Exploring restaurants in Tokyo

Abhinav Kumar Goswami

05 April 2020

1. Introduction

Background

Changing demographics and lifestyles are driving the growth of food service businesses. Busy customers have little time to cook or desire to do so. They want fresh bread flavour, without baking hassle. They want to wash delicious, healthy meals without dishes. Working parents, elderly people and bachelors prefer to buy their meals. Many restaurants fail during their first year, frequently due to a lack of planning.

Problem

The ideal place to open a restaurant would be a place where there are more customers and less competition. The proximity of the restaurant to other businesses and the scope of development of the area can also affect sales. This project aims to find the best location to open a restaurant.

Conditions

The restaurant should be far from bakeries and other restaurants. The brand value of the restaurants that are near should be less. It is not a good idea to set up a restaurant in a residential area; however, if the restaurant offers food delivery services, it might be a good idea to open the restaurant at a considerable distance. The restaurant should be accessible by students and office workers.

2. Data

Overview

To solve the problem the following data is needed

- List of neighbourhoods of Tokyo
- Their coordinates
- Venue data

Data Sources

List of neighbourhoods was downloaded from Wikipedia (https://en.wikipedia.org/wiki/Special_wards_of_Tokyo). Geocoder was used to get the coordinates. Foursquare was used to get the venue data.

Data Cleaning

A dataframe of neighbourhoods was created using pandas.read_html("https://en.wikipedia.org/wiki/Special_wards_of_Tokyo")[3]

```
datahtml = pd.read_html('https://en.wikipedia.org/wiki/Special_wards_of_Tokyo')
datahtml[3].head()
```

	No.	Flag	Name	Kanji	Population(as of October 2016	Density(/km2)	Area(km2)	Major districts
0	01	NaN	Chiyoda	千代田区	59441	5100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...
1	02	NaN	Chūō	中央区	147620	14460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...
2	03	NaN	Minato	港区	248071	12180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
3	04	NaN	Shinjuku	新宿区	339211	18620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
4	05	NaN	Bunkyō	文京区	223389	19790	11.29	Hongō, Yayoi, Hakusan

The data consists of 24 rows and 8 columns.

```
[4] data.shape
(24, 8)
```

Only the columns- names, population, density, area and major districts were kept. All other columns were discarded. The last row consisting of the total area, density and population was discarded.

```
[8] data.shape

(23, 5)

[9] data.head()
```

	Name	Population(as of October 2016	Density(/km2)	Area(km2)	Major districts
0	Chiyoda	59441	5100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...
1	Chūō	147620	14460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...
2	Minato	248071	12180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
3	Shinjuku	339211	18620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
4	Bunkyō	223389	19790	11.29	Hongō, Yayoi, Hakusan

Columns were renamed for convenience.

Renaming the columns

```
[10] data.columns=(['name', 'population', 'density', 'area','major_districts'])

[11] data.head()
```

	name	population	density	area	major_districts
0	Chiyoda	59441	5100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...
1	Chūō	147620	14460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...
2	Minato	248071	12180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
3	Shinjuku	339211	18620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
4	Bunkyō	223389	19790	11.29	Hongō, Yayoi, Hakusan

Geocoder was used to find the longitude and latitude of the wards in Tokyo.

```
[12] geolocator = Nominatim(user_agent="tokyoexp")

     latitude=[]
     longitude=[]

     for i in range (data.shape[0]):
       location = geolocator.geocode(data['name'][i])
       latitude.append(location.latitude)
       longitude.append(location.longitude)
       print((location.latitude,location.longitude))
```

[14]

