

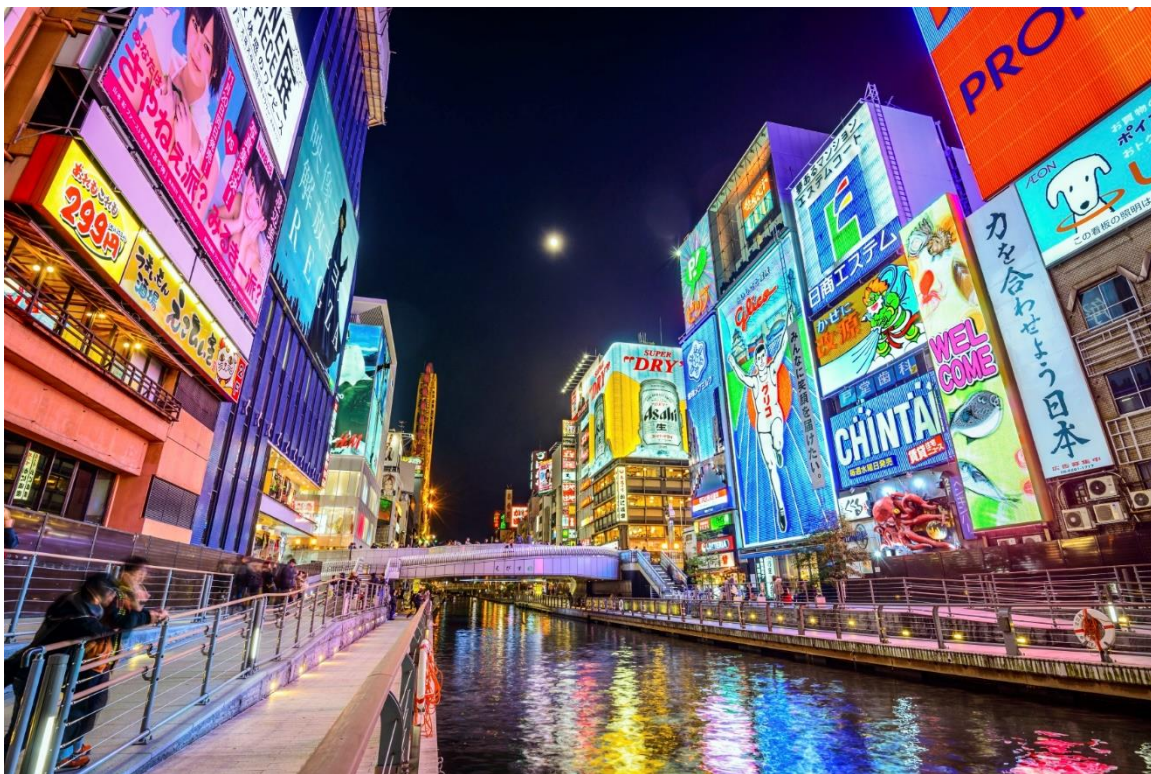
CAPSTONE PROJECT – The Battle of Neighborhoods

APPLIED DATA SCIENCE

To open a Vietnamese restaurant in Osaka, Japan

By Le Mai Minh Phong

Jan 2020



Introduction

Osaka is the third largest city in Japan, with a population of over 2.5 million people in its greater metropolitan area. It is the central metropolitan city of the Kansai region and the largest of the Osaka-Kobe-Kyoto trio. Osaka is also known as the “Nation’s Kitchen,” and is well known for delicious foods. You can also find a variety of cuisines in Osaka street, from Western to South East Asia. In recent years, the number of foreign people especially Asian in Osaka keep raising every year that I not event mention the number of foreigners come to Osaka for travelling. In December 2019, they recorded over 150 thousand foreigners who are working and living in Osaka. With the trend of exchange cultures, the majority of Japanese people nowadays, beside their neutralism flavor foods, are searching for more deep flavor in South East Asian cuisine.

