Minato City Restaurant Clustering

Determining prime locations for new restaurants.

Abstract

We explore Minato City areas based on the top venues and cluster them to the determine what kinds of restaurants to start and in which areas to start them.

Dexter Valley

Introduction

A. Background:

Minato City is located in the centre of Tokyo. It has an active economy, as numerous businesses locate their offices in the city, many of which serve as headquarter offices. Minato City also has a large number of cultural and art facilities. It plays the role of the centre of politics, economics, culture and art.

In Minato City, you can find popular places like Daiba, Roppongi, Shiodome, Azabujuban, and distinctive streetscapes in areas like Aoyama, Azabu, Akasaka, Takanawa and Shibaura. Modern and state-of-the-art buildings coexist with historical and traditional cultural properties from the Edo Period and the Meiji Era. Minato City is an attractive city where one encounters harmony between tradition and modernity.

Minato City has easy access to and from both Haneda Airport and Narita International Airport. Also, it is connected with regional Japanese cities through Shinagawa Station's Shinkansen and the Port of Tokyo. Minato City is the centre of the land, sea and air transportation network. In Tokyo, Minato City has the largest number of guest rooms in hotels and inns that accommodate people visiting for sightseeing and business.

We will be exploring the similarities population share within their respective area and the overall population, and how they differ between themselves and the overall population.The using this information to identify key markets for restaurant growth. As well as identify which restaurant would be most successful in those markets.

B. Business Problem

The following exploratory data analysis could be helpful in business. For example, say a client would like to open a new restaurant, however they do not know what kind of restaurant they would like to open, let alone where to open it. In this article, we will determine the optimal area in Minato to open a restaurant, as well as the perfect business for that area.

C. Data

In order to answer the business question, data will need to be gathered data from a couple sources. The first sources will be "**address-zipcode.tokyo**". From "**address-zipcode.tokyo**", we will identify the area of Minato. Next, collect the approximate coordinates of each area. With the coordinates and the area names collected, we will next be using the Foursquare API to collect venues information for each area, within a designated radius. The venue data collected from Foursquare will then be used to determine the density of food restaurant in each area and what type of food is popular in each area.

Once the business information is gathered, the next step will be to cluster areas in Minato based on Venues Categories. This information will allow us to cluster the popularity of venues in areas into discrete groups which will help us understand sentiment. Then moving on to identifying area within the clusters which are prime candidates for a new restaurant, as well as identifying which specific business would be the most lucrative.

Methodology

D. Data Collection

- **Web Scraping**

  The information needed to drive the entire project begins with collecting neighborhood names and postal codes. "Address-zipcode.tokyo" website is a good sources becasue they have a table dedicated to the postal code and name of each neighborhood. Some cleaning needs to be done to properly populate our data frame with the correct information. (see Figure 1)

[6]:

Postal Code	Minato Neighborhood
1070052	Akasaka(Tsuginobiruonozoku)
1060045	Azabujuban
1060041	Azabudai
1060043	Azabunagasaka-Cho
1060042	Azabumamiana-Cho
1050002	Atago
1050022	Kaigan(1-2-Chome)
1080022	Kaigan(3-Chome)
1070061	Kitaaoyama
1080075	Konan(Tsuginobiruonozoku)
1050014	Shiba(1-3-Chome)
1080014	Shiba(4-5-Chome)
1050023	Shibaura(1-Chome)
1080023	Shibaura(2-4-Chome)
1050011	Shibakoen
1050012	Shibadaimon
1080072	Shirogane
1080071	Shirokanedai
1050004	Shimbashi
1350091	Daiba
1080074	Takanawa
1050001	Toranomon
1060031	Nishiazabu
1050003	Nishishimbashi
1050013	Hamamatsu-Cho
1060044	Higashiazabu
1050021	Higashishimbashi
1080073	Mita
1070062	Minamiaoyama
1060047	Minamiazabu
1070051	Motoakasaka
1060046	Motoazabu
1060032	Roppongi

(Figure 1)

