

# The Battle of Neighborhoods

Tokyo sushi restaurant

# Introduction: Business Problem

- ▶ There are many restaurants in Tokyo, Japan. Our customers open a sushi restaurant in Yokohama, Japan. They want to open a sushi restaurant in Tokyo in next year. However, they don't know where to open a sushi restaurant in Tokyo. We analyze where customers should open a sushi restaurant.
- ▶ Many people use trains in Japan. Therefore, many people eat any foods near the station. We recommend to customers a station with a large number of users, a station with few sushi restaurants and a close atmosphere to Yokohama.
- ▶ The target is those who want to open a sushi restaurant in Tokyo.

# Data sources

This demonstration will make use of the following data sources:

- List of stations in Tokyo.

This wiki page has list of number of daily station entries for each stations in Tokyo.

([https://en.wikipedia.org/wiki/List\\_of\\_East\\_Japan\\_Railway\\_Company\\_stations](https://en.wikipedia.org/wiki/List_of_East_Japan_Railway_Company_stations))

- Stations in Tokyo location data retrieved using Google maps API.

Data coordinates o stations will be retrieved using google API.

- Venues near stations in Tokyo from FourSquare API

(FourSquare website: [www.foursquare.com](http://www.foursquare.com))

# List of stations in Tokyo with daily station entries using BeautifulSoup

↕	JR East Lines	Other Lines	Code ↕	Daily Station Entries ↕
	<ul style="list-style-type: none"><li>■ Jōban Line</li><li>■ Jōban Line (Local)</li><li>■ Jōban Line (Rapid)</li><li>■ Narita Line</li></ul>			29,989
	■ Takasaki Line			40,395
	■ Yokohama Line			10,099
	■ Narita Line			3,137
	■ Itō Line			913
	<ul style="list-style-type: none"><li>■ Keihin-Tōhoku Line</li><li>■ Saikyō Line</li><li>■ Shōnan-Shinjuku Line</li><li>■ Takasaki Line</li><li>■ Utsunomiya Line (Tōhoku Main Line)</li></ul>	<ul style="list-style-type: none"><li>■ Saitama Rapid Railway Line (Akabane-iwabuchi)</li><li>■ Tokyo Metro Namboku Line (Akabane-iwabuchi)</li></ul>	ABN	87,346
	■ Jōban Line			5,532
	■ Itsukaichi Line			7,310

The dataset we use was information for stations in Tokyo with daily station entries from Wikipedia.

From the original data, we will only use station name and daily station entries. We got this table data from Wikipedia using BeautifulSoup. After, we removed columns for JR East Lines and Code Source Year, we sorted by Daily Station Entries.

# Get stations in Tokyo and Yokohama station location data retrieved using Google maps API

	Station	Daily Station Entries	Latitude	Longitude
0	Shinjuku	734154	35.689596	139.700478
1	Ikebukuro	544762	35.730445	139.708519
2	Shibuya	402766	35.659391	139.701917
3	Tokyo	380997	35.681236	139.767125
4	Shinagawa	323893	35.629368	139.739273
5	Shimbashi	243890	35.666111	139.759721
6	Akihabara	230689	35.698557	139.773142
7	Takadanobaba	199741	35.712640	139.703874
8	Kita-Senju	194136	35.748916	139.804754
9	Ueno	174832	35.711964	139.777839
10	Yūrakuchō	162252	35.675541	139.763291
11	Tachikawa	155868	35.724463	139.404766
12	Hamamatsuchō	151480	35.655133	139.756914
13	Tamachi	148346	35.645647	139.747451
14	Kichijōji	137555	35.702702	139.579537
15	Kamata	133593	35.563256	139.715321
16	Ebisu	128555	35.646580	139.710215
17	Gotanda	127996	35.627267	139.722595
18	Ōsaki	127838	35.619627	139.728050
19	Nakano	122846	35.705591	139.665471
20	Yokohama	394900	35.444991	139.636768

We add coordinated of these station using google map api to our dataframe. We use top 20 stations in Tokyo of number of Daily Station Entries and Yokohama station for our analysis.

# Generate Map by folium





# Methodology

- ▶ Using the Foursquare API, the explore API function was used to get the most common venue categories in each neighborhood, and then used this feature to group the stations into clusters.
- ▶ The k-means clustering algorithm was used for the analysis.
- ▶ Finally, the Folium library is used to visualize the recommended stations and their emerging clusters.

# Retrieving FourSquare Places of interest

We can get 878 venues near top 20 stations in Tokyo and Yokohama station. The data we got by FourSquare have 165 unique venues.

	Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Shinjuku	35.689596	139.700478	Lumine the Yoshimoto (レミネのよしもと)	35.689490	139.701115	Comedy Club
1	Shinjuku	35.689596	139.700478	SUSHI TOKYO TEN	35.688184	139.700285	Sushi Restaurant
2	Shinjuku	35.689596	139.700478	Verve Coffee Roasters	35.688269	139.701198	Coffee Shop
3	Shinjuku	35.689596	139.700478	Sarutahiko Coffee & TiKiTaKa Ice Cream (篠田 蔭珈琲と...)	35.689201	139.699075	Coffee Shop
4	Shinjuku	35.689596	139.700478	AKOMEYA TOKYO	35.688908	139.702209	Organic Grocery

After we got these venues, we executed one hot encoding by each venue.

[illegible]



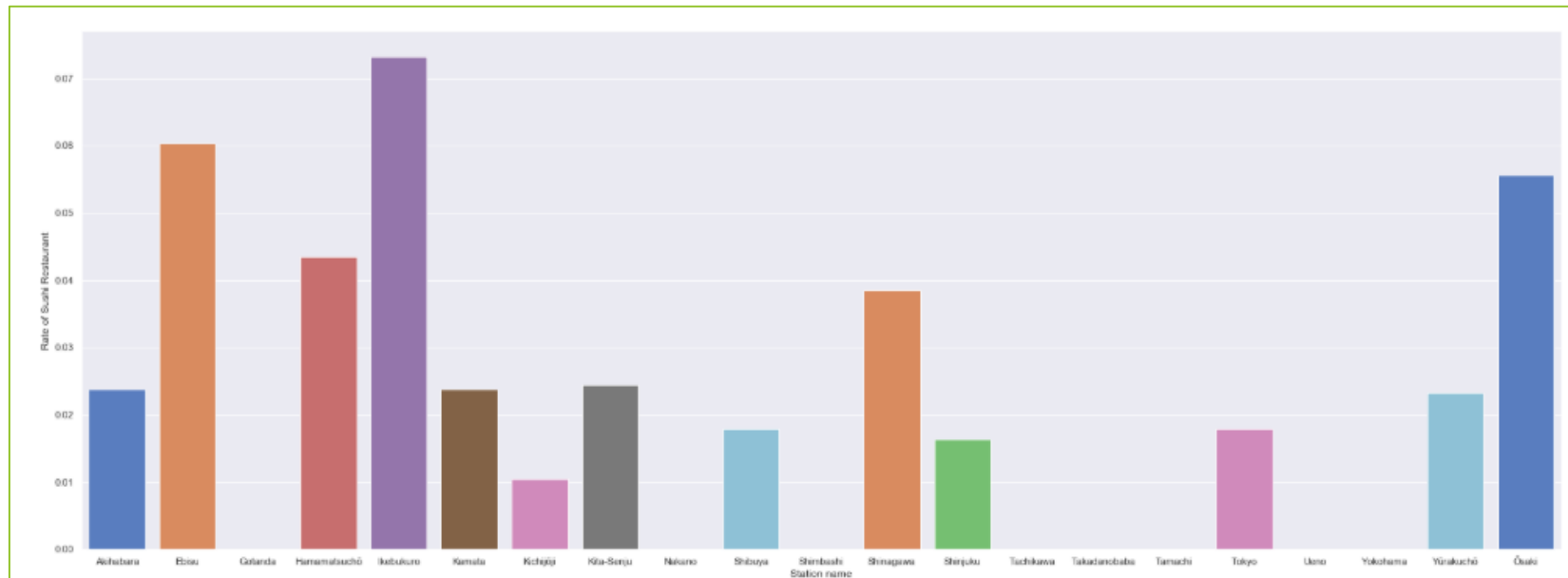
# Regrouping and Category Statistics

- After one hot encoding, we grouped this data by each station.

