**Capstone Project - The Battle of Neighborhoods**

**1. Discussion and Background of the Business Problem:**

**Problem Statement:** Prospects of a Lunch Eating Joint close to College Areas in Tokyo, Japan.

Tokyo is one of the most populous metropolitan cities in the world. Currently ranked in the top 5 of global economic power index. So, this makes Tokyo as one of the best places to start with a new business opportunity.

During the daytime, especially in the morning and during the lunch hours, campus areas around colleges provide huge opportunities for Eating joints business. Reasonably priced shops are usually always full during the breakfast and lunch hours (11 am - 3 pm) and, given this scenario, we will go through the pros and cons of opening a breakfast cum lunch eating joint in highly dense college places. Usually the profit margin for a decent food joint lies within 15−20% range but, it can be higher, as we will be discussing here more in detail. The core of Tokyo is made of 23 wards (municipalities) but, will be later choosing the 6 busy business wards of Tokyo — 'Odaiba','Shinjuku','Kinshicho','Shinagawa','Shibuya','Ikebukuro', to target the college students.

We will discuss all the points one by one. First, I would like to outline the audience that would be interested in this case study, followed by the data which would be used to complete the analysis.

**Target Audience**

These days most of the information is available over the internet but those answers are mostly based on the personal preferences and experiences of the people. This piece of information is more data-driven based analysis and with better recommendations. So, Which kind of clients/people or a group of people would be interested in this project -

- Business personnel’s who want to invest in a business or open a restaurant. This analysis will be a comprehensive guide to start or expand restaurants targeting the large pool of students in Tokyo during lunch hours.

- Freelancers who love to open their own restaurant as the side business. This analysis will give them an idea, how good it is to open a restaurant along with the pros and cons of this restaurant business.

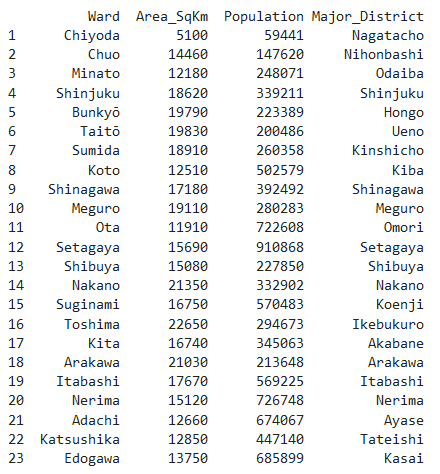
- New comers in the city/Students/Freshers, to find reasonable lunch/breakfast place closer to their colleges.

- Students/Professionals who are interested in Data Science and want to implement some of the most used Data Analysis methods to obtain necessary data, analyse it, and conclude it.

**2. Data Preparation:**

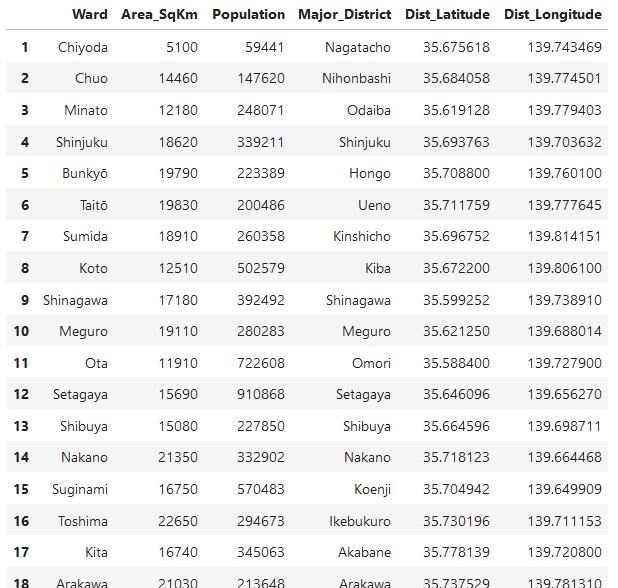
**2.1 Scrapping Tokyo Wards Table from Wikipedia**

Special Wards of Tokyo page from Wiki (<https://en.wikipedia.org/wiki/Special_wards_of_Tokyo>) is used to scrap the table to create a data-frame. Requests and Beautifulsoup4 libraries will be used to scrap and create a data-frame with Ward names in Tokyo, Area, Population and 1st Major District.



