

# Coursera Capstone

## IBM Applied Data Science Capstone

### MAKOTO Takamatsu

#### Introduction

In this project, we determine which areas of Tokyo and what business are most effective in attracting customers when starting a business in Tokyo. Tokyo is one of the major cities in the world. The main difference between this city and other cities is that the area of the city is very large. Tokyo is made up of 23 regions. Therefore, it is interesting to note that Tokyo has a variety of regional characteristics. For example, Tokyo-Shibuya(渋谷区) and Shinjuku(新宿区) are known as a town for young people. Also, Chiyoda(千代田区) is the political center of the city and has the National Diet Building and other facilities; Minato(港区) is the center of IT and other businesses. In order to start a business that effectively attracts customers, it is important to first decide which region to start in and which business fits the character of that region.

#### Business Problem

The goal of this capstone project is to know the best business that fits the location. If the business is best suited to the location, the store's profit margins will increase; if it's not a good fit, the time to close will be short.

#### Target Audience of this project

- Sales-people learn more about the area characteristics and do the promotion fit the city effectively.
- Investors can decide what businesses in that area are worth investing in.
- Learn which businesses a newcomer should start to increase their chances of success.

#### Data

Tokyo was selected as one of the cities for this project. Here are the datasets used in the project and why we chose them.

1. Tokyo region table from Wikipedia[1]: Scrap the Tokyo region table from Wikipedia, and obtain population, area name and the latitude and longitude coordinates of 23 major areas.
2. Land Values in Tokyo [2] : Scraping this website, we have obtained the land values for different areas of Tokyo.
3. Foursquare API [3] : Get information on all the businesses in the area you want to select with this API.

#### Methods

## Data Preparation

### Extracted information from Wikipedia

Scraping the coordinates of the area name, population, and latitude and longitude of the main area of the 23 wards of Tokyo from the Wikipedia page of Tokyo, using beautiful soup, and formatting the table data. Then, using Geopy, it acquires the latitude and longitude information of the area in the 23 wards.

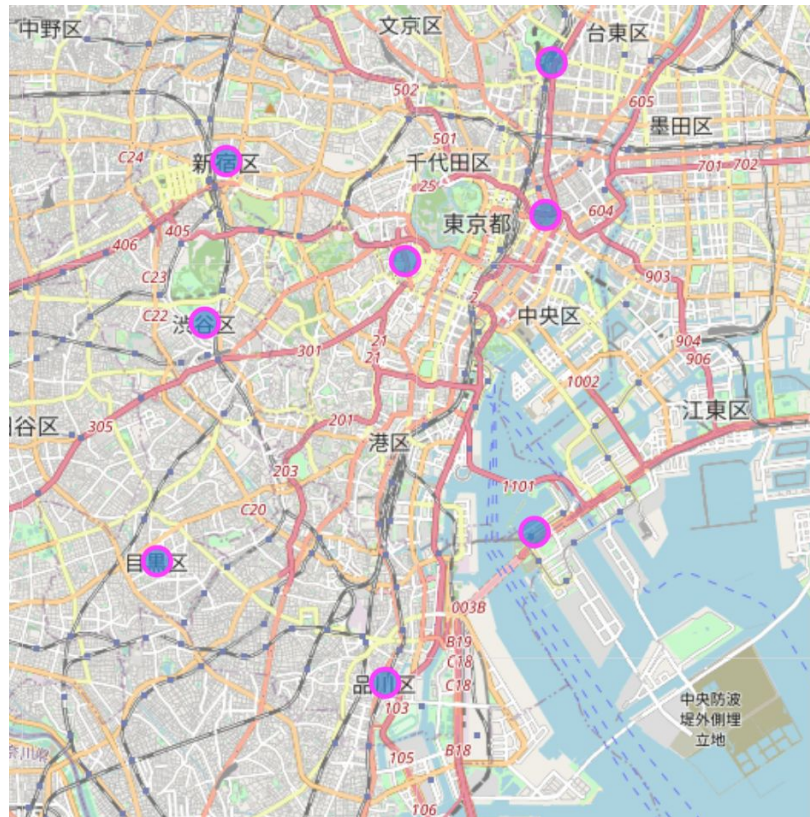


Fig1. Narrowing down the target area

### Determine the major areas of Tokyo from data

The data on land prices in Tokyo show that places with high land prices are important because they are expected to attract many customers for business. The average price quotes for each ward in Tokyo were obtained from Land Values in Tokyo [2], and the 8 wards (top 30%) were selected in descending order of average market price. As shown in the figure, the eight wards in Tokyo with the highest average land prices are Chiyoda, Chuoh, Minato, Shibuya, Taitoh, Shinjuku, Meguro, and Shinagawa.