	name	population	density	area	major_districts	latitude	longitude
0	Chiyoda	59441	5100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.693810	139.753216
1	Chūō	147620	14460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...	35.666255	139.775565
2	Minato	248071	12180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055
3	Shinjuku	339211	18620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...	35.693763	139.703632
4	Bunkyō	223389	19790	11.29	Hongō, Yayoi, Hakusan	35.718810	139.744732
5	Taitō	200486	19830	10.11	Ueno, Asakusa	35.717450	139.790859
6	Sumida	260358	18910	13.77	Kinshichō, Morishita, Ryōgoku	35.700429	139.805017
7	Kōtō	502579	12510	40.16	Kiba, Ariake, Kameido, Tōyōchō, Monzennakachō,...	35.649154	139.812790
8	Shinagawa	392492	17180	22.84	Shinagawa, Gotanda, Ōsaki, Hatanodai, Ōimachi,...	35.599252	139.738910
9	Meguro	280283	19110	14.67	Meguro, Nakameguro, Jiyugaoka, Komaba, Aobadai	35.621250	139.688014
10	Ōta	722608	11910	60.66	Ōmori, Kamata, Haneda, Den-en-chōfu	35.561206	139.715843
11	Setagaya	910868	15690	58.05	Setagaya, Shimokitazawa, Kinuta, Karasuyama, T...	35.646096	139.656270
12	Shibuya	227850	15080	15.11	Shibuya, Ebisu, Harajuku, Daikanyama, Hiroo, S...	35.664596	139.698711
13	Nakano	332902	21350	15.59	Nakano	35.718123	139.664468
14	Suginami	570483	16750	34.06	Kōenji, Asagaya, Ogikubo	35.699493	139.636288
15	Toshima	294673	22650	13.01	Ikebukuro, Komagome, Senkawa, Sugamo	35.736156	139.714222
16	Kita	345063	16740	20.61	Akabane, Ōji, Tabata	-0.220164	-78.512327
17	Arakawa	213648	21030	10.16	Arakawa, Machiya, Nippori, Minamisenju	35.737529	139.781310
18	Itabashi	569225	17670	32.22	Itabashi, Takashimadaira	35.774143	139.681209
19	Nerima	726748	15120	48.08	Nerima, Ōizumi, Hikarigaoka	35.748360	139.638735
20	Adachi	674067	12660	53.25	Ayase, Kitasenju, Takenotsuka	35.783703	139.795319
21	Katsushika	447140	12850	34.80	Tateishi, Aoto, Kameari, Shibamata	35.751733	139.863816
22	Edogawa	685899	13750	49.90	Kasai, Koiwa	35.678278	139.871091

The coordinates of Kita are wrong, so I used geolocator to find the coordinates of the first major_district in kita ward. I wrote a function that would replace the coordinates of the ward with the coordinates of the first major district if the latitude of ward is less than 35.

```
for i in range(data.shape[0]):
  if data.loc[i,'latitude']<35:
    district=data.loc[i,'major_districts'].split(',')[0]
    location = geolocator.geocode(district)
    data.loc[i,'latitude']=location.latitude
    data.loc[i,'longitude']=location.longitude
```

In order to visualize the density of districts on the map, I added a new column "radius", consisting of normalized value of density.

```
[18] div=data['density'].max()/13
     data['radius']=data['density']/div
     data.head()
```

	name	population	density	area	major_districts	latitude	longitude	radius
0	Chiyoda	59441	5100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.693810	139.753216	2.927152
1	Chūō	147620	14460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...	35.666255	139.775565	8.299338
2	Minato	248071	12180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	6.990728
3	Shinjuku	339211	18620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...	35.693763	139.703632	10.686976
4	Bunkyō	223389	19790	11.29	Hongō, Yayoi, Hakusan	35.718810	139.744732	11.358499

I used Foursquared to get the neighbourhood data and made a list of top 10 most visited places in that ward and to make a heat map of food industries.

	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	Akabane, Ōji, Tabata	Sake Bar	Convenience Store	Ramen Restaurant	Bar	Shopping Mall	Italian Restaurant	Coffee Shop	Supermarket	BBQ Joint	Soba Restaurant
1	Arakawa, Machiya, Nippori, Minamisenju	Convenience Store	Park	Grocery Store	Intersection	Noodle House	Café	Ramen Restaurant	Indian Restaurant	Theater	Donburi Restaurant
2	Ayase, Kitasenju, Takenotsuka	Convenience Store	Japanese Restaurant	Supermarket	Restaurant	Drugstore	Park	Pharmacy	Flea Market	Gaming Cafe	Furniture / Home Store
3	Hongō, Yayoi, Hakusan	Japanese Restaurant	Park	Intersection	Steakhouse	Road	Museum	Spa	BBQ Joint	Botanical Garden	Szechuan Restaurant
4	Ikebukuro, Komagome, Senkawa, Sugamo	Ramen Restaurant	Music Store	Chinese Restaurant	Convenience Store	Japanese Restaurant	Yoshoku Restaurant	Hobby Shop	Sake Bar	Rock Club	Recording Studio

