Battle of the Neighborhoods

-A guide map for Tokyo Special Wards –

I. Introduction

1.1 Background

Tokyo is the capital, as well as the political, economic, and cultural center of Japan. Tokyo is a prosperous and bustling Metropolis, which includes 23 special wards, among some of them are very famous, even for foreigners are also familiar with, such like Shinjuku, Shibuya, or Ginza. As one of the world's top economic giants, theoretically, there should be unlimited business opportunities in Tokyo, but at the same time, expensive prices in local and a trend towards saturated market which are not easy for new business starters to achieve their ambitions. In addition, the unique characteristics of society and humanities in Japan make it fairly difficult to grasp the pulse of Japanese tastes or preferences.

Therefore, this project aims to start with exploring the 23 special wards in the Tokyo Metropolis, analyzing the geographical, demographical, and social datasets, summarizing their characteristics for drawing a rough image sketch as well as getting a general understanding of the situation there. In addition, we use Foursquare to get the most common venues in each special ward and encapsulate the ranking of best popular venues to provide relevant and valuable information guide for the target groups.

1.2 Overview of project

- **Description of Business Problem**

There exist already countless cuisines or gastronomy in the flourishing areas of Tokyo. Nevertheless, some business opportunities are still waiting for being discovered. Tokyo is a massive market, if you want to start a new business in Tokyo, such as a restaurant, a Café or a bakery, the first task is to find a suitable location, such as business districts, tourist attractions, office building gathering blocks or universities, etc. This project is although preferentially fit for developing the gastronomic business, but it also has reference value for other types of businesses.

To say it briefly:

Find suitable location to start gastronomic business in the range of 23 special wards of Tokyo

- **Target groups**

New in Japan and Tokyo? Interested in searching a cool apartment in the top neighborhoods of Tokyo? Want to start a new business (e.g. shops, gastronomy etc.) in the heart of Tokyo? Or just need some tips for exploring and sightseeing in Tokyo?

So, if you are:

- A business starter seeking for the business opportunity
- A backpacker who is excited for your first adventure
- A newcomer who decided to begin a brand new life
- A data scientist who wants to learn about the Tokyo society from statistical perspective
- Or any other reasons related to Tokyo

Then, here is a statistic based guide map which contains useful information that may help you finding the right direction.

II. Data collection, pre-processing

2.1 Data source

- Main data

The information collected by Foursquare for the most common venues in all 23 special wards of Tokyo, which is used to determine the most popular categories of venues in each ward, building the foundation for future business.

- Supplementary data

Data on the geographical, demographic, social, and economic conditions in special wards to understand statistic information such as local income levels, consumption levels, and consumption preferences etc. , which are helping to define and refine the business target population in Tokyo.

2.2 Data pre-processing

2.2.1 Retrieve data from below original data source

① Wikipedia webpage for introduction of Tokyo Special Wards

Data source : https://en.wikipedia.org/wiki/Special_wards_of_Tokyo

Here you can find basic information about Tokyo's 23 special wards (population, population density, area, main neighborhoods in each area, etc. This will help us get a rough idea about basic information of each special wards in Tokyo.

② Demographic data about population in Tokyo Special Wards

Data source : https://www.toukei.metro.tokyo.lg.jp/jsuikei/2020/js203a0000.xls

The demographic information of Wikipedia above is 2016, which is relatively outdated. To obtain the most accurate information, we queried the website from the Tokyo Metropolitan Bureau of Statistics and found the actual demographic data of 2020. The latest version for 2020 will also be used in further processing.

③ Average monthly rental market prices in Tokyo Special Wards

Data source : https://www.daiwahouse.co.jp/chintai/tokyo/souba/

Here you can find the information about average monthly rent price of each district in Tokyo. It should be noted that the price used here is the average market price of all room types, that is, other factors such as room size and specific location are not considered.

④ Education level (graduation rate of university) in Tokyo Special Wards

Data source : http://wildhog.hatenablog.com/entry/2018/05/01/170000

This website provides the university graduation rate of each district in Tokyo in 2018 (the information for 2019 and later is temporarily not found), from which we can know the total number of college students in each special ward, the number of university graduates, the university graduation rate and so on.

⑤ Average annual income in Tokyo Special Wards

Data source :

https://www.nenshuu.net/prefecture/shotoku/shotoku_pre.php?prefecture=%E6%9D%B1%E4%BA%AC%E9%83%BD

Here you can find information on the annual income per capita of each district in Tokyo. The data is very simple, without considering the specific industry, type of work, gender, and other factors.

⑥ Geographical data

GeoPy : https://geopy.readthedocs.io/en/stable/

GeoPy is a Python module used for locating the coordinates of addresses, cities, countries, and landmarks worldwide. This project is used to obtain the geographical coordinates of each special ward in Tokyo.

⑦ Foursquare

Foursquare : https://de.foursquare.com/

Foursquare is a location-based recommendation service for restaurants and other places. It is used to obtain information on the most popular venues (restaurants, cafes, sightseeing spots, etc.) in various districts of Tokyo. It is also the main source of information for this project.