- **Gathering Coordinates**

With the neighborhood identified and loaded into our data frame, the next objective is to collect the coordinates for each area. For this, we will be utilizing Google search engine by

searching all 33 neighborhoods in Minato. For example, a search for Akasaka coordinates into Google search engine manually. After collecting the correct coordinates of each neighborhood I manually preset the coordinates of each neighborhood in their respected order into a dataframe (Figure 2a). After collecting all the correct data, I configure the two dataframes into a CSV.file to combine both dataframe and to implement appropriate column names (Figure 2b) and then converted back into a Panda dataframe so data can be applied into the Foursquare API to find the Venues in each Neighborhood (Figure 2c).

[5]:

Latitude	Longitude
35.6710	139.7345
35.6547	139.7352
35.6607	139.7416
35.6581	139.7378
35.6591	139.7395
35.6648	139.7488
35.6747	139.7317
35.6400	139.7601
35.6734	139.7205
35.6313	139.7499
35.6516	139.7505
35.6480	139.7486
35.6486	139.7543
35.6572	139.7486
35.6571	139.7534
35.6444	139.7282
35.6388	139.7256
35.6648	139.7563
35.6323	139.7754
35.6343	139.7345
35.6653	139.7460
35.6584	139.7231
35.6651	139.7512
35.6571	139.7556
35.6561	139.7416
35.6646	139.7614
35.6472	139.7409
35.6655	139.7179
35.6500	139.7307
35.6765	139.7282
35.6546	139.7307
35.6619	139.7345

(Figure 2a)

```
[6]: import pandas
     df = pandas.read_csv('Minato_Coordinates.csv')
     print(df)
```

	Postal Code	Area	Latitude	Longitude
0	1070052	Akasaka	35.6710	139.7345
1	1060045	Azabujuban	35.6547	139.7352
2	1060041	Azabudai	35.6607	139.7416
3	1060043	Azabunagasaka-Cho	35.6581	139.7378
4	1060042	Azabumamiana-Cho	35.6591	139.7395
5	1050002	Atago	35.6648	139.7488
6	1050022	Kaigan(1-2-Chome)	35.6747	139.7317
7	1080022	Kaigan(3-Chome)	35.6400	139.7601
8	1070061	Kitaaoyama	35.6734	139.7205
9	1080075	Konan(Tsuginobiruonozoku)	35.6313	139.7499
10	1050014	Shiba(1-3-Chome)	35.6516	139.7505
11	1080014	Shiba(4-5-Chome)	35.6480	139.7486
12	1050023	Shibaura(1-Chome)	35.6486	139.7543
13	1080023	Shibaura(2-4-Chome)	35.6572	139.7486
14	1050011	Shibakoen	35.6571	139.7534
15	1050012	Shibadaimon	35.6444	139.7282
16	1080072	Shirogane	35.6388	139.7256
17	1080071	Shirokanedai	35.6648	139.7563
18	1050004	Shimbashi	35.6323	139.7754
19	1350091	Daiba	35.6343	139.7345
20	1080074	Takanawa	35.6653	139.7460
21	1050001	Toranomon	35.6584	139.7231
22	1060031	Nishiazabu	35.6651	139.7512
23	1050003	Nishishimbashi	35.6571	139.7556
24	1050013	Hamamatsu-Cho	35.6561	139.7416
25	1060044	Higashiazabu	35.6646	139.7614
26	1080073	Mita	35.6472	139.7409
27	1070062	Minamiaoyama	35.6655	139.7179
28	1060047	Minamiazabu	35.6500	139.7307
29	1070051	Motoakasaka	35.6765	139.7282
30	1060046	Motoazabu	35.6546	139.7307
31	1060032	Roppongi	35.6619	139.7345

(Figure 2b)

[7]

