IBM Applied Data Science Capstone

Opening a Restaurant for Clients with Certain Food Restrictions in Tokyo

By: Mahmoud Mohamed
April 2021

1. Introduction

As of 2021, Tokyo prefecture has an estimated population of 13,960,236. The Greater Tokyo Area is the most populous metropolitan area in the world, with more than 37.393 million residents as of 2020.

Although there are over 148,000 (Annual Report on Food Sanitation 2016) restaurants in Tokyo, most of them focus on serving the Japanese client like Ramen restaurants, Tonkatsu restaurants, ... etc. With the increasing number of foreigners in Tokyo in the past years, few restaurants started to focus on serving clients with food restrictions, but still not enough.

The aim of this report is to explore best locations for opening a restaurant for clients with certain food restrictions in Tokyo. For opening such a unique restaurant, picking the proper place is a key factor for such risky investment as it affects its chances of success or failure. Depending on the available data, I would like to give some insights about it.

1.1. Business Problem

The objective of this Capstone project is to analyse and select the best locations in Tokyo to open a new Halal/ Kosher/ Vegan restaurant. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question: Where in Tokyo, should the investor open such a restaurant.

1.2. Target Audience

- ✓ Developers and investors looking to open or invest in a niche restaurant in Tokyo as there are many foreigners visiting and living in Japan and some of them have some food restrictions.
- ✓ Business Analysts or Data Scientists, who wish to analyze Tokyo neighbourhoods using python, jupyter notebook and some machine learning techniques.
- ✓ Someone curious about data that want to have an idea, how beneficial it is to open a niche restaurant and what are the pros and cons of this business.

2. Data

First, we need some information about the wards of Tokyo, we can scrap Wikipedia to get that information:

Data source:
https://en.wikipedia.org/wiki/Special_wards_of_Tokyo#List_of_special_wards

```
In [2]:  df = pd.read_html('https://en.wikipedia.org/wiki/Special_wards_of_Tokyo#List_of_special_wards')[3]
         df.head()
```

Out[2]:

	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

We can then clean the data by getting rid of useless columns and rows and also by renaming the column into easy and meaningful names like the next sample:

	Romaji	Kanji	Population	Pop. density per km2	Area in km2
0	Chiyoda	千代田区	59441	5100	11.66
1	Chūō	中央区	147620	14460	10.21
2	Minato	港区	248071	12180	20.37
3	Shinjuku	新宿区	339211	18620	18.22
4	Bunkyō	文京区	223389	19790	11.29
5	Taitō	台東区	200486	19830	10.11

Note that Romaji columns represents the Latin spelling of the ward name, while Kanji column represents the name in Chinese characters.

Second, we need the latitudes and longitudes of the wards, we can get it using *geopy.geocoders.Nominatim* from *geopy* python library. "Kanji" column is used to retrieve locations from geopy. A sample of the new table after merging the coordinates follows:

	Romaji	Kanji	Population	Pop. density per km2	Area in km2	Latitude	Longitude
0	Chiyoda	千代田区	59441	5100	11.66	35.693810	139.753216
1	Chūō	中央区	147620	14460	10.21	35.666255	139.775565
2	Minato	港区	248071	12180	20.37	35.643227	139.740055
3	Shinjuku	新宿区	339211	18620	18.22	35.693763	139.703632
4	Bunkyō	文京区	223389	19790	11.29	35.718810	139.744732
5	Taitō	台東区	200486	19830	10.11	35.717450	139.790859

Third, by using Foursquare API we will get all the venues in each neighbourhood. We can filter these venues to get only restaurants like the next sample:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Chiyoda	35.69381	139.753216	Kanda Tendonya (神田天丼家)	35.695765	139.754682	Tempura Restaurant
2	Chiyoda	35.69381	139.753216	Sushi Masa (九段下 寿司政)	35.695234	139.752227	Sushi Restaurant
3	Chiyoda	35.69381	139.753216	Jimbocho Kurosu (神保町 黒須)	35.695539	139.754851	Ramen 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

Now we are ready to start analysing our data and exploring different possibilities for the potential new restaurants.