

# Optimal Location to Open a Chinese Restaurant in Tokyo, Japan

Cherrie He

June 10, 2020

# Introduction

## Background

- To be successful with any restaurant opening, market research is the best thing to start off.
- At one time, market researches before running a business were implemented in traditional ways, like visiting local communities or interviewing potential customers. Whereas nowadays, the researches are largely facilitated since the machine learning power were added.
- Data analyze and machine learning techniques play very important role in helping businesspeople grab market share and become profitable when opening a restaurant in a smart and scientific way.

# Introduction

## Business Problem

- A client seeks to establish a Chinese restaurant in Tokyo, Japan.
- This capstone report will determine the optimal and most strategic location for running the restaurant by exploring the following questions:
  - **Is the market saturated?**
  - **Who are the potential customers?**
  - **Who are the local competitors?**

# Introduction

## Interests

- This report will be targeted to stakeholders interested in opening a Chinese restaurant in Tokyo, Japan. Various factors such as over-saturation or no demand ,for the type of restaurant that the customer wants to open, effect the success or failure of the restaurant. Hence, customers can bolster their decisions using the descriptive and predictive capabilities of data science.

# Data Acquisition and Wrangling



Data Sources are mainly scrapped from Wiki websites and downloaded from [Statistics of Tokyo](#)



Tokyo Wards Table from Wikipedia

I first make use of Special Wards of Tokyo page from [to scrap the table to create a data-frame. Simply use pandas.read\\_html and then extract the target table.](#)



Tokyo land price are downloaded from



For the demographical data, which will help analyse the potential customers are downloaded from [TOKYO STATISTICAL YEARBOOK 2018](#)

# Methodology

## Exploratory Data Analysis

- **Distributions of Restaurants in Tokyo**

By filtering all restaurants and other places serve foods out of the venues Dataframe, extract the information of local restaurant distributions. There are 195 unique categories. I will choose the places that contain "Restaurant" or seemingly serve food.