	Ward	Average_Price(SqKm)
0	Chiyoda-Ku	2705898
1	Chuo-Ku	2699719
2	Minato-Ku	2121252
3	Shinjuku-Ku	939286
4	Bunkyo-Ku	945155
5	Taito-Ku	1064759
6	Sumida-Ku	617190
7	Koto-Ku	658047
8	Shinagawa-Ku	767398
9	Meguro-Ku	863712
10	Ota-Ku	560106
11	Setagaya-Ku	664106
12	Shibuya-Ku	1360332
13	Nakano-Ku	596154
14	Suginami-Ku	559122
15	Toshima-Ku	731392
16	Kita-Ku	547695
17	Arakawa-Ku	492675
18	Itabashi-Ku	436736
19	Nerima-Ku	402240
20	Adachi-Ku	295750
21	Katsushika-Ku	324876
22	Edogawa-Ku	332511
23	Hachioji-Shi	143350
24	Tachikawa-Shi	259136

Fig2. Scrapping the average price quotes in each area

### Analyze business trends in key areas

Use the Foursquare API to get a list of the 100 most active stores in the area within a mile of your target 8th district. The top businesses in Tokyo include Sake, BBQ joints and Japanese restaurants.

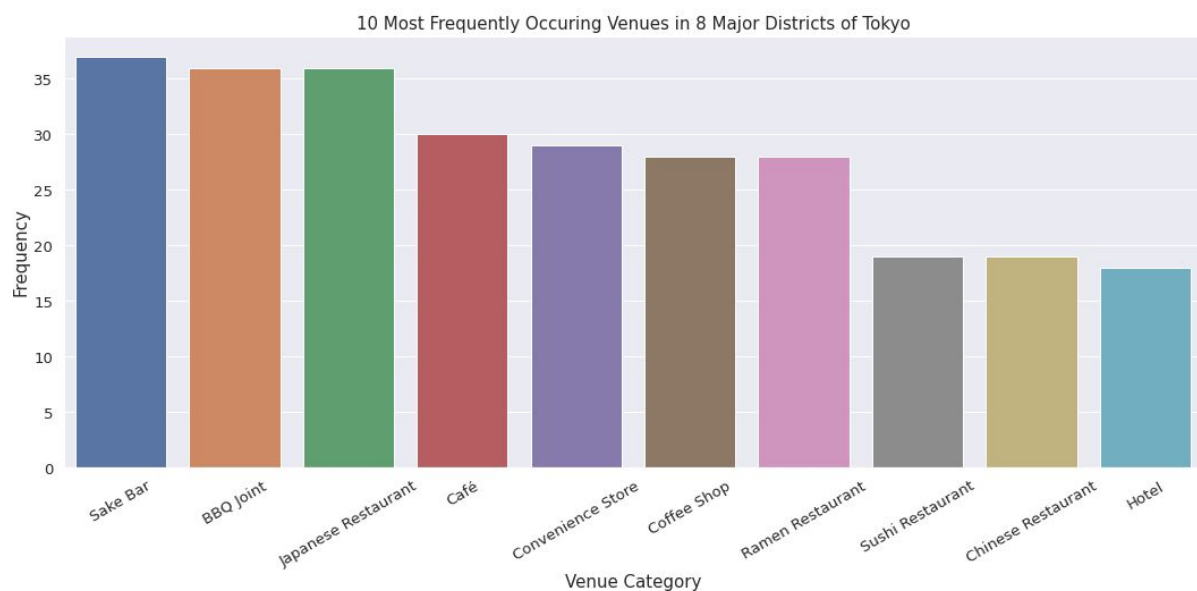


Fig3. The average of the top businesses in each region

Here's a list of the top 10 businesses in each district selected from Foursquare data. Figure 4 lists the top ten business rankings for each district.