3. Methodology

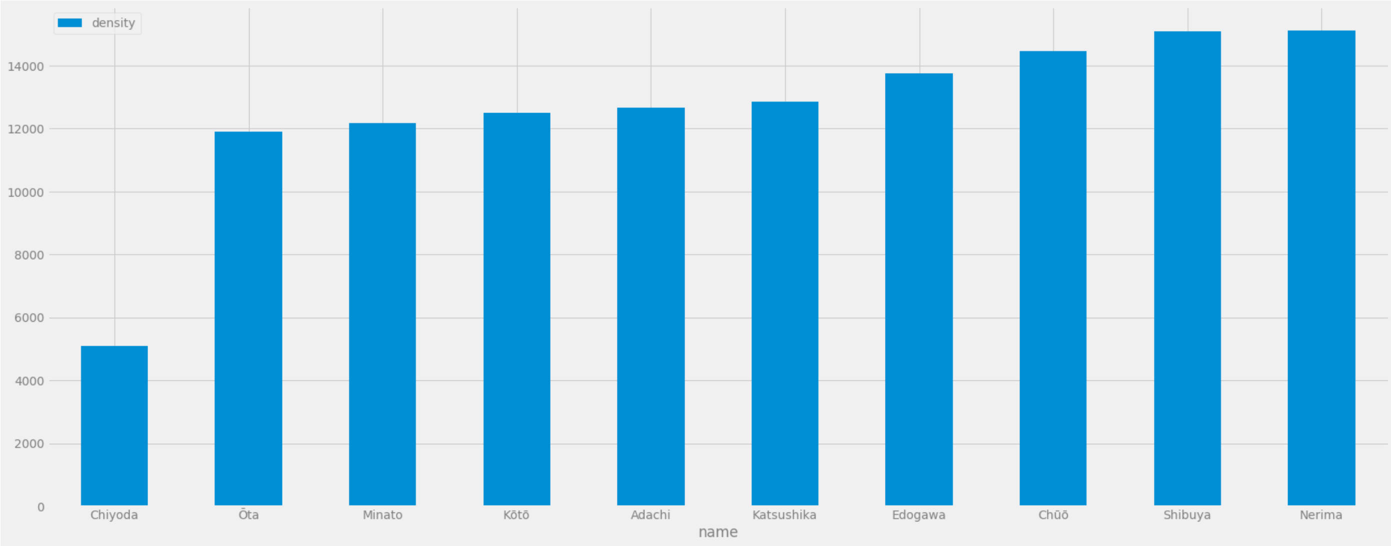
Map

I used folium to create the map, the radius of circle markers in the map is directly proportional to the population density of that ward.

From the above map, it is evident that Chidoya has the least population density, whereas the population density of Toshima is the highest. Hence, it is not a good idea to set up a new restaurant in Chidoya.

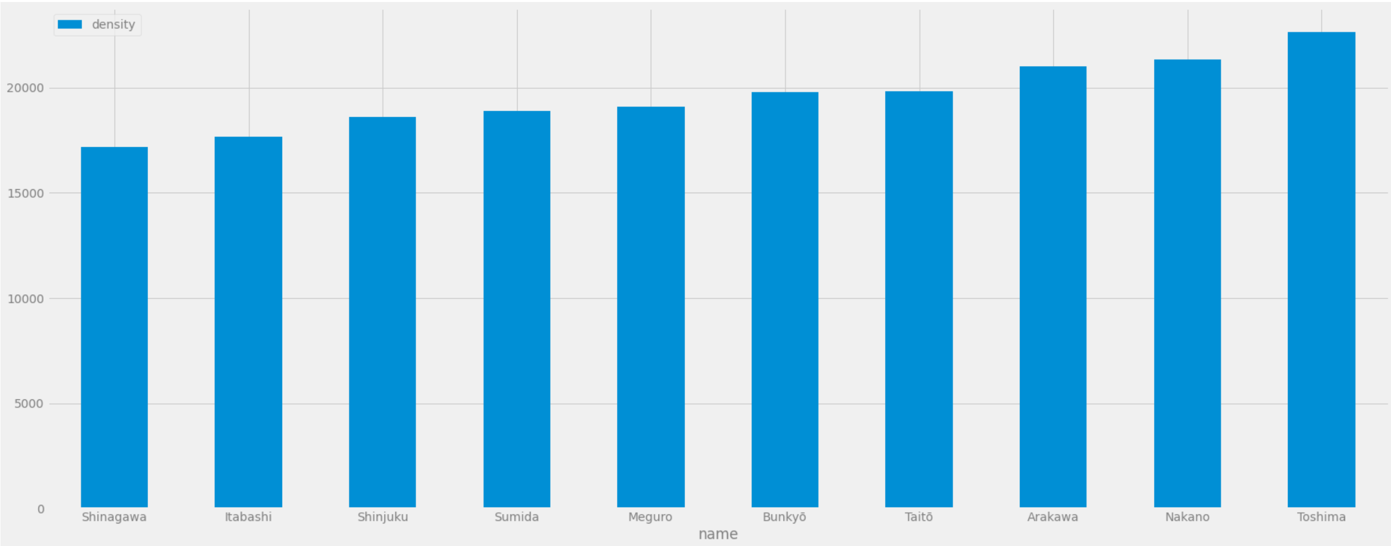
Plot (population density)

On arranging the data in ascending order of density and plotting the first 10 elements, we get the following.



