The Battle of Tokyo Wards

Planning Food Tours in Tokyo Using K-Means Clustering Algorithm

Chonlawit Sirikupt

The link to my notebook and the code:

https://github.com/chonlawitsirikupt/Coursera_Capstone/blob/main/Battle%20of%20Neighborhoods%20(Week%202).ipynb

For IBM Data Science Professional Certificate

Part of the "Battle of Neighborhoods" Capstone Project

February 2021

Table of Contents

Introduction: The Business Problem

As the world's largest and most populated metropolis, Tokyo has a lot to offer in a myriad number of ways. One of the most exciting aspects is the dining experience. Tokyo features a full spectrum of both local and regional Japanese cuisine in addition to all types of international fare. From cheap hole-in-the-wall joints in the alleyways to expensive high-class restaurants on Roppongi Hills, delicious food can be found in every corner of the city with virtually every budget in between.

A fictitious start-up tour company called ***2 Rice 1 Sake*** reached out to me last week about an idea to launch an exciting travel concept organized around "food tours." The client plans to attract young and older foreign tourists who want to go on a fun food escapade in and around Tokyo. The firm was recently established and has few contacts on the ground. ***2 Rice 1 Sake*** has requested an initial exploratory study on Tokyo's different wards and their culinary landscape so that it can choose where to focus their food tours in in the later stages of project formulation. Specifically, the client wants to obtain a broad picture of the kinds of restaurants that are popular and most frequented in different wards. The client is not interested in other recreational venues like parks, game centers, or sports facilities.

To fulfill the client's demands, I leverage Foursquare location data and deploy K-Means clustering method to group Tokyo's 23 districts into their categories based on their restaurant venues information. K-Means clustering is suitable for this project because it takes unlabeled data and categorizes them based on similar features that provide key insights into underlying patterns about restaurant venues in Tokyo. To provide an additional depth to the picture that would contribute to the client's planning process, I also utilize K-Means clustering to analyze an publicly available dataset on celebrity visits to Tokyo with the aim of grouping the districts into categories based on restaurant venues where prominent celebrities such as Tyler the Creator and Anthony Bourdain have visited.

Data Requirements

- **List of Tokyo wards (districts) including their coordinates (longitude and latitude)**
  Ward data will be from Wikipedia to produce a DataFrame that shows key details about each of the 23 wards that exists in Tokyo. Using the Pandas library, the coordinates of the 23 major wards will be obtained from the *geocoder class* of Geopy client.

  Source: https://en.wikipedia.org/wiki/Special_wards_of_Tokyo#List_of_special_wards

- **Popular restaurant categories in each ward**
  The Foursquare API key will be used to access all possible venues located in each ward. Only restaurant venues will be filtered and analyzed given that our client prioritizes food tours above all else. We will then use K-Means clustering to cluster them into different groups for analysis.

- **Prominent celebrities and locations of their dine-outs in Tokyo**
  After the culinary features of each ward has been determined, this dataset will be imported from Kaggle and the exploratory analysis will be replicated. The dataset, which contain details about which celebrities visited Tokyo and the locations they had dined out during their stay, will be merged with the previous DataFrame. We will filter out the celebrities, the restaurants they dined at, and the locations of the latter.
  Source https://www.kaggle.com/alnguyen22/celebrities-in-tokyo

Structure of the Exploratory Study

This exploratory study consists of two parts.

Part 1 maps out the types of popular restaurant venues in Tokyo after leveraging data on Tokyo wards from Wikipedia and geocoder class of the Geopy client, as well as Foursquare API. After the initial exploratory analysis, I adopt K-means clustering as a method for grouping types of popular restaurants into their respective categories. The result is a greater understanding of the primary clusters of restaurant venues in Tokyo that are most frequented. I provide observations and show how the clusters would help the client focus on a specific niche when going forward to formulating a marketing strategy.

Part 2 replicates the same methodology in Part 1 with the data set on celebrity visits to Tokyo obtained from Kaggle. The result of this secondary analysis complements the findings in Part 1 by providing the client with a greater understanding of the popularity of some Tokyo wards that managed to attract prominent celebrities. I reflect on the findings and show why the client might want to select areas or restaurant venues where celebrities have visited when going forward to formulate his or her marketing strategy.

The study ends with final ruminations about the exploratory study and advances a set of recommendations for how the client should plan his or her food tours in Tokyo.

Methodology

Part 1: Mapping Popular Restaurant Venues in Tokyo

I: Data Importation and Wrangling

In Japan, cities are administratively subdivided into "wards," which are local entities directly controlled by the municipal government. They handle administrative functions such as registration, health insurance, and property taxation. For Tokyo, the metropolis is subdivided into "special wards," which are city-level wards with municipal autonomy largely comparable to other forms of municipalities. For this exploratory study, we focus on Tokyo, which has a total of 23 special wards. We begin by creating a function that scrapes the names of the special wards from the listed table that is available on Wikipedia and subsequently import them into a DataFrame, which will be cleaned and prepared for exploratory analysis.

```
import pandas as pd
import numpy as np
import wikipedia as wp
import requests
```

```
df_tokyowards = pd.read_html('https://en.wikipedia.org/wiki/Special_wards_of_Tokyo#List_of_special_wards')[3]
df_tokyowards
```

```
clean_tokyo_wards.drop(['Flag', 'Major districts'], axis=1, inplace=True)
```

```
clean_tokyo_wards.head(10)
```

	No.	Name	Kanji	Population(as of October 2016	Density(/km2)	Area(km2)
0	01	Chiyoda	千代田区	59441	5100	11.66
2	03	Minato	港区	248071	12180	20.37
3	04	Shinjuku	新宿区	339211	18620	18.22
4	05	Bunkyō	文京区	223389	19790	11.29
5	06	Taitō	台東区	200486	19830	10.11
6	07	Sumida	墨田区	260358	18910	13.77
7	08	Kōtō	江東区	502579	12510	40.16
8	09	Shinagawa	品川区	392492	17180	22.84
9	10	Meguro	目黒区	280283	19110	14.67
10	11	Ōta	大田区	722608	11910	60.66

```
clean_tokyo_wards.shape
```

```
(22, 6)
```

II: Obtaining the Longitude and Latitude of Each Ward

Once we have imported the wards into a Pandas DataFrame and cleaned them, we proceed to locate the coordinates of each ward. I use the geocoder class from the Geopy client to extract the coordinates of each special ward in Tokyo. The longitude and latitude values of each ward will be appended to the DataFrame we just created. We will use these new columns to help plot the locations on the map and locate nearby values.

```
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent = "Tokyo_explorer")
```

```
clean_tokyo_wards['city_coord'] = clean_tokyo_wards['Name'].apply(geolocator.geocode).apply(lambda x: (x.latitude, x.longitude))
clean_tokyo_wards
```