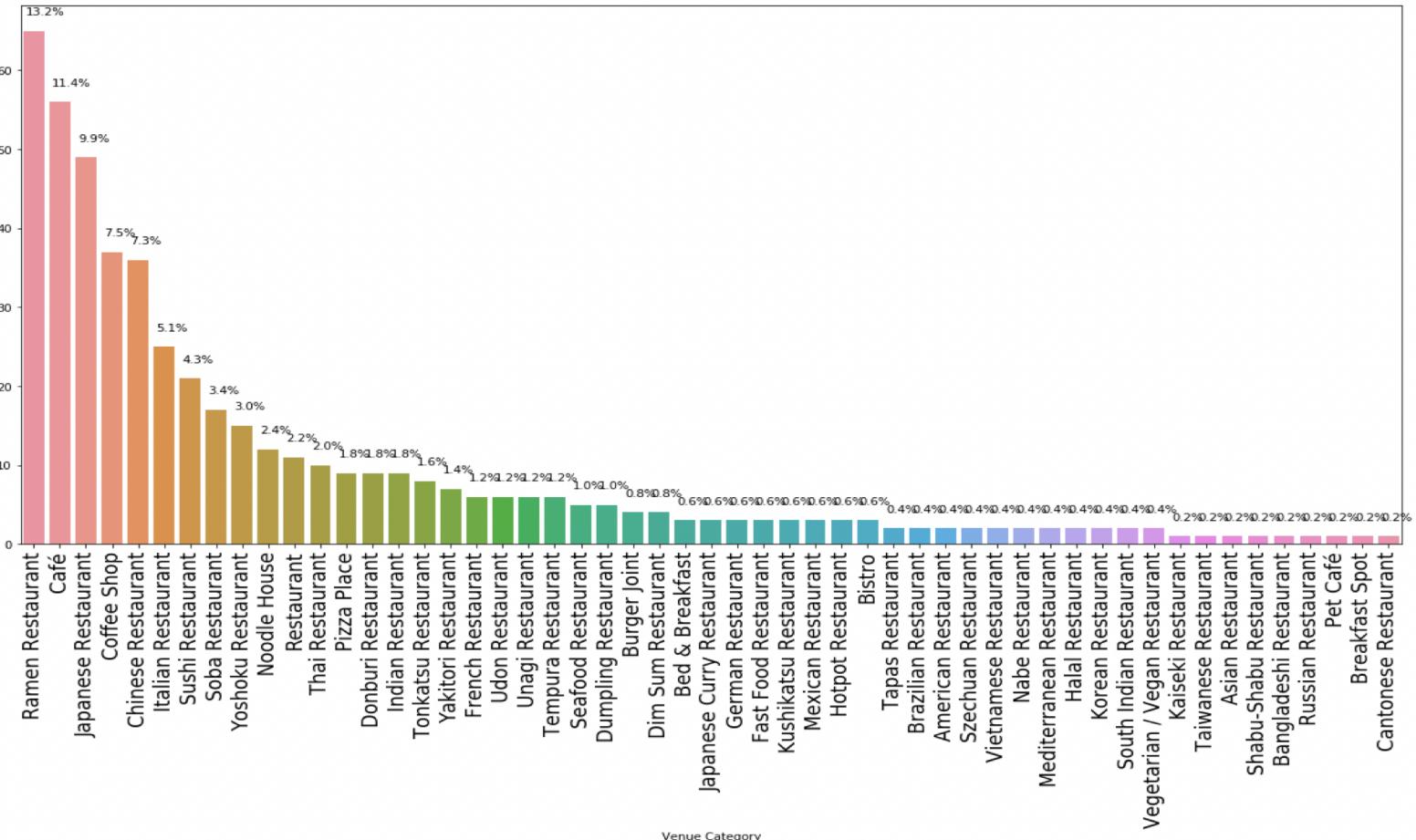
- **Data Preview**

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Nagatacho	35.675618	139.743469	Nagatacho Kurosawa (永田町 黒澤)	35.674699	139.741737	Japanese Restaurant
2	Nagatacho	35.675618	139.743469	Tully's Coffee	35.674594	139.743007	Coffee Shop
3	Nagatacho	35.675618	139.743469	Shinamen Hashigo (支那麵 はしご)	35.672184	139.741576	Ramen Restaurant
4	Nagatacho	35.675618	139.743469	All Day Dining Origami (オールデイ ダイニング ORIGAMI)	35.673815	139.741104	Restaurant
5	Nagatacho	35.675618	139.743469	Sushi Isshin (鮓一新)	35.672589	139.739399	Sushi Restaurant