	Postal Code	Area	Latitude	Longitude
0	1070052	Akasaka	35.6710	139.7345
1	1060045	Azabujuban	35.6547	139.7352
2	1060041	Azabudai	35.6607	139.7416
3	1060043	Azabunagasaka-Cho	35.6581	139.7378
4	1060042	Azabumamiana-Cho	35.6591	139.7395
5	1050002	Atago	35.6648	139.7488
6	1050022	Kaigan(1-2-Chome)	35.6747	139.7317
7	1080022	Kaigan(3-Chome)	35.6400	139.7601
8	1070061	Kitaaoyama	35.6734	139.7205
9	1080075	Konan(Tsuginobiruonozoku)	35.6313	139.7499
10	1050014	Shiba(1-3-Chome)	35.6516	139.7505
11	1080014	Shiba(4-5-Chome)	35.6480	139.7486
12	1050023	Shibaura(1-Chome)	35.6486	139.7543
13	1080023	Shibaura(2-4-Chome)	35.6572	139.7486
14	1050011	Shibakoen	35.6571	139.7534
15	1050012	Shibadaimon	35.6444	139.7282
16	1080072	Shirogane	35.6388	139.7256
17	1080071	Shirokanedai	35.6648	139.7563
18	1050004	Shimbashi	35.6323	139.7754
19	1350091	Daiba	35.6343	139.7345
20	1080074	Takanawa	35.6653	139.7460
21	1050001	Toranomon	35.6584	139.7231
22	1060031	Nishiazabu	35.6651	139.7512
23	1050003	Nishishimbashi	35.6571	139.7556
24	1050013	Hamamatsu-Cho	35.6561	139.7416
25	1060044	Higashiazabu	35.6646	139.7614
26	1080073	Mita	35.6472	139.7409
27	1070062	Minamiaoyama	35.6655	139.7179
28	1060047	Minamiazabu	35.6500	139.7307
29	1070051	Motoakasaka	35.6765	139.7282
30	1060046	Motoazabu	35.6546	139.7307
31	1060032	Roppongi	35.6619	139.7345

(Figure 2c)

- **Identifying Business in Each Neighborhood**

Moving on, the next crucial step is to collect information about the available businesses in each area. To do this, we will utilize the Foursquare API. Using this crowdsources data, we will populate our data frame with all businesses with a specified radius. For this study, a 800-meter radius was used to return enough results to perform our

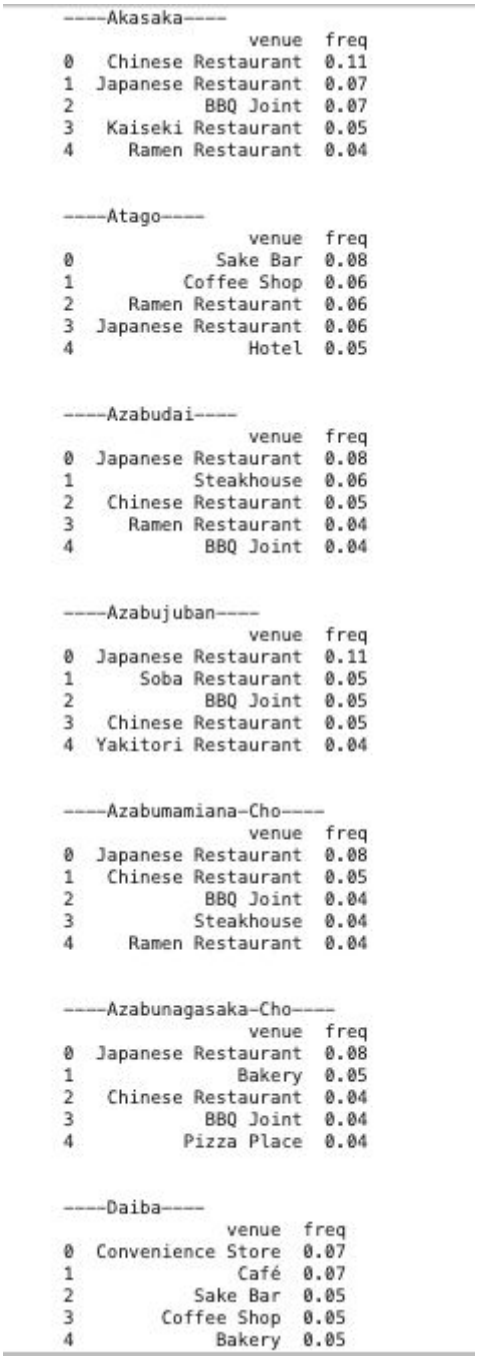
data analysis and clustering. Once the data is gathered, we populate the data frame (Figure 3a). Then configure a one hot encoding to find the percentage density of the top 5 venues in each Neighborhood (Figure 3b - Figure 3g).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Akasaka	35.671	139.7345	Kaisu	35.669896	139.734565	Café
1	Akasaka	35.671	139.7345	Kaisuian (會水庵)	35.670152	139.734413	Kaiseki Restaurant
2	Akasaka	35.671	139.7345	Akasaka Kikunoi (赤坂 菊乃井)	35.669690	139.734194	Kaiseki Restaurant
3	Akasaka	35.671	139.7345	Muromachi Sunaba (室町砂場)	35.671166	139.736184	Soba Restaurant
4	Akasaka	35.671	139.7345	トラットリア ピッツェリア エッセ ドゥエ	35.669884	139.734827	Italian Restaurant

```
# The Number of Venue Categories
print('There are {} uniques categories.'.format(len(Minato_venues_15['Venue Category'].unique())))
```