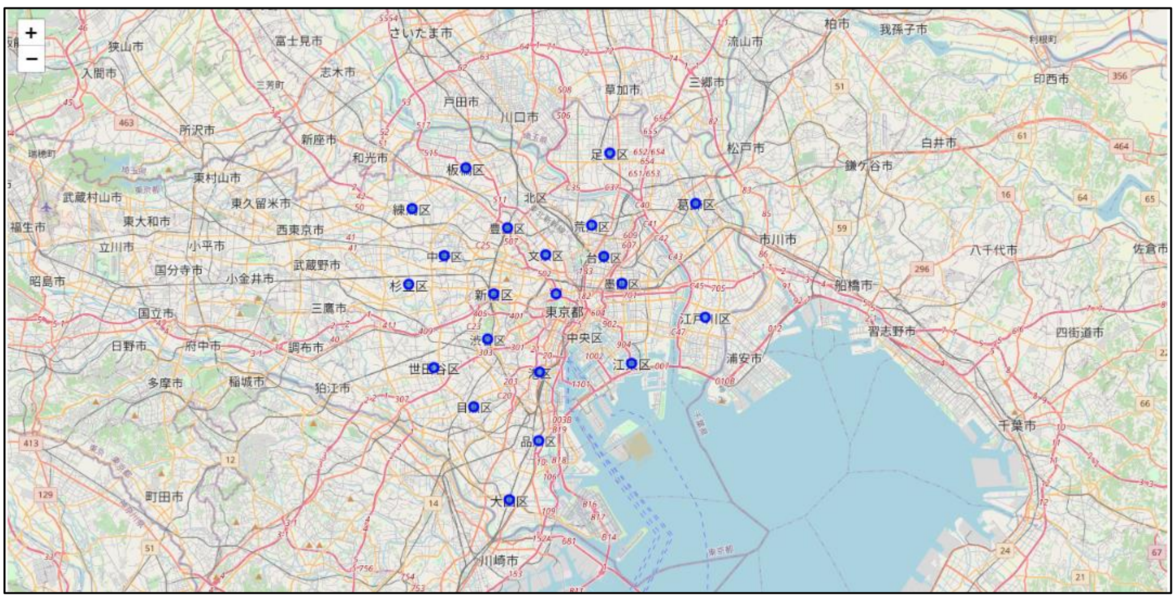
	No.	Ward	Kanji	Population(as of October 2016	Density(/km2)	Area(km2)	Latitude	Longitude
0	01	Chiyoda	千代田区	59441	5100	11.66	35.693810	139.753216
2	03	Minato	港区	248071	12180	20.37	35.643227	139.740055
3	04	Shinjuku	新宿区	339211	18620	18.22	35.693763	139.703632
4	05	Bunkyō	文京区	223389	19790	11.29	35.718810	139.744732
5	06	Taitō	台東区	200486	19830	10.11	35.717450	139.790859
6	07	Sumida	墨田区	260358	18910	13.77	35.700429	139.805017
7	08	Kōtō	江東区	502579	12510	40.16	35.649154	139.812790
8	09	Shinagawa	品川区	392492	17180	22.84	35.599252	139.738910
9	10	Meguro	目黒区	280283	19110	14.67	35.621250	139.688014
10	11	Ōta	大田区	722608	11910	60.66	35.561206	139.715843
11	12	Setagaya	世田谷区	910868	15690	58.05	35.646096	139.656270
12	13	Shibuya	渋谷区	227850	15080	15.11	35.664596	139.698711
13	14	Nakano	中野区	332902	21350	15.59	35.718123	139.664468
14	15	Suginami	杉並区	570483	16750	34.06	35.699493	139.636288
15	16	Toshima	豊島区	294673	22650	13.01	35.736156	139.714222
16	17	Kita	北区	345063	16740	20.61	-0.220164	-78.512327
17	18	Arakawa	荒川区	213648	21030	10.16	35.737529	139.781310
18	19	Itabashi	板橋区	569225	17670	32.22	35.774143	139.681209
19	20	Nerima	練馬区	726748	15120	48.08	35.748360	139.638735
20	21	Adachi	足立区	674067	12660	53.25	35.783703	139.795319
21	22	Katsushika	葛飾区	447140	12850	34.80	35.751733	139.863816
22	23	Edogawa	江戸川区	685899	13750	49.90	35.678278	139.871091

III: Explore Tokyo Wards

Now that we have the coordinates for each ward, we will generate visualizations of different Tokyo wards according to their values using Folium, which is a library used for visualizing geospatial data. Next, using Foursquare API, we will first examine Setagaya Ward as a vector for getting a glimpse into the types of venues in the Tokyo before going on to obtain nearby venues in all of the city's wards. Finally, because our client has prioritized restaurants over other venues, I filter out restaurant only venues and examine the extent to which they cluster together.

```
for lat, lng, label in zip(clean_tokyo_wards['Latitude'], clean_tokyo_wards['Longitude'], clean_tokyo_wards['Ward']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_tokyo)

map_tokyo
```



After visualizing the different special wards in Tokyo, we briefly explore the types of venues at are in Setagaya, the largest of the 23 wards. Known for its reputation as an upscale residential district that features a mix of greenery and funk shopping and nightlife areas, Setagaya is highly popular among students and young adults. We begin by calling the Foursquare API key that allows us to access venues data. Using the same DataFrame, we then filter out Setagaya and extract the first 100 venues that are in the ward within a radius of 500 meters.

Shimokitazawa, Setagaya Ward

```
clean_tokyo_wards.loc[11, 'Ward']
```

```
'Setagaya'
```

```
ward_latitude = clean_tokyo_wards.loc[11, 'Latitude']
ward_longitude = clean_tokyo_wards.loc[11, 'Longitude']

ward_name = clean_tokyo_wards.loc[11, 'Ward']

print('Latitude and longitude values of {} are {}, {}.'.format(ward_name,
                                                              ward_latitude,
                                                              ward_longitude))
```

```
Latitude and longitude values of Setagaya are 35.646096, 139.65627.
```

Get the Top 100 Venues that are in Setagaya Ward Within a Radius of 500 Meters

```
LIMIT = 100
radius = 500
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    ward_latitude,
    ward_longitude,
    radius,
    LIMIT)
url
```

After obtaining the top 100 venues, we then create a function that extract the categories of venues in Setagaya. From this initial exploratory analysis of Setagaya, we see that Foursquare provides useful glimpses into the ward's diverse culinary landscape that help explain its reputation as a popular spot for students and young adults.

```
# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']
```

```
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues)

filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

```
print (nearby_venues['categories'].value_counts()[0:20])
```

Convenience Store	7
Café	3
Bakery	2
Tram Station	2
Ramen Restaurant	2
Tea Room	1
Supermarket	1
Intersection	1
Beer Bar	1
Park	1
Liquor Store	1
Unagi Restaurant	1
History Museum	1
Drugstore	1
Japanese Family Restaurant	1
Yoshoku Restaurant	1
Pizza Place	1
Used Bookstore	1
Fast Food Restaurant	1
Sake Bar	1

Name: categories, dtype: int64

After our surface tour of Setagaya, we will extend our exploration to all Tokyo wards following the same steps that were conducted in the previous section. Again, create a function that extract the categories of nearby venues in all Tokyo wards. We also create a new DataFrame and filter out only restaurant venues.

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        results = requests.get(url).json()["response"]['groups'][0]['items']

        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