Business Problem

The ultimate goal of this capstone project is to analyze and to provide the best locations in Osaka in order to open a new Vietnamese restaurant. In the scope of this capstonse project, I don’t want to mention about a particular high-end, exclusive restaurant like the image of a Italian or France restaurant, but I’m talking about a small family restaurant in general. In Japan, when looking for a place to open this kind of business, the first thing and also the most important thing you have to put in consideration is the accessibility. Most of people in Japan transport by public transportations, such as train, subway, bus. Other aspect should be considered is the target customers. From the perspective of sharing the similarities in cuisine, I think the main customer will be South East Asian.

By applying what I have learned through out this course like data science methodology and machine learning techniques, I will provide a solution to answer the business question: if someone is looking to open a new Vietnamese restaurant in Osaka area, where should they consider to open it?

Data

We will need data from reliable sources for analysis. To understand our problem and quantify result we will use the following data.

- List of wards/boroughs in Osaka.
- Latitude and longitude coordinates of those wards.
- Venue data, particularly data related to restaurant.
- Foreign population, particularly Vietnamese in Osaka.

For the list of wards in Osaka, we can find them in this Wikipedia page <https://en.wikipedia.org/wiki/Osaka> that contains a list of wards in Osaka, with a total of 24 wards. In order to pull the data out of the html page, we will use BeautifulSoup that is a Python library provides a beautiful web scraping technique. Then we will get the geographical coordinates of the wards using Python Geocoder package.

For the venue data, we will apply to a Foursquare Developer Account at <https://foursquare.com/> that allows us to get the venue data for those wards via Foursquare API. The Foursquare API will provide many categories of the venue data, especially the restaurant category in general and the Vietnamese restaurant category in particular.

For the foreign population, it's amazing that the Osaka office provides us the information in real detail at their official website below.

[https://www.city.osaka.lg.jp/shimin/cmsfiles/contents/00000006/6893/\(0109\)07_kubetu_kokusekibetu_gaikokujin.xls](https://www.city.osaka.lg.jp/shimin/cmsfiles/contents/00000006/6893/(0109)07_kubetu_kokusekibetu_gaikokujin.xls)

We can find an Excel file with data for number of foreign people in each ward of Osaka in separate nationalities.

Methodology

In this project we will direct our efforts on detecting areas of Osaka that have low Vietnamese restaurant density. We will limit our analysis to area ~2km around each Ward's center.

In first step we have collected the required data: location and type (category) of every restaurant within 2km from each Ward's center (24 wards). We have also identified Vietnamese restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Osaka - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Vietnamese restaurants in vicinity) and focus our attention on those areas.

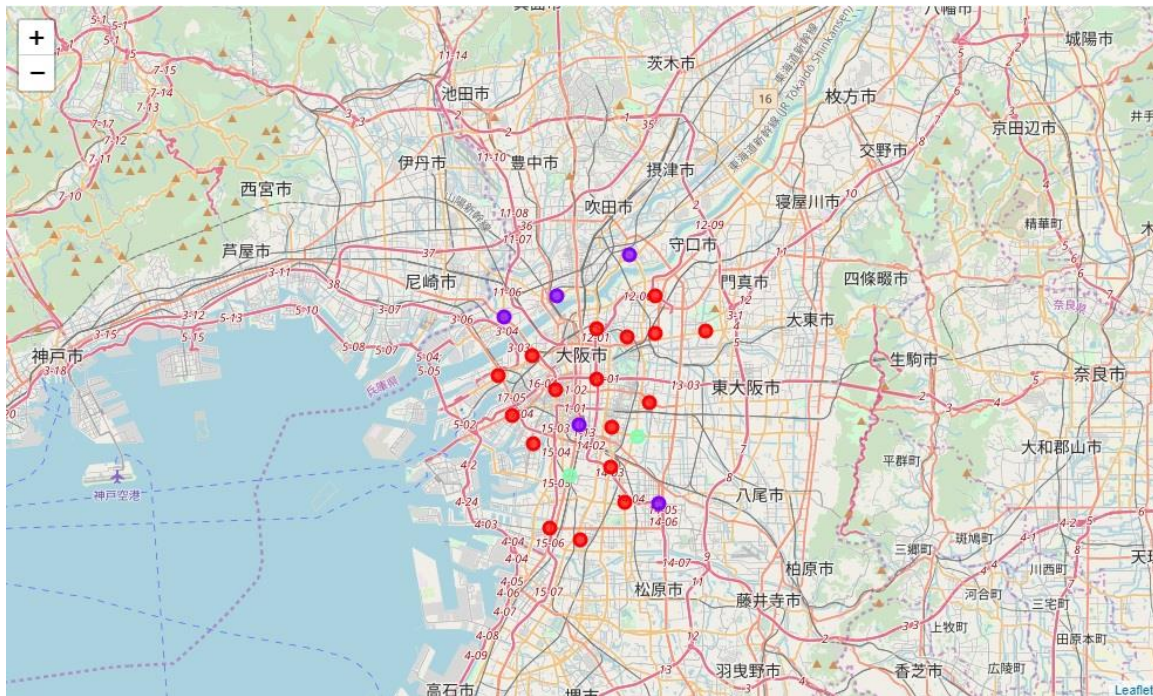
In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with high Vietnamese density, and we want locations without Vietnamese restaurants in radius of 200 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Result

