Best Neighborhood for Starting Restaurant in Tokyo

Hasan uz zaman

May 17, 2020

1. Introduction

Japan is one of the leading technologically advanced and civilized countries of the world. And Tokyo is the heart of Japan. Tokyo is a diverse city with a blend of different cultures, history and races. Beside japanese , many people from outside japan are living and working and leading a healthy life. Tokyo is the largest metropolitan area in the world. Around 35 million people live here. Tokyo is famous for many things such as anime, electronics, video games, food items etc. In this report , we will talk about food. Food is an essential part for the residents of Tokyo. Famous food items that Tokyo is known for are Edomae-zushi, Monjayaki, Ramen, Tempura, Tendon, Soba etc . There are also food items like indian, chinese, steak, fries, kababs, fried chicken, burgers etc can be found in Tokyo. All these food items are made and presented by different japanese restaurants, indian restaurants, global food chain shops like starbucks, mcdonalds.

Problem: There are many neighborhoods in Tokyo. Different types of restaurants and food shops number varys Neighborhood to Neighborhood. In some neighborhoods, there are many japanese restaurants compared to other types of cuisine restaurants and there also exist other neighborhoods where the picture is the exact opposite. So it's not easy to determine where to start a new restaurant in the right neighborhood where the restaurant will make good business and make good profit. This problem is for businessmen or shop owners or new startup platforms who want to find a suitable neighborhood to start their restaurant business.

Solution To Problem: So to start a new restaurant in a particular neighborhood or to start a specific type of restaurant, we have to analyze the neighborhoods data related to its restaurants and food shops with machine learning and data science to determine best possible suited neighborhoods to start a restaurant.

2. Data Collection and analyze

To solve our given problem, the 2nd step is to collect data or data acquisition. For this step, we need to collect data about different neighborhoods of Tokyo. Tokyo is a big city with over 50 plus neighborhoods, So first thing we have to do is to collect the names of neighborhoods in Tokyo city. As there is no prepared list or table containing the names of Tokyo's neighborhoods on the internet or any website, we have to make a table containing neighborhoods of Tokyo city. We create a google sheet and on that sheet we will list the neighborhoods as a table under label Neighborhood. The names of the neighborhoods are taken from the link : https://en.wikipedia.org/wiki/Category:Neighborhoods_of_Tokyo. It's a wikipedia page containing all the neighborhood's names. After putting neighborhoods names on the table, we need two more columns to add. The columns will contain each neighborhood's corresponding latitude and longitude. We can obtain the coordinates of neighborhoods using Google or Foursquare location service. We have to put the right coordinates to the right neighborhood. After finishing the task , finally our neighborhood table is ready. Since we can not use excel sheets directly on our machine learning process, we first need to convert the excel file into a csv file. We will use google's colaboratory for machine learning or data analysis environment to solve our given problem. In colaboratory, we will first import our essential libraries like pandas, numpy, matplotlib, sklearn, etc to make our environment ready for work. Then we will create a dataframe and in that dataframe, we will load our csv file that we made on Tokyo's neighborhoods data. Now our data frame is ready for further work.

```
[ ] downloaded = drive.CreateFile({'id':id})
    downloaded.GetContentFile('Filename.csv')
    df = pd.read_csv('Filename.csv')
```

