Where should I stay in Tokyo?

A. Introduction

A.1. Description & Disscusion of the Background

I am planning a trip to Tokyo and looking for a neighborhood to stay. I hope it would be easy to find local food around, have many café and music clubs where I can linger around in my free time, and have plenty of shops where I can find interesting gifts for my friends.

I read an article about 11 neighborhoods to visit for the first-time traveler and want to choose a neighborhood among them. The neighborhoods are Shinjuku, Shibuya, Akihabara, Harajuku, Ginza, Ueno, Shimokitazawa, Kenji, Nakameguro, Roppongi, and Odaiba. Each is described exciting and attractive to me.

Since the popular kinds of venues in a neighborhood present its style, I will use Foursquare to find out the most popular venues in these 11 neighborhoods and cluster them to see which one most fits my style.

A.2. Data Description

To consider the problem we can list the datas as below:

- I read an travelling article recommending 11 neighborhoods to visit for the first-time traveler on Elite and used python **geopy** library to get the location of the neighborhoods.
- I used **Foursquare API** to get the most common venues in the 11 neighborhoods.

B. Methodology

As a database, I used GitHub repository in my study. My master data which has the main components *Neighborhood, Latitude* and *Longitude* informations of Tokyo.

	Neighborhood	Latitude	Longitude
0	Shinjuku	35.693763	139.703632
1	Shibuya	35.664596	139.698711
2	Akihabara	35.699736	139.771250
3	Harajuku	35.668705	139.705336
4	Ginza	35.669516	139.764306
5	Ueno	35.711759	139.777645
6	Shimokitazawa	35.659756	139.668811
7	Kenji	48.105209	-1.678935
8	Nakameguro	35.644140	139.698832
9	Roppongi	35.662819	139.731338
10	Odaiba	35.619128	139.779403

I used python **folium** library to visualize geographic details of Tokyo and its neighborhoods and I created a map of Tokyo with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:

I utilized the Foursquare API to explore the neighborhoods and segment them. I designed the limit as **100 venue** and the radius **500 meter** for each neighborhood from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

The total number of venues are 990

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Shinjuku	35.693763	139.703632	Thermae Yu (テルマー湯)	35.694526	139.705139	Spa
1	Shinjuku	35.693763	139.703632	Michishirube (道しるべ)	35.694744	139.703705	Yakitori Restaurant
2	Shinjuku	35.693763	139.703632	Rock Bar Mother	35.694661	139.702795	Bar
3	Shinjuku	35.693763	139.703632	Metal Bar GODZ	35.694921	139.702918	Bar
4	Shinjuku	35.693763	139.703632	Acacia (アカシア)	35.693116	139.701570	Yoshoku Restaurant

In the table below, we can see all the neighborhoods except for Kenji and Odaiba reached the limit of 100 venues. The result does not mean that inquiry run all the possible results in boroughs. We can improve the results by running with more pairs of latitude and longitude.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Akihabara	100	100	100	100	100	100
Ginza	100	100	100	100	100	100
Harajuku	100	100	100	100	100	100
Kenji	53	53	53	53	53	53
Nakameguro	100	100	100	100	100	100
Odaiba	37	37	37	37	37	37
Roppongi	100	100	100	100	100	100
Shibuya	100	100	100	100	100	100
Shimokitazawa	100	100	100	100	100	100
Shinjuku	100	100	100	100	100	100
Ueno	100	100	100	100	100	100

In summary of this graph **183** unique categories were returned by Foursquare, then I created a table which shows list of top 5 venue category for each neighborhood in below table.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Akihabara	Electronics Store	Hobby Shop	Café	Ramen Restaurant	Arcade
1	Ginza	Coffee Shop	Café	Japanese Restaurant	Boutique	Sushi Restaurant
2	Harajuku	Café	Clothing Store	Boutique	Coffee Shop	Thai Restaurant
3	Kenji	Hotel	French Restaurant	Creperie	Sushi Restaurant	Cupcake Shop
4	Nakameguro	Japanese Restaurant	BBQ Joint	Italian Restaurant	Coffee Shop	Sushi Restaurant
5	Odaiba	Convenience Store	Bus Stop	Intersection	Japanese Restaurant	Hot Spring
6	Roppongi	Japanese Restaurant	Steakhouse	Ramen Restaurant	Chinese Restaurant	Art Gallery
7	Shibuya	Café	Record Shop	Ramen Restaurant	Nightclub	Rock Club
8	Shimokitazawa	Café	Sake Bar	Coffee Shop	Bar	Ramen Restaurant
9	Shinjuku	Bar	Sake Bar	Ramen Restaurant	BBQ Joint	Record Shop
10	Ueno	Sake Bar	Japanese Restaurant	Ramen Restaurant	Chinese Restaurant	Bed & Breakfast

We have some common venue categories in neighborhoods. Therefore, I used unsupervised learning **K-means algorithm** to cluster the neighborhoods. K-Means algorithm is one of the most common cluster method of unsupervised learning.

I will run K-Means to cluster the neighborhoods into **3** clusters.