	Station	Acai House	Adult Boutique	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	BBQ Joint
0	Akihabara	0.000000	0.011905	0.000000	0.000000	0.035714	0.000000	0.00000	0.000000	0.011905	0.000000
1	Ebisu	0.000000	0.000000	0.000000	0.012048	0.000000	0.000000	0.00000	0.012048	0.000000	0.072289
2	Gotanda	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000
3	Hamamatsuchō	0.000000	0.000000	0.043478	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.130435
4	Ikebukuro	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.048780
5	Kamata	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.023810
6	Kichijōji	0.000000	0.000000	0.000000	0.000000	0.010526	0.000000	0.00000	0.010526	0.000000	0.010526
7	Kita-Senju	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.024390
8	Nakano	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.080000
9	Shibuya	0.017857	0.000000	0.000000	0.000000	0.000000	0.017857	0.00000	0.000000	0.000000	0.053571
10	Shimbashi	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.02439	0.000000	0.000000	0.024390
11	Shinagawa	0.000000	0.000000	0.000000	0.019231	0.000000	0.000000	0.00000	0.000000	0.000000	0.019231
12	Shinjuku	0.000000	0.000000	0.000000	0.000000	0.016393	0.000000	0.00000	0.000000	0.000000	0.000000
13	Tachikawa	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000
14	Takadanobaba	0.000000	0.000000	0.000000	0.000000	0.042553	0.000000	0.00000	0.000000	0.000000	0.021277
15	Tamachi	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000
16	Tokyo	0.000000	0.000000	0.000000	0.000000	0.000000	0.017857	0.00000	0.000000	0.000000	0.000000
17	Ueno	0.000000	0.000000	0.000000	0.090909	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000
18	Yokohama	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000
19	Yūrakuchō	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.023256
20	Ōsaki	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000

# Visualization about Rate of Sushi Restaurant

We create bar plot for rate of sushi restraint of each station using seaborn.