```
tokyo_venues = getNearbyVenues(names=clean_tokyo_wards['Ward'],
                                   latitudes=clean_tokyo_wards['Latitude'],
                                   longitudes=clean_tokyo_wards['Longitude']
```

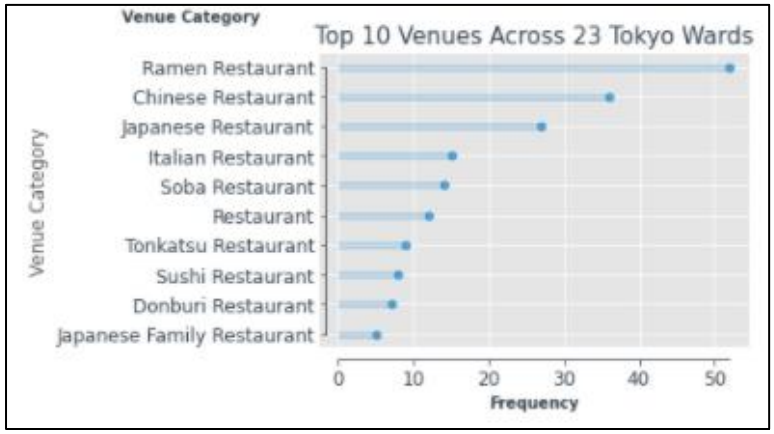
```
tokyo_venues_restaurant = tokyo_venues[tokyo_venues['Venue Category'].str.contains('Restaurant')].reset_index(drop=True)

tokyo_venues_restaurant.index = np.arange(1, len(tokyo_venues_restaurant )+1)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Chiyoda	35.69381	139.753216	Jimbocho Kurosu (神保町 黒須)	35.695539	139.754851	Ramen Restaurant
2	Chiyoda	35.69381	139.753216	Kanda Tendonya (神田天丼家)	35.695765	139.754682	Tempura Restaurant
3	Chiyoda	35.69381	139.753216	Sushi Masa (九段下 寿司政)	35.695234	139.752227	Sushi Restaurant
4	Chiyoda	35.69381	139.753216	Bondy (欧風カレー ボンディ)	35.695544	139.757356	Japanese Curry Restaurant
5	Chiyoda	35.69381	139.753216	たいよう軒	35.696454	139.754809	Chinese Restaurant

After creating the new DataFrame, we use Matplotlib to visualize the most frequently visited restaurant category in Tokyo. Our findings so far indicate that Ramen restaurants are the most popular and frequently visited, followed by Chinese restaurants and Japanese restaurants (I assume that these could be Japanese restaurants that do not exclusively offer Ramen but also other non-Ramen dishes).

	Venue_Category	Frequency
0	Ramen Restaurant	52
1	Chinese Restaurant	35
2	Japanese Restaurant	27
3	Soba Restaurant	15
4	Italian Restaurant	14
5	Tonkatsu Restaurant	10
6	Restaurant	10
7	Sushi Restaurant	9
8	French Restaurant	7
9	Donburi Restaurant	7



IV: Cluster the Wards Using K-Means

K-Means clustering is an unsupervised machine learning algorithm which can be deployed to locate clusters of information that share similar characteristics and to classify these groups into special categories. The algorithm works by creating a determined set of clusters from the data points defined by the user. It subsequently iteratively tries to find the optimum centroid to classify the data points into. From our exploratory data analysis of all restaurant venues across Tokyo's 23 special wards, we can potentially spot the formation of some major clusters. We can deploy K-Means clustering to find restaurant categories and wards where some of the popular venues are concentrated. This would provide helpful guidance for the client when making decisions about selecting restaurant categories that fit the concept of Tokyo food tours. We first pre-process the data using "One Hot Encoding," which creates a binary column for each category and returns a sparse matrix or dense array. After we find the mean value for the frequency of visits to a particular venue, we then create new DataFrame and display the top 10 venues for each ward.

```
# one hot encoding
tokyo_onehot = pd.get_dummies(tokyo_venues_restaurant[['Venue Category']], prefix="", prefix_sep="")

# add neighbourhood column back to dataframe
tokyo_onehot['Neighborhood'] = tokyo_venues_restaurant['Neighborhood']

# move neighbourhood column to the first column
fixed_columns = [tokyo_onehot.columns[-1]] + list(tokyo_onehot.columns[:-1])
tokyo_onehot = tokyo_onehot[fixed_columns]

tokyo_onehot.head()
```

	Neighborhood	Asian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	Fast Food Restaurant	French Restaurant	...	Teishoku Restaurant	Tempura Restaurant	Thai Restaurant	Tonkatsu Restaurant	Udon Restaurant	Unagi Restaurant	Vegetarian / Vegan Restaurant
0	Adachi	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Arakawa	0.000000	0.000000	0.000000	0.250000	0.125000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Bunkyō	0.000000	0.000000	0.000000	0.333333	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Chiyoda	0.000000	0.000000	0.000000	0.121212	0.000000	0.0000	0.000000	0.000	0.090909	...	0.000	0.030303	0.030303	0.060606	0.000000	0.000000	0.000000
4	Edogawa	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Itabashi	0.000000	0.000000	0.000000	0.500000	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Katsushika	0.000000	0.000000	0.000000	0.000000	0.285714	0.0000	0.142857	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Kita	0.000000	0.000000	0.076923	0.000000	0.000000	0.0000	0.000000	0.000	0.076923	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.076923
8	Kōtō	0.000000	0.000000	0.000000	0.666667	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	Meguro	0.000000	0.000000	0.000000	0.200000	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	Minato	0.000000	0.000000	0.000000	0.100000	0.000000	0.0000	0.000000	0.000	0.100000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	Nakano	0.000000	0.000000	0.000000	0.125000	0.125000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.125000	0.000000	0.000000	0.000000
12	Nerima	0.000000	0.000000	0.000000	1.000000	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	Setagaya	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.000000	0.125	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.125000	0.000000
14	Shibuya	0.047619	0.047619	0.000000	0.142857	0.047619	0.0000	0.000000	0.000	0.095238	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	Shinagawa	0.000000	0.000000	0.000000	0.285714	0.142857	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	Shinjuku	0.000000	0.030303	0.000000	0.090909	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.060606	0.060606	0.030303	0.030303	0.000000
17	Suginami	0.000000	0.000000	0.000000	0.181818	0.000000	0.0000	0.090909	0.000	0.000000	...	0.000	0.000000	0.000000	0.090909	0.000000	0.000000	0.000000
18	Sumida	0.000000	0.000000	0.000000	0.222222	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.111111	0.000000	0.111111	0.000000
19	Taitō	0.000000	0.000000	0.000000	0.076923	0.000000	0.0000	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.000000	0.038462	0.000000
20	Toshima	0.000000	0.000000	0.000000	0.125000	0.062500	0.0625	0.000000	0.000	0.000000	...	0.000	0.000000	0.000000	0.000000	0.062500	0.000000	0.000000
21	Ōta	0.000000	0.000000	0.000000	0.100000	0.000000	0.0000	0.050000	0.000	0.000000	...	0.025	0.025000	0.000000	0.075000	0.050000	0.025000	0.000000

22 rows × 49 columns

```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]

num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Restaurant'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Restaurant'.format(ind+1))

# create a new dataframe
districts_venues_sorted = pd.DataFrame(columns=columns)
districts_venues_sorted['Neighborhood'] = tokyo_grouped['Neighborhood']

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

districts_venues_sorted
```

```
districts_venues_sorted.rename(columns={"Neighborhood": "Ward"}, inplace=True)
districts_venues_sorted.head()
```

	ClusterLabel	Ward	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
0	0	Adachi	Japanese Restaurant	Japanese Family Restaurant	Asian Restaurant	Sushi Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant
1	1	Arakawa	Ramen Restaurant	Chinese Restaurant	Donburi Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Taiwanese Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant
2	2	Bunkyo	Italian Restaurant	Chinese Restaurant	Szechuan Restaurant	Taiwanese Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant
3	1	Chiyoda	Ramen Restaurant	Chinese Restaurant	Japanese Curry Restaurant	French Restaurant	Restaurant	Indian Restaurant	Soba Restaurant	Japanese Restaurant	Tonkatsu Restaurant	Italian Restaurant
4	1	Edogawa	Ramen Restaurant	Italian Restaurant	Yakitori Restaurant	Vietnamese Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant