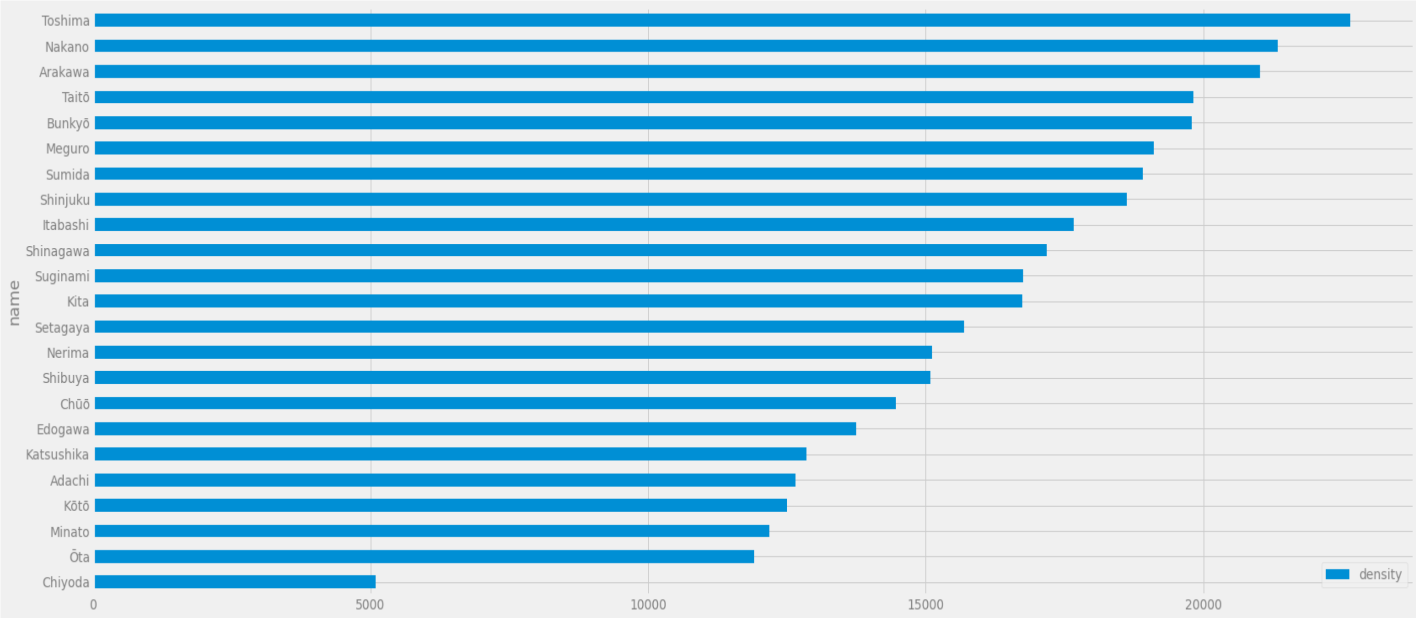
From the above plot it is evident that the population density (population/km2) of all wards in Tokyo is more than 10,000 (except for Chiyoda).

The following is a plot of districts with high population density.



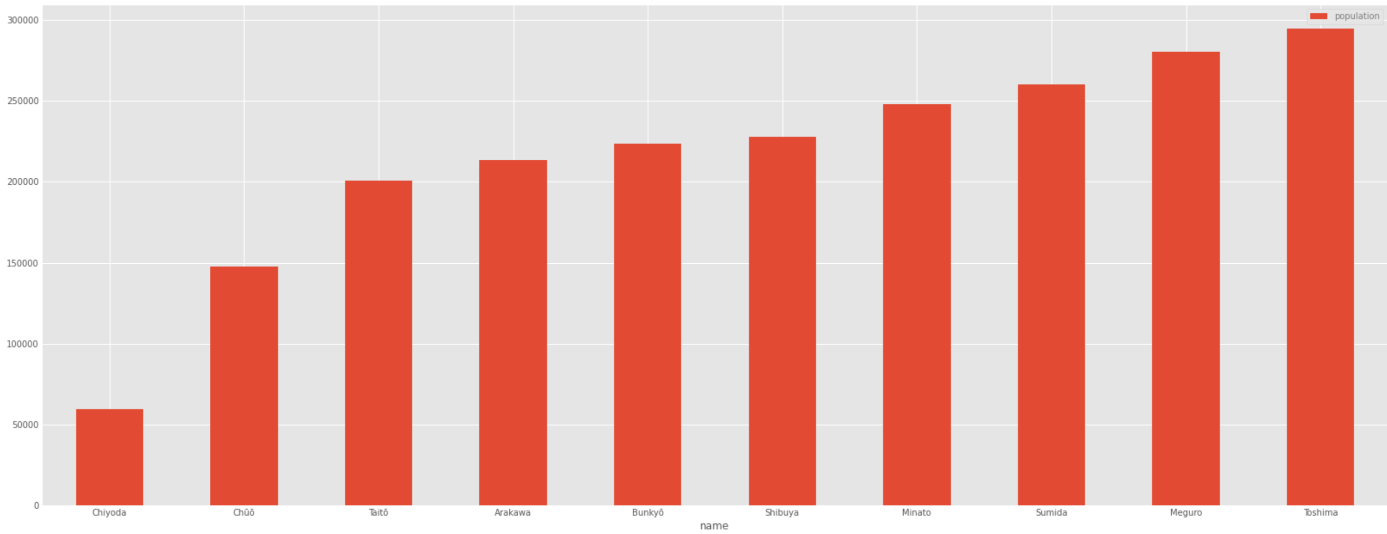
We can see that the population density of Toshima, Nakano and Arakawa exceeds 20,000.