----- Ueno -----			----- Shinjuku -----		
	Venue	Freq		Venue	Freq
0	Sake Bar	0.12	0	Sake Bar	0.08
1	Tonkatsu Restaurant	0.06	1	BBQ Joint	0.07
2	BBQ Joint	0.06	2	Bar	0.07
3	Chinese Restaurant	0.04	3	Ramen Restaurant	0.06
4	Café	0.04	4	Japanese Restaurant	0.04
5	Japanese Restaurant	0.04	5	Pub	0.04
6	Coffee Shop	0.03	6	Rock Club	0.03
7	Ramen Restaurant	0.03	7	Movie Theater	0.02
8	Science Museum	0.03	8	Thai Restaurant	0.02
9	Sushi Restaurant	0.02	9	Hotel	0.02

----- Meguro -----			----- Odaiba -----		
	Venue	Freq		Venue	Freq
0	Convenience Store	0.11	0	Shopping Mall	0.04
1	Coffee Shop	0.07	1	Japanese Restaurant	0.04
2	Café	0.06	2	Exhibit	0.04
3	Japanese Restaurant	0.06	3	Theme Park	0.04
4	Grocery Store	0.06	4	Discount Store	0.03
5	Ramen Restaurant	0.04	5	Park	0.03
6	Chinese Restaurant	0.04	6	Italian Restaurant	0.03
7	Park	0.04	7	Convenience Store	0.03
8	Bakery	0.03	8	Coffee Shop	0.03
9	Sake Bar	0.03	9	Plaza	0.03

----- Nagatacho -----			----- Shibuya -----		
	Venue	Freq		Venue	Freq
0	Japanese Restaurant	0.09	0	Café	0.10
1	BBQ Joint	0.08	1	Record Shop	0.07
2	Coffee Shop	0.07	2	Coffee Shop	0.05
3	Chinese Restaurant	0.05	3	Sushi Restaurant	0.03
4	Hotel	0.05	4	Nightclub	0.03
5	Szechuan Restaurant	0.04	5	French Restaurant	0.03
6	Korean Restaurant	0.03	6	Sake Bar	0.03
7	Dessert Shop	0.03	7	Bar	0.03
8	Italian Restaurant	0.03	8	Concert Hall	0.02
9	Ramen Restaurant	0.03	9	Seafood Restaurant	0.02

----- Nihonbashi -----			----- Shinagawa -----		
	Venue	Freq		Venue	Freq
0	Japanese Restaurant	0.06	0	Convenience Store	0.15
1	Café	0.06	1	Sake Bar	0.07
2	BBQ Joint	0.05	2	Ramen Restaurant	0.07
3	Hotel	0.04	3	BBQ Joint	0.05
4	Dessert Shop	0.04	4	Grocery Store	0.04
5	Hobby Shop	0.04	5	Sushi Restaurant	0.04
6	Bakery	0.04	6	Yakitori Restaurant	0.03
7	Sushi Restaurant	0.04	7	Park	0.03
8	Soba Restaurant	0.03	8	Donburi Restaurant	0.03
9	French Restaurant	0.03	9	Restaurant	0.02

Fig4. Top businesses in each region

Figure 5 lists the relative frequency of each business. The following table shows the distinctive trends in each district by business discipline. Looking at this figure, the restaurant trends are characteristic of the districts of Nihonbashi (Chuo-Ku) and Odaiba (Minato-Ku), where restaurants with a strong connection to Japanese cuisine abound, and Odaiba, where Italian restaurants are more compatible with shopping malls. A lot of noodles and BBQ.

