Alvon Chua Kang Jin

Abstract

IDENTIFY LOCATION FOR NEW RESTAURANT OPENING IN SINGAPORE

Clustering the Neighborhood of Singapore

Capstone Project - The Battle of Neighborhoods 10 January 2021

Table of Contents

[Introduction 2](#_Toc61192774)

[The Business Problem 2](#_Toc61192775)

[Approach 2](#_Toc61192776)

[Data 3](#_Toc61192777)

[Data Sources 3](#_Toc61192778)

[Assumptions 3](#_Toc61192779)

[Methodology 4](#_Toc61192780)

[Results 5](#_Toc61192781)

[Discussion 6](#_Toc61192782)

[Shortlist & Drill Down Neighborhood 7](#_Toc61192783)

[Recommendation 9](#_Toc61192784)

[Conclusion 10](#_Toc61192785)

# Introduction

Singapore, a small but developed country, is well-known to be a food paradise in the Asian region, with huge and diverse cuisines available, owing to its unique multi-racial society and cosmopolitan culture. In Singapore, at most food centers or shopping malls, one can typically find cuisines with the likes of Chinese, Malay, India, Western, Japanese, Korean, Thai, Vietnamese and even Mexican! It is no surprise then, that numerous Food & Beverage (F&B) businesses have ventured into the Singapore F&B market, taking advantage of the country’s rich food culture and the people’s love for food.

## The Business Problem

In this project, we are playing a role of a Data Science Consultant advising our client, who is trying to venture into the Singapore F&B market by setting up their very first Fusion Food restaurant in Singapore.

However, given the highly competitive food market in Singapore, there is a need to identify a strategic location in Singapore where our new Fusion Food restaurant will be best positioned to attract customers. We aim to identify a location in Singapore, where restaurant competition can be minimized while ensuring adequate traffic for the business.

The identified location will be our recommendation to the client on which Neighborhood to set up their restaurant business.

## Approach

We will leverage the power of Machine Learning, i.e. using K-means Clustering technique, on the dataset to process the Neighborhood Location Data into multiple clusters, where in each cluster, we will then explore the neighborhood to unravel any useful patterns and subsequently, identify a neighborhood that can best set up our restaurant. The tools to be used for this analysis will be Python and Foursquare.

# Data

## Data Sources

Based on our Business Problem, the data that we will require is:

* Frequency of restaurants in each neighborhood
* Frequency of restaurant substitute in each neighborhood, i.e. Food Court, Coffee Shop

We will use the Singapore Neighborhood *sg.csv* dataset downloaded from <https://simplemaps.com/data/sg-cities> to identify the geo-coordinates of each neighborhood. Following each neighborhood, we will use the Foursquare API to retrieve nearby venues information centred around the neighborhood.

## Assumptions

We will make the following assumptions for this project:

1. Scope of neighborhood location will be restricted to the sg.csv dataset as the client is not interested in the Southern part of Singapore due to higher rental cost. Any Southern Neighborhood location data listed in the sg.csv dataset will be ignored
2. List of neighborhood locations in sg.csv dataset are assumed to show only locations in Singapore that are available for restaurant space rental
3. Scope of analysis will only cover the top three most common ‘nearby’ venues for each neighborhood as other venues not in the top three will be deemed as insignificant in impacting a restaurant’s competitiveness
4. ‘Nearby’ is defined as within a radius of 500m around the neighborhood, as any venues beyond 500m is deemed as insignificant in impacting a restaurant’s competitiveness

# Methodology

The focus of this project is to first cluster the data to understand any patterns found in each cluster of neighborhood and then identify the cluster(s) that best serve the setting up of a new restaurant. Within the cluster, we will drill down and identify the neighborhood that has the smallest restaurants’ or its substitutes presence. This is done as follows:

1. Firstly, we obtain the data by downloading it from the web in csv format and then process it to be ready for analysis.
2. Secondly, with the help of Foursquare API, we explore the neighborhood data by getting the nearby venue information of each neighborhood such as Venue Location and Venue Category. This will allows us to explore what are the typical venues found in a neighborhood.
3. Thirdly, we will use clustering technique to cluster the neighborhood data to find meaningful neighborhood venue patterns, followed by a data visualization of the different clusters. This technique is used because of the exploratory nature of the project, specifically, we want to explore and identify a pattern in each neighborhood, and determine whether this pattern is compatible to keeping a restaurant’s competition low.
4. Fourthly, for each cluster, we will drill down to analyze the count of the top three venue categories to determine the competition landscape of a restaurant.
5. Lastly, we will then identify a cluster that has the most compatible venue pattern with setting up of a new restaurant, i.e. low restaurant competition. This will be our main basis in recommending the location to our client.

# Results

The clustering of our neighborhood data resulted in 5 different clusters of various sizes as illustrated below. Refer to the Jupyter Notebook for more in-depth analysis of each cluster.

|  |  |  |  |
| --- | --- | --- | --- |
| **Cluster** | **Cluster Name** | **Cluster Description** | **Color** |
| 0 | ‘Bus Pooling’ Cluster | This cluster represents a neighborhood where one can commonly find Bus Stations and Swimming Pools as the most common venue found. There are restaurants found in the neighborhood, but they are not the most common venues. | Blue |
| 1 | Not meaningful to assign a name | Not meaningful to analyze this cluster since there is only 1 neighborhood. | Black |
| 2 | ‘Restaurant’ Cluster | This is the biggest cluster in the dataset, which indicates the most common venue pattern found in Singapore neighborhood. Typically, a neighborhood in this cluster has many restaurants followed by coffee shops and food court. | Red |
| 3 | ‘Restaurant Lite’ Cluster | A neighborhood that has many restaurants but fewer than Cluster 2. It is mostly populated with substitutes like Food Courts and Coffee Shops. | Purple |