The below plot compares the population density of all districts in Tokyo.



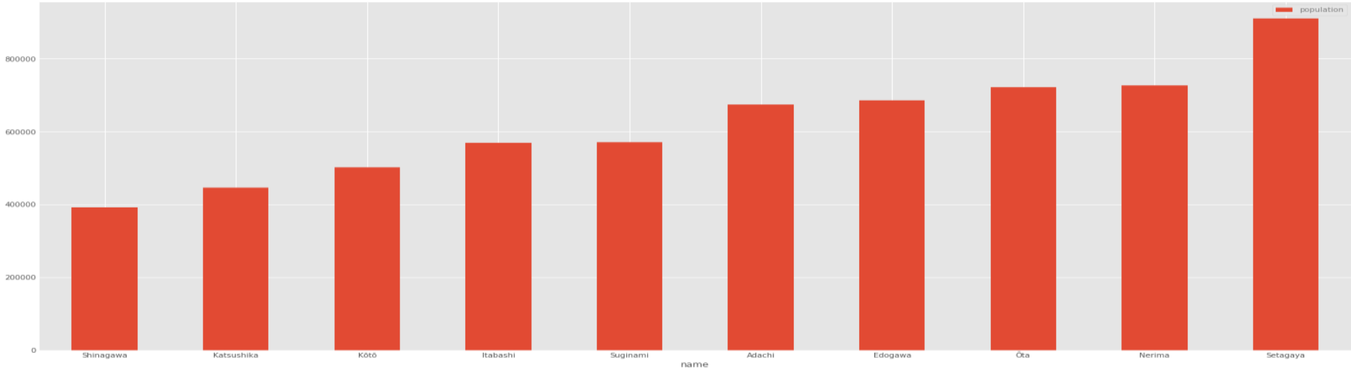
Plot (population)

Sorting the data by its population gives us the following plot.