**2.2. Getting Coordinates of the 1st Major Districts: Geopy Client**

The next step is to capture the coordinates of these 1st major districts using geocoder class of Geopy client.



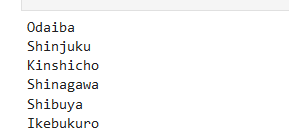
**2.3. Average Land Price in Major Wards of Tokyo: Web Scrapping**

Another web-page (<https://utinokati.com/en/details/land-market-value/area/Tokyo/>) would be used to scrap the table and capture the average the average land price of these Major districts in a dataframe. This factor will later help us in selecting the best districts to open a restaurant.



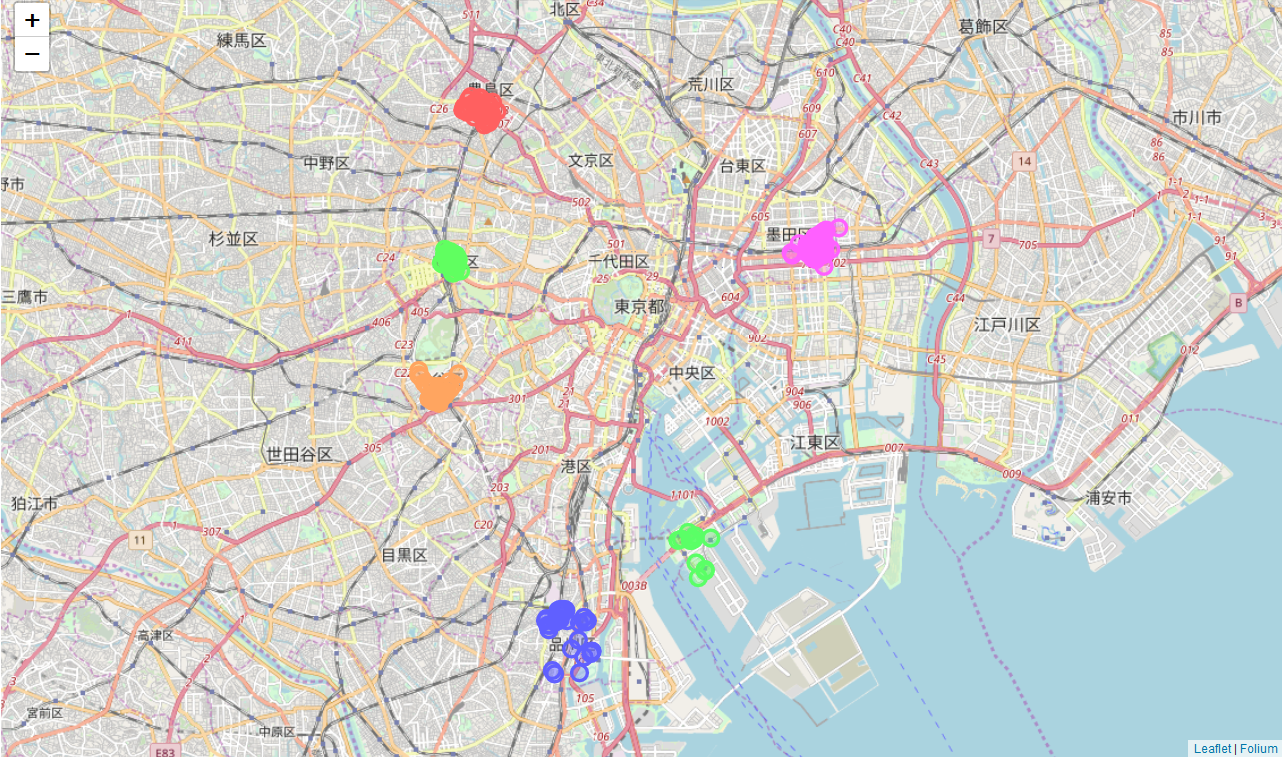
**2.4. Using Foursquare Location Data:**

Foursquare (<https://developer.foursquare.com/>) data is very comprehensive and it powers location data for Apple, Uber etc. For this business problem, Foursquare API is used to retrieve information about the popular spots around these 6 Major Districts of Tokyo. The popular spots returned depends on the highest foot traffic and thus it depends on the point of time when the call is made. So, we may get different popular venues depending upon different time of the day. The API call returns a JSON file and in-turn it is stored as dataframe.