*Table 1.0 – Summary of clusters formed by K-Means*

Map

Description automatically generated

*Figure 1.0 – Map of Singapore illustrating the neighborhood clusters in different colors*

# Discussion

As per the results shown in Table 1.0, Cluster 0, ‘Bus Pooling’ cluster seems to be the most appropriate cluster for setting up a new restaurant with the least competition in the neighborhood. Therefore, let’s dive deeper into Cluster 0.

|  |  |  |  |
| --- | --- | --- | --- |
| **Neighborhood** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** |
| Kampong Cutforth | Golf Course | Pool | Restaurant |
| Kampong Sungai Tengah | Breakfast Spot | Light Rail Station | Supermarket |
| Tay Keng Loon Estate | Bus Station | Factory | Food Court |
| Kampong Beremban | Pet Store | Resort | Farm |
| Kampong Bukit Panjang | Bus Station | Supermarket | Karaoke Bar |
| Yio Chu Kang Estate | Restaurant | Hotel Pool | Pool |
| Yio Chu Kang | Bus Station | Restaurant | Pool |
| Kampong Amoy Quee | Pool | Food Court | Tennis Court |
| Singapore United Plantation | Pool | College Cafeteria | BBQ Joint |
| Saint Michael’s Estate | Café | Soccer Stadium | Coffee Shop |
| Teacher’s Housing Estate | Coffee Shop | Bus Station | Office |
| Saga | Hotel | Pier | Factory |
| Kampong Kopit | Hotel | Pier | Factory |
| Kampong Siren | Military Base | Pool | Bus Station |
| Kampong Sungai Blukar | Soccer Field | Gym | Beer Garden |
| Yew Tee | Bus Station | Building | Office |
| Punggol | High School | Bus Line | Chinese Restaurant |
| Kampong Pasir Ris | Warehouse Store | Furniture / Home Store | Moving Target |
| Kampong Java Teban | Ski Area | Planetarium | Playground |
| Kampong Sungai Pandan | Baseball Stadium | Steakhouse | Food & Drink Shop |
| West Coast Village | Baseball Stadium | Steakhouse | Food & Drink Shop |

*Table 2.0 – Cluster 0’s top three most common venue categories for each neighborhood*

|  |  |
| --- | --- |
|  | Restaurant (including Hotel which assumed to have Restaurants) |
|  | Restaurant Substitutes such as Food Court, Coffee Shops, Café and Bars |

## Shortlist & Drill Down Neighborhood

From Table 2.0, we can observe that there are four neighborhoods that are not highlighted:

* Kampong Siren
* Yew Tee
* Kampong Pasir Ris
* Kampong Java Teban

Now, let’s try to further differentiate these four neighborhood in order to single out the best neighborhood for our client.

Starting from Kampong Siren, it is worth noting that this is a location that is near a Military Air Base, where we would expect lower traffic in the area. Hence, it will not be wise to set up a restaurant in here.

Map

Description automatically generated

*Figure 2.0 – A map of ‘Kampong Siren’ neighborhood*

Moving to the second neighborhood, ‘Yew Tee’, at the west side of the neighborhood we have a pretty densely populated residential and schools area, while on the east side, there is a Train Depot. This looks like a promising candidate for our neighborhood of choice. We will keep this as an option.

Map

Description automatically generated

*Figure 2.1 – A map of ‘Kampong Siren’ neighborhood*

Moving to the third neighborhood, ‘Kampong Pasir Ris’, this neighborhood consists of a mixture of factory, offices, residential buildings, schools as well as a highway. (indicated by the red road). This means that there is much more activity going on in this neighborhood and to take it further, we even have a highway connected to this neighborhood, making it more accessible by the people. We would expect many traffic in this area. This will be our number one option for now.

Application, map

Description automatically generated

*Fig. 2.2 – A map of ‘Kampong Pasir Ris’ neighborhood*

To the last neighborhood, ‘Kampong Java Teban’, we can see from the map below that it is mostly a recreational park area with some residential buildings on the east. We would not expect many visitors in a park, hence, this neighborhood will not be considered as an option.

Diagram, map

Description automatically generated

*Fig. 2.3 – A map of ‘Kampong Java Teban’ neighborhood*

## Recommendation

Based on the discussion above, the recommended neighborhood to start a new restaurant is ‘Kampong Pasir Ris’. It has a small presence of restaurants or other food centers; there are many social activities going on in the area, example, Factories, Offices and Schools, as well as an enhanced accessibility due to the connection of the highway.

Application, map

Description automatically generated

*Fig. 2.4 – A map of ‘Kampong Pasir Ris’ neighborhood (Recommended Neighborhood)*

# Conclusion

The power of Machine Learning has enabled us to perform cluster analysis, which segmentizes and explores our neighborhood data efficiently as deemed fit by the K-Means model. On the basis of restaurant presence alone, we identify Cluster 0 as the most appropriate cluster for setting up a new restaurant. Subsequently, further analysis of Cluster 0 is conducted to determine the most appropriate neighborhood. This is done by specifically analyzing each of the neighborhood, in terms of, restaurant presence and finally the volume of traffic.

All in all, identifying a location for a restaurant is no simple task as it actually requires more rigorous analysis of the data and many other factors should be taken into account such as taste preferences of the neighborhood, income distribution of the neighborhood, rental cost, customer traffic etc. However, this project serves as a good foundation for us to shortlist neighborhood in Singapore for a new restaurant set up, solely based on the restaurant competition landscape.