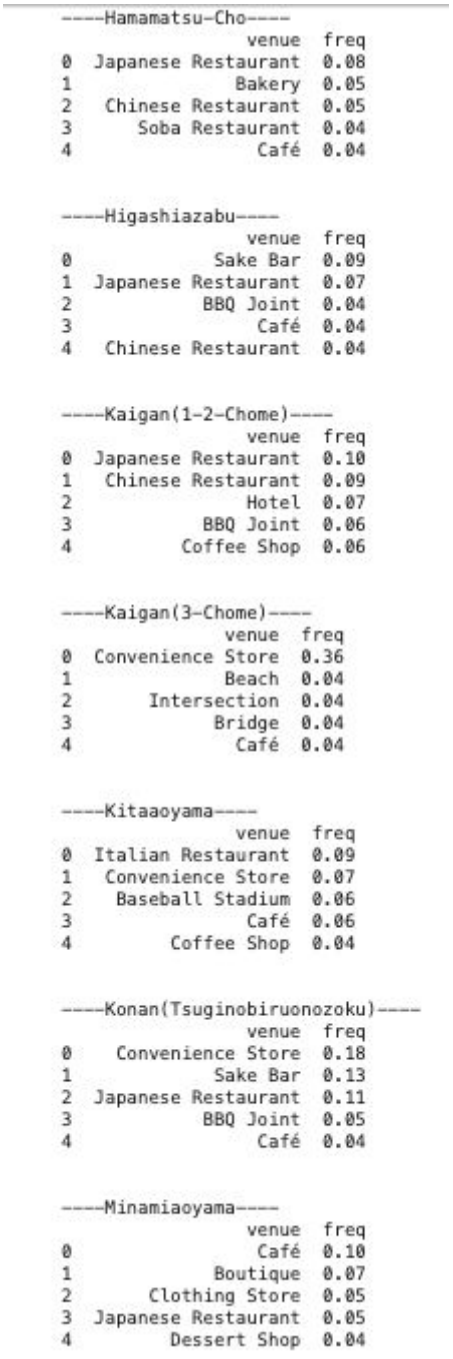
```
There are 205 uniques categories.
```