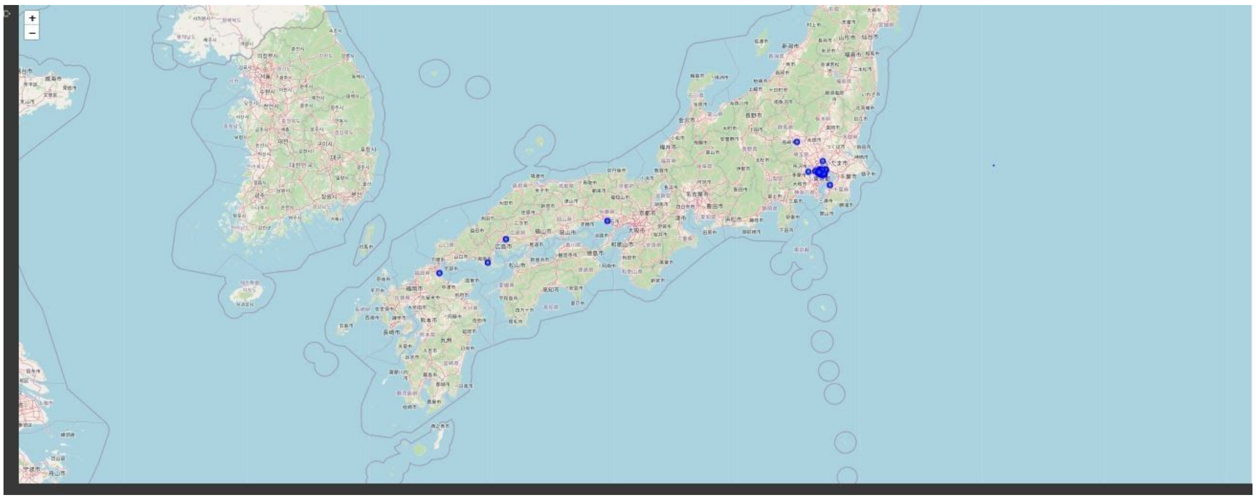
```
[ ] df
```

	Neighborhood	Latitude	Longitude
0	Agariyashiki	33.9779	132.0887
1	Akihabara	35.7023	139.7745
2	Aoyama	35.6695	139.7192
3	Asagaya	35.7100	139.6376
4	Daikanyama	35.6505	139.7042
5	Ebisu, Shibuya	35.6461	139.7154
6	Ebisuminami, Shibuya	35.6450	139.7084
7	Ebisunishi	35.6486	139.7058

We will use location service Foursquare to extract data about the venues of all the neighborhoods using latitude and longitude of Neighborhoods and using those data, we will find our solution to the given problem. We will discuss the later process in the next section.

3. Methodology section And exploratory data analysis

In this section we will use our machine learning approach to find a solution from our dataset. We already created our data frame. Now we are gonna work with our data frame.There’s two types of approaches we can take to solve our problem. These are supervised and unsupervised learning. We will use unsupervised learning as we dont know our outcome labels. Out of many techniques of unsupervised learning, we will use the clustering technique. Because clustering is used in statistical data and **clustering** analysis gives some valuable insights from our data by seeing what groups the data points fall into when we apply a **clustering** algorithm. Here we will use k-means clustering to cluster or Group our neighborhoods based on their similarities and the number of groups represented by the variable **K.** First let our neighborhoods of Tokyo mapped in with markers using our Foursquare credentials with their respective coordinates.



After mapping, we are now going to get our venue information of all the neighborhoods using Foursquare api. There are 54 neighborhoods in our Data frame. Getting all the venue info of neighborhoods, we will count them to find how many venues each neighborhood has.

```
Tokyo_venues.groupby('Neighborhood').count()
```

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agariyashiki	3	3	3	3	3	3
Akihabara	100	100	100	100	100	100
Aoyama	40	40	40	40	40	40
Asagaya	23	23	23	23	23	23
Daikanyama	100	100	100	100	100	100
Ebisu, Shibuya	100	100	100	100	100	100
Ebisuminami, Shibuya	100	100	100	100	100	100

Now using dummy or indicator variables, we will label categories of each neighborhood.

```python
Tokyo_grouped = Tokyo_onehot.groupby('Neighborhood').mean().reset_index()
Tokyo_grouped
```

	Neighborhood	ATM	Accessories Store	American Restaurant	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium	Australian Restaurant
0	Agariyashiki	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
1	Akihabara	0.000000	0.00	0.000000	0.00	0.010000	0.000000	0.020000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
2	Aoyama	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
3	Asagaya	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
4	Daikanyama	0.000000	0.00	0.000000	0.01	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
5	Ebisu, Shibuya	0.000000	0.00	0.020000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.000000	0.00
6	Ebisuminami, Shibuya	0.000000	0.00	0.010000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.00