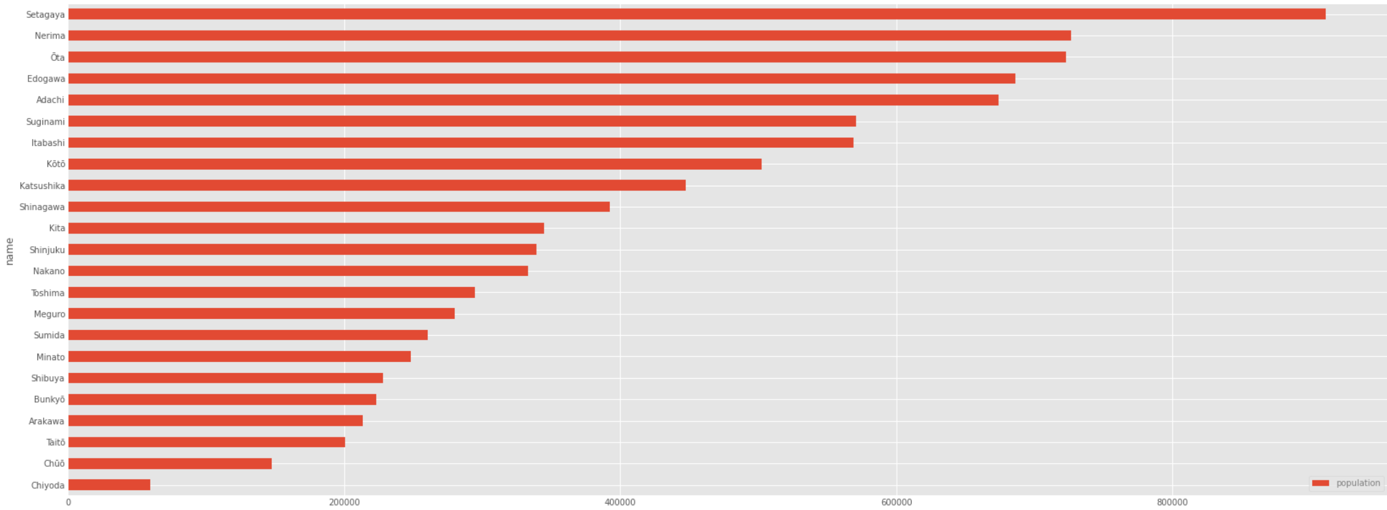


The above plot compares districts with low population.

Districts with high population can be seen in the following plot.



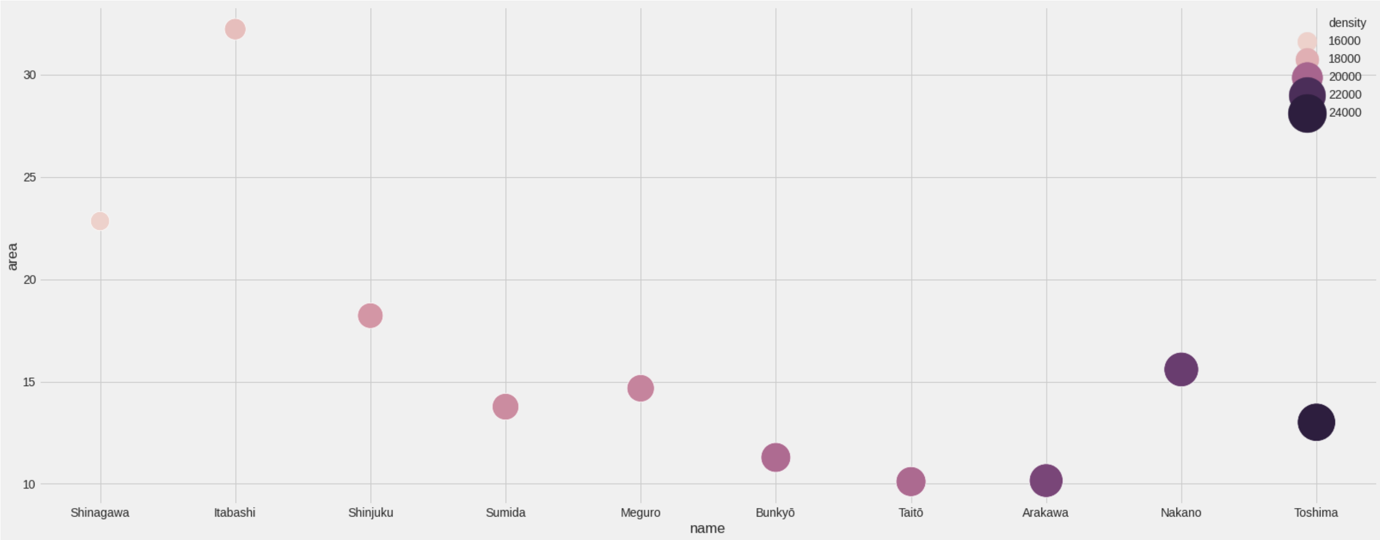
And the below plot is for overall comparison.