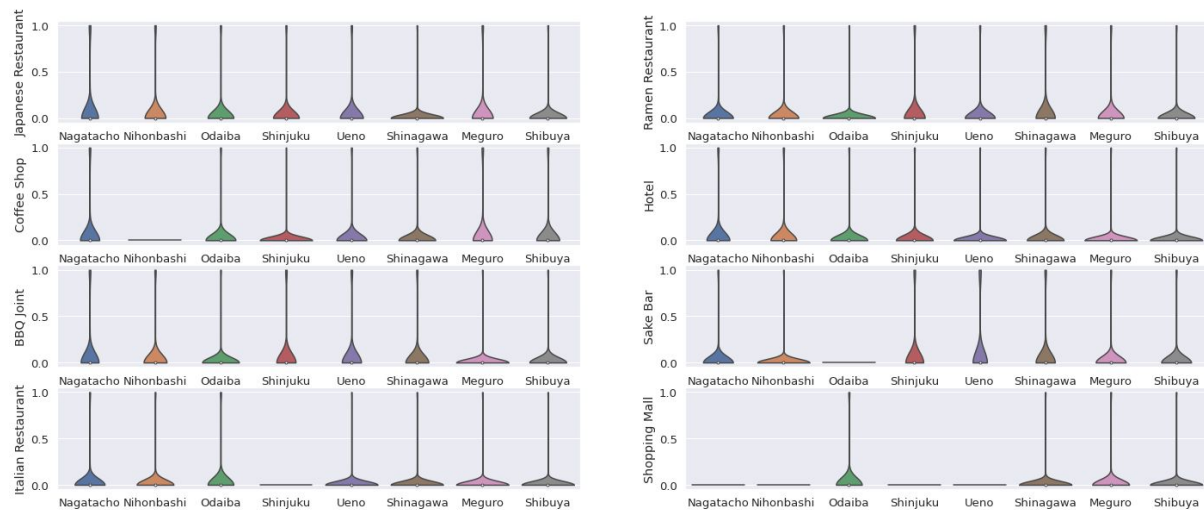


Fig5. The relative frequency of business in each region

### Clustering business trends in each district

Finally, we cluster these eight districts based on the category frequency of business in each district and use K-means clustering. Thus, our district-by-district business trend analysis is clustered based on the similarity of the categories in each district. Using the K-Means algorithm from the Scikit-learn library, we obtain the five clusters shown in Figure 6 below.

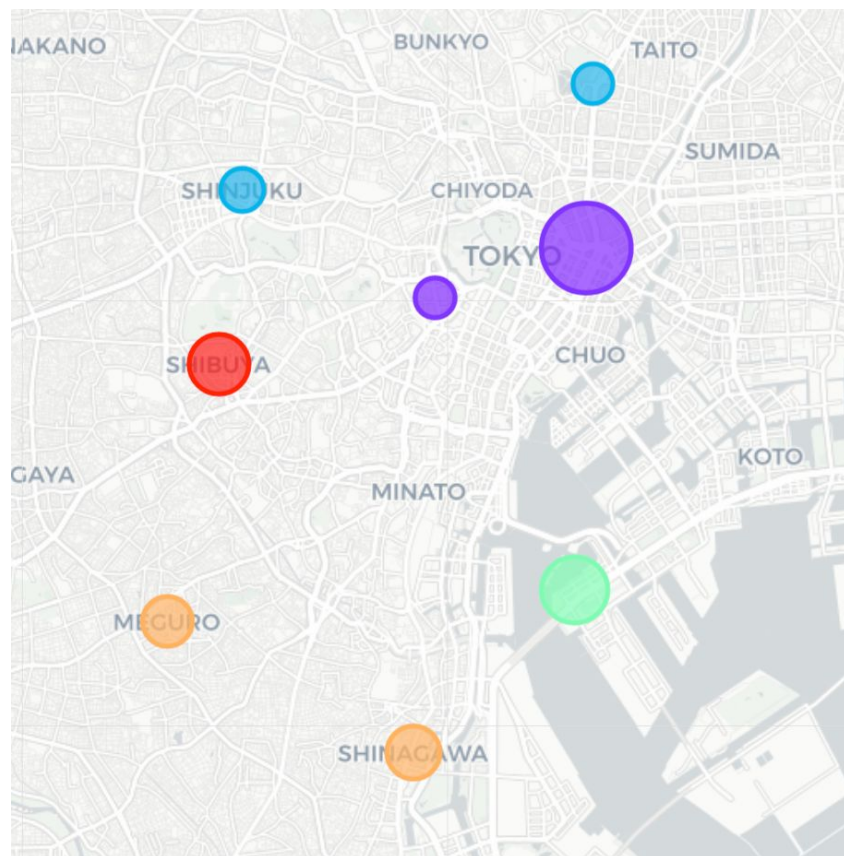


Fig6. Clustering of business trends

The radius of the cluster indicates the relative size of the businesses in each district. The similarities between the five clusters give us an idea of the characteristics and tendencies of the inhabitants of the area. Cluster 0 (Shibuya) is a young neighborhood dominated by record shops and nightclubs.