Next, it’s time to make a dataframe by labeling top 10 common places for each neighborhood and it’s also done by Foursquare api.

```python
# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Tokyo_grouped['Neighborhood']

for ind in np.arange(Tokyo_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Tokyo_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

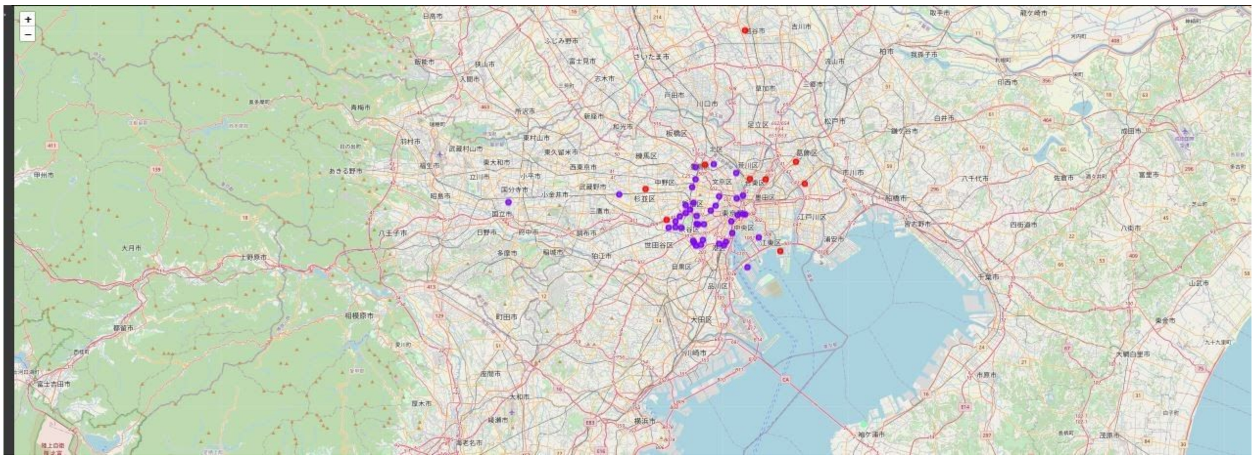
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agariyashiki	River	Park	Japanese Restaurant	Yoshoku Restaurant	Frozen Yogurt Shop	Food Court	Food Stand	Food Truck	Fountain	French Restaurant
1	Akihabara	Ramen Restaurant	Café	Hobby Shop	Electronics Store	Sake Bar	Japanese Restaurant	Chinese Restaurant	Comic Shop	Burger Joint	Soba Restaurant
2	Aoyama	Italian Restaurant	Sushi Restaurant	French Restaurant	Rock Club	Ramen Restaurant	Japanese Restaurant	Coffee Shop	Japanese Curry Restaurant	Sri Lankan Restaurant	Breakfast Spot
3	Asagaya	Convenience Store	Ramen Restaurant	Sake Bar	Used Bookstore	Bus Stop	Grocery Store	Café	Bookstore	Park	Pet Store
4	Daikanyama	Japanese Restaurant	Café	Boutique	Italian Restaurant	Bakery	Coffee Shop	BBQ Joint	Bar	Rock Club	French Restaurant

After mapping and getting venue info, we will use our clustering algorithm to cluster our neighborhoods. We will use 3 for the value of k as it’s the suitable number for our clustering.