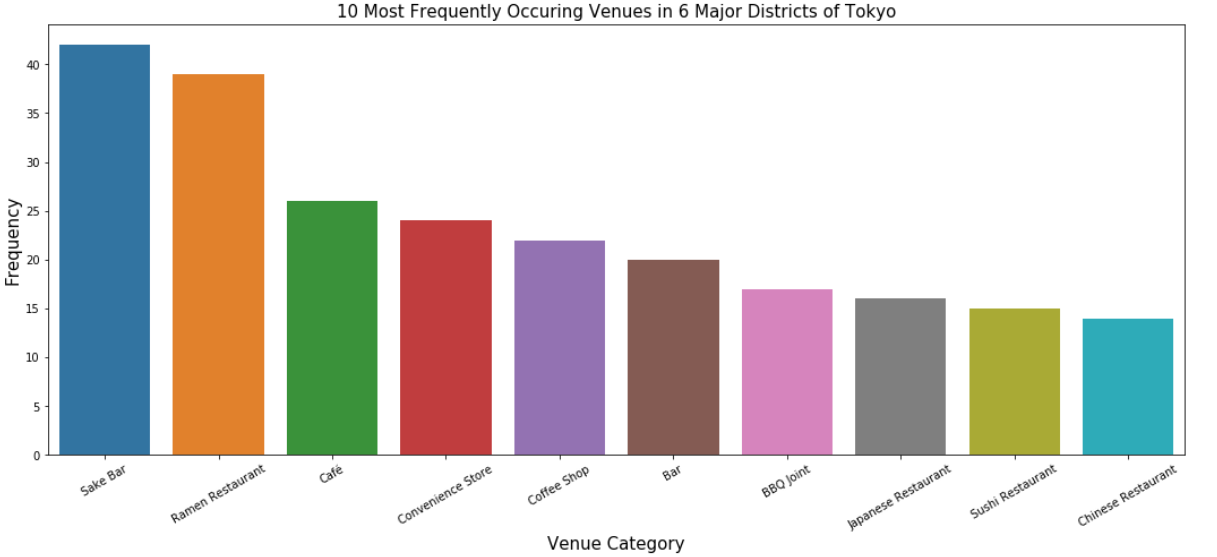


**3. Exploring through the Data :**

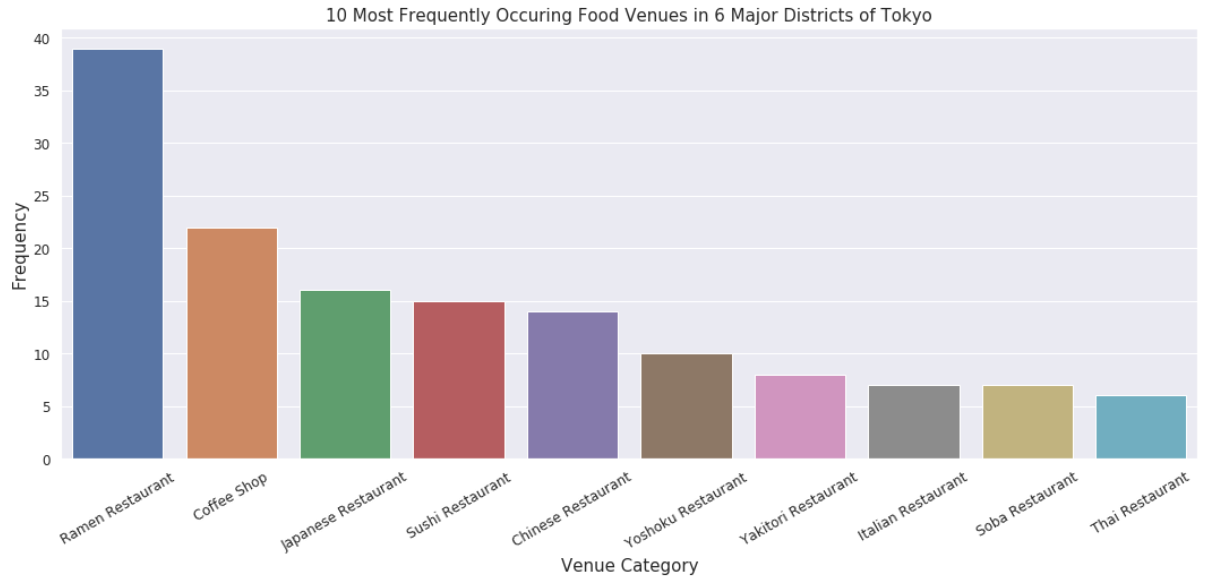
As we explore through the data using Foursquare API to explore the 6 major districts around the 6 wards chosen for detailed analysis , we understand that Foursqaure API talks about 592 top venues. Out of which, 212 venues are only related to Restaurants or Coffee Shops. The below leaflet map, created by using Folium library, which highlights the distribution of 212 food only joints across 6 key districts - Odaiba, Shinjuku , Kinshicho , Shinagawa , Shibuya , Ikebukuro



