Applied Data Science Capstone

Capstone Project

Opening a Japanese restaurant in Bangkok

Introduction

Thailand is now a very popular tourist spot for the travellers around the globe because of food, shopping malls, markets and outdoor activities. Most people will go to Bangkok in their first Thai experience due to its convenience. Also, ASEAN countries are growing quickly due to external environments. Because of the underlying economic growth and cultural exchange, Thai would start to experience and explore other countries' culture as well. Moreover, more Japanese nationals are moving to Thailand according to Embassy of Japan in Thailand. Opening a restaurant providing Japanese cuisine is a great chance.

Business Problem

The aim of this project is to help businessmen find the best place to open a Japanese restaurant in Bangkok in which popularity and competitiveness are considered concurrently. Locating in a very popular district faces higher competitiveness. In contrast, opening it far away from popular district faces lower return. Therefore, clustering analysis, one of the most popular machine learning methodologies, will be adopted to let them know where has higher business opportunity and lower risk.

Targeted audience

The targeted audience is for whom looking forward to opening a Japanese restaurant. As it involves massive fixed cost and investment, the project can help them start or expand their Japanese catering business in order to seek for the Thai and Japanese customers to minimize their risk and maximize their return.

Data

To solve the problem, the following data is required

- Bangkok's districts
- Districts' latitude and longitude
- Venue data around each district

The sources of data are from Wikipedia and Foursquare. List of districts of Bangkok and districts' latitude and longitude can be found in Wikipedia (https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok). It contains 50 districts with corresponding latitude and longitude. Scraping is used to extract the data from Wikipedia via Python making use of "request" and "Beautiful Soup". However, not all the districts are equipped with longitude and longitude data. Those districts will be removed in this analysis during data cleansing. Then, Foursquare location API is used to explore the venue data near to each district. After clustering analysis, the data should provide us with what are popular venues in each cluster. Finally the result is visualized and analyzed to solve the business problem.

To sum up, the analysis will be done from data scraping from Wikipedia, calling Foursquare location APIs, data cleansing, data wrangling, k-means clustering analysis and data visualization. Finally, the result should be capable to provide an insight for whom they want to open a Japanese restaurant in Bangkok.