After merging this Dataframe with the previously cleaned Dataframe, the K-value to 3, which means we want to group our data points (popular restaurant categories) into 3 major clusters, we get the following results:

	No.	Ward	Kanji	Population(as of October 2016	Density(/km2)	Area(km2)	Latitude	Longitude	ClusterLabel	1st Most Common Restaurant	...	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
0	01	Chiyoda	千代田区	59441	5100	11.66	35.693810	139.753216	1	Ramen Restaurant	...	Japanese Curry Restaurant	French Restaurant	Restaurant	Indian Restaurant	Soba Restaurant	Japanese Restaurant	Tonkatsu Restaurant	Italian Restaurant
2	03	Minato	港区	248071	12180	20.37	35.643227	139.740055	1	Soba Restaurant	...	Yakitori Restaurant	Chinese Restaurant	French Restaurant	Kosher Restaurant	Indian Restaurant	Kebab Restaurant	Japanese Restaurant	Asian Restaurant
3	04	Shinjuku	新宿区	339211	18620	18.22	35.693763	139.703632	1	Ramen Restaurant	...	Chinese Restaurant	Shabu-Shabu Restaurant	Kushikatsu Restaurant	Tonkatsu Restaurant	Thai Restaurant	Russian Restaurant	Seafood Restaurant	Brazilian Restaurant
4	05	Bunkyo	文京区	223389	19790	11.29	35.718810	139.744732	2	Italian Restaurant	...	Szechuan Restaurant	Taiwanese Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant
5	06	Taito	台東区	200486	19830	10.11	35.717450	139.790859	1	Ramen Restaurant	...	Sushi Restaurant	Italian Restaurant	Nabe Restaurant	Chinese Restaurant	Soba Restaurant	Sukiyaki Restaurant	South Indian Restaurant	Korean BBQ Restaurant

5 rows × 21 columns

Part 1 Results and Discussion

My findings after conducting K-Means clustering show that the three main clusters of the most frequented restaurant venues are 1) Ramen restaurants 2) international restaurants and 3) Japanese restaurants. Among international restaurants, Chinese restaurants appear to be the most frequented.

Cluster 1: Japanese Restaurants

	Ward	Area(km2)	Latitude	Longitude	ClusterLabel	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
6	Sumida	13.77	35.700429	139.805017	0	Japanese Restaurant	Chinese Restaurant	Ramen Restaurant	Sushi Restaurant	Unagi Restaurant	Tonkatsu Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant
8	Shinagawa	22.84	35.599252	139.738910	0	Chinese Restaurant	Japanese Family Restaurant	Donburi Restaurant	Soba Restaurant	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant
9	Meguro	14.67	35.621250	139.688014	0	Japanese Restaurant	Italian Restaurant	Chinese Restaurant	Sushi Restaurant	Taiwanese Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant
20	Adachi	53.25	35.783703	139.795319	0	Japanese Restaurant	Japanese Family Restaurant	Asian Restaurant	Sushi Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant

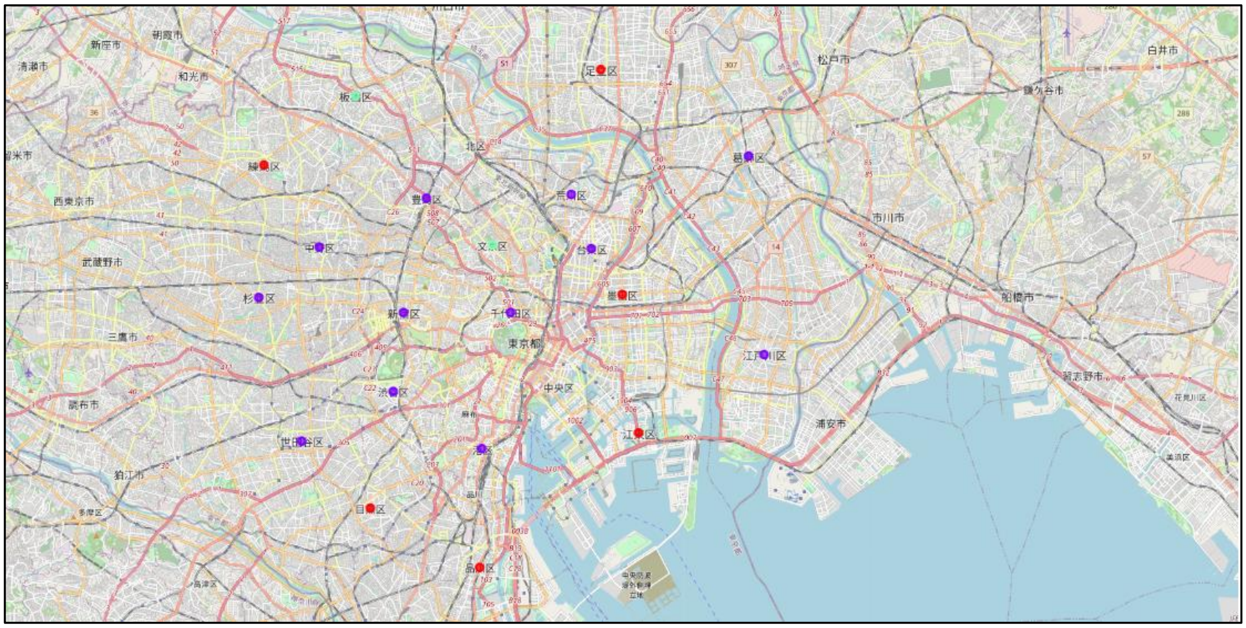
Cluster 2: Ramen Restaurants

	Ward	Area(km2)	Latitude	Longitude	ClusterLabel	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
0	Chiyoda	11.66	35.693810	139.753216	1	Ramen Restaurant	Chinese Restaurant	Japanese Curry Restaurant	French Restaurant	Restaurant	Indian Restaurant	Soba Restaurant	Japanese Restaurant	Tonkatsu Restaurant	Italian Restaurant
2	Minato	20.37	35.643227	139.740055	1	Soba Restaurant	Kaiseki Restaurant	Yakitori Restaurant	Chinese Restaurant	French Restaurant	Kosher Restaurant	Indian Restaurant	Kebab Restaurant	Japanese Restaurant	Asian Restaurant
3	Shinjuku	18.22	35.693763	139.703632	1	Ramen Restaurant	Japanese Restaurant	Chinese Restaurant	Shabu-Shabu Restaurant	Kushikatsu Restaurant	Tonkatsu Restaurant	Thai Restaurant	Russian Restaurant	Seafood Restaurant	Brazilian Restaurant
5	Taito	10.11	35.717450	139.790859	1	Ramen Restaurant	Japanese Restaurant	Sushi Restaurant	Italian Restaurant	Nabe Restaurant	Chinese Restaurant	Soba Restaurant	Sukiyaki Restaurant	South Indian Restaurant	Korean BBQ Restaurant
10	Ōta	60.66	35.561206	139.715843	1	Ramen Restaurant	Chinese Restaurant	Japanese Restaurant	Tonkatsu Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Udon Restaurant	Italian Restaurant	Korean Restaurant	Taiwanese Restaurant
11	Setagaya	58.05	35.646096	139.656270	1	Ramen Restaurant	Yoshoku Restaurant	Japanese Restaurant	Unagi Restaurant	Fast Food Restaurant	Szechuan Restaurant	Japanese Family Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant
12	Shibuya	15.11	35.664596	139.698711	1	Ramen Restaurant	Chinese Restaurant	French Restaurant	Japanese Restaurant	Asian Restaurant	Italian Restaurant	Nabe Restaurant	Brazilian Restaurant	Mexican Restaurant	South Indian Restaurant
13	Nakano	15.59	35.718123	139.664468	1	Ramen Restaurant	Chinese Restaurant	Donburi Restaurant	Soba Restaurant	Tonkatsu Restaurant	Indian Restaurant	Italian Restaurant	Asian Restaurant	Szechuan Restaurant	Russian Restaurant
14	Suginami	34.06	35.699493	139.636288	1	Italian Restaurant	Chinese Restaurant	Soba Restaurant	Ramen Restaurant	Shabu-Shabu Restaurant	Dumpling Restaurant	Tonkatsu Restaurant	Szechuan Restaurant	Russian Restaurant	Seafood Restaurant
15	Toshima	13.01	35.736156	139.714222	1	Ramen Restaurant	Chinese Restaurant	Udon Restaurant	Middle Eastern Restaurant	Soba Restaurant	Korean Restaurant	Japanese Family Restaurant	Yoshoku Restaurant	Donburi Restaurant	Dongbei Restaurant
16	Kita	20.61	-0.220164	-78.512327	1	Restaurant	South American Restaurant	Cajun / Creole Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	French Restaurant	Asian Restaurant	Szechuan Restaurant	Russian Restaurant	Shabu-Shabu Restaurant
17	Arakawa	10.16	35.737529	139.781310	1	Ramen Restaurant	Chinese Restaurant	Donburi Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Taiwanese Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant
21	Katsushika	34.80	35.751733	139.863816	1	Donburi Restaurant	Soba Restaurant	Dumpling Restaurant	Korean Restaurant	Ramen Restaurant	Asian Restaurant	Szechuan Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant
22	Edogawa	49.90	35.678278	139.871091	1	Ramen Restaurant	Italian Restaurant	Yakitori Restaurant	Vietnamese Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant

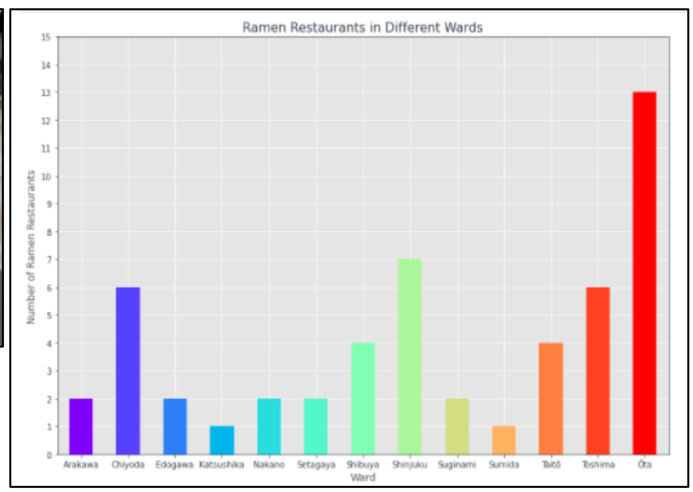
Cluster 3: International Restaurants (Particularly Chinese)

	Ward	Area(km2)	Latitude	Longitude	ClusterLabel	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
7	Koto	40.16	35.649154	139.812790	3	Chinese Restaurant	Indian Restaurant	Asian Restaurant	Szechuan Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant
19	Nerima	48.08	35.748360	139.638735	3	Chinese Restaurant	Asian Restaurant	Szechuan Restaurant	Restaurant	Russian Restaurant	Seafood Restaurant	Shabu-Shabu Restaurant	Soba Restaurant	South American Restaurant	South Indian Restaurant

The visualization of the three main clusters show that Ramen restaurants (Cluster 1; blue) are spread out across the city and are also concentrated in major entertainment and shopping districts such as Shinjuku and Shibuya. Japanese restaurants (Cluster 2; red) that might offer additional dishes beyond Ramen seem to in smaller entertainment districts around the edges of the center that have less foot traffic. Some of these districts are also suburban and residential areas. International restaurants (Cluster 3; light green) are primarily in Itabashi and Bunkyo.



Further analysis of ramen restaurants showed that these venues are mostly concentrated in Ota, followed by Shinjuku, Chiyoda, Toshima, and Shibuya wards.



Since Ramen restaurants form the largest cluster, the client should consider organizing his or her food tours in Tokyo around them. The client might also be interested in selecting a few Japanese restaurants in the second cluster to provide customers with a wider range of local dining experience. However, client might also find it more appealing to base the food tours exclusively on ramen.

Part 2: Mapping Celebrity Visits to Tokyo (and Where They Dined)

I: Data Importation

Tokyo's flare has without a doubt has drawn in significant international attention. From global movie premieres to concerts, the city is no stranger to welcoming a myriad group of renowned celebrities. Films that millennials have grown up with, such as Lost in Translation, Kill Bill, The Last Samurai Resident Evil: After Life, and Inception, also feature scenes that were shot in Tokyo. It has also featured live performances from top-chart artists such as ASAP Rocky and Tyga. Given that celebrities have large individual followings, it might be a useful exercise to consider the restaurants where they had dined during their visits to Tokyo and in which areas these venues are primarily concentrated in. This would provide the client with additional options when deciding on the scope of his or her food tours.

Since the dataset on celebrity visits to Tokyo is available as a CSV file that is accessible via Kaggle, we do not need to perform any data wrangling. We first begin by importing the CSV file and using Pandas to create a DataFrame. After creating the DataFrame, we filter out only restaurant venues since the dataset also contains non-restaurant venues such as bars and parks.

	Celebrity	Type	Establishment	Location	Ward	Venue	Month
14	Camilla Belle	Actor	Afuri Ramen	Harajuku	Shibuya-ku	Restaurant	April
15	Pharrell Williams	Musician	Afuri Ramen	Harajuku	Shibuya-ku	Restaurant	August
22	Heidi Klum	Model	Aoi Marushin	Asakusa	Taito-ku	Restaurant	April
23	Camilla Belle	Actor	Appia Alta	Nishi-Azabu	Minato-ku	Restaurant	April
24	Kate Hudson	Actor	Aronia De Takazawa	Akasaka	Minato-ku	Restaurant	November
...	...	...	...	...	...	...	...
335	Rihanna	Musician	Tsukiji Fish Market	Tsukiji	Chuo-ku	Restaurant	NaN
337	David Beckham	Athlete	Umi	Omotesando	Minato-ku	Restaurant	May
344	Anthony Bourdain	TV Personality	XEX Morimoto	Roppongi	Minato-ku	Restaurant	NaN
347	Aziz Ansari	Actor	Yakumo Saryo	Meguro	Meguro-ku	Restaurant	May
354	Shawn Mendez	Musician	zero Tokyo	Ginza	Chuo-ku	Restaurant	February

126 rows × 7 columns

II: Obtaining the Longitude and Latitude of Each District (Within a Ward)

Like in Part 1, we obtain the longitude and latitude values of each location. However, this time we will extract the coordinates of each district within a ward. For example, Shibuya Ward contains well-known commercial and residential districts such as Harajuku, Omotesando, Ebisu, Sendagaya, and Ebisu.