(Figure 3a)



```
----Akasaka----
                venue  freq
0   Chinese Restaurant  0.11
1  Japanese Restaurant  0.07
2            BBQ Joint  0.07
3   Kaiseki Restaurant  0.05
4     Ramen Restaurant  0.04


----Atago----
                venue  freq
0           Sake Bar  0.08
1         Coffee Shop  0.06
2     Ramen Restaurant  0.06
3  Japanese Restaurant  0.06
4               Hotel  0.05


----Azabudai----
                venue  freq
0  Japanese Restaurant  0.08
1           Steakhouse  0.06
2   Chinese Restaurant  0.05
3     Ramen Restaurant  0.04
4            BBQ Joint  0.04


----Azabujuban----
                venue  freq
0  Japanese Restaurant  0.11
1      Soba Restaurant  0.05
2            BBQ Joint  0.05
3   Chinese Restaurant  0.05
4  Yakitori Restaurant  0.04


----Azabumamiana-Cho----
                venue  freq
0  Japanese Restaurant  0.08
1   Chinese Restaurant  0.05
2            BBQ Joint  0.04
3           Steakhouse  0.04
4     Ramen Restaurant  0.04


----Azabunagasaka-Cho----
                venue  freq
0  Japanese Restaurant  0.08
1               Bakery  0.05
2   Chinese Restaurant  0.04
3            BBQ Joint  0.04
4          Pizza Place  0.04


----Daiba----
              venue  freq
0  Convenience Store  0.07
1              Café  0.07
2          Sake Bar  0.05
3       Coffee Shop  0.05
4            Bakery  0.05
```

(Figure 3b)



```
----Hamamatsu-Cho----
                venue  freq
0  Japanese Restaurant  0.08
1              Bakery  0.05
2   Chinese Restaurant  0.05
3      Soba Restaurant  0.04
4                Café  0.04


----Higashiazabu----
                venue  freq
0           Sake Bar  0.09
1  Japanese Restaurant  0.07
2           BBQ Joint  0.04
3                Café  0.04
4   Chinese Restaurant  0.04


----Kaigan(1-2-Chome)----
                venue  freq
0  Japanese Restaurant  0.10
1   Chinese Restaurant  0.09
2                Hotel  0.07
3           BBQ Joint  0.06
4         Coffee Shop  0.06


----Kaigan(3-Chome)----
             venue  freq
0  Convenience Store  0.36
1              Beach  0.04
2       Intersection  0.04
3             Bridge  0.04
4               Café  0.04


----Kitaaoyama----
                venue  freq
0  Italian Restaurant  0.09
1   Convenience Store  0.07
2    Baseball Stadium  0.06
3                Café  0.06
4         Coffee Shop  0.04


----Konan(Tsuginobiruonozoku)----
                venue  freq
0    Convenience Store  0.18
1           Sake Bar  0.13
2  Japanese Restaurant  0.11
3           BBQ Joint  0.05
4                Café  0.04


----Minamiaoyama----
                venue  freq
0                Café  0.10
1            Boutique  0.07
2      Clothing Store  0.05
3  Japanese Restaurant  0.05
4        Dessert Shop  0.04
```

(Figure 3c)

```
----Minamiazabu----
                venue  freq
0  Japanese Restaurant  0.08
1            BBQ Joint  0.08
2      Soba Restaurant  0.07
3              Bakery  0.06
4                Café  0.04


----Mita----
                venue  freq
0    Convenience Store  0.10
1  Japanese Restaurant  0.10
2          Coffee Shop  0.07
3      Soba Restaurant  0.07
4   Chinese Restaurant  0.07


----Motoakasaka----
                venue  freq
0               Hotel  0.08
1  Japanese Restaurant  0.07
2   Chinese Restaurant  0.07
3     Ramen Restaurant  0.05
4                Café  0.05


----Motoazabu----
                venue  freq
0  Japanese Restaurant  0.08
1      Soba Restaurant  0.05
2              Bakery  0.04
3   Italian Restaurant  0.04
4          Coffee Shop  0.04


----Nishiazabu----
                venue  freq
0           Sake Bar  0.11
1  Japanese Restaurant  0.07
2     Ramen Restaurant  0.05
3          Coffee Shop  0.05
4            BBQ Joint  0.05


----Nishishimbashi----
                venue  freq
0  Japanese Restaurant  0.10
1     Ramen Restaurant  0.08
2            BBQ Joint  0.06
3               Hotel  0.05
4  Tonkatsu Restaurant  0.05


----Roppongi----
                venue  freq
0  Japanese Restaurant  0.09
1           Steakhouse  0.06
2               Bakery  0.05
3     Ramen Restaurant  0.04
4            BBQ Joint  0.04
```