Here is my merged table with cluster labels for each neighborhood.

|	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
0	Shinjuku	35.693763	139.703632	Bar	Sake Bar	Ramen Restaurant	BBQ Joint	Record Shop	0
1	Shibuya	35.664596	139.698711	Café	Record Shop	Ramen Restaurant	Nightclub	Rock Club	1
2	Akihabara	35.699736	139.771250	Electronics Store	Hobby Shop	Café	Ramen Restaurant	Arcade	1
3	Harajuku	35.668705	139.705336	Café	Clothing Store	Boutique	Coffee Shop	Thai Restaurant	1
4	Ginza	35.669516	139.764306	Coffee Shop	Café	Japanese Restaurant	Boutique	Sushi Restaurant	1

C. Results

Let's explore each cluster:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
0	Shinjuku	Bar	Sake Bar	Ramen Restaurant	BBQ Joint	Record Shop	0
5	Ueno	Sake Bar	Japanese Restaurant	Ramen Restaurant	Chinese Restaurant	Bed & Breakfast	0
8	Nakameguro	Japanese Restaurant	BBQ Joint	Italian Restaurant	Coffee Shop	Sushi Restaurant	0
9	Roppongi	Japanese Restaurant	Steakhouse	Ramen Restaurant	Chinese Restaurant	Art Gallery	0
10	Odaiba	Convenience Store	Bus Stop	Intersection	Japanese Restaurant	Hot Spring	0

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
1	Shibuya	Café	Record Shop	Ramen Restaurant	Nightclub	Rock Club	1
2	Akihabara	Electronics Store	Hobby Shop	Café	Ramen Restaurant	Arcade	1
3	Harajuku	Café	Clothing Store	Boutique	Coffee Shop	Thai Restaurant	1
4	Ginza	Coffee Shop	Café	Japanese Restaurant	Boutique	Sushi Restaurant	1
6	Shimokitazawa	Café	Sake Bar	Coffee Shop	Bar	Ramen Restaurant	1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
7	Kenji	Hotel	French Restaurant	Creperie	Sushi Restaurant	Cupcake Shop	2

So the first cluster is about restaurants and bars, while the second cluster has many coffee shops, restaurants, and stores where I can find gifts, and the third cluster has many hotels and restaurants.
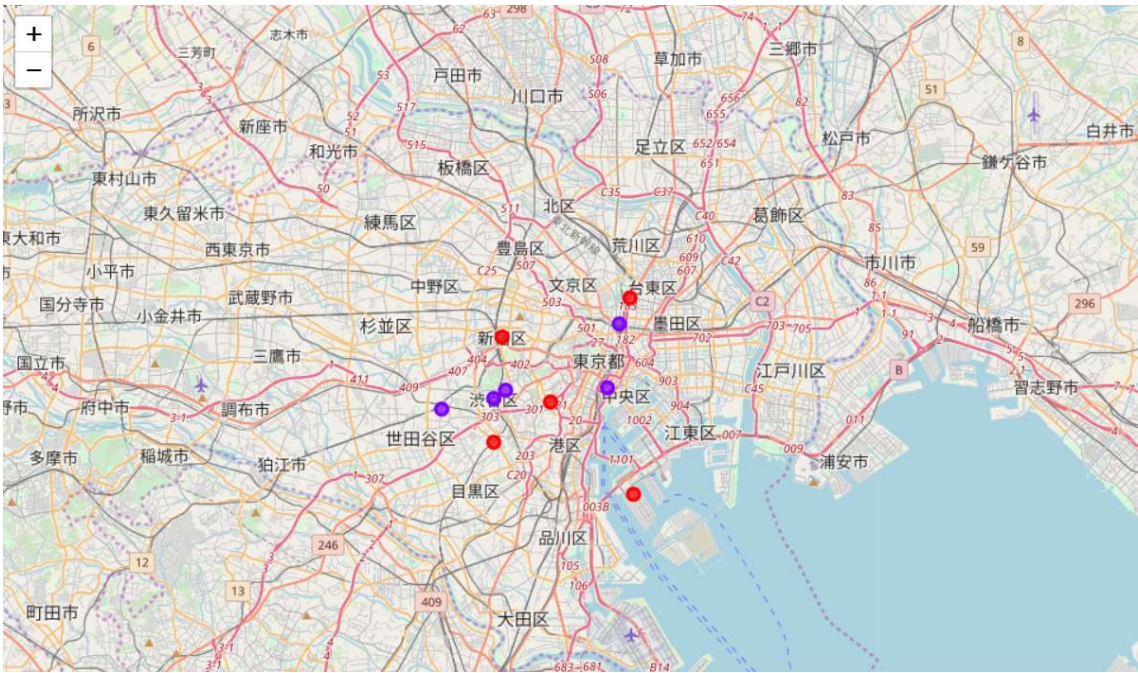
D. Discussion

As I mentioned before, I want to look for a neighborhood that is easy to find food and coffee shop around, and has plenty of interesting shops to buy gifts for my friends. However, Tokyo is a big city and its neighborhoods vary in styles. Since I believe that a neighborhood's style is presented by its most common venues, I gathered the venue data of the neighborhoods in Tokyo and clustered them to see which best fits my style. The venues are limited in the radius of 500m, my desirable walking distance, around the latitude – longitude of each neighborhood.

I used the Kmeans algorithm as part of this clustering study. From the study's result, neighborhoods in cluster 2, including Shibuya, Akihabara, Harajuku, Ginza, and Shimokitazawa, seem to fit my style the most. For more detailed and accurate results, the data set can be expanded and the details of the neighborhood or street can also be drilled.

I ended the study by visualizing the data and clustering information on the Tokyo map. In future studies, I can add information of attractions and room rates in these 5 neighborhoods to find the best neighborhood to stay during my trip.

F. Conclusion

Although this study is mainly for my interest, it also provides useful information for first-time travelers planning a trip to Tokyo. With more information on attractions and costs for rooms and food, we can mitigate lots of planning time for travelers.