2.2.2 Data converting and cleaning

After collecting the original data from the above data sources, imported all of them into the project notebook, and then import the required tools and libraries. From now, we can start the data pre-processing. Firstly, convert the original data into pandas data frames. Then filter and clean the data to remove outliers and unnecessary information.

- Demonstration of processed datasets

  (Note: here only shows the head of each data frame i.e. first 5 rows)

① Wikipedia webpage for introduction

② Updated demographic data

Table 1. Basic information and updated demographic data of Tokyo

	Name	Kanji	Major_districts	Population(2020)	Area(km²)	Density(/km²)
0	Chiyoda	千代田区	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	66080.0	11.66	5667.0
1	Chūō	中央区	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...	168553.0	10.21	16509.0
2	Minato	港区	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	260535.0	20.37	12790.0
3	Shinjuku	新宿区	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...	349101.0	18.22	19160.0
4	Bunkyō	文京区	Hongō, Yayoi, Hakusan	236043.0	11.29	20907.0

③ Average monthly rental market prices (MM_yen: million yen)

Table 2. Rental market prices

	Kanji	RMP/(MM_yen)
0	千代田区	22.2
1	中央区	13.9
2	港区	19.4
3	新宿区	14.5
4	文京区	12.9

④ Education level (graduation rate of university)

Table 3. Level of education

	Kanji	Total	Graduates	Uni_Grad	Uni_ratio
0	千代田区	41978	38922	14290	36.7%
1	杉並区	432766	396403	143649	36.2%
2	中央区	109813	104866	37566	35.8%
3	文京区	182238	163476	58188	35.6%
4	港区	179914	170019	53193	31.3%

⑤ Average annual income in Tokyo Special Wards (MM_yen: million yen)

Table 4. Average annual imcome

	Kanji	AvAn/(MM_yen)
0	港区	1217
1	千代田区	1081
2	渋谷区	872
3	中央区	690
4	目黒区	637

⑥ Geographical data

Table 5. Geographical data

	Name	Kanji	Major_districts	Population(2020)	Area(km²)	Density(/km²)	MD_Latitude	MD_Longitude
0	Chiyoda	千代田区	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	66080.0	11.66	5667.0	35.693810	139.753216
1	Chūō	中央区	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...	168553.0	10.21	16509.0	35.666255	139.775565
2	Minato	港区	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	260535.0	20.37	12790.0	35.643227	139.740055
3	Shinjuku	新宿区	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...	349101.0	18.22	19160.0	35.693763	139.703632
4	Bunkyō	文京区	Hongō, Yayoi, Hakusan	236043.0	11.29	20907.0	35.718810	139.744732

⑦ Foursquare (only demonstrate a small part of the original entire data frame)

Table 6. List of venues

	Major_districts	MD_Latitude	MD_Longitude	Venue	VN_Latitude	VN_Longitude	Category
0	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.69381	139.753216	Bondy (欧風カレーボンディ)	35.695544	139.757356	Japanese Curry Restaurant
1	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.69381	139.753216	Nippon Budokan (日本武道館)	35.693356	139.749865	Stadium
2	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.69381	139.753216	National Museum of Modern Art (東京国立近代美術館)	35.690541	139.754694	Art Museum
3	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.69381	139.753216	Kitanomaru Park (北の丸公園)	35.691653	139.751201	Park
4	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.69381	139.753216	Kanda Coffee	35.697455	139.754686	Café

Finally, we have got the required data sets save the processed data for further analysis.

III. Methodology

3.1 First glimpse of Tokyo

Before we actually starting our search for suitable locations for business developing, let us firstly have a glimpse of the overview of local economic, cultural, and other social situations of Tokyo, which can definitely help us to get clearer idea for the decision making later. For example, here are several typical problems from analytical and statistical perspectives, which are also mostly being interested at the same time.

- Income vs. expenses (here referred to rental prices)