The results from the k-means clustering show that we can categorize the wards into 3 clusters based on the frequency of occurrence for Vietnamese restaurant and number of Vietnamese:

- Cluster 0: wards with moderate number of Vietnamese restaurant and low density of Vietnamese
- Cluster 1: wards with low number of Vietnamese restaurant and moderate density of Vietnamese
- Cluster 2: wards with low number of Vietnamese restaurant and high density of Vietnamese

The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, and cluster 2 in mint green color.



Conclusion and further discussion

Purpose of this project was to identify Osaka areas close to center with low number of restaurants (particularly Vietnamese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Vietnamese restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general wards that justify further analysis (生野区 and 西成区), and then generated extensive collection of locations which satisfy some basic requirements regarding density of Vietnamese. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of wards and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location, accessibility to train station, real estate availability, prices, social and economic dynamics of every ward etc.



References

Osaka general information from Wikipedia, the free encyclopedia

<https://en.wikipedia.org/wiki/Osaka>

Foursquare Developers Documentation

<https://developer.foursquare.com/docs>

Vietnamese population in Osaka from Osaka Office Website

[https://www.city.osaka.lg.jp/shimin/cmsfiles/contents/0000006/6893/\(0109\)07_kubetu_kokusekibetu_gaikokujin.xls](https://www.city.osaka.lg.jp/shimin/cmsfiles/contents/0000006/6893/(0109)07_kubetu_kokusekibetu_gaikokujin.xls)

Appendix

Cluster 0

- | | | |
|--------------------------------------|-------------------------------|------------------------------|
| ➤ <u>Abeno-ku</u>
阿倍野区 | ➤ <u>Joutou-ku</u>
城東区 | ➤ <u>Suminoe-ku</u>
住之江区 |
| ➤ <u>Asahi-ku</u>
旭区 | ➤ <u>Kita-ku</u>
北区 | ➤ <u>Sumiyoshi-ku</u>
住吉区 |
| ➤ <u>Chuo-ku</u>
中央区 | ➤ <u>Konohana-ku</u>
此花区 | ➤ <u>Taisho-ku</u>
大正区 |
| ➤ <u>Fukushima-ku</u>
福島区 | ➤ <u>Minato-ku</u>
港区 | ➤ <u>Tennouji-ku</u>
天王寺区 |
| ➤ <u>Higashinari-ku</u>
東成区 | ➤ <u>Miyakojima-ku</u>
都島区 | ➤ <u>Tsurumi-ku</u>
鶴見区 |
| ➤ <u>Higashisumiyoshi-ku</u>
東住吉区 | ➤ <u>Nishi-ku</u>
西区 | |

Cluster 1

- | | | |
|-------------------------------------|-----------------------------------|-----------------------------|
| ➤ <u>Higashiyodogawa-ku</u>
東淀川区 | ➤ <u>Naniwa-ku</u>
浪速区 | ➤ <u>Yodogawa-ku</u>
淀川区 |
| ➤ <u>Hirano-ku</u>
平野区 | ➤ <u>Nishiyodogawa-ku</u>
西淀川区 | |

Cluster 2

- | | |
|--------------------------|------------------------------|
| ➤ <u>Ikuno-ku</u>
生野区 | ➤ <u>Nishinari-ku</u>
西成区 |
|--------------------------|------------------------------|