# Methodology

## Exploratory Data Analysis

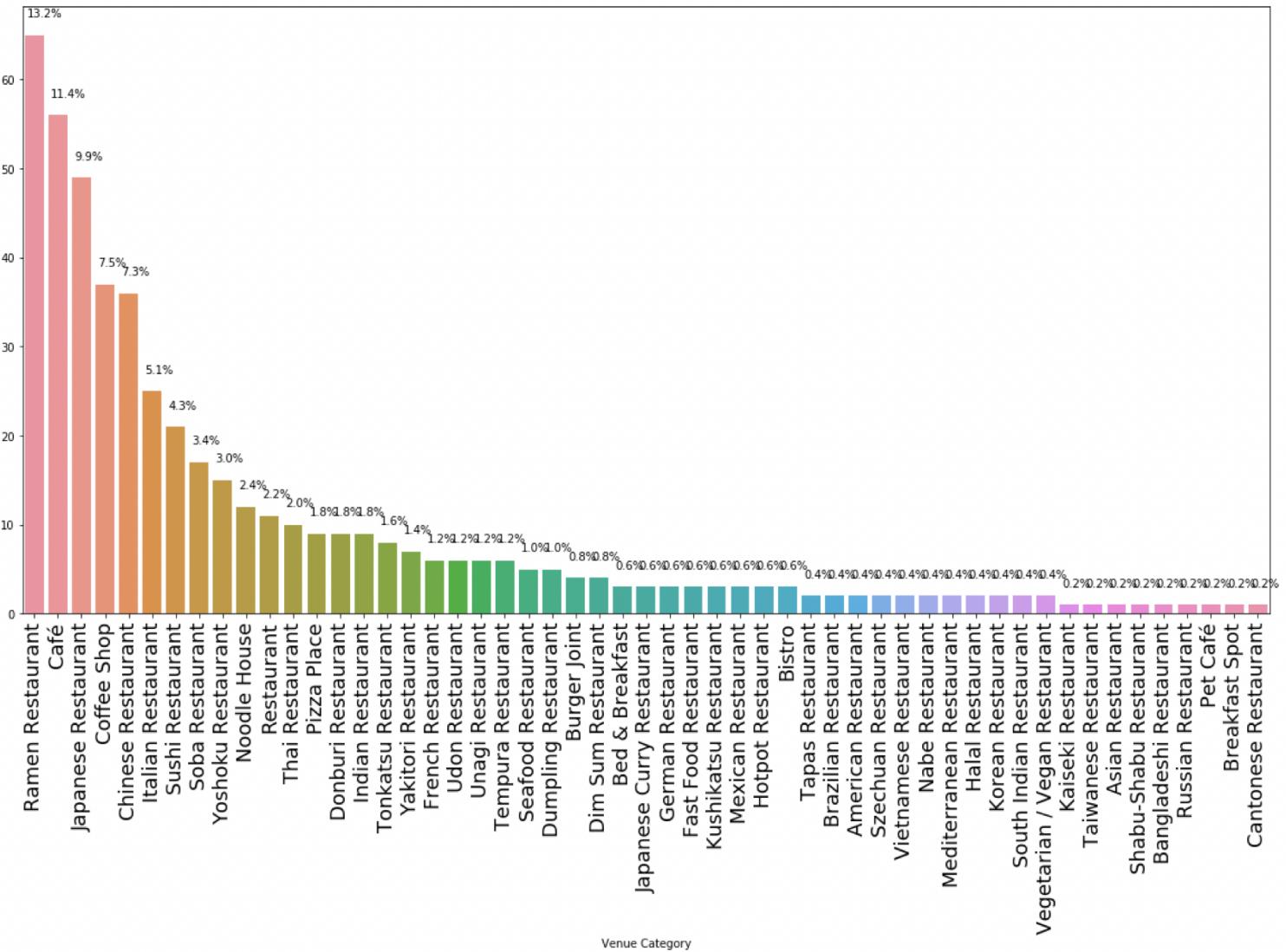
- There are 493 out of 1201 nearby venues serve food. By plotting a bar chart and annotating the frequency of their distributions, it is clearly that the most popular restaurant is "Ramen Restaurant" which has the top frequency of distribution at 13.2%. Followed by "Cafe" at 11.4%, "Japanese Restaurant" at 9.9%, "Chinese Restaurant" at 7.5% and "Coffee Shop" at 7.3%.