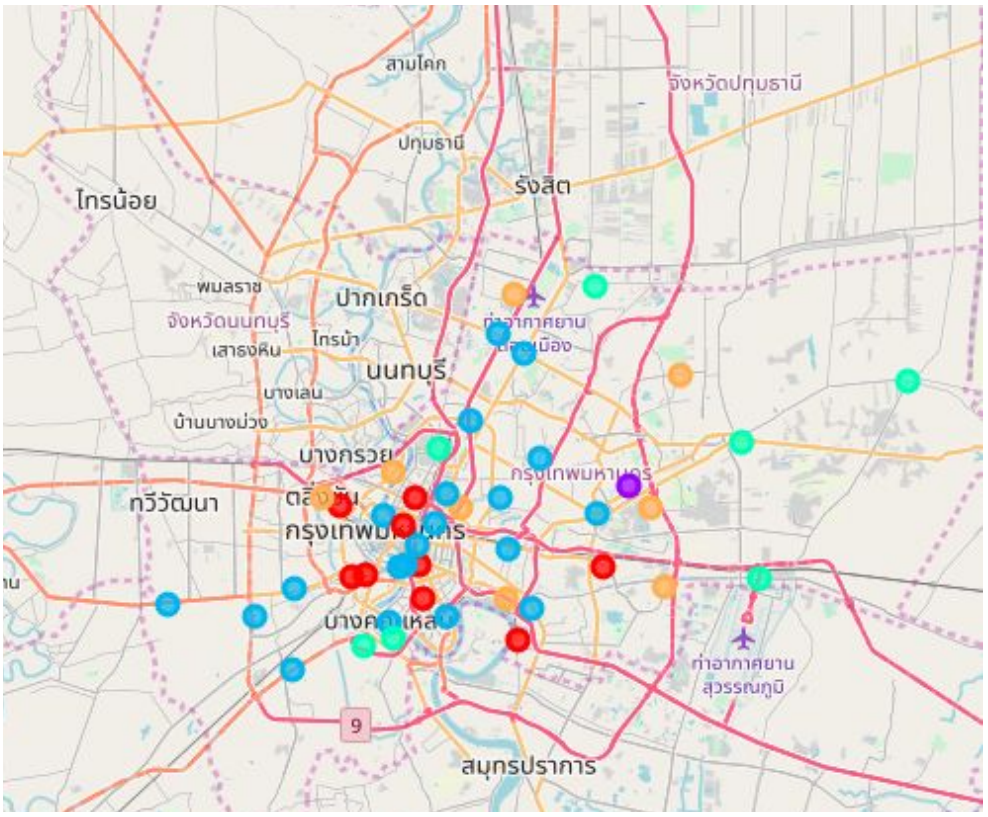
Methodology

Firstly, we will surf the internet to get the related data such as Bangkok's districts and corresponding latitude and longitude. In Wikipedia, 50 main districts with location data in Bangkok can be found. We can scrape the table making use of 'request' and 'Beautiful Soup' library and then transform them into a dataframe for further analysis. However, 5 districts are missing their latitude and longitude data. Those districts will be removed during data cleansing. After the data is cleaned, we can visualize how the districts are located in Bangkok with the help of 'Folium' library. At the same time, we can make sure the districts' latitude and longitude are correct.

Secondly, Foursquare APIs are utilized to explore the districts and segment them. 4 types of data are required in APIs such as venue name, venue latitude, venue longitude and venue category. Moreover, 100 venues and 500 radius for each district are explored. If there is higher radius, it may cause the same venue appearing in two districts' respectively which may lower accuracy of the clustering analysis result. Then, 1000 venues and 149 unique categories are returned. To understand the characteristics and what services or shops are popular in each district, venue categories dummies are built and rows are grouped by district, latitude and longitude simultaneously and take the mean of the frequency of occurrence of each category. Then, k-means clustering algorithm is performed to group the districts with similar features based on the top 10 most common venues. K-means clustering algorithm is a popular unsupervised machine learning technique. It allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is a suitable algorithm in this business problem given input vectors without referring to known, or labelled, outcomes. In order to avoid the results being too concentrated or too loose, 5 clusters are adopted. We can visualize the result of k-means clustering on the map then. More importantly, the result is capable to tell what are the characteristics among clusters, say what is the consumers preference there. More importantly, it should be able to answer where is the best place to open a Japanese restaurant.

Results

The districts are clustered in 5 groups considering the top 10 most common venues.



The table shows the relationship between cluster and color as well as districts count for each cluster. More importantly, the result shows the characteristics for the clusters.

Color	Cluster	No. of districts in the cluster
Red	1	9
Purple	2	1
Blue	3	20
Green	4	7
Orange	5	8

Cluster 1

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Bang Na	0	Asian Restaurant	Noodle House	Intersection	Coffee Shop	Café	Seafood Restaurant	Satay Restaurant	Thai Restaurant	Electronics Store	Farmers Market
7	Bang Rak	0	Noodle House	Hotel	Chinese Restaurant	Hotel Bar	Café	Clothing Store	Coffee Shop	Som Tum Restaurant	Massage Studio	Buffet
9	Bangkok Noi	0	Noodle House	Thai Restaurant	Park	Snack Place	Café	Chinese Restaurant	Coffee Shop	Dessert Shop	Asian Restaurant	Electronics Store
10	Bangkok Yai	0	Noodle House	Asian Restaurant	Dessert Shop	Farmers Market	Coffee Shop	Photography Studio	Train Station	BBQ Joint	Soup Place	Thai Restaurant
16	Dusit	0	Noodle House	Asian Restaurant	Dessert Shop	Convenience Store	Som Tum Restaurant	Coffee Shop	Market	Wings Joint	Floating Market	Food Truck
32	Pom Prap Sattru Phai	0	Noodle House	Café	Asian Restaurant	Thai Restaurant	Chinese Restaurant	Dim Sum Restaurant	Convenience Store	American Restaurant	Snack Place	Shopping Mall
39	Sathon	0	Noodle House	Asian Restaurant	Thai Restaurant	Chinese Restaurant	Convenience Store	Coffee Shop	Dessert Shop	Seafood Restaurant	Bakery	Bar
40	Suan Luang	0	Noodle House	Chinese Restaurant	Thai Restaurant	Coffee Shop	Asian Restaurant	Convenience Store	Som Tum Restaurant	German Restaurant	Café	Farmers Market
42	Thon Buri	0	Noodle House	Bus Stop	Fried Chicken Joint	Asian Restaurant	Pharmacy	Train Station	Deli / Bodega	Spa	Seafood Restaurant	Food Court

Cluster 2

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Bueng Kum	1	Food Truck	Miscellaneous Shop	Convenience Store	Gym / Fitness Center	Wings Joint	Fast Food Restaurant	Food Stand	Food Court	Food & Drink Shop	Floating Market