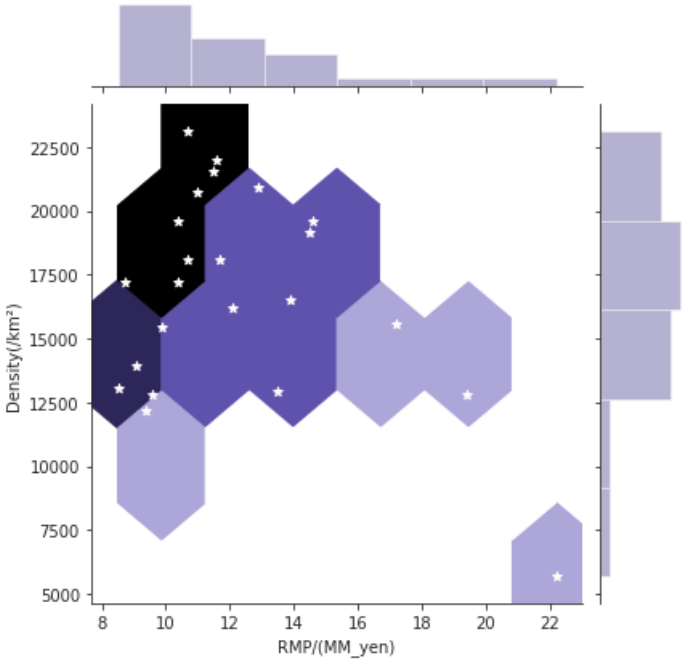

Fig.1 Relationship between population density and rental prices

From the plot, the majority of the markers are concentrated in the upper left corner of the figure, indicating that districts with higher population density have mostly relative lower rental prices, which is in line with the market discipline. Wards with low population density, such as Chiyoda and Minato, have the highest rental prices on the contrary. Interestingly, Edogawa and Katsushika, which are with relative less population, also have the lower monthly rent. As a result, there is basically no strong linear correlation between population density and monthly rent in Tokyo 23 special wards.

- Education level vs. income level

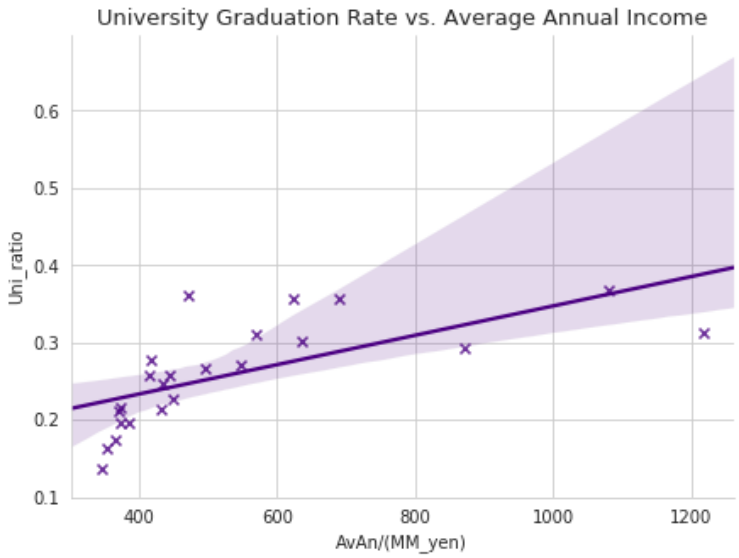

Fig.2 Relationship between education level and average annual income

Here we can see the top 3 players: Chiyoda, Minato, and Shibuya in terms of average annual income ranking(in the right half of plot), are not very prominent in university graduation rate compared to other wards, and even slightly lower than some of other wards. As there is no significant difference among the overall university graduation rate in each special ward. while considering other factors (e.g. individual real estate, untaxed income etc.) could also influence the average annual income, so we can conclude that no strong linear correlation is confirmed between the university graduation rate and average annual income.

- Population density vs. rent prices

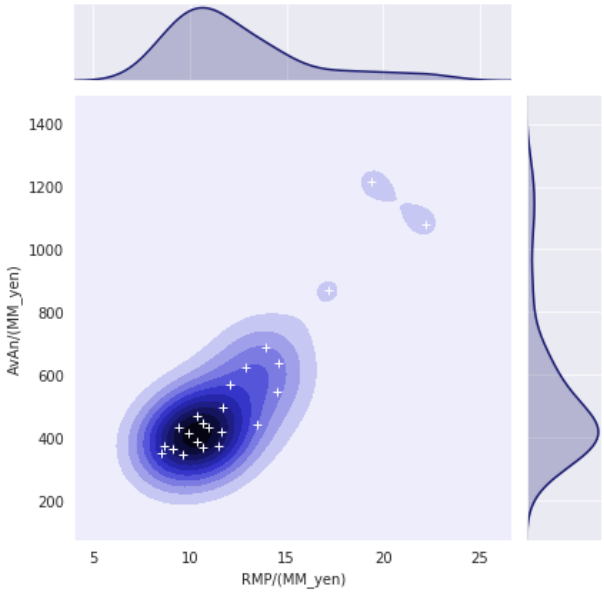

Fig.3 Relationship between average annual income and rental prices

The density curves of annual income and monthly rent price have roughly the same trends, that is to say, the ratio of income and price is positively correlated in most wards. Moreover, the income and price ratio in the majority of wards are relatively similar. From the above plot, quite a lot of wards that have average annual income of around 4 million yen, while their monthly rent is mainly around the level of 100,000 yen (markers gathering the cluster in the under left corner). To calculate in this way, the average annual rent expenditure is about 1.2 million yen, accounting for almost one-third of the entire income.

This is actually not a good signal. Although it is far from enough to draw a convictive conclusion only from observing the rent prices perspective, but it still strongly shows that prices in Tokyo are way too expensive, and the ratio of income and expenditure is obviously unbalanced.

As always, the best 3 outstanding players are still Chiyoda, Minato, and Shibuya.

3.2 Foursquare

With the above background research, now we are able to start our investigation of the existing venues in Tokyo. First, to create a map of where all special wards locate.

