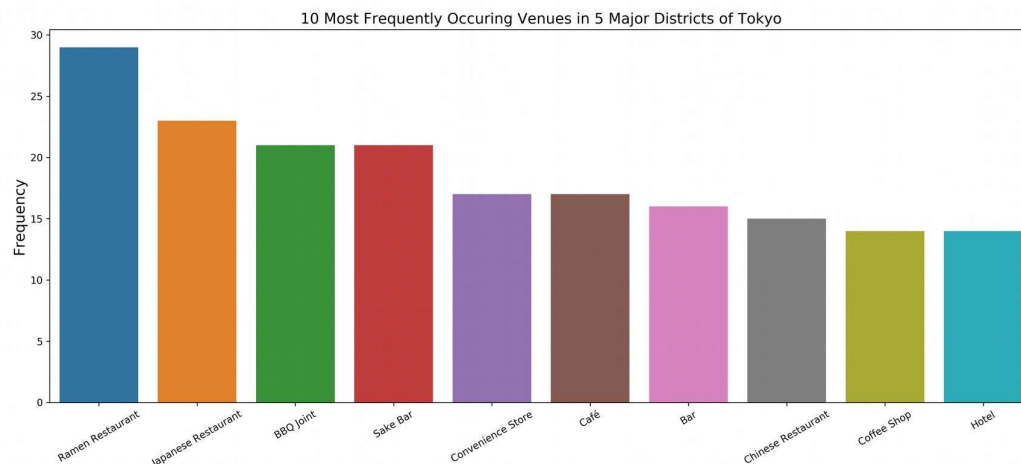
From the Foursquare data, we could see that there are \$134\$ unique categories, but for this analysis

I mostly later on concentrated in Restaurant category. As the focus is on 5 major business districts (Nagatacho, Nihombashi, Shibuya, Shinjuku, and Shinagawa), we found that there are 193 restaurants (searching for keyword Restaurant) among the 500 top venues in these 5 districts. I have used [Folium](#) library to plot a leaflet map of only these restaurants in these 5 major districts of Tokyo which is as shown below, where the colors representations are the following-- Nihombashi- Green, Nagatacho- Red, Shibuya- Orange, Shinjuku- Magenta, Shinagawa- Blue.



Here we have found out that

- **Ramen restaurants top the charts of most common venues in the 5 districts, followed by Japanese restaurants and BBQ joints.**
- A plot of the ten most frequent venues in these 5 districts are as below



4. Conclusion

Finally, to conclude this project, we have got a small glimpse of how real-life data-science projects look like. I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major districts of Tokyo and saw the results of segmentation of districts using Folium leaflet map. Potential for this kind of analysis in a real-life business problem is discussed in great detail. Also, some of the drawbacks and chance for improvements to represent even more realistic pictures are mentioned. Finally, since my analysis were mostly concentrated exploring the neighborhoods of Tokyo to obtain the most frequently occurring venues, some of the results obtained are listed as per the data given at the time of analysis.