

Coursera Capstone

IBM Applied Data Science Capstone

**Opening a New Vegetarian/Vegan Restaurant in
Singapore**

By: Gunawan

July 2020



Introduction

Long gone are the days where you'll get a confused look accompanied by "huh?" or "what?" when you talk about the vegan lifestyle. Vegans abstain from any animal products whereas vegetarians still consume dairy and eggs. While some people give up animal products for moral or environmental reasons, many do so to adopt a healthier lifestyle. Scientific research has shown that vegetarians are generally less prone to 4 major ailments: Cancer, Diabetes, Meat-based illnesses and Cardiovascular diseases. Over the last few years, climate change has been a hot topic and the push towards veganism has been sprouting alongside it.

Veganism is gaining ground in Singapore. Singapore was named sixth most vegan-friendly city in the world by HappyCow. There are currently over 700 restaurants offering vegan and vegetarian dining options in Singapore, according to the HappyCow app, which lists vegan and vegetarian restaurants and health food stores around the world.

With PETA Asia declaring Singapore the second most vegan-friendly city in the region, just losing out to Taiwan's capital Taipei in the top spot, and beating Bangalore, Bangkok and Bali, which all made the top 10. In addition to a growing appetite for plant-based innovations on the island, the boom in vegan, alternative and inclusive eateries in the city-state makes it a natural pathway for those wishing to enter the Asia market. We believe there is room for opportunity to start a new vegetarian / vegan restaurant in Singapore.

Business Problem

The objective of this capstone project is to analyse and select the best districts in Singapore to open a new vegetarian/vegan restaurant.

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question:

In Singapore, if entrepreneurs are looking to open a new vegetarian/vegan restaurant, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to entrepreneurs looking to open a new vegetarian / vegan restaurant in Singapore.

Data

To solve the problem, we will need the following data:

- List of districts in Singapore.
- Latitude and longitude coordinates of those districts. This is required in order to plot the map and to get the venue data.
- Venue data, particularly data related to vegetarian/vegan restaurants. We will use this data to perform clustering on the districts.

Sources of data and methods to extract them

This keylocations page (<https://keylocation.sg/singapore/districts-map>) contains a list of districts in Singapore, with a total of 28 districts. We will use web scraping techniques to extract the data from the keylocations page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the districts using Python Geocoder package which will give us the latitude and longitude coordinates of the districts.

After that, we will use Foursquare API to get the venue data for those districts. Foursquare has one of the largest database of 105+ million places and is used by over 150,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the vegetarian/vegan restaurant category in order to help us to solve the business problem put forward.

This is a project that will make use of many data science skills, from web scraping (keylocations page), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

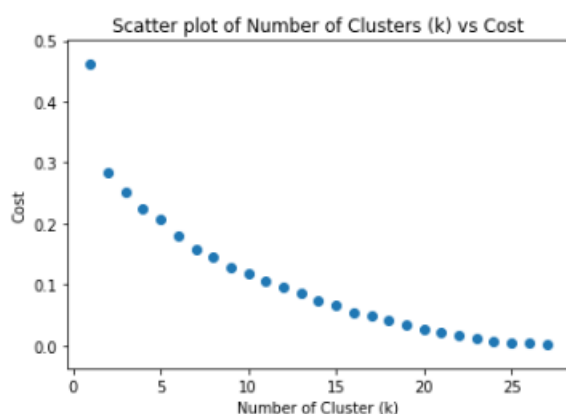
In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

Methodology

Firstly, we need to get the list of districts in Singapore. We will be using the list from This keylocations page (<https://keylocation.sg/singapore/districts-map>). We will do web scrapping using Python requests and beautifulsoup packages to extract the list of district data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the districts in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Singapore.

Next, we will use Foursquare API to get the top 200 venues that are within a radius of 4000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare by passing in the geographical coordinates of the districts in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude, and venue longitude. With the data, we can check how many venues are returned for each district and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each district by grouping the rows by district and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. We will be using top 20 venue categories of the district to be used for the clustering. We also extract the mean of the frequency of occurrence of "Vegetarian / Vegan Restaurant" as additional column for our analysis.

Lastly, we will implement the "elbow" method of selecting the optimal number of clusters for K-means clustering. K-means is a simple unsupervised machine learning algorithm that groups data into a specified number (k) of clusters while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project.



We will use elbow = 7 to cluster the districts into 7 clusters based on their frequency of occurrence for top 20 venues. The results will allow us to identify which cluster have higher / lower concentration of "Vegetarian / Vegan Restaurant". Based on the occurrence of "Vegetarian / Vegan Restaurant" in the districts, it will help us to answer the question as to which districts are most suitable to open new "Vegetarian / Vegan Restaurant"

Results

The results from the k-means clustering show that we can categorize the districts into 7 clusters based on the frequency of occurrence for top 20 most common venues category

Cluster 0 (tourist district in Singapore surrounded by hotels, shopping malls, and different types of restaurants including Vegetarian / Vegan)

```
In [311]: # Cluster 0 (tourist district in Singapore surrounded by hotels, shopping malls, and different types of restaurants including Vegetarian / Vegan)
sg_merged.loc[sg_merged['Cluster Labels'] == 0, sg_merged.columns[0] + [1] + list(range(5, sg_merged.shape[1]))]]
```

```
Out[311]:
```