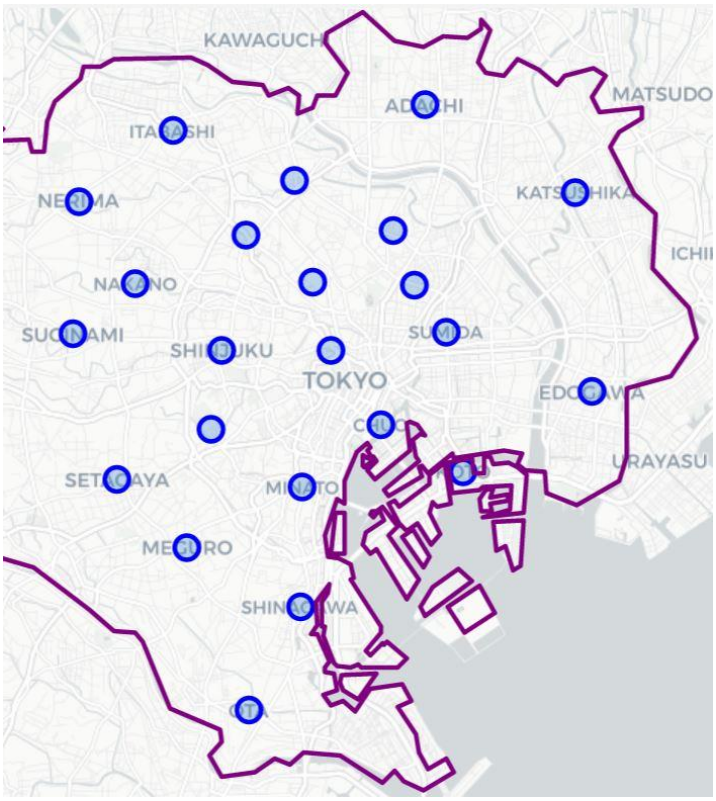

Fig.4 A map of Tokyo special wards

Next, using Foursquare to search the venues of each district in the database. As the result, we obtained is a total of number of 2300 venues in 234 different unique categories. In order to present the density of venues in each ward intuitively, we apply a heatmap to show the distribution of all venues. According to the created heatmap, we can clearly define the 'hottest' places are located in the upper left corner.

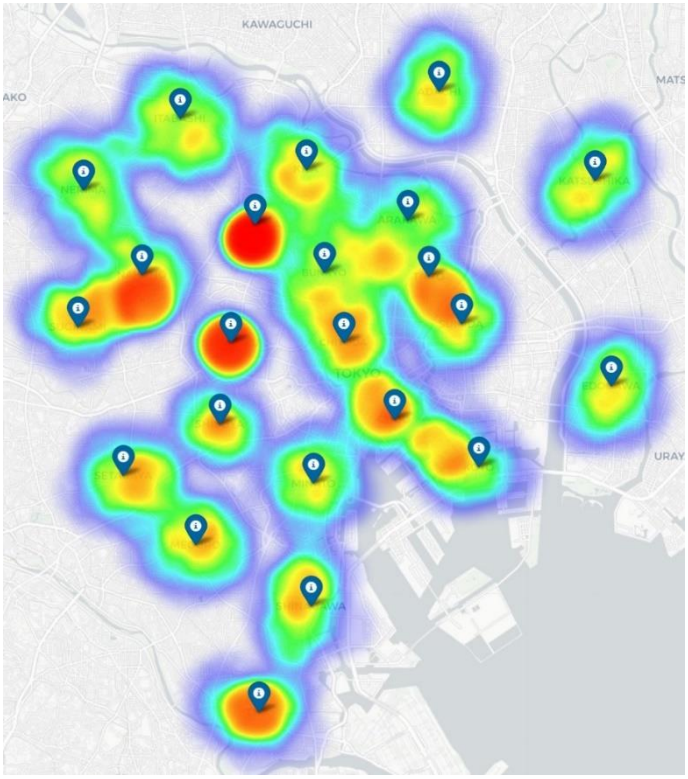

Fig.5 Heatmap of venues density

3.3 Battle of neighborhoods

Now, we use various data analysis methods and visualization tools to process the collected data further and to find ideal ‘candidate’ areas in Tokyo 23 special wards where are suitable for starting new (gastronomic) business.

Round 1 : Location selection from perspective of population density

Let us first zoom in the above heatmap and have a closer look at areas where venues are most densely concentrated, i.e. the hottest corner on the map.