# Methodology

## Exploratory Data Analysis

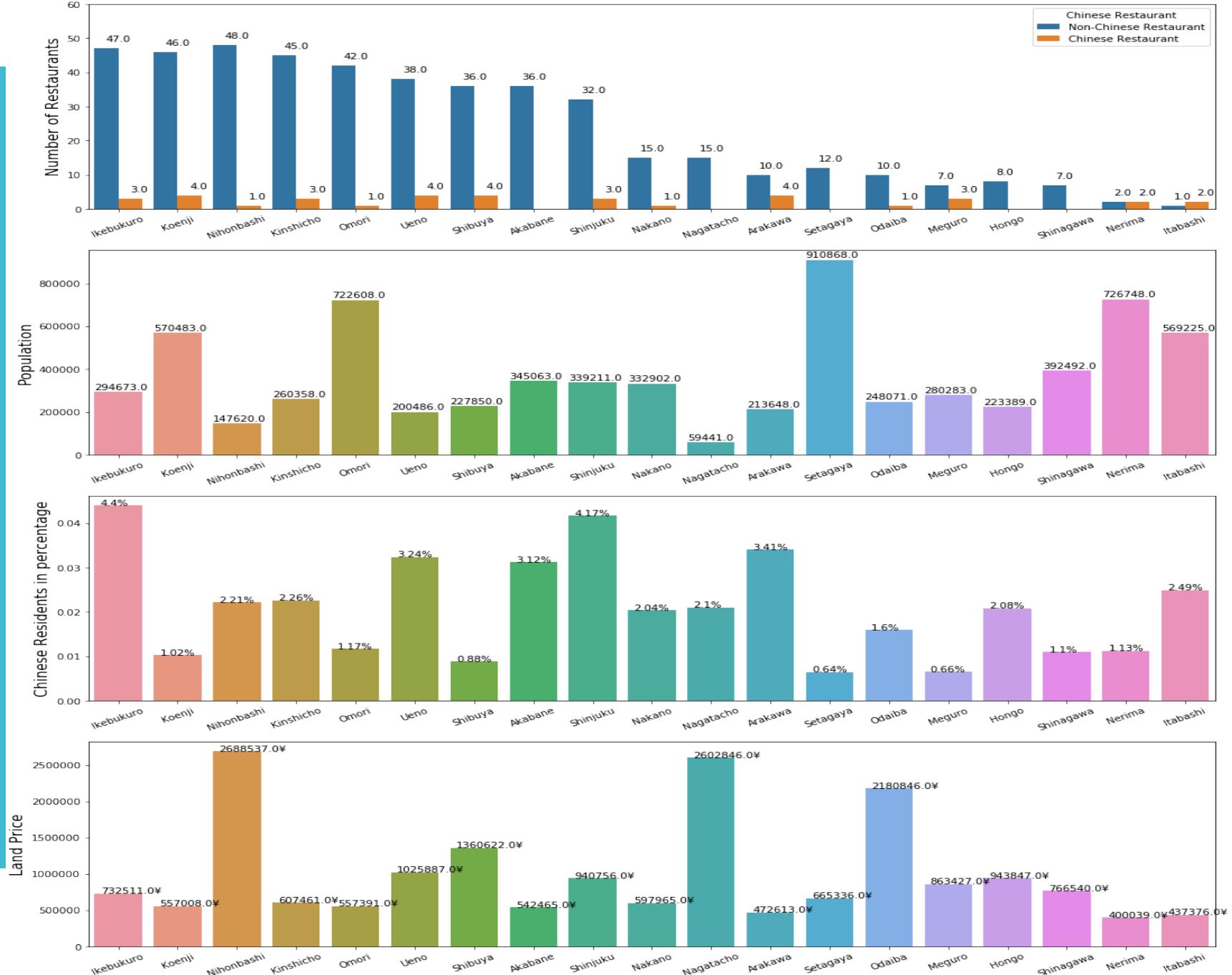
The number of Chinese restaurants are as half as that of Ramen restaurants, and it is 1% less than Japanese restaurant. Whereas the restaurants ranked below No.5 took at most half of the frequency of Chinese restaurants.