When we further drill down the Venue categories, there are total 160 unique categories out which the below Bar Graph talks about the top 10 most frequently occurring venues in above mentioned 6 major districts of Tokyo.



When talking about the only Food related venue categories, there are 39 unique categories out of which the below Bar Graph shoes the top 10 most frequently occurring food venues in the 6 major discussed districts of Tokyo.



As we can very well conclude that, Ramen food is the most favourite choice of people among the eating joints, followed by Coffee shops and then Japanese food.

The next categorization is based on the top 5 frequently visited venues, in each of the 6 districts and the results are as captured below :

%%%%%%%%%Ikebukuro%%%%%%%%

Venue Freq

0 Sake Bar 0.12

1 Café 0.08

2 Japanese Restaurant 0.06

3 Ramen Restaurant 0.05

4 Coffee Shop 0.05

%%%%%%%%%Kinshicho%%%%%%%%

Venue Freq

0 Ramen Restaurant 0.12

1 Sake Bar 0.11

2 Coffee Shop 0.05

3 Japanese Restaurant 0.03

4 Thai Restaurant 0.03

%%%%%%%%%Odaiba%%%%%%%%

Venue Freq

0 Convenience Store 0.09

1 Intersection 0.07

2 Plaza 0.07

3 Coffee Shop 0.05

4 Shopping Mall 0.03

%%%%%%%%%Shibuya%%%%%%%%

Venue Freq

0 Café 0.12

1 Record Shop 0.05

2 Sushi Restaurant 0.05

3 Coffee Shop 0.05

4 Ramen Restaurant 0.05

%%%%%%%%%Shinagawa%%%%%%%%

Venue Freq

0 Convenience Store 0.16

1 Ramen Restaurant 0.11

2 Sake Bar 0.07

3 BBQ Joint 0.05

4 Park 0.03

%%%%%%%%%Shinjuku%%%%%%%%

Venue Freq

0 Bar 0.10

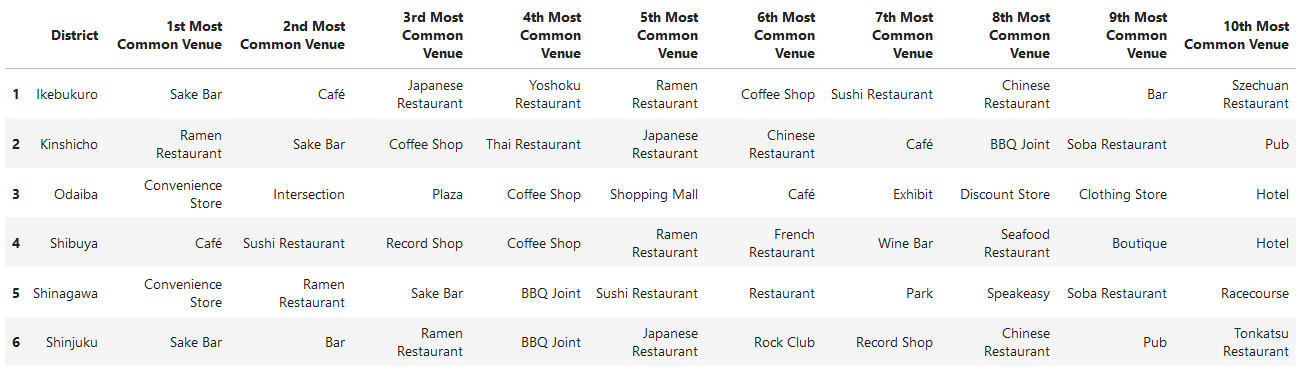
1 Sake Bar 0.10

2 Ramen Restaurant 0.05

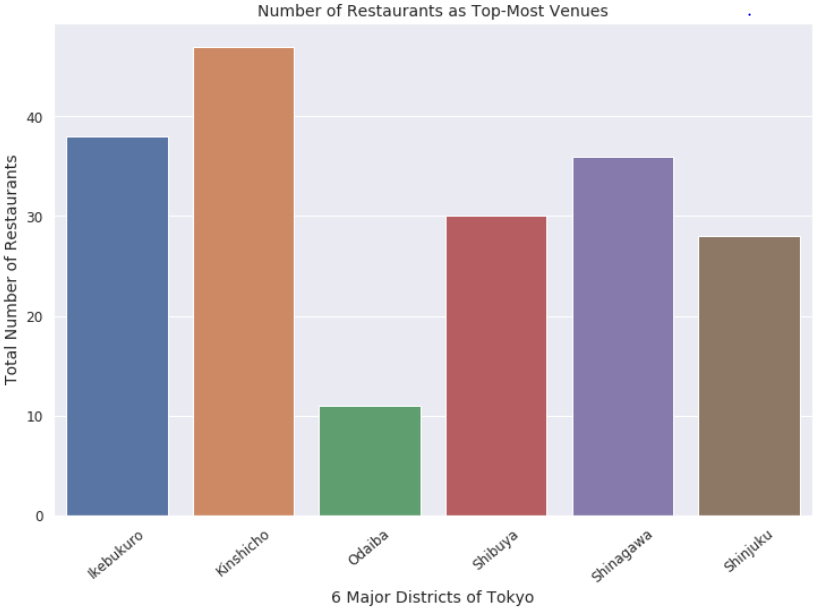
3 BBQ Joint 0.05

4 Japanese Restaurant 0.04

The above data is captured in the form of Dataframe as well, which gives more clear picture the 10 most commonly visited venues for all the 6 Districts among all the Venue categories:

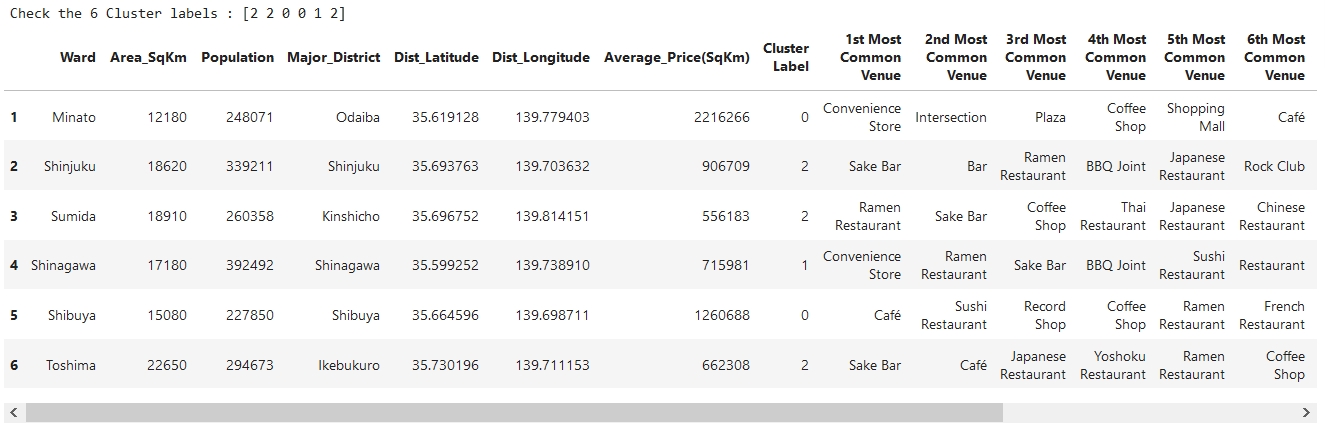


There are some more topics which are explored further and one of them is to find the Total number of restaurants for each district which are marked as most visited ones.