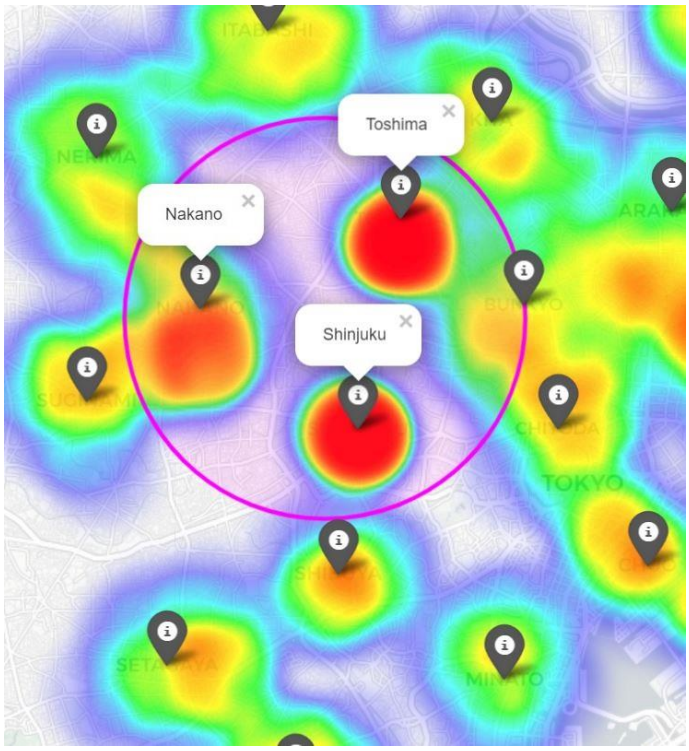

Fig.6 Wards with most concentrated venues

We can easily learn about that the most venue-concentrated regions are around Shinjuku, Toshima, and Nakano. As one of the three major sub-centers in Tokyo, it is not surprising that Shinjuku has very high-density of venues. But why are Nakano and Toshima?

Table 7. Wards with highest population density

	Name	Density(/km²)
4	Bunkyō	20907.0
5	Taitō	20718.0
13	Nakano	21977.0
15	Toshima	23106.0
17	Arakawa	21538.0

The answer is found by filtering the data based on a population density of more than 20,000 / km², the population density of Toshima and Nakano is among the top ones. Therefore, with no surprise to see why the venues are concentrated in these two wards as well.

From above analysis, it seems that areas with high population density and at the same time also already existing a large number of venues are maybe not ideal 'candidates', as the actual situation is more likely towards the "Red Sea". That is, the market is getting saturated and the competition would be cutthroat, which is not very friendly for new or small business starters.

But how about other wards where the population density is high but have not equipped with too many venues yet? Let us take a view at the remaining three wards: Bunkyo, Arakawa, and Taito.

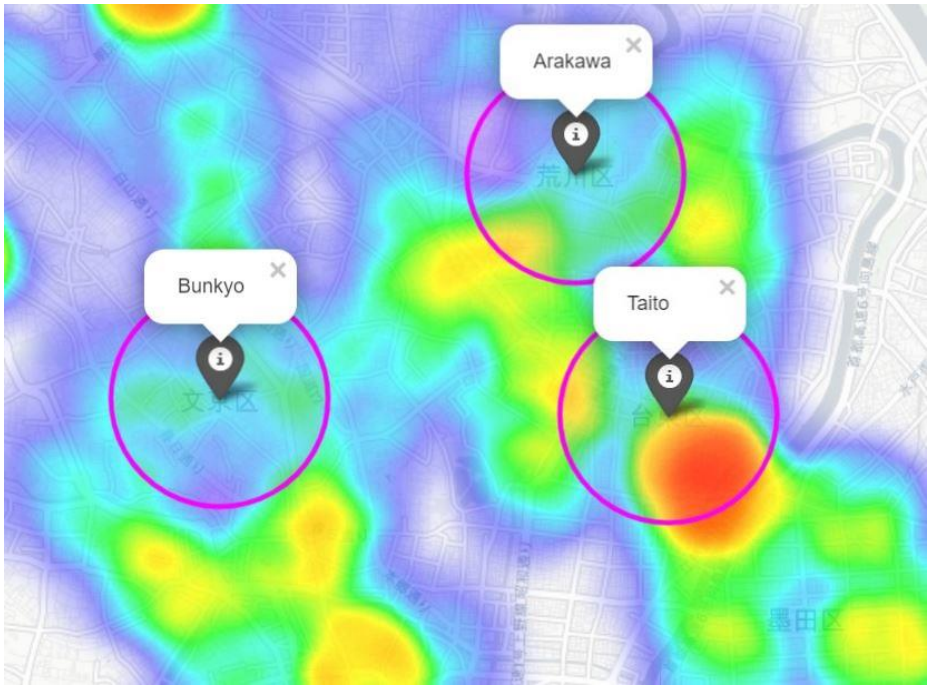

Fig.7 Bunkyo, Arakawa, and Taito

Except that Taito seems a little bit crowded, Bunkyo and Arakawa look quite good. By the way, Bunkyo is a region where many universities in Tokyo gather together, and the most famous of them is of course the University of Tokyo. Moreover, places surrounding the universities are always able to keep a stable foot traffic (except for the summer and winter holidays), as well as better public environmental security. Therefore, they are ideal choices for starting business.

The winners of round 1 : Bunkyo, Arakawa

Round 2 : Location selection from perspective of average annual incomes

Except for population density, other factors such as rental price or local average annual income also play important roles involved in the decision making. The rent price greatly influences the business operating cost while local average income determining the consumption level (i.e. the selling price). If the pricing is too far from the local average income level, it is more likely that the business may not be attractive enough for target groups to come.

In order to solve this problem, firstly selecting the wards where the people earn the highest income. Here using the filtering condition with: average annual income > 6 million Yen (≈ USD 56000), while also observing the market rental prices accordingly.