```python
neighborhoods_venues_sorted.insert(0,'Cluster Label',  kmeans.labels_)

Tokyo_merged = mydata

# merge tokyo_grouped with tokyo_data to add latitude/longitude for each neighborhood
Tokyo_merged = Tokyo_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

Tokyo_merged.head() # check the last columns!
```

	Neighborhood	Latitude	Longitude	Cluster Label	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 Mos Commo Venu
0	Agariyashiki	33.9779	132.0887	2	River	Park	Japanese Restaurant	Yoshoku Restaurant	Frozen Yogurt Shop	Food Court	Food Stand	Food Truck	Fountain	Frenc Restaurar
1	Akihabara	35.7023	139.7745	1	Ramen Restaurant	Café	Hobby Shop	Electronics Store	Sake Bar	Japanese Restaurant	Chinese Restaurant	Comic Shop	Burger Joint	Sob Restaurar
2	Aoyama	35.6695	139.7192	1	Italian Restaurant	Sushi Restaurant	French Restaurant	Rock Club	Ramen Restaurant	Japanese Restaurant	Coffee Shop	Japanese Curry Restaurant	Sri Lankan Restaurant	Breakfas Spo
3	Asagaya	35.7100	139.6376	0	Convenience Store	Ramen Restaurant	Sake Bar	Used Bookstore	Bus Stop	Grocery Store	Café	Bookstore	Park	Pet Stor
4	Daikanyama	35.6505	139.7042	1	Japanese Restaurant	Café	Boutique	Italian Restaurant	Bakery	Coffee Shop	BBQ Joint	Bar	Rock Club	Frenc Restaurar

We can also see our clusters in our map using folium and Foursquare credentials.



Let's see our clusters

Our first cluster, t0

```python
[39] t0 = Tokyo_merged.loc[Tokyo_merged['Cluster Label'] == 0, Tokyo_merged.columns[[0] + list(range(3, Tokyo_merged.shape[1]))]]
     t0
```

	Neighborhood	Cluster Label	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	Agariyashiki	0	Park	Japanese Restaurant	River	Yoshoku Restaurant	Forest	Fishing Spot	Flower Shop	Food & Drink Shop	Food Court	Food Stand
1	Akihabara	0	Ramen Restaurant	Café	Hobby Shop	Electronics Store	Sake Bar	Chinese Restaurant	Japanese Restaurant	Kebab Restaurant	Bar	Burger Joint
2	Aoyama	0	Italian Restaurant	French Restaurant	Sushi Restaurant	Japanese Curry Restaurant	Coffee Shop	Ramen Restaurant	Rock Club	Japanese Restaurant	Furniture / Home Store	Pizza Place
4	Daikanyama	0	Japanese Restaurant	Café	Italian Restaurant	Boutique	Bakery	Coffee Shop	BBQ Joint	Rock Club	French Restaurant	Bar
5	Ebisu, Shibuya	0	Italian Restaurant	Japanese Restaurant	Ramen Restaurant	Sake Bar	Soba Restaurant	BBQ Joint	Coffee Shop	Pizza Place	Bar	Chinese Restaurant

2nd cluster, t1

```
[40] t1 = Tokyo_merged.loc[Tokyo_merged['Cluster Label'] == 1, Tokyo_merged.columns[[0] + list(range(3, Tokyo_merged.shape[1]))]]
     t1
```

	Neighborhood	Cluster Label	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
3	Asagaya	1	Convenience Store	Japanese Restaurant	Ramen Restaurant	Shopping Mall	Chinese Restaurant	Pharmacy	Pet Store	Grocery Store	Café	BBQ Joint
18	Imado	1	Convenience Store	Ramen Restaurant	Wagashi Place	Intersection	Jewelry Store	Baseball Field	Food & Drink Shop	Candy Store	Performing Arts Venue	Park
21	Kami-ikebukuro	1	Convenience Store	Park	Intersection	Drugstore	Music Venue	Kebab Restaurant	Persian Restaurant	Factory	Bus Stop	Pizza Place
22	Kanda, Tokyo	1	Ramen Restaurant	Convenience Store	Chinese Restaurant	Gourmet Shop	Pharmacy	Shipping Store	Supermarket	Fried Chicken Joint	French Restaurant	Fountain
26	Miyamoto-cho, Tokyo	1	Convenience Store	Chinese Restaurant	Drugstore	Bath House	Forest	Fishing Spot	Flower Shop	Food & Drink Shop	Food Court	Food Stand

And finally our 3rd cluster, t2