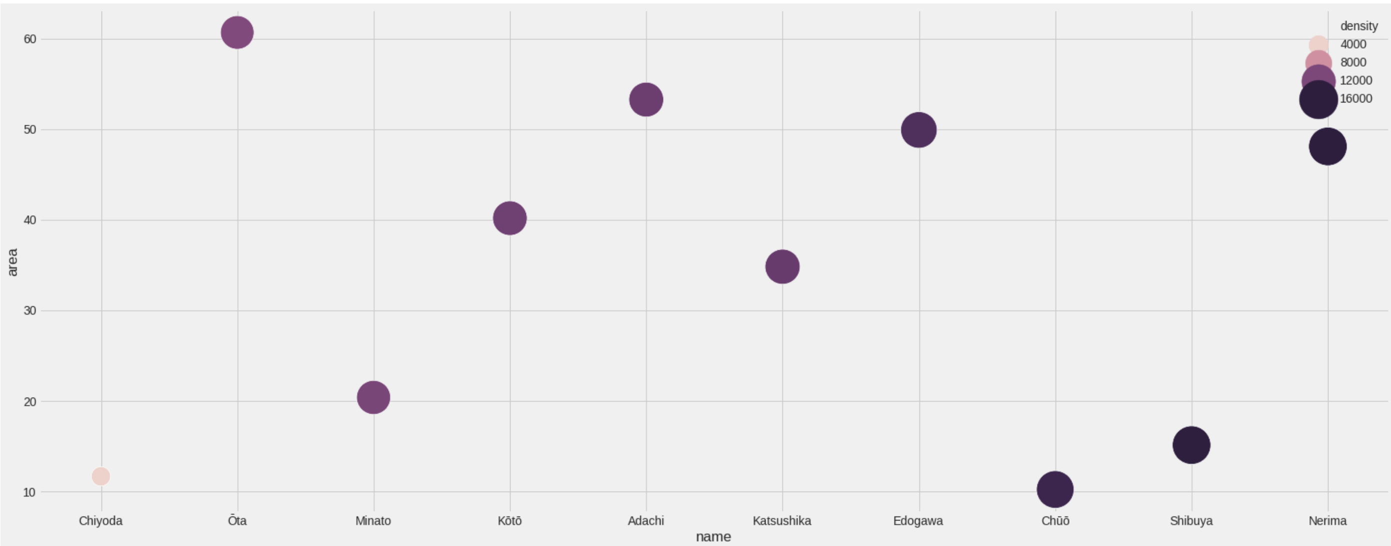
This data was fed to foursquare to get a list of top venues and nearby restaurants.

Scatter plot (density and area)

Both area and density of a district can be compared using the following plot. Bigger bubbles and darker shades imply high population density, whereas y axis depicts the area (in km2).

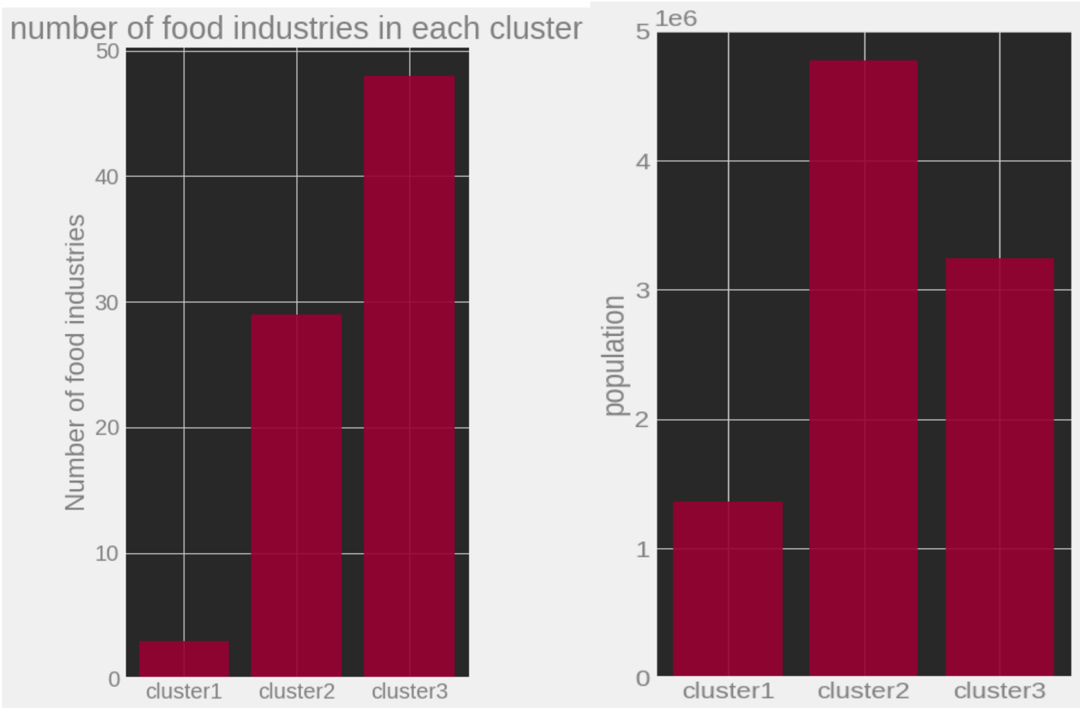


The above plot shows the variation of area with density in high population density districts, whereas the plot below shows the same thing for low population density regions.



Clustering

I did clustering on the data which had a list of top 10 venues. I divided my districts into 3 clusters. Clustering the data based on top venues will help me analyse the neighbourhood. Neighbourhoods that are famous for their restaurants will be clustered separately. Thus giving me an idea of the impact it will have on the restaurant business.



From the left plot it is evident that the number of food industries (in top 10 venues) in cluster 1 is the least; whereas no. of food industries in cluster 3 is highest. The right plot shows that the population of cluster 2 is more than that of cluster 3, depicting that cluster 2 is a residential area.