Table 8. Wards with highest average annual income

	Name	AvAn/(MM_yen)	RMP/(MM_yen)
0	Chiyoda	1081	22.2
1	Chūō	690	13.9
2	Minato	1217	19.4
4	Bunkyō	623	12.9
9	Meguro	637	14.6
12	Shibuya	872	17.2

Among the above winners, they can be further divided into 2 groups. Chiyoda, Minato, and Shibuya are in the 'elite group' among the winners as always, which are occupying the top three positions. Their average annual income is over 8 million yen. The members in second echelon also achieve impressive average annual income of more than 6 million Yen. In above 6 wards concentrate the majority of the potential target groups with high consumption level in Tokyo.

Nevertheless, if taking a look at the monthly rent price again, we will find that the average monthly rent of the 'elite group' is extremely expensive, even far exceeding the prices in other parts of Tokyo. This is not a small burden which also brings a lot of pressure for the new business starters. Therefore, they are not suitable choices. Under the condition of ensuring that target customers have a certain level of consumption, what is more, reducing fixed expenses(i.e. cost) is an efficient means for smoother business development as well. Thus, personally speaking, the three wards in the second echelon (Chuo, Bunkyo, and Meguro) are more suitable for location selection.

Then, to visualize the 3 players in second echelon on the heatmap.

Fig.8 Bunkyo, Chuo, and Meguro

It seems that the venues density of Chuo and Meguro is at a similar level (Chuo may be slightly higher than Meguro). Chuo has one of Tokyo's most famous districts - Ginza, and it is also a financial center where the Central Bank of Japan and the stock exchange are located. Meguro is relatively 'low-profile' compared with Chuo, where mostly equipped with high-end residential neighborhoods and Konzerns or enterprises as the mainstay.

Among those winners, one player that caught our eyes: yes, Bunkyo. As one of the winners from last battle, this time, Bunkyo has won the competition again.

The winners of round 2 : Bunkyo, Chuo, Meguro

Round 3 : Most popular venues in Tokyo

Similarly, the choice of business type is closely related to the quantity of existing business portfolios. Therefore, in the third round we are going to study the number and categories (business type) of existing venues by now. The venues data collected from Foursquare are sorted, grouped, and categorized as further processing, and then, finally we gained the top 10 common venues in each special ward.

Table 9. Top 10 common venues in each ward

	Major_districts	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4 th Most Common Venue	5 th Most Common Venue	6 th Most Common Venue	7 th Most Common Venue	8 th Most Common Venue	9 th Most Common Venue	10 th Most Common Venue
0	Akabane, Ōji, Tabata	Ramen Restaurant	Sake Bar	Japanese Restaurant	Park	Shopping Mall	Museum	Grocery Store	BBQ Joint	Bath House	Chinese Restaurant
1	Arakawa, Machiya, Nippori, Minamisenju	Ramen Restaurant	Japanese Restaurant	BBQ Joint	Sake Bar	Park	Coffee Shop	Chinese Restaurant	Bakery	Supermarket	Grocery Store
2	Ayase, Kitasenju, Takenotsuka	Convenience Store	Ramen Restaurant	Discount Store	Supermarket	Italian Restaurant	Restaurant	Steakhouse	Café	Clothing Store	Donburi Restaurant
3	Hongō, Yayoi, Hakusan	Café	Ramen Restaurant	BBQ Joint	Japanese Restaurant	Chinese Restaurant	Garden	Steakhouse	Sake Bar	Dessert Shop	Park
4	Ikebukuro, Komagome, Senkawa, Sugamo	Ramen Restaurant	Sake Bar	Café	Chinese Restaurant	Coffee Shop	Japanese Restaurant	Yoshoku Restaurant	Steakhouse	Szechuan Restaurant	Hobby Shop

And visualizing the ranking to see the most popular venues in Tokyo.