```
[59] t2 = Tokyo_merged.loc[Tokyo_merged['Cluster Label'] == 2, Tokyo_merged.columns[[0] + list(range(3, Tokyo_merged.shape[1]))]]
     t2
```

	Neighborhood	Cluster Label	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
51	Ushigome	2	Beach	Japanese Restaurant	Yoshoku Restaurant	Fountain	Flower Shop	Food & Drink Shop	Food Court	Food Stand	Food Truck	Forest

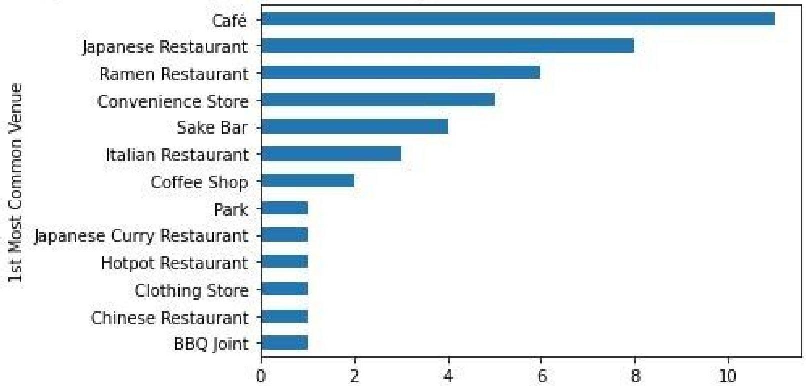
We will discuss the outcome of our clustering in the result function.

4. Result Section

Now we are gonna discuss the outcome of our work. In order to do that, We have to analyze our clusters. We have created 3 dataframe with our 3 clusters and their names are t0, t1 and t2. We won't use the t2 cluster as it has only 1 element.Now for t0 and t1, we will see the 1st most common places of these two clusters using bar chart :

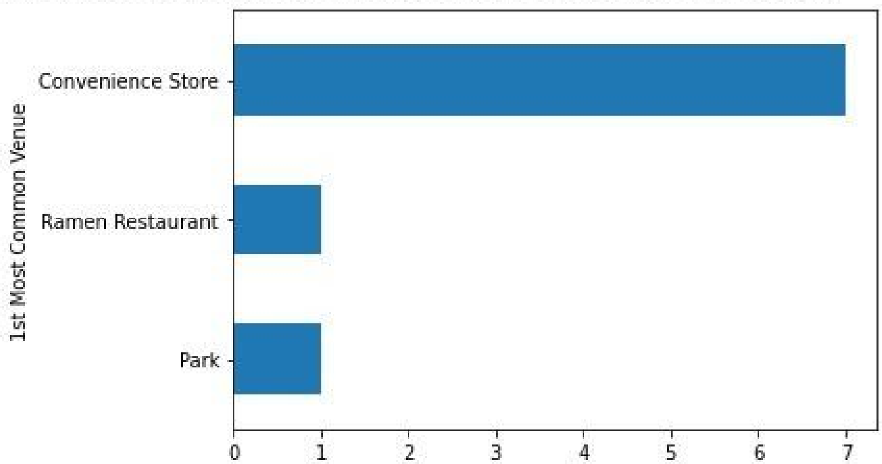
```
[106]  x1 = tx01.groupby('1st Most Common Venue')['Frequency'].mean().sort_values()
       x1.plot(kind='barh')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fb9183f4630>




```
x2 = tx11.groupby('1st Most Common Venue')['Frequency'].mean().sort_values()
x2.plot(kind='barh')
```