# Methodology

What about the restaurant distribution of each ward?

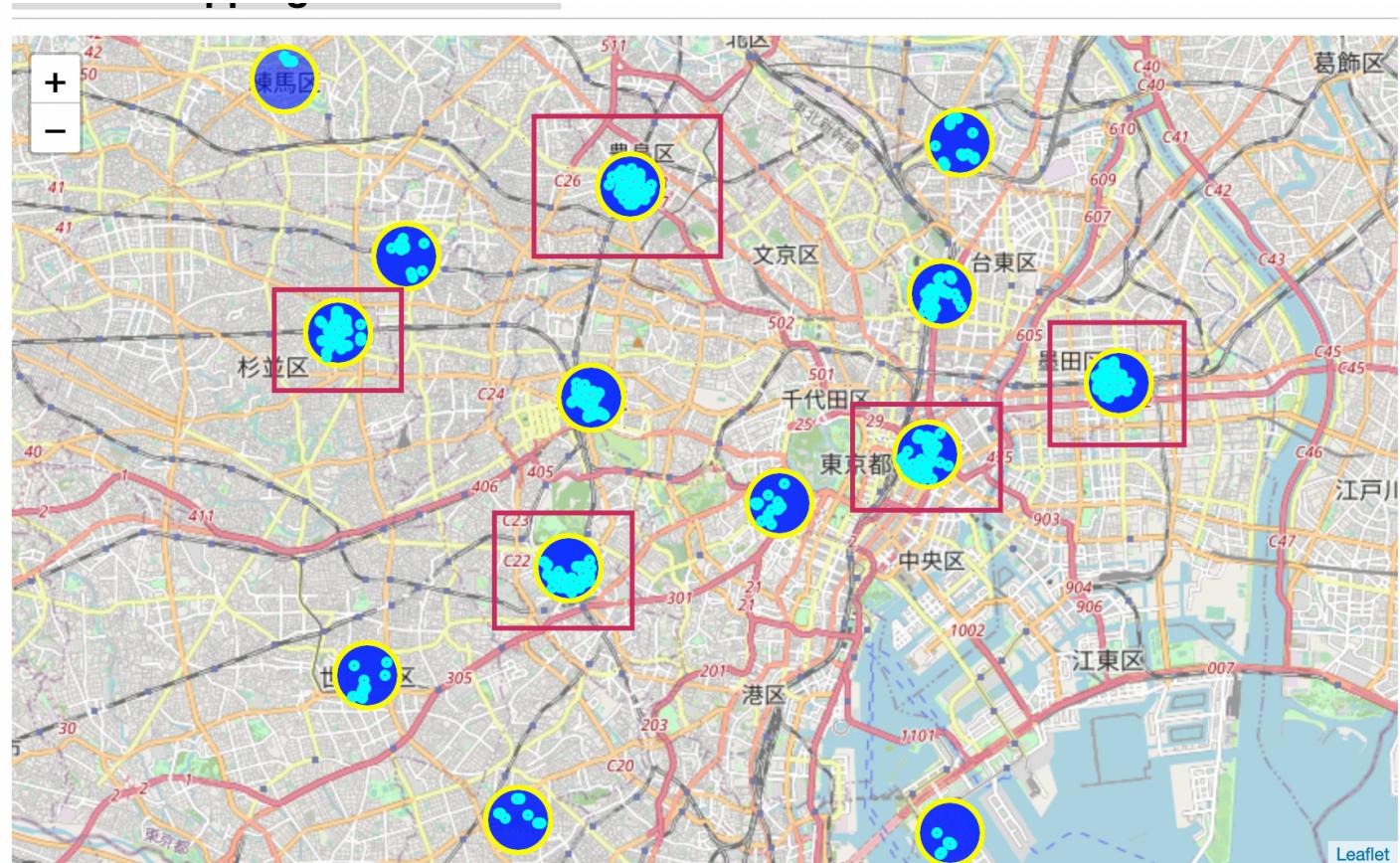
The number of restaurants and Tokyo population are presented as follows. This will help analyze the relationship between market requirements and market size.