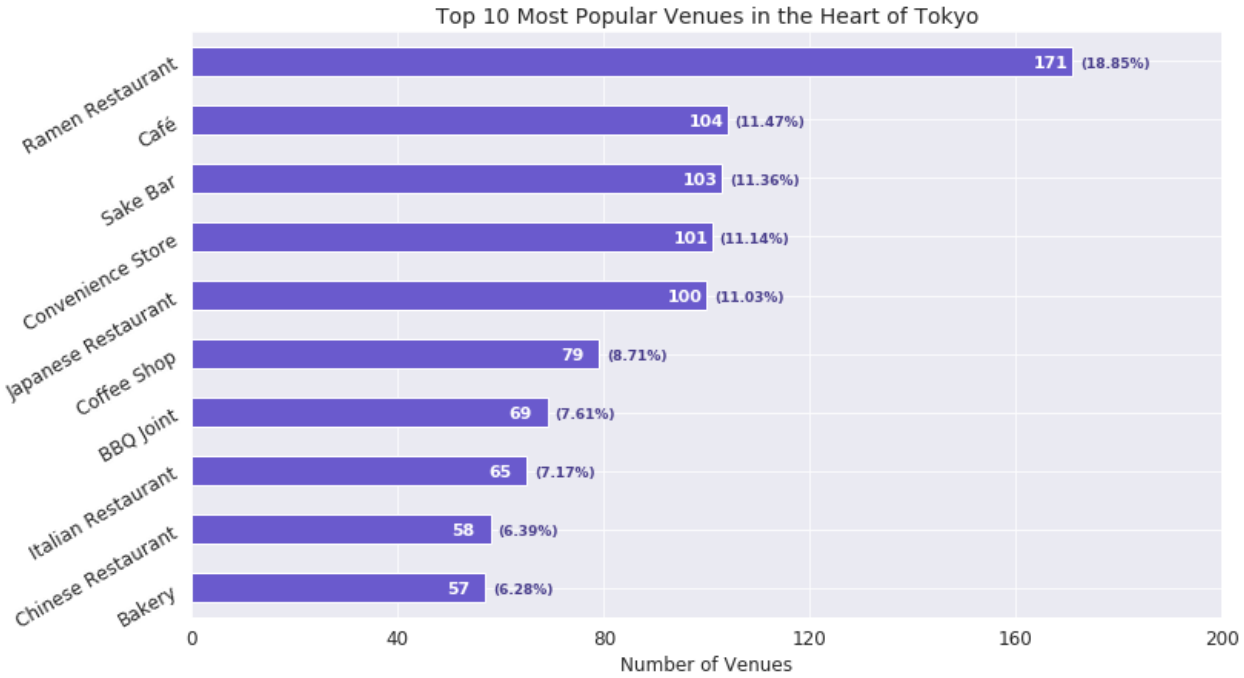

Fig.9 Top 10 popular venues in Tokyo

Living up to our expectations, Japanese ramen has won the championship in this competition in the end. In the result, we can also feel the 'power' of Japanese ramen. The number of ramen bars is around 170, which is nearly 1.5 times that of the second-ranked Cafe (104). From 2nd place to 5th place, the number of gastronomies is basically around 100, no significant difference.

It is worth noting that the fourth-ranked ‘convenience store’ is a kind of retail business model with very specific ‘East Asian’ characteristics. It can be regarded as a reduced-size supermarket that sells all necessities of life, but the range of products available for selection is limited. The existence of convenience stores brings great comfort to people's lives and saves a lot of time.

The winners of round 3 : Japanese ramen, Café, Sake bar

3.4 Clustering

Last but not least, we learned about the similarity of special wards by clustering them with K-means algorithm. The method for determining adaptive ‘k-cluster’ is called ‘the elbow method’.

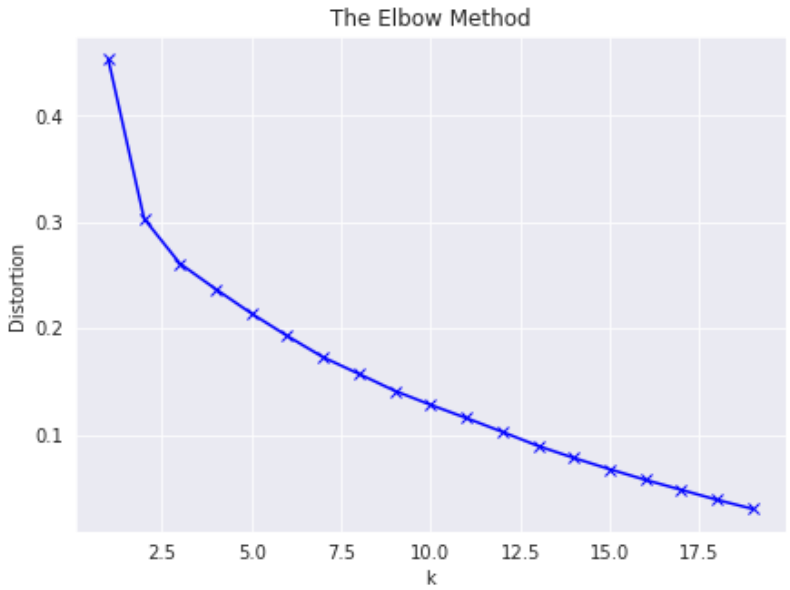

Fig.10 The elbow method

Although the suggestion given by computer is that the value of k = 3 or k = 4, in the subsequent experiments, we used 3 and 4 for the calculation, and found that the results were not very satisfactory. So we also have tried k = 5, and the result was slightly better.

Because of that we also created clustering map with k = 5.