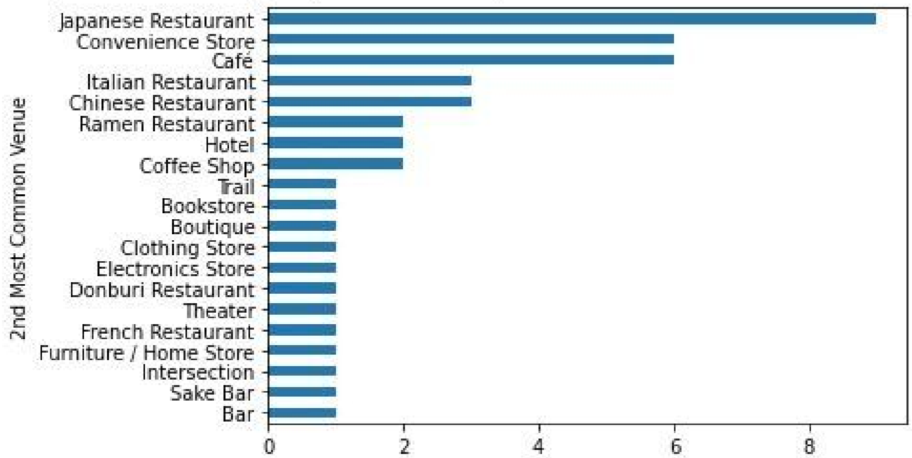
<matplotlib.axes._subplots.AxesSubplot at 0x7fb9182b7d68>



By analyzing these two bar charts, we can clearly see that in cluster t0, there are many chinese and japanese restaurants where in cluster t1, there is no major restaurant. Similarly if we see the same comparison for 2nd most common places,

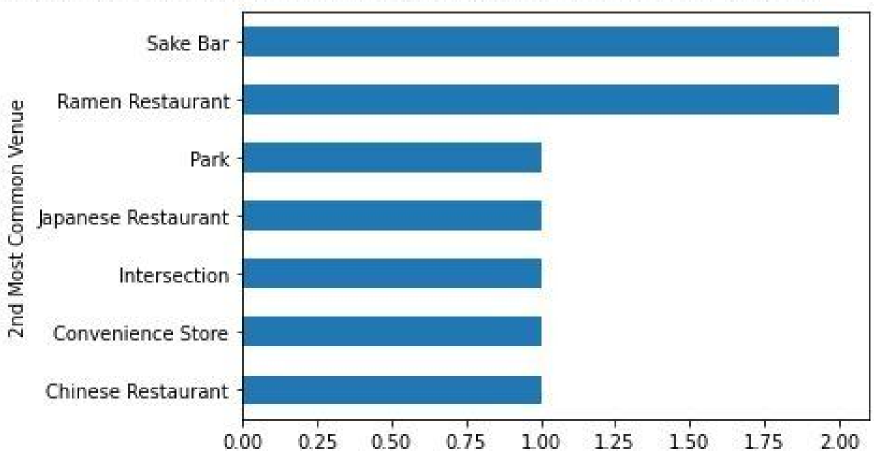
```
[110]  x3 = tx02.groupby('2nd Most Common Venue')['Frequency'].mean().sort_values(
       x3.plot(kind='barh')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb9182b0e10>
```




```
[112]  x4= tx12.groupby('2nd Most Common Venue')['Frequency'].mean().sort_values()
       x4.plot(kind='barh')
```