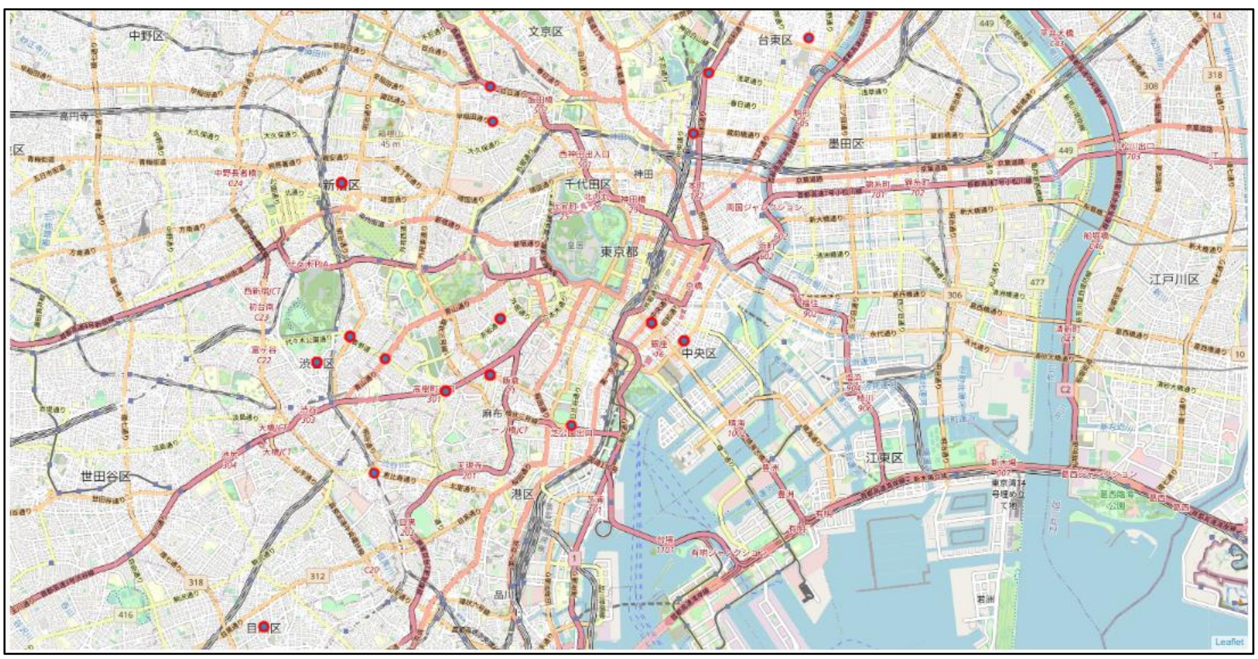
```
df_celeb_restaurant['city_coord'] = df_celeb_restaurant['Location'].apply(geolocator.geocode).apply(lambda x: (x.latitude, x.longitude))
df_celeb_restaurant
```

```
df_celeb_restaurant.drop(['city_coord'], axis=1, inplace=True)
df_celeb_restaurant
```

	Celebrity	Type	Establishment	Location	Ward	Venue	Month	Latitude	Longitude
14	Camilla Belle	Actor	Afuri Ramen	Harajuku	Shibuya-ku	Restaurant	April	35.668705	139.705336
15	Pharrell Williams	Musician	Afuri Ramen	Harajuku	Shibuya-ku	Restaurant	August	35.668705	139.705336
22	Heidi Klum	Model	Aoi Marushin	Asakusa	Taito-ku	Restaurant	April	35.717597	139.797563
23	Camilla Belle	Actor	Appia Alta	Nishi-Azabu	Minato-ku	Restaurant	April	35.659869	139.724688
24	Kate Hudson	Actor	Aronia De Takazawa	Akasaka	Minato-ku	Restaurant	November	35.671679	139.735622
...	...	...	...	...	...	...	...	...	...
335	Rihanna	Musician	Tsukiji Fish Market	Tsukiji	Chuo-ku	Restaurant	NaN	35.668101	139.772583
337	David Beckham	Athlete	Umi	Omotesando	Minato-ku	Restaurant	May	35.665170	139.712435
344	Anthony Bourdain	TV Personality	XEX Morimoto	Roppongi	Minato-ku	Restaurant	NaN	35.662457	139.733498
347	Aziz Ansari	Actor	Yakumo Saryo	Meguro	Meguro-ku	Restaurant	May	35.621250	139.688014
354	Shawn Mendez	Musician	zero Tokyo	Ginza	Chuo-ku	Restaurant	February	35.670910	139.766021

126 rows × 9 columns

After obtaining the coordinates for each district, we then visualize their locations using Folium. We can already some clusters forming, particularly those more in the major entertainment and shopping districts and more residential areas around the edges of these centers.



We first pre-process the celebrity visits data using One Hot Encoding and then find the mean value for the frequency of visits to each district. Subsequently, we create new DataFrame and display the top 3 venues for each district.

	Location	Ali Larter	Anne Hathaway	Ansel Elgort	Anthony Bourdain	Ariana Grande	Arnold Schwarzenegger	Ashton Kutcher	Aziz Ansari	Barack Obama	...	Shay Mitchell	Steven Spielberg	Steven Tyler	Sting	The Weeknd	Tim Burton	Tom Cruise	Tom Hanks	Tony Hawk	Zedd
14	Harajuku	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
15	Harajuku	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
22	Asakusa	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
23	Nishi-Azabu	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
24	Akasaka	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 80 columns

	Location	Ali Larter	Anne Hathaway	Ansel Elgort	Anthony Bourdain	Ariana Grande	Arnold Schwarzenegger	Ashton Kutcher	Aziz Ansari	Barack Obama	...	Shay Mitchell	Steven Spielberg	Steven Tyler	Sting	The Weeknd	Tim Burton	Tom Cruise	Tom Hanks	Tony Hawk	Zedd
0	Akasaka	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.20	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
1	Akihabara	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	1.0	0.0	0.000000
2	Aoyama	0.000000	0.00	0.000000	0.200000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
3	Asakusa	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
4	Ebisu	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
5	Edogawabashi	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	1.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
6	Ginza	0.000000	0.05	0.000000	0.050000	0.000000	0.05	0.000000	0.100000	0.050000	...	0.000000	0.05	0.000	0.000000	0.000000	0.000000	0.050000	0.0	0.0	0.000000
7	Harajuku	0.045455	0.00	0.000000	0.000000	0.045455	0.00	0.000000	0.000000	0.000000	...	0.045455	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.045455
8	Hatanodai	0.000000	0.00	0.000000	1.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
9	Kagurazaka	0.000000	0.00	0.000000	1.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
10	Kotobuki	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
11	Meguro	0.000000	0.00	0.000000	0.333333	0.000000	0.00	0.000000	0.333333	0.000000	...	0.333333	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
12	Nishi-Azabu	0.000000	0.00	0.000000	0.125000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.125	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
13	Omotesando	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
14	Roppongi	0.000000	0.00	0.000000	0.055556	0.000000	0.00	0.055556	0.111111	0.055556	...	0.000000	0.00	0.000	0.055556	0.000000	0.000000	0.055556	0.0	0.0	0.000000
15	Shibakoen	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
16	Shibuya	0.000000	0.00	0.000000	0.111111	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
17	Shinjuku	0.000000	0.00	0.055556	0.111111	0.000000	0.00	0.000000	0.055556	0.000000	...	0.000000	0.00	0.000	0.000000	0.055556	0.055556	0.000000	0.0	0.0	0.000000
18	Tsukiji	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
19	Ueno	0.000000	0.00	0.000000	1.000000	0.000000	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000