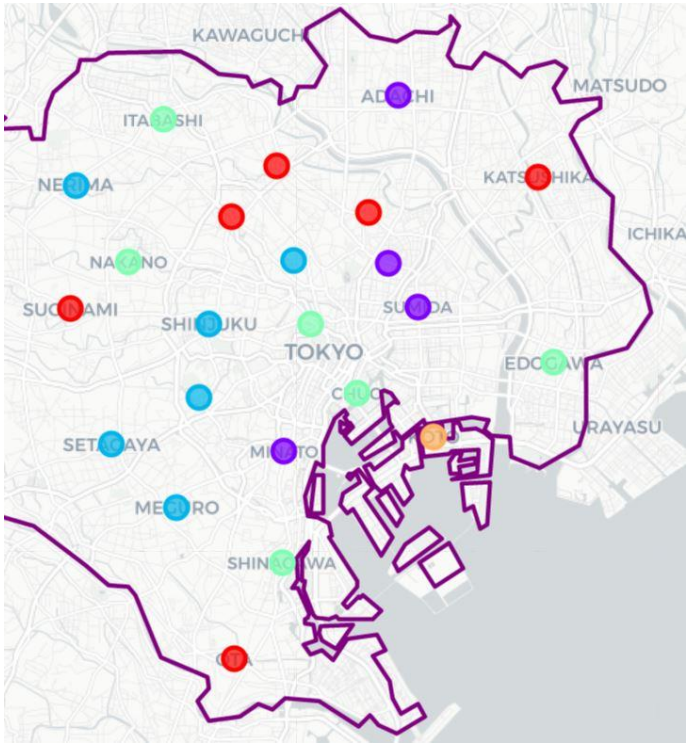

Fig.11 Clustering map

As still quite a few of samples seemed not to be correctly categorized, we need do some manual corrections as below:

Katsushika: Cluster 1 → Cluster 4

Suginami: Cluster 1 → Cluster 3

Sumida: Cluster 2 → Cluster 1

Adachi: Cluster 2 → Cluster 4

Shibuya: Cluster 3 → Cluster 2

Nerima: Cluster 3 → Cluster 1

Chiyoda: Cluster 4 → Cluster 3

Chuo: Cluster 4 → Cluster 2

Shinagawa: Cluster 4 → Cluster 1

Nakano: Cluster 4 → Cluster 1

Koto: Cluster 5 → Cluster 4

(as cluster 5 has only one element and is also very similar to cluster 4, so we merged them into one cluster)

Now we get updated clusters as below:

Cluster 1: “Japanese Ramen”

Ota, Toshima, Kita, Arakawa, Sumida, Nerima, Shinagawa, Nakano

Cluster 2: “Japanese Restaurant” (incl. Sushi)

Minato, Taito, Shibuya, Chuo

Cluster 3: “Café”, “Bakery”, “Bar”

Suginami, Shinjuku, Bunkyo, Meguro, Setagaya, Chiyoda

Cluster 4: “Convenience Store”

Katsushika, Adachi, Itabashi, Edogawa, Koto

IV. Result & Discussion

4.1 Summary of analysis results

Judging from the results of the two battles, Bunkyo has undoubtedly become the final champion, as it stood out in both battles. The environment in Bunkyo where gathered by universities is relatively friendly and suitable for carrying out various types of business. Other wards such as Chuo, Meguro and Arakawa are also very good choices.

Moreover, unsurprisingly, Japanese ramen, which represents and inherits Japanese culture, is indeed the most popular venue in Tokyo. Actually, Japanese ramen is not only a 'must' for foreign tourists who come to Japan, but also a beloved traditional specialty food from Japanese natives. In addition, Cafés and Sake bars are also very popular.

Finally, it is undeniable that due to the extreme high similarity of the Tokyo special wards, the clustering process using K-means worked not very successful. Nearly half of the wards are not correctly clustered. Of course, this has a lot to do with the design of the experiment. In future research, we will choose more suitable methods for further analysis.

4.2 Further discussion

If planning to start business in Tokyo, it is a good direction to avoid some of the most frequent types of venues that already exist. For example, in Bunkyo, Cafés and ramen shops are the two types with the largest number of venues, so avoiding these two kinds of businesses may have more scope for longtime development.

For the pricing problem of the future business, a good tip is referring to the average annual income level of each ward. If the product price is relatively high, or there might have certain requirements for the target customer groups, the better choices are the wards with higher or even top average annual income, so the income level should be the first factor of consideration. Meanwhile, if pricing is more affordable, then population density can be the main consideration.

V. Conclusion

Throughout the course of the entire project, we found that the most popular venues in each special ward are closely related to the economy, demography, social environment, and other factors of the ward itself. For instance, Chiyoda, Shinjuku, and Bunkyo, which belong to the same Cluster 3, have the most Western-style venues (Cafés, Bars, etc.) in these three wards, but the reasons behind them are sophisticated. Chiyoda, as the political center of Tokyo, has a large number of government institutions and embassies from various countries, thus the level of internationalization is the most influential factor. Shinjuku is a very famous tourist spot, and the number of foreigners in Shinjuku ranks in the first place of Tokyo. Bunkyo is equipped with many universities, so for a large number of students, the demand for Cafés should also satisfy the needs consequently. This shows that we need to carry out adequate and diversified investigations before really starting our business, also to understand the local customs and cultures from multiple perspectives. These will definitely benefit us with much more confidence, as well as handling a butcher's cleaver with ease in the future.

Last but not least, this project has only done a pretty rough analysis based on tiny database about the situation in the Tokyo special wards. In practical, specific problems should be analyzed together with sufficient amount of data for drawing more accurate and appropriate conclusions.