(Figure 3d)

```
----Shiba(1-3-Chome)----
                venue  freq
0  Japanese Restaurant  0.12
1      Soba Restaurant  0.07
2   Chinese Restaurant  0.06
3            BBQ Joint  0.06
4             Sake Bar  0.05


----Shiba(4-5-Chome)----
                venue  freq
0    Convenience Store  0.10
1   Chinese Restaurant  0.10
2  Japanese Restaurant  0.09
3             Sake Bar  0.06
4     Ramen Restaurant  0.06


----Shibadaimon----
                venue  freq
0    Convenience Store  0.11
1            BBQ Joint  0.07
2  Japanese Restaurant  0.06
3                Café  0.06
4   Italian Restaurant  0.05


----Shibakoen----
                venue  freq
0  Japanese Restaurant  0.09
1     Ramen Restaurant  0.08
2            BBQ Joint  0.06
3   Chinese Restaurant  0.05
4  Tonkatsu Restaurant  0.05


----Shibaura(1-Chome)----
                venue  freq
0   Convenience Store  0.14
1  Chinese Restaurant  0.06
2     Soba Restaurant  0.05
3           BBQ Joint  0.05
4            Sake Bar  0.05


----Shibaura(2-4-Chome)----
                venue  freq
0  Japanese Restaurant  0.07
1     Ramen Restaurant  0.07
2            BBQ Joint  0.06
3   Chinese Restaurant  0.05
4                Hotel  0.05


----Shimbashi----
              venue  freq
0     Clothing Store  0.05
1        Coffee Shop  0.05
2               Café  0.04
3      Shopping Mall  0.04
4  Convenience Store  0.03
```

(Figure 3e)

```
----Shiba(1-3-Chome)----
                venue  freq
0  Japanese Restaurant  0.12
1      Soba Restaurant  0.07
2   Chinese Restaurant  0.06
3            BBQ Joint  0.06
4             Sake Bar  0.05


----Shiba(4-5-Chome)----
                venue  freq
0    Convenience Store  0.10
1   Chinese Restaurant  0.10
2  Japanese Restaurant  0.09
3             Sake Bar  0.06
4     Ramen Restaurant  0.06


----Shibadaimon----
                venue  freq
0    Convenience Store  0.11
1            BBQ Joint  0.07
2  Japanese Restaurant  0.06
3                Café  0.06
4   Italian Restaurant  0.05


----Shibakoen----
                venue  freq
0  Japanese Restaurant  0.09
1     Ramen Restaurant  0.08
2            BBQ Joint  0.06
3   Chinese Restaurant  0.05
4  Tonkatsu Restaurant  0.05


----Shibaura(1-Chome)----
                venue  freq
0   Convenience Store  0.14
1  Chinese Restaurant  0.06
2     Soba Restaurant  0.05
3           BBQ Joint  0.05
4            Sake Bar  0.05


----Shibaura(2-4-Chome)----
                venue  freq
0  Japanese Restaurant  0.07
1     Ramen Restaurant  0.07
2            BBQ Joint  0.06
3   Chinese Restaurant  0.05
4                Hotel  0.05


----Shimbashi----
              venue  freq
0     Clothing Store  0.05
1        Coffee Shop  0.05
2               Café  0.04
3      Shopping Mall  0.04
4  Convenience Store  0.03
```

(Figure 3f)