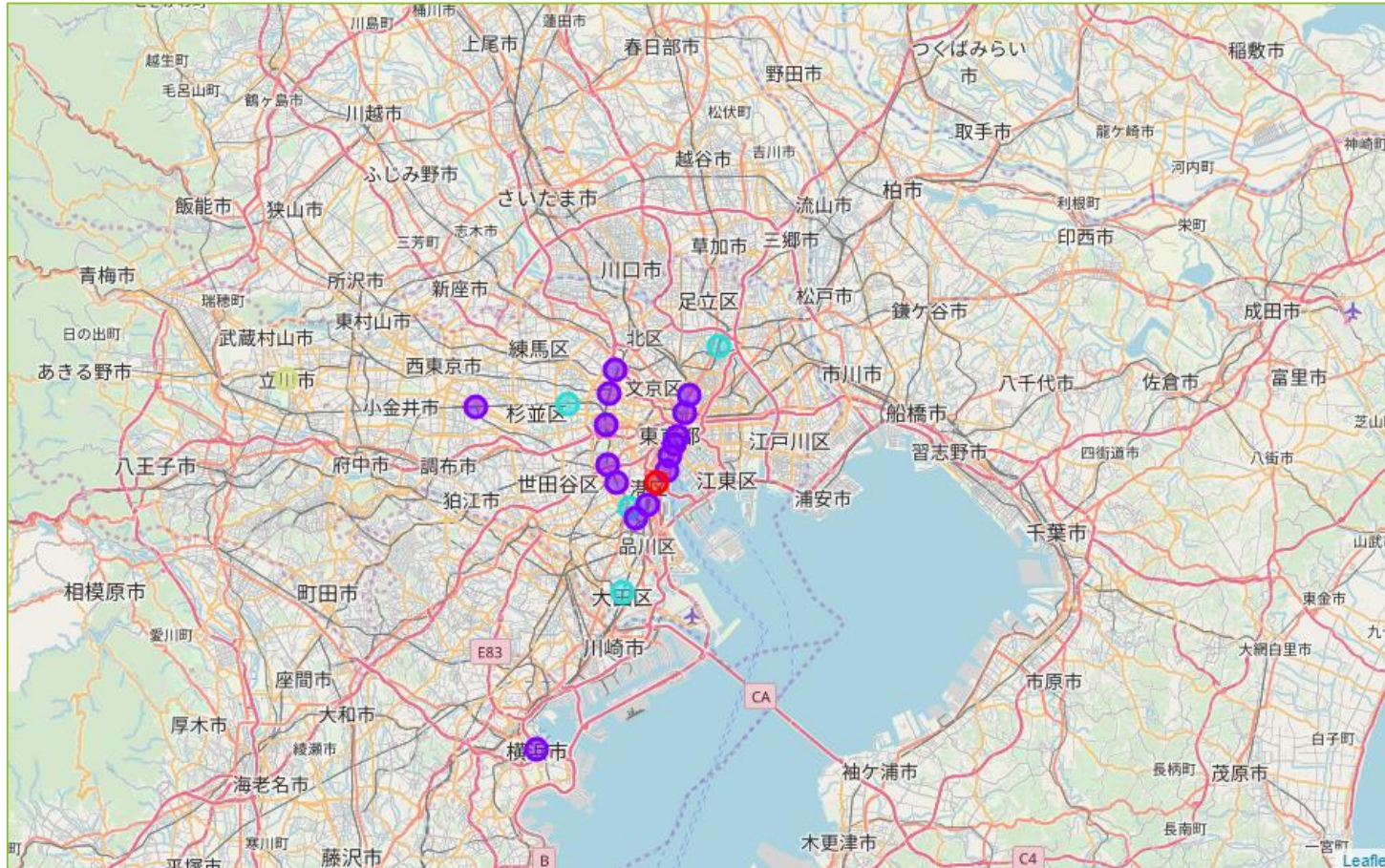


# Clustering of Stations

	Station	Cluster Labels	Sushi Restaurant	Daily Station Entries	Latitude	Longitude
0	Akihabara	1	0.023810	230689	35.698557	139.773142
1	Ebisu	1	0.060241	128555	35.646580	139.710215
2	Gotanda	2	0.000000	127996	35.627267	139.722595
3	Hamamatsuchō	1	0.043478	151480	35.655133	139.756914
4	Ikebukuro	1	0.073171	544762	35.730445	139.708519
5	Kamata	2	0.023810	133593	35.563256	139.715321
6	Kichijōji	1	0.010526	137555	35.702702	139.579537
7	Kita-Senju	2	0.024390	194136	35.748916	139.804754
8	Nakano	2	0.000000	122846	35.705591	139.665471
9	Shibuya	1	0.017857	402766	35.659391	139.701917
10	Shimbashi	1	0.000000	243890	35.666111	139.759721
11	Shinagawa	1	0.038462	323893	35.629368	139.739273
12	Shinjuku	1	0.016393	734154	35.689596	139.700478
13	Tachikawa	3	0.000000	155868	35.724463	139.404766
14	Takadanobaba	1	0.000000	199741	35.712640	139.703874
15	Tamachi	0	0.000000	148346	35.645647	139.747451
16	Tokyo	1	0.017857	380997	35.681236	139.767125
17	Ueno	1	0.000000	174832	35.711964	139.777839
18	Yokohama	1	0.000000	394900	35.444991	139.636768
19	Yūrakuchō	1	0.023256	162252	35.675541	139.763291
20	Ōsaki	1	0.055556	127838	35.619627	139.728050

- The next thing to do now, is to create clusters of the station using the k-means to cluster the station into 4 clusters.
- After we executed clustering, we added cluster labels to dataframe.
- Finally, our dataframe have station name, cluster label, rate of sushi restaurant.

Generate map which represented  
diffluent cluster labels by folium





# Filter by same Cluster Labels of Yokohama and sort by rate of Sushi Restaurant

	Station	Cluster Labels	Sushi Restaurant	Daily Station Entries	Latitude	Longitude
10	Shimbashi	1	0.000000	243890	35.666111	139.759721
14	Takadanobaba	1	0.000000	199741	35.712640	139.703874
17	Ueno	1	0.000000	174832	35.711964	139.777839
6	Kichijōji	1	0.010526	137555	35.702702	139.579537
12	Shinjuku	1	0.016393	734154	35.689596	139.700478
9	Shibuya	1	0.017857	402766	35.659391	139.701917
16	Tokyo	1	0.017857	380997	35.681236	139.767125
19	Yūrakuchō	1	0.023256	162252	35.675541	139.763291
0	Akihabara	1	0.023810	230689	35.698557	139.773142
11	Shinagawa	1	0.038462	323893	35.629368	139.739273
3	Hamamatsuchō	1	0.043478	151480	35.655133	139.756914
20	Ōsaki	1	0.055556	127838	35.619627	139.728050
1	Ebisu	1	0.060241	128555	35.646580	139.710215
4	Ikebukuro	1	0.073171	544762	35.730445	139.708519

# Results

- ▶ The aim of this project is to help our customers who want to open sushi restaurant in Tokyo.
- ▶ A station which has many users and few sushi restaurants and a close atmosphere to Yokohama station is "Shinbashi" station.
- ▶ We recommend Shinbashi station to our customer.

# Discussion and Conclusion

- ▶ Using Foursquare API, we haven't collected amount of venues in Tokyo.
- ▶ I think many Japanese have not used foursquare and are using another service.
- ▶ If we want to collect venue information, we should use other API. However, most famous API for collecting venue information is available Japanese language.