	District	Area	Vegetarian / Vegan Restaurant	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	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue	16th Most Common Venue	17th Most Common Venue	18th Most Common Venue	19th Most Common Venue	20th Most Common Venue	
7	0	Little India	0.00	0	Hotel	Shopping Mall	Japanese Restaurant	Coffee Shop	Whisky Bar	Cosmetics Shop	Movie Theater	Café	Cocktail Bar	Performing Arts Venue	Sushi Restaurant	Park	Dessert Shop	Indian Restaurant	Wine Bar	Ice Cream Shop	Art Gallery	Yoga Studio	Italian Restaurant	Bubble Tea Shop	
8	9	Orchard	0.01	0	Hotel	Japanese Restaurant	Shopping Mall	Cosmetics Shop	Park	Whisky Bar	Cocktail Bar	Indian Restaurant	Sushi Restaurant	Dessert Shop	Coffee Shop	Café	Clothing Store	Art Gallery	Yoga Studio	Wine Bar	Performing Arts Venue	Italian Restaurant	Shopping Mall	Seahood Restaurant	
9	10	Tanglin	0.00	0	Hotel	Japanese Restaurant	Shopping Mall	Sushi Restaurant	Bakery	Clothing Store	Cocktail Bar	Café	Cosmetics Shop	Coffee Shop	Garden	Wine Bar	Seahood Restaurant	Dessert Shop	Movie Theater	Dumpling Restaurant	Yoga Studio	Bar	Chinese Restaurant	Modern European Restaurant	
10	11	Newton	0.01	0	Hotel	Japanese Restaurant	Shopping Mall	Bakery	Cocktail Bar	Sushi Restaurant	Chinese Restaurant	Clothing Store	Cocktail Bar	Cosmetics Shop	Park	Bar	Dessert Shop	Dumpling Restaurant	Coffee Shop	Movie Theater	French Restaurant	Garden	Boutique	Supermarket	Hotel
11	12	Toa Payoh	0.01	0	Hotel	Coffee Shop	Supermarket	Japanese Restaurant	Shopping Mall	Bakery	Clothing Store	Cocktail Bar	Cosmetics Shop	Ice Cream Shop	Sushi Restaurant	Dessert Shop	Pool	Chinese Restaurant	Seahood Restaurant	Flower Shop	Snack Place	Indian Restaurant	Whisky Bar	Boutique	

Cluster 1 (east area of singapore surrounded by coffee shop, cafe, and different types of restaurants including Vegetarian / Vegan)

[312]: # Cluster 1 (east area of singapore surrounded by coffee shop, cafe, and different types of restaurants including Vegetarian / Vegan)

sg_merged.loc[sg_merged['Cluster Labels'] == 1, sg_merged.columns[0] + [1] + list(range(5, sg_merged.shape[1]))]]

Out[312]:

District	Area	Vegetarian / Vegan Restaurant	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	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue	16th Most Common Venue	17th Most Common Venue	18th Most Common Venue	19th Most Common Venue	20th Most Common Venue	
12	13	Central East	0.02	1	Chinese Restaurant	Hotel	Food Court	BBQ Joint	Cafe	Coffee Shop	Ice Cream Shop	Dessert Shop	Asian Restaurant	Soup Place	Snack Place	Clothing Store	Thai Restaurant	Supermarket	Climbing Gym	Pool	Convenience Store	Seahood Restaurant	Park	
13	14	Eunos	0.02	1	Chinese Restaurant	Coffee Shop	Noodle House	Food Court	Bar	BBQ Joint	Asian Restaurant	Dessert Shop	Italian Restaurant	Hotel	Vegetarian / Vegan Restaurant	Bakery	Climbing Gym	Japanese Restaurant	Beach	Vietnamese Restaurant	Cafe	French Restaurant	Middle Eastern Restaurant	Mexican Restaurant
14	15	East Coast	0.02	1	Chinese Restaurant	Coffee Shop	Noodle House	Bar	Dessert Shop	Hotel	BBQ Joint	Food Court	Italian Restaurant	Asian Restaurant	Beach	Vegetarian / Vegan Restaurant	Cafe	Yoga Studio	Bakery	Sporting Goods Shop	Japanese Restaurant	Multiplex	Park	Vietnamese Restaurant
15	16	Upper East Coast	0.01	1	Chinese Restaurant	Coffee Shop	Dessert Shop	Cafe	Beach	Food Court	Bakery	Ice Cream Shop	Asian Restaurant	Seahood Restaurant	Indian Restaurant	Park	Italian Restaurant	Hotel	Pizza Place	Supermarket	American Restaurant	Sporting Goods Shop	Yoga Studio	Thai Restaurant
16	19	North East	0.03	1	Chinese Restaurant	Coffee Shop	Cafe	Dessert Shop	Bakery	Asian Restaurant	Noodle House	Ice Cream Shop	Trail	Thai Restaurant	Supermarket	Park	Vegetarian / Vegan Restaurant	Food Court	Italian Restaurant	Clothing Store	Diner	Breakfast Spot	Fish & Chips Shop	Pizza Place

Cluster 2 (west area of singapore surrounded by bakery, cafe, park, and different types of restaurants including Vegetarian / Vegan)

```
[313]: # Cluster 2 (west area of singapore surrounded by bakery, cafe, park, and different types of restaurants including Vegetarian / Vegan)
sg_merged.loc[sg_merged['Cluster Labels'] == 2, sg_merged.columns[0] + [1] + list(range(5, sg_merged.shape[1]))]]
```

Out[313]:

District	Area	Vegetarian / Vegan Restaurant	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	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue	16th Most Common Venue	17th Most Common Venue	18th Most Common Venue	19th Most Common Venue	20th Most Common Venue	
2	3	Central