	Neighborhood	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
0	Akasaka	Chinese Restaurant	Japanese Restaurant	BBQ Joint	Kaiseki Restaurant	Hotel	French Restaurant	Ramen Restaurant	Steakhouse	Soba Restaurant	Coffee Shop
1	Atago	Sake Bar	Japanese Restaurant	Ramen Restaurant	Coffee Shop	Chinese Restaurant	Hotel	Soba Restaurant	Bistro	Yoshoku Restaurant	Seafood Restaurant
2	Azabudai	Japanese Restaurant	Chinese Restaurant	Steakhouse	Ramen Restaurant	BBQ Joint	Nightclub	Hotel	Bakery	Yakitori Restaurant	Coffee Shop
3	Azabujuban	Japanese Restaurant	Chinese Restaurant	Soba Restaurant	BBQ Joint	Yakitori Restaurant	Bakery	Korean Restaurant	Sake Bar	Ramen Restaurant	Café
4	Azabumamiana-Cho	Japanese Restaurant	Chinese Restaurant	Steakhouse	BBQ Joint	Pizza Place	Bakery	Udon Restaurant	Ramen Restaurant	Ice Cream Shop	Korean Restaurant
5	Azabunagasaka-Cho	Japanese Restaurant	Bakery	Chinese Restaurant	Pizza Place	BBQ Joint	Korean Restaurant	Steakhouse	Yakitori Restaurant	Udon Restaurant	Soba Restaurant
6	Daiba	Convenience Store	Café	Sake Bar	Hotel	Bakery	Coffee Shop	Steakhouse	Japanese Restaurant	Italian Restaurant	Bar
7	Hamamatsu-Cho	Japanese Restaurant	Bakery	Chinese Restaurant	Soba Restaurant	Café	Sake Bar	Park	Pizza Place	Korean Restaurant	Italian Restaurant
8	Higashiazabu	Japanese Restaurant	Sake Bar	BBQ Joint	Seafood Restaurant	Hotel	Chinese Restaurant	Ramen Restaurant	Café	Art Gallery	Tonkatsu Restaurant
9	Kaigan(1-2-Chome)	Japanese Restaurant	Hotel	Chinese Restaurant	Coffee Shop	BBQ Joint	French Restaurant	Café	Szechuan Restaurant	Ramen Restaurant	Soba Restaurant
10	Kaigan(3-Chome)	Convenience Store	Restaurant	Trail	Bridge	Fried Chicken Joint	Rest Area	Music Venue	Café	Canal	Grocery Store
11	Kitaaoyama	Italian Restaurant	Convenience Store	Café	Baseball Stadium	Japanese Restaurant	Coffee Shop	Hotel	Chinese Restaurant	French Restaurant	Rock Club

(Figure 3g)