Cluster 3

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bang Kapi	2	Convenience Store	Noodle House	Thai Restaurant	Flea Market	Coffee Shop	Market	Steakhouse	Massage Studio	Buffet	Museum
1	Bang Khae	2	Japanese Restaurant	Wings Joint	Gourmet Shop	Shopping Mall	Noodle House	Coffee Shop	Fast Food Restaurant	Convenience Store	Supermarket	Grocery Store
2	Bang Khen	2	Asian Restaurant	Coffee Shop	Convenience Store	Bakery	Garden Center	Noodle House	Som Tum Restaurant	Vietnamese Restaurant	Duty-free Shop	Electronics Store
3	Bang Kho Laem	2	Noodle House	Thai Restaurant	Chinese Restaurant	Coffee Shop	Sushi Restaurant	Hotpot Restaurant	Shopping Mall	Vietnamese Restaurant	Seafood Restaurant	Fast Food Restaurant
4	Bang Khun Thian	2	Thai Restaurant	Japanese Restaurant	Noodle House	Pizza Place	Café	Bakery	Restaurant	Juice Bar	Fast Food Restaurant	Building
12	Chatuchak	2	Coffee Shop	Flea Market	Thai Restaurant	Beer Bar	Food Truck	Cocktail Bar	Convenience Store	Noodle House	Massage Studio	Gym
17	Huai Khwang	2	Asian Restaurant	Som Tum Restaurant	Hotel	Noodle House	Thai Restaurant	Restaurant	Chinese Restaurant	Grocery Store	Supermarket	Hotpot Restaurant
19	Khlong San	2	Hotel Bar	Clothing Store	Café	Hotel	Department Store	Dessert Shop	Coffee Shop	Noodle House	Japanese Restaurant	Park
21	Lak Si	2	Thai Restaurant	Coffee Shop	Japanese Restaurant	Hotel	Supermarket	Shopping Mall	Food Court	Café	Canal	Soccer Stadium
23	Lat Phrao	2	Noodle House	Café	Som Tum Restaurant	Asian Restaurant	Coffee Shop	Gift Shop	Hotel	Organic Grocery	Convenience Store	Sake Bar
26	Nong Khaem	2	Thai Restaurant	Diner	Dessert Shop	Department Store	Noodle House	Pet Store	Convenience Store	Flea Market	Market	Coffee Shop
27	Pathum Wan	2	Convenience Store	Noodle House	Chinese Restaurant	Asian Restaurant	Dessert Shop	Thai Restaurant	Sporting Goods Shop	Japanese Restaurant	Seafood Restaurant	Som Tum Restaurant
28	Phasi Charoen	2	Coffee Shop	Japanese Restaurant	Fast Food Restaurant	BBQ Joint	Steakhouse	Department Store	Mobile Phone Shop	Clothing Store	Shabu-Shabu Restaurant	Pizza Place
29	Phaya Thai	2	Café	Thai Restaurant	Coffee Shop	Japanese Restaurant	Som Tum Restaurant	Noodle House	Bar	Sushi Restaurant	Bakery	BBQ Joint
30	Phra Khanong	2	Convenience Store	Fast Food Restaurant	Shopping Mall	Hotel	Italian Restaurant	Japanese Restaurant	Spa	Snack Place	Café	Supermarket
31	Phra Nakhon	2	Noodle House	Hostel	Café	Hotel	Thai Restaurant	Bakery	Asian Restaurant	Halal Restaurant	Massage Studio	Snack Place
35	Ratchathewi	2	Coffee Shop	Café	Steakhouse	Hostel	Hotel	Convenience Store	Food Court	Restaurant	Gym / Fitness Center	Jazz Club
37	Samphanthawong	2	Art Gallery	Hotel Bar	Hostel	Hotel	Chinese Restaurant	Coffee Shop	Restaurant	Dive Spot	Cocktail Bar	Seafood Restaurant
43	Watthana	2	Café	Coffee Shop	Thai Restaurant	Chinese Restaurant	Noodle House	Cocktail Bar	Hotel	BBQ Joint	Furniture / Home Store	Japanese Restaurant
44	Yan Nawa	2	Fast Food Restaurant	Thai Restaurant	Brewery	Hotpot Restaurant	Chinese Restaurant	Café	Intersection	Ice Cream Shop	Harbor / Marina	Japanese Restaurant