	Major_District	Dist_Latitude	Dist_Longitude	Average_Price(SqKm)	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Shibuya	35.664596	139.698711	1360332	0	Café	Record Shop	Coffee Shop	Nightclub	French Restaurant	Bar	Sushi Restaurant	Sake Bar	Stationery Store	Bookstore

Fig7. Red cluster 0

The purple cluster 1 (Nihongobashi, Nagatacho) is dominated by a Japanese restaurant.

	Major_District	Dist_Latitude	Dist_Longitude	Average_Price(SqKm)	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Nagatacho	35.675618	139.743469	2705898	1	Japanese Restaurant	BBQ Joint	Coffee Shop	Hotel	Chinese Restaurant	Szechuan Restaurant	Yakitori Restaurant	Ramen Restaurant	Korean Restaurant	Theater
2	Nihonbashi	35.684068	139.774503	2699719	1	Café	Japanese Restaurant	BBQ Joint	Dessert Shop	Hotel	Sushi Restaurant	Hobby Shop	Bakery	Gift Shop	French Restaurant

Fig8. Purple cluster 1

The light blue cluster 2 (Uno, Shinjuku) is dominated by Sake, which shows that business is more active in the evening and at night.

	Major_District	Dist_Latitude	Dist_Longitude	Average_Price(SqKm)	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Shinjuku	35.693763	139.703632	939286	2	Sake Bar	BBQ Joint	Bar	Ramen Restaurant	Japanese Restaurant	Pub	Rock Club	Department Store	Movie Theater	Thai Restaurant
5	Ueno	35.711795	139.776075	1064759	2	Sake Bar	BBQ Joint	Tonkatsu Restaurant	Japanese Restaurant	Chinese Restaurant	Café	Coffee Shop	Ramen Restaurant	Science Museum	Bed & Breakfast

Fig9. Light blue cluster 2

Green Cluster 3 (Odaiba), with its theme park and shopping mall, is a gathering place for young families.

	Major_District	Dist_Latitude	Dist_Longitude	Average_Price(SqKm)	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Odaiba	35.626722	139.772101	2121252	3	Theme Park	Japanese Restaurant	Shopping Mall	Exhibit	American Restaurant	Discount Store	Park	Italian Restaurant	Coffee Shop	Plaza

Fig10. Green cluster 3

Cluster 4 in Orange (Meguro, Shinagawa) is dominated by a convenience store, which is intended for office workers and people to live and work during the daytime.

	Major_District	Dist_Latitude	Dist_Longitude	Average_Price(SqKm)	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Shinagawa	35.599252	139.738910	767398	4	Convenience Store	Ramen Restaurant	Sake Bar	BBQ Joint	Sushi Restaurant	Grocery Store	Park	Donburi Restaurant	Yakitori Restaurant	Bar
7	Meguro	35.621250	139.688014	863712	4	Convenience Store	Coffee Shop	Japanese Restaurant	Grocery Store	Café	Chinese Restaurant	Ramen Restaurant	Park	Dessert Shop	Bakery

Fig11. Orange cluster 4

## Result and Discussion

A glimpse of Tokyo's business data may provide interesting insights that may be useful to investors, business starters, and tourists alike.

- Tokyo's top ten businesses are dominated by restaurants of various categories, except for convenience stores and hotels.

- The areas with the highest average market prices in Tokyo (Chiyoda, Chuo, Minato, and Shibuya) are located in a circle. It belongs to the Yamanote Line. Areas away from the Yamanote Line (Koto, Sumida, Ota) have lower average market prices.

Clustering is based on the most common locations taken from Foursquare data.

- In the five clusters, people come together in the area for different purposes, with many different restaurant genres. For example, the dominant cluster of convenience stores is the district where salaried workers work. The clusters, dominated by record shops and nightclubs, are the areas where young people tend to congregate.

However, our analysis ignores other factors such as the distance from the nearest train station to the venue, the price range of the restaurant, and the Michelin restaurant, among others. This is due to the lack of such data in a small exploratory study such as ours. Therefore, our analysis helps travelers to get a rough overview of the distribution of restaurants in the 23 main districts of Tokyo. Additionally, this result may be different if other clustering methods, such as DBSCAN, are used.

## **Conclusion**

To complete this project, we got a little glimpse of what a real-world data science project might look like. Similarly with real-world problems, we find solutions to those problems by analyzing big data and visualizing it in a visual way.

We saw the results of scraping web data using the Python library, exploring the major districts of Tokyo using the Foursquare API, and dividing up the districts using the Folium leaflet map.