E. Exploratory Data Analysis

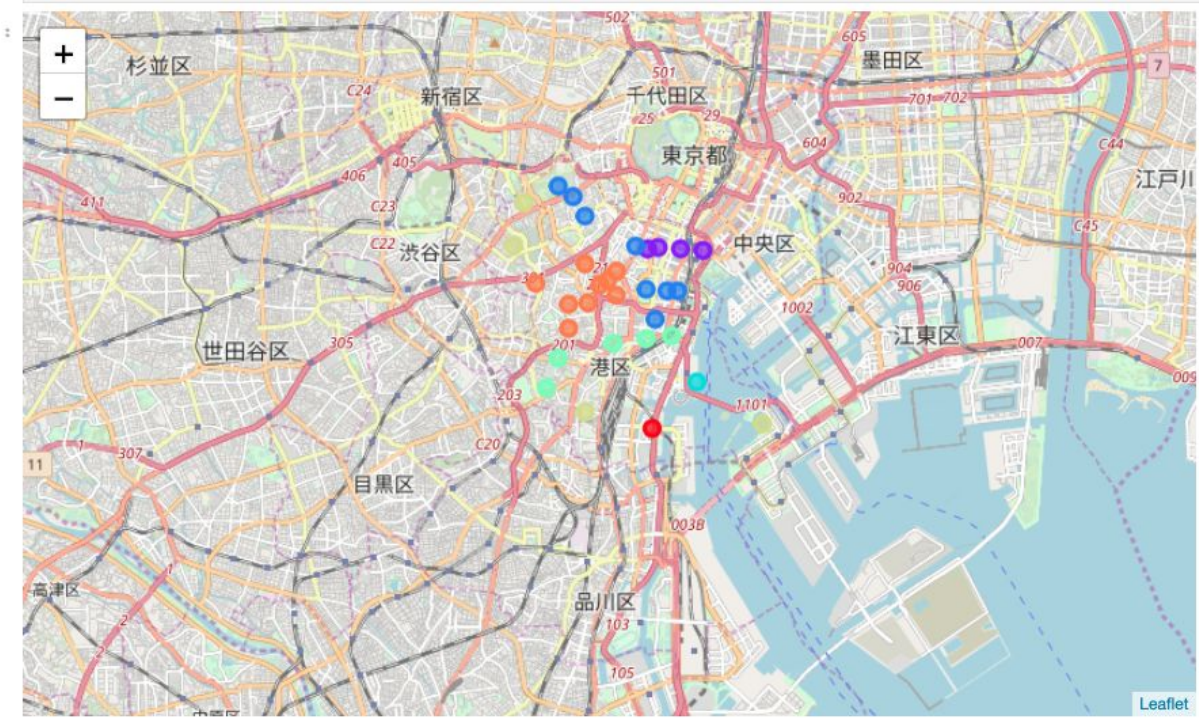
- **Clustering Areas**

With the data now gathered, we can begin clustering the data. To do this we will employ k-means Clustering. K-means Clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation into k cluster in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. (k-means clustering, n.d.) However, it should be noted that there exist drawbacks to using this approach for clustering. A key limitation of k-means is its cluster model.

The concept is based on spherical clusters that are separable so that the mean converges towards the clusters center. The clusters are expected to be of similar size, so that the assignment to the nearest cluster center is the correct assignment. Since data is split halfway between cluster means, this can lead to suboptimal splits. Therefores, to lower the risk of poor k-Means clustering, we will select the best k to fit our model.By looping through

various k’s,we find that the best k is k = 7.From here we set the k of our model to 7 and proceed to fit the model.

We assign the labels to each area and add the column labels to our data frame. To visualize the clustered areas, we generate a map of Minato with areas color coded for each cluster(see figure 4). Using the labels, we now separate our data by cluster and determine the top three most popular businesses by cluster. This information is then used to explore intra-cluster differences.



(Figure 4)

- **Selecting Areas**

To do this we find the intra cluster distance for each area and osrt the areas in each cluster based on distance to the centroid. The areas with the highest distance are selected. These are the areas we explore for starting a business. The reason for this is that these area have a high inter cluster distance, meaning that they are very different from the areas within other clusters. And high intra-cluster distance, meaning they are more like the areas within their own cluster. However, when compared to the intra cluster areas, theses areas are the weakest (furthest apart) So, we choose them to bring these sample closer to the centroid.

By comparing the restaurants available in each of the selected areas with the top three restaurants across the cluster we can determine what restaurants to open and which areas to open them.

Results

F. Recommendations

Cluster 2 and Cluster 6 is the most popular venue among the clusters. These clusters are consistently known for their Restaurants. If you are considering developing restaurant with the least amount of competition. I recommend Cluster 0, Kaigan(3-Chome) and Cluster 3, Kita Aoyama,Shibadaimon,Shirogane, and Daiba. These are prime candidates to develop a restaurant as these clusters are not known for their food as Cluster 2 and 6. They have limited food options and they are not known for their food.

If you want to be in a more densely area with diverse venues. I recommend developing a restaurant in Cluster 2.The food options are limitless, just to name a few their is Japanese,Chinese,BBQ Joint, Soba, Sushi, and Italian. This would be a great opportunity to

place International or American restaurant that served a little of everything to have an upperhand on the competition being that majority of the competition is selling the exact same food.

Conclusion

In summation, by utilizing a K-Means clustering approach to classifying cities and towns we can determine area that are prime candidates for new startup or business expansion, as well as weak areas. In addition, we can also identify which businesses are most likely to succeed in the selected areas through sentiment analysis of the top ten businesses in those areas. However, there are disadvantages to using this approach which largely arise from not enough data. Certain markets might not be right for certain businesses due to local factors that not obvious when exploring top ten trends across segments. Therefore, it is important to conduct further analysis of each area to confirm the predictions of the model.