20 rows × 80 columns

	Location	1st Most Common Celebrity	2nd Most Common Celebrity	3rd Most Common Celebrity	4th Most Common Celebrity	5th Most Common Celebrity	6th Most Common Celebrity	7th Most Common Celebrity	8th Most Common Celebrity	9th Most Common Celebrity	10th Most Common Celebrity
0	Akasaka	Pharrell Williams	Kate Hudson	Samuel L. Jackson	Lady GaGa	Steven Spielberg	Lucy Hale	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon
1	Akihabara	Tom Hanks	Ali Larter	Lorde	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale
2	Aoyama	Anthony Bourdain	G-Eazy	Holly Madison	Camilla Belle	Hugh Jackman	Ali Larter	Matt Damon	Nicolas Cage	Nicholas Hoult	Nicholas Cage
3	Asakusa	Heidi Klum	Ali Larter	Julianne Moore	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale
4	Ebisu	Keanu Reeves	Ali Larter	Lorde	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale
5	Edogawabashi	Steven Tyler	Ali Larter	Lorde	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale
6	Ginza	Milla Jovovich	Aziz Ansari	Lucy Hale	Anne Hathaway	Hugh Jackman	Drew Barrymore	David Beckham	Nicholas Cage	Jodie Foster	Shawn Mendez
7	Harajuku	Kim Kardashian	Ali Larter	Chloe Moretz	Shay Mitchell	Selena Gomez	Pharrell Williams	Paul McCartney	Lorde	Kourtney Kardashian	Justin Bieber
8	Hatanodai	Anthony Bourdain	Ali Larter	Lorde	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale
9	Kagurazaka	Anthony Bourdain	Ali Larter	Lorde	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale
10	Kotobuki	Hugh Jackman	Julianne Moore	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale	Lorde
11	Meguro	Shay Mitchell	Anthony Bourdain	Aziz Ansari	Ali Larter	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie
12	Nishi-Azabu	Camilla Belle	Anthony Bourdain	Johnny Depp	Jodie Sweetin	Quentin Tarantino	Steven Tyler	Lady GaGa	Scott Weinger	Ali Larter	Nicholas Hoult
13	Omotesando	David Beckham	Julianne Moore	Paul McCartney	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale
14	Roppongi	Aziz Ansari	Demi Moore	Cameron Diaz	Lady GaGa	Margot Robbie	Matt Damon	Ian Ziering	Nicolas Cage	Red Zeppelin	Drew Barrymore
15	Shibakoen	Camilla Belle	Ali Larter	Pete Wentz	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale
16	Shibuya	David Beckham	Julianne Moore	Anthony Bourdain	Prabal Gurung	Leonardo DiCaprio	Holly Madison	James Hetfield	Charles Bukowski	Milla Jovovich	Matt Damon
17	Shinjuku	Anthony Bourdain	Josh Groban	Scott Weinger	John Mayer	Milla Jovovich	Pete Wentz	Pharrell Williams	Drew Barrymore	Keanu Reeves	Coldplay
18	Tsukiji	Josh Groban	Kristen Stewart	Drew Barrymore	Nicholas Hoult	Rihanna	Jeff Bezos	Ian Ziering	Kourtney Kardashian	Lady GaGa	Leonardo DiCaprio
19	Ueno	Anthony Bourdain	Ali Larter	Lorde	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale

For our analysis, we will only on the 1st most common celebrity (first column) who visited the venues given that many of the values for the 2nd to 10th most common celebrity in every cluster were actually

0 (meaning that they never visited these locations). I set the number of most common celebrity to 10 to see how the data would look like if our dataset on celebrity visits was larger than the current one.

III: Explore and Cluster Districts using K-Means

After merging the DataFrame with the previously cleaned DataFrame and setting the K-value to 3, we get the following results:

	index	Establishment	Location	Ward	Venue	Month	Latitude	Longitude	1st Most Common Celebrity	2nd Most Common Celebrity	3rd Most Common Celebrity	4th Most Common Celebrity	5th Most Common Celebrity	6th Most Common Celebrity	7th Most Common Celebrity	8th Most Common Celebrity	9th Most Common Celebrity	10th Most Common Celebrity	Cluster_Labels
0	14	Afuri Ramen	Harajuku	Shibuya-ku	Restaurant	April	35.668705	139.705336	Kim Kardashian	Ali Larter	Chloe Moretz	Shay Mitchell	Selena Gomez	Pharrell Williams	Paul McCartney	Lorde	Kourtney Kardashian	Justin Bieber	0
1	15	Afuri Ramen	Harajuku	Shibuya-ku	Restaurant	August	35.668705	139.705336	Kim Kardashian	Ali Larter	Chloe Moretz	Shay Mitchell	Selena Gomez	Pharrell Williams	Paul McCartney	Lorde	Kourtney Kardashian	Justin Bieber	0
2	22	Aoi Marushin	Asakusa	Taitō-ku	Restaurant	April	35.717597	139.797563	Heidi Klum	Ali Larter	Julianne Moore	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale	2
3	23	Appia Alta	Nishi-Azabu	Minato-ku	Restaurant	April	35.659869	139.724688	Camilla Belle	Anthony Bourdain	Johnny Depp	Jodie Sweetin	Quentin Tarantino	Steven Tyler	Lady GaGa	Scott Weinger	Ali Larter	Nicholas Hoult	0
4	24	Aronia De Takazawa	Akasaka	Minato-ku	Restaurant	November	35.671679	139.735622	Pharrell Williams	Kate Hudson	Samuel L. Jackson	Lady GaGa	Steven Spielberg	Lucy Hale	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
121	335	Tsukiji Fish Market	Tsukiji	Chuo-ku	Restaurant	NaN	35.668101	139.772583	Josh Groban	Kristen Stewart	Drew Barrymore	Nicholas Hoult	Rihanna	Jeff Bezos	Ian Ziering	Kourtney Kardashian	Lady GaGa	Leonardo DiCaprio	0
122	337	Umi	Omotesando	Minato-ku	Restaurant	May	35.665170	139.712435	David Beckham	Julianne Moore	Paul McCartney	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	Lucy Hale	0
123	344	XEX Morimoto	Roppongi	Minato-ku	Restaurant	NaN	35.662457	139.733498	Aziz Ansari	Demi Moore	Cameron Diaz	Lady GaGa	Margot Robbie	Matt Damon	Ian Ziering	Nicolas Cage	Red Zeppelin	Drew Barrymore	0
124	347	Yakumo Saryo	Meguro	Meguro-ku	Restaurant	May	35.621250	139.688014	Shay Mitchell	Anthony Bourdain	Aziz Ansari	Ali Larter	Nicolas Cage	Nicholas Hoult	Nicholas Cage	Milla Jovovich	Matt Damon	Margot Robbie	0
125	354	zero Tokyo	Ginza	Chuo-ku	Restaurant	February	35.670910	139.766021	Milla Jovovich	Aziz Ansari	Lucy Hale	Anne Hathaway	Hugh Jackman	Drew Barrymore	David Beckham	Nicholas Cage	Jodie Foster	Shawn Mendez	0