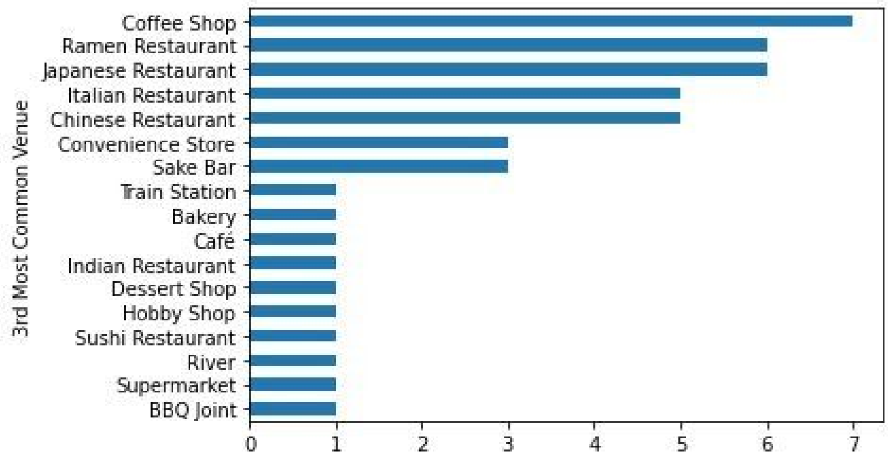
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb9182366d8>
```



The outcome is the same as for 1st common venues, There are plenty restaurants in the neighborhoods of cluster t0, where there is less number of restaurants in t1 cluster restaurants. Let’s see for 3rd common venues

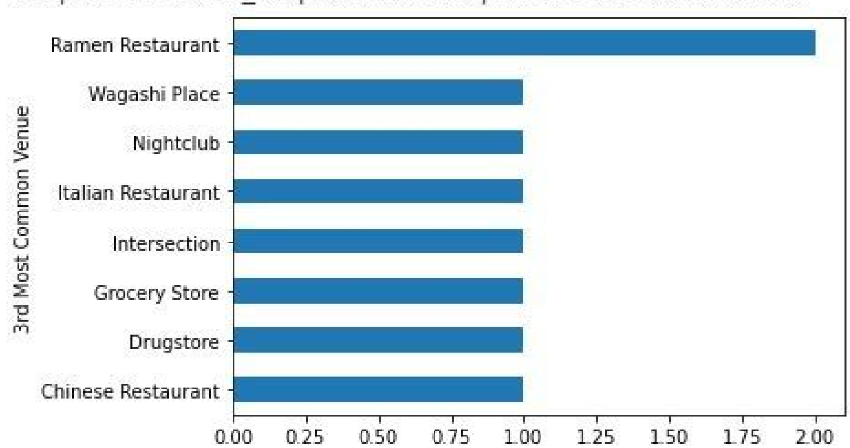
```
[114]  x5 = tx03.groupby('3rd Most Common Venue')['Frequency'].mean().sort_values()
       x5.plot(kind='barh')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fb918156dd8>




```
[116]  x6 = tx13.groupby('3rd Most Common Venue')['Frequency'].mean().sort_values()
       x6.plot(kind='barh')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fb9180eca20>



The result is the same as the first two as cluster t0 has more restaurants than cluster t1. Analyzing the three common venue columns for both clusters, it can be said that the t0 cluster neighborhoods has more food places than t1 cluster neighborhoods. As a result , the t1 cluster neighborhoods are an ideal place for starting a new restaurant.

5. Discussion

The purpose of this project is to figure out a suitable neighborhood in Tokyo city for starting a new restaurant. To solve our task, we used k-means clustering machine learning algorithm and Foursquare api to find out optimum results. We managed to distinguish neighborhoods with higher numbers of restaurants and lower numbers of restaurants. Neighborhoods like Akihabara, Daikanyama, Ebisu, Ebisunishi have plenty of restaurants like chinese , japanese, raman shop, italian. On the other hand neighborhoods like Asagaya, Imado, Kanda, Shin-Koiwa have less number of restaurants. So Neighborhoods with fewer restaurants are the best place to start new food places or restaurants as there is less competition among restaurants to draw customers towards them and a way to make large profits.

6. Conclusion

In this project, bar charts and Foursquare location api are used to distinguish neighborhoods with similar characteristics. By that way, we find neighborhoods with most common venues and from there, we are able to identify the neighborhoods with less number of food places as they are the best suitable place for starting new restaurants. We mainly did our work in Google colaboratory with the help of pandas, numpy , matplotlib etc libraries. We used folium for our mapping. All these tools are great for doing online projects.