Cluster 4

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Bang Sue	3	Thai Restaurant	Coffee Shop	Noodle House	Hotpot Restaurant	Badminton Court	Bar	Seafood Restaurant	Wings Joint	Food Stand	Food Court
13	Chom Thong	3	Thai Restaurant	Toll Plaza	Fast Food Restaurant	Coffee Shop	Electronics Store	Food Stand	Food Court	Food & Drink Shop	Floating Market	Flea Market
22	Lat Krabang	3	Thai Restaurant	Steakhouse	Asian Restaurant	Café	Bar	Restaurant	Boat or Ferry	Noodle House	Flea Market	Wings Joint
24	Min Buri	3	Intersection	Department Store	Coffee Shop	Thai Restaurant	Dog Run	Donburi Restaurant	Dumpling Restaurant	Duty-free Shop	Electronics Store	Dive Spot
25	Nong Chok	3	Thai Restaurant	Gym / Fitness Center	Flea Market	Shopping Mall	Other Repair Shop	Park	Dessert Shop	Convenience Store	Duty-free Shop	Electronics Store
34	Rat Burana	3	Thai Restaurant	Coffee Shop	Hotpot Restaurant	Noodle House	Chinese Restaurant	Bistro	Food Stand	Food Court	Food & Drink Shop	Floating Market
36	Sai Mai	3	Thai Restaurant	Deli / Bodega	Noodle House	Bar	Farmers Market	Food Stand	Food Court	Food & Drink Shop	Floating Market	Flea Market

Cluster 5

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Bang Phlat	4	Convenience Store	Dog Run	Cocktail Bar	Café	Restaurant	Fast Food Restaurant	Bus Station	Massage Studio	Coffee Shop	Food Stand
14	Din Daeng	4	Convenience Store	Thai Restaurant	Hotel	Recreation Center	Sports Club	Park	Stadium	Farmers Market	Food Court	Food & Drink Shop
15	Don Mueang	4	Restaurant	Spa	Thai Restaurant	Convenience Store	Buffet	Gym	Food Truck	Hotel	Dumpling Restaurant	Donburi Restaurant
18	Khlong Sam Wa	4	Pub	Japanese Restaurant	Coffee Shop	Convenience Store	Restaurant	Thai Restaurant	Dim Sum Restaurant	Dog Run	Donburi Restaurant	Dive Spot
20	Khlong Toei	4	Recording Studio	Bar	Canal	Food Court	Bistro	Thai Restaurant	Karaoke Bar	Convenience Store	Farmers Market	Electronics Store
33	Prawet	4	Convenience Store	Comfort Food Restaurant	Food Stand	Halal Restaurant	Farmers Market	Food Court	Food & Drink Shop	Floating Market	Flea Market	Fast Food Restaurant
38	Saphan Sung	4	Thai Restaurant	Japanese Restaurant	Restaurant	Convenience Store	Stadium	Electronics Store	Food Court	Food & Drink Shop	Floating Market	Flea Market
41	Taling Chan	4	Coffee Shop	Soccer Field	Floating Market	Convenience Store	Farmers Market	Food Stand	Food Court	Food & Drink Shop	Flea Market	Fast Food Restaurant

Discussion

Most of the districts fall into cluster 3 followed by cluster 1, 4 and 5. Just 1 district falls into cluster 2. Therefore, we can believe cluster 3 is the most popular cluster followed by cluster 1, 4 and 5. The least popular cluster is cluster 2. It is a good opportunity to run the business in this clustering sequence: 3, (1, 4, 5), 2.

As most of the districts fall into cluster 3, let's take a closer look at cluster 3 first (see Cluster 3 result). We can see it is full of restaurants with different style cuisine. It should be very competitive when we run a Japanese restaurant here.

Then move on to cluster 1, it is found that it is full of noodle house, Asian restaurant, Chinese restaurant, Thai restaurant and cafe (see Cluster 1 result). People right there should love eating and Asian food. However, just a few Japanese restaurants are opened there. Therefore, cluster 1 is a good choice to open a Japanese restaurant.

Venues in cluster 4 and cluster 5 are too diverse with different categories. It poses more risk than cluster 1 as we are not sure if people right there love eating (see Cluster 4 & 5 results).

Cluster 2 just has one district so it should not be a good start due to low popularity (see Cluster 2 result).

To sum up, it is a good choice to open a Japanese restaurant in districts in cluster 1.

Conclusion

To answer where is the best place to open a Japanese restaurant in Bangkok in which popularity and competitiveness are considered concurrently for the targeted audience, we can see districts in cluster 1 are the best places to open a Japanese restaurant.

References

Understanding K-means Clustering in Machine Learning
https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1