126 rows × 19 columns

Part 2 Results and Discussion

Cluster 1: Major Entertainment and Shopping Districts

	Establishment	Location	Ward	Venue	Month	Latitude	Longitude	1st Most Common Celebrity
0	Afuri Ramen	Harajuku	Shibuya-ku	Restaurant	April	35.668705	139.705336	Kim Kardashian
1	Afuri Ramen	Harajuku	Shibuya-ku	Restaurant	August	35.668705	139.705336	Kim Kardashian
3	Appia Alta	Nishi-Azabu	Minato-ku	Restaurant	April	35.659869	139.724688	Camilla Belle
4	Aronia De Takazawa	Akasaka	Minato-ku	Restaurant	November	35.671679	139.735622	Pharrell Williams
5	Birdland	Ginza	Chuo-ku	Restaurant	December	35.670910	139.766021	Milla Jovovich
...	...	...	...	...	...	...	...	...
121	Tsukiji Fish Market	Tsukiji	Chuo-ku	Restaurant	NaN	35.668101	139.772583	Josh Groban
122	Umi	Omotesando	Minato-ku	Restaurant	May	35.665170	139.712435	David Beckham
123	XEX Morimoto	Roppongi	Minato-ku	Restaurant	NaN	35.662457	139.733498	Aziz Ansari
124	Yakumo Saryo	Meguro	Meguro-ku	Restaurant	May	35.621250	139.688014	Shay Mitchell
125	zero Tokyo	Ginza	Chuo-ku	Restaurant	February	35.670910	139.766021	Milla Jovovich

122 rows × 18 columns

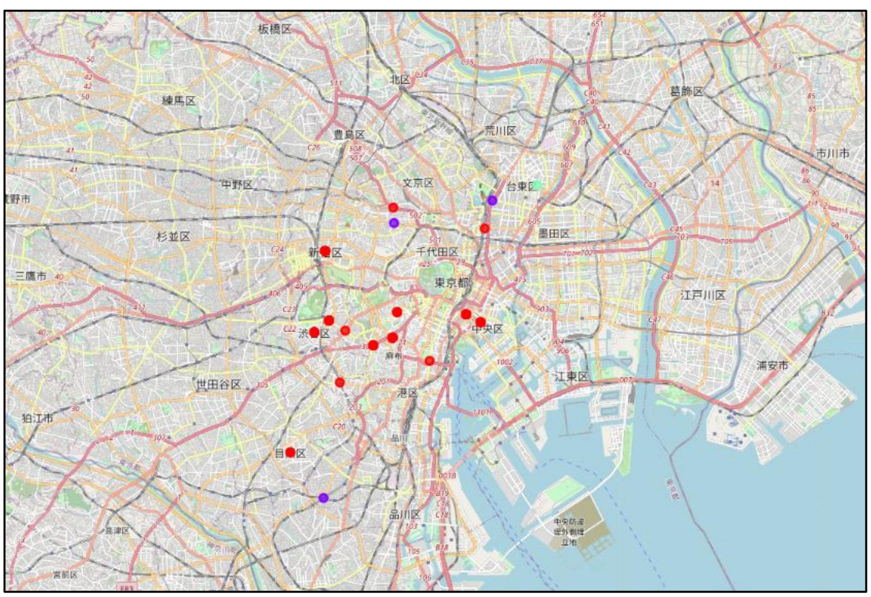
Cluster 2: Smaller Entertainment Districts

	Establishment	Location	Ward	Venue	Month	Latitude	Longitude	1st Most Common Celebrity
7	Chanko Kuroshio	Kagurazaka	Shinjuku-ku	Restaurant	NaN	35.703930	139.734135	Anthony Bourdain
8	Daitoryo	Ueno	Taitō-ku	Restaurant	NaN	35.711821	139.777569	Anthony Bourdain
117	Toriki	Hatanodai	Shinagawa-ku	Restaurant	NaN	35.604876	139.702668	Anthony Bourdain

Cluster 3: Smaller Entertainment Districts

	Establishment	Location	Ward	Venue	Month	Latitude	Longitude	1st Most Common Celebrity
2	Aoi Marushin	Asakusa	Taitō-ku	Restaurant	April	35.717597	139.797563	Heidi Klum

When we visualize the three main clusters, we can see that foreign celebrities tend to visit restaurant venues located in the major entertainment and shopping districts such as Shibuya and Shinjuku (Cluster 1; blue). Cluster 2 (red) and 3 (light green) are similar in a sense that districts such as Ueno, Asakusa, and Hatanodai are smaller entertainment districts. These exceptions are Anthony Bourdain. We might speculate why Bourdain preferred these venues since his TV program focused on exploring local cultures and their best kept secrets.



Asakusa, Taito Ward

Conclusion and Recommendations

This exploratory sought to provide marketing solutions for **2 Rice 1 Sake**, an emerging client in the global tourism industry who plans to launch a novel traveling concept that is built on taking customers on exciting food tours in and around Tokyo, Japan. This study utilized a combination of exploratory data analysis and K-Means clustering methods to analyze data on popular restaurant categories and popular celebrity locations in Tokyo with the aim of into the city's dining-out scene and diverse culinary landscape. The main findings from this study can be summarized into the following points:

- Ramen restaurants form the largest cluster among the most visited restaurant types, followed by Japanese and international restaurants respectively. Among the international restaurants, Chinese restaurants are the most popular.
- Ramen restaurants are mostly concentrated in Ota, followed by Shinjuku, Chiyoda, Toshima, and Shibuya wards.
- Most prominent foreign celebrities tend to visit restaurant venues located in the major entertainment and shopping districts such as Shibuya and Shinjuku. Only a handful visited venues in smaller entertainment districts.

There are some limitations that should be noted in this study. Before discussing the specific limitations intrinsic to first and second part of the study, it is worthwhile to point out the methodological shortcomings of the overall study. The defined number of clusters (K value) for Part 1 and Part 2 was arbitrarily determined, which might explain why the results returned in Part 2 showed two clusters that came to be interpreted as "Smaller Entertainment Districts" instead of just one cluster of the same values. To avoid this kind of arbitrary selection of K value, this study should have transformed the data to fit a standard normal distribution and deployed the "elbow method," which runs the K-Means algorithm for a range of possible K-values to find the optimal number of clusters.

The limitations central to Part 1 and 2 result from the comprehensiveness of the data at hand. First, the Foursquare data should be observed with some caution; it should be pointed out that there may be more Ramen restaurants that may have not been mentioned or "checked-in" by Foursquare users that might lead to an undercounting of some of these restaurants across Tokyo. Second, Foursquare did not have data on each restaurant's menus, which limits our understanding of what selection or dishes they have to offer. Third, because we only looked at most frequently visited types of restaurants, the ratings of each restaurant's food quality which could have provided the client with more specific insights were outside the scope of this study. Similarly, because Part 2 sought out to map out locations where the celebrities had dined during their stay in Tokyo, it left out analyses of the menus, prices, and ratings of each restaurant's food quality.

Notwithstanding these notable limitations, the findings from this exploratory study provide key insights into Tokyo's culinary landscape that would help the client develop a more comprehensive marketing strategy. To enhance the client's planning and future decision-making process, the following recommendations should be taken in consideration:

- The client should seek out Ramen spots in the following wards: Ota, Shinjuku, Chiyoda, Toshima, and Shibuya.
- The client should also consider Japanese restaurants that offer more than Ramen dishes.
- The client should diversify his or her food tour experience in ways that complement the hunt for the most fulfilling ramen by selecting additional venues where prominent celebrities have dined out. This would be added value for the client given that celebrities have a huge following. Going forward, the client should conduct additional research on which celebrities best align with his or her target niche so that they would be attracted to joining the tour.
- The client should select venues that are both in major and lesser entertainment and shopping districts areas to maximize the flavor and character of the food tour.