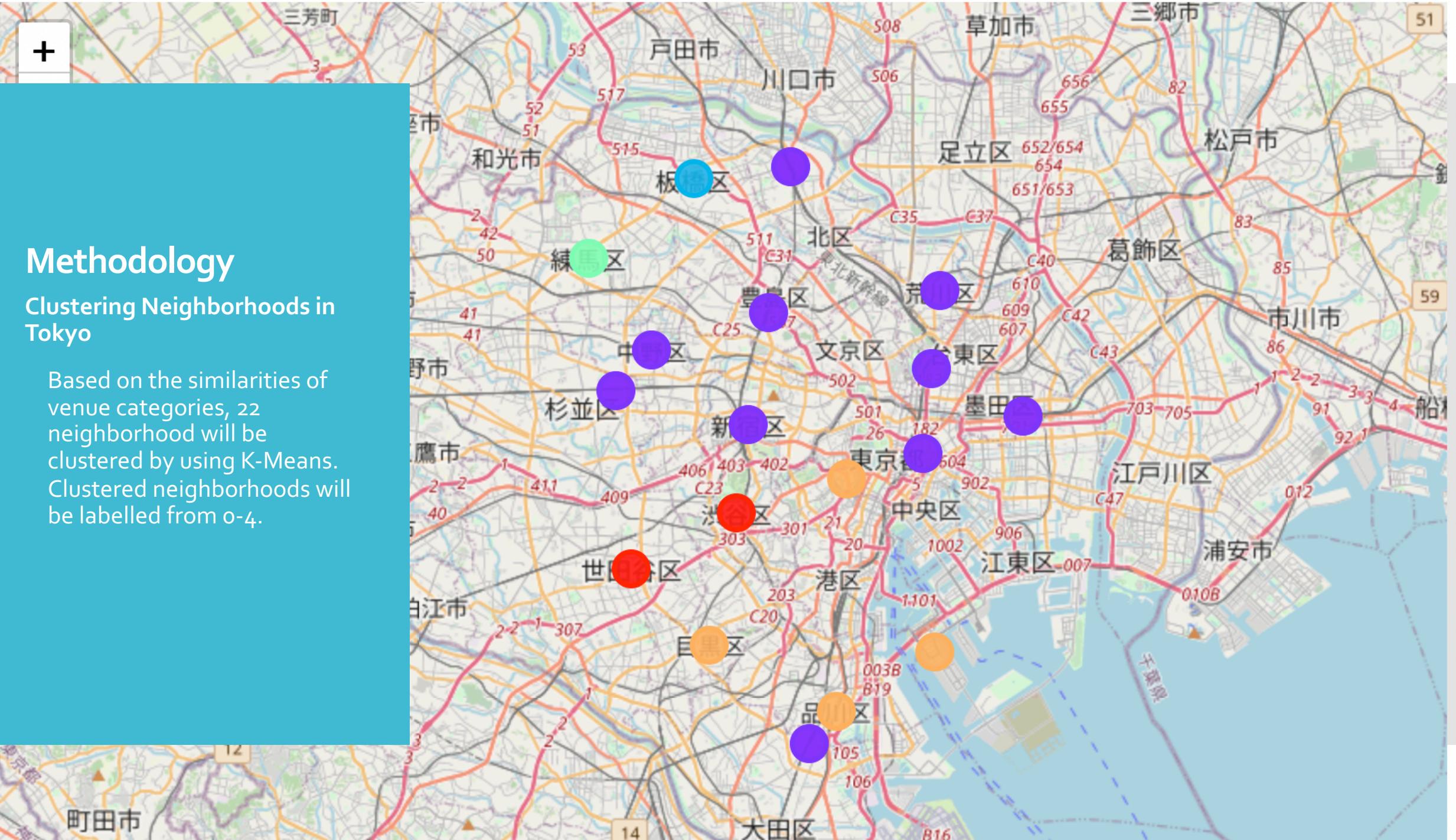


# Methodology

## Folium Mapping Visualization

- Open a restaurant in following highlighted areas would face competitive environment, although the crowd of the restaurants indicates high volume of customers





# Methodology

# Clustering Neighborhoods in Tokyo

Based on the similarities of venue categories, 22 neighborhood will be clustered by using K-Means. Clustered neighborhoods will be labelled from 0-4.

# Results

As stated at the beginning of this project, this report will answer the following questions by exploring the geographical and market data with the machine learning power deployed.

According to my analysis, Cluster 1, where Chinese restaurants with low ranking of popularity, but with high Chinese resident proportions and fairly low land prices, are the most optimal locations to open a Chinese restaurant.

## **Is the market saturated?**

Ramen restaurants are the most frequently distributed among all districts, which is followed by Cafe and Japanese restaurants, whereas Chinese restaurants are almost 50% less than the Ramen restaurants distribution frequency. It indicates: 1. Opening a Chinese restaurant is easier than that for a Ramen restaurant. Because the market for Ramen restaurant are more likely to be saturated, which would bring fierce market competition. 2. Chinese restaurants might not as popular as Ramen and Japanese Restaurants. 3. Districts like Ikebukuro, Shinjuku, Ueno and Itabashi, that have higher Chinese residents' proportions and relatively low land prices are most likely to be selected as the most strategical places to open a Chinese restaurant. Those districts are labelled as Cluster 1.

## **Who are the potential customers?**

Both Chinese residents and Non-Chinese residents are potential customers. The percentage of Chinese residents in Arakawa and Itabashi, where take the Chinese restaurant as their 1st Most Common Venue, are not the highest ones. This might point the Chinese restaurants are popular to not just Chinese Residents.

## **Who are the local competitors?**

Although Chinese restaurants are the 1st or 2nd popular restaurants for some districts, however there are at most 4 Chinese restaurants out of the popular restaurants of each districts. Chinese restaurants performed lower market occupancy than Ramen and Cafe, but they are much frequently distributed than all other kind of restaurants. Therefore, the main competitors of Chinese Restaurants are Japanese restaurants, especially Ramen restaurants.

# Discussion and Conclusion

- Since the clustering was based only on the most common venues of each district, further discussions can be pivoting on other clustering such as DBSCAN. Some of the districts are not recorded in [TOKYO STATISTICAL YEARBOOK 2018](#). This kind of information might be found in Japanese websites. For the number of Chinese residents in Kasai and Tateshi could be discovered by who are interested and may understand Japanese.
- This project offers a glimpse of how real life data-science projects look like. Despite the importance of traditional ways of doing marketing research, machine learning and data analyze techniques are largely facilitated the conveniences of solving business problems. My analysis were mainly based on the acquirable data from English websites, more interesting analytics can be continuously conducted as more multilingual resources are surely available and will be discovered by interested readers.