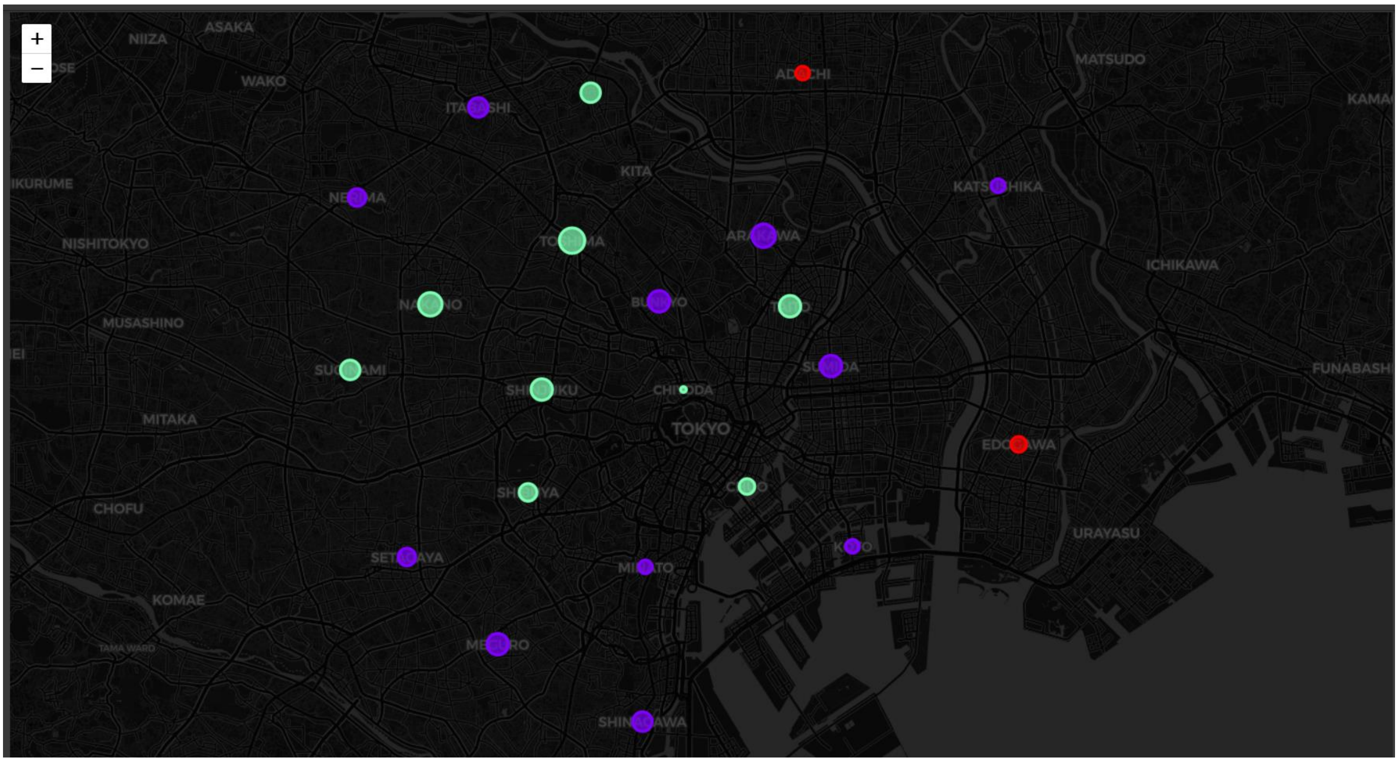
Heatmap

A heat map of all the food industries in Tokyo can be seen below.

4. Results

Cluster3 is famous for its restaurants, and Cluster2 has the highest population. The population of cluster 1 is the least, and it doesn't have good restaurants.

On making a map of clusters we can see how districts are clustered.



Cluster1=red Cluster2=blue Cluster3=green

Cluster1- Districts in cluster1 have low population density, there aren't many restaurants in cluster 1. The most common venues in cluster1 are convenience stores.

Cluster2- Cluster2 has high density, the most common venue in cluster2 is Convenience Store, there are good restaurants in cluster2.

Cluster3- Cluster3 has high population density and is famous for its restaurants.

5. Discussion

While opening a restaurant we need to keep the following things in mind

- Anticipated sales volume.
- Accessibility to potential customers.
- The rent-paying capacity of your business.
- Traffic density.
- Proximity to other businesses
- Future development

6. Conclusion

Opening a restaurant in cluster3 where population density is between 15,000 and 20,000 would be a good idea.