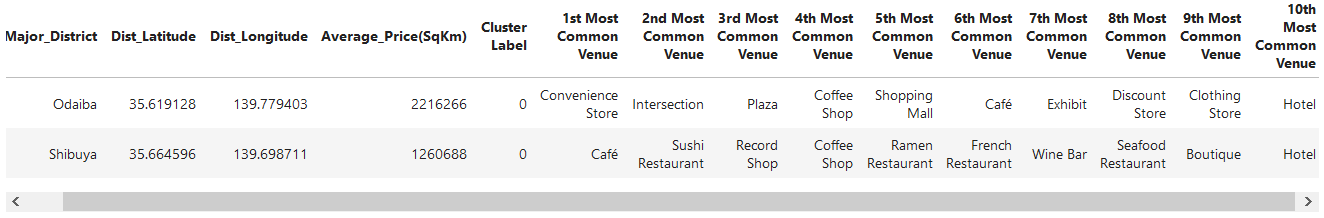


**4. Clustering**

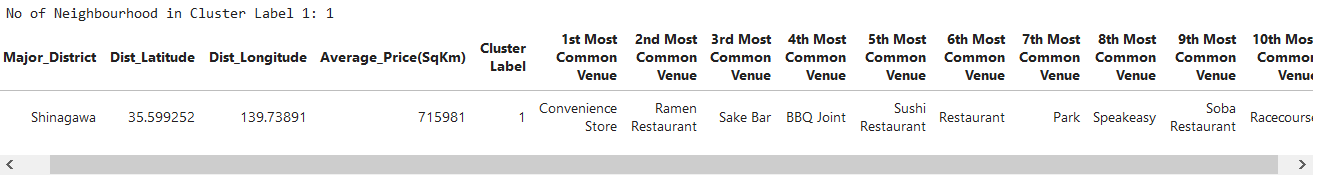
Finally, we will try to cluster these 6 districts based on the frequency of venue categories and will be using K-Means clustering algorithm. So, as per the expected output of this algorithm, these 6 districts would be clustered based on the similarities of venue categories. Using K-Means algorithm, 3 clusters are obtained as shown below -



**Cluster 0:**



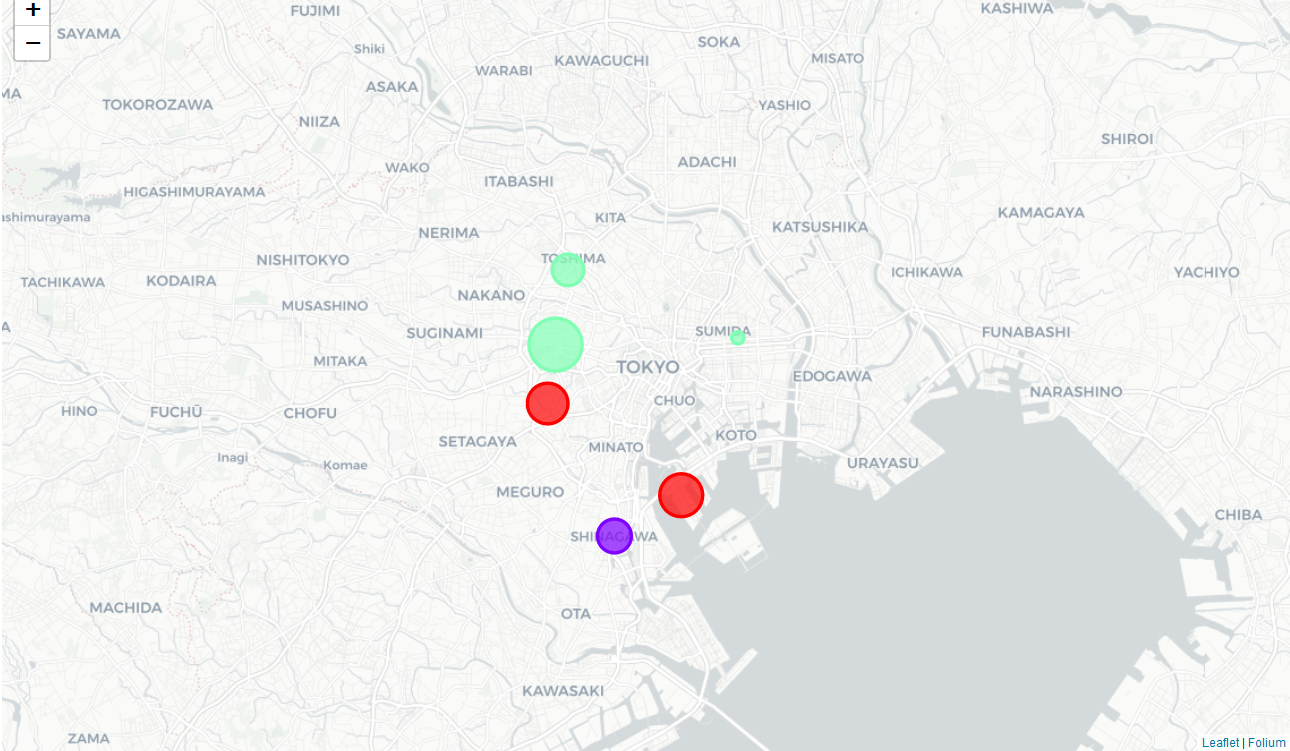
**Cluster 1:**



**Cluster 2:**



Here the radius of the circles represents the number of restaurants as most common venue for the corresponding district and, we have seen before that it is maximum for Kinshicho district and minimum for Odaiba.



From the most common venues this clustering makes a complete sense as Ikebukuro, Shinjuku are dominated by pubs, bars and cafe falls under the purple cluster, whereas Nagatacho, Nihombashi dominated by Cafes, Coffee shops and Shushi restaurants falls under red cluster and Shinagawa stands alone (green cluster) with frequently visited food points as Ramen restaurant, Bar and BBQ Joints.

**5. Results**

The results of The Battle of Neighbourhoods concluded with data-driven analysis and clustering, are summarized as below--

* Except Odaiba, Ramen restaurants is present in top 5 position of most common venues in remaining of the 5 districts.
* Coffee Shops in Odaiba, tops the most frequently visited in the food joints category with limited visits to the restaurants.
* Ikebukuro district in Toshima ward is dominated by Japanese, Yoshoku & Ramen restaurants as the most common venues and Kinshicho district in Sumida ward is dominated by Ramen, Thai & Japanese restaurants
* Shibuya area is dominated by bars, pubs, clubs and cafe as most common venues.
* Kinshicho has maximum number of restaurants as the most common venue whereas has Odaiba area has the least. But, Coffee Shops and Cafe are found to be among the most visited destinations in this area.
* Since the clustering was based only on the most common venues o each district, Odaiba, Shibuya fall under the same cluster and, Shinjuku, Kinshicho and Ikebukuro fall under another cluster. Shinagawa is separated from both clusters as; convenient stores stand out as the most common venue (with a very high frequency).

**6. Discussion**

According to this analysis, **Odaiba followed by Shinagawa Districts will provide least competition for an upcoming lunch restaurant** as the Convenience Store is the most common venue in these areas and the frequency of restaurants as common venue are very low compared to the remaining districts.

Also, as it can be seen from the web-scrapped data, **the average land price in and around Shinagawa is much cheaper compared to Odaiba and other Districts close to central Tokyo.** *So, Shinagawa region could potentially be a most favourable deal to start with the restaurant business.*

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, the conclusion of this analysis could vary depending on other factors as well. However, it gives us some kick-start and important piece of information on the potentials of doing restaurant business in and around the major districts of Tokyo.

**7. Conclusion**

As a conclusion to this project, it gives us an over-all idea about the real-life data-science projects. Many frequently used python libraries have been used in this analysis like BeautifulSoup to scrap web-data into data-frames, Foursquare API to explore the major districts of Tokyo and saw the results of segmentation of districts using Folium leaflet map and other results using Bar graph. Analysis of this kind of real-life business problem is of great potential and brings a lot of scope for improvement in business related. Also, some of the drawbacks and chance for improvements to represent even more realistic pictures are mentioned. Finally, this analysis is mostly concentrating on the possibilities of opening a restaurant targeting the huge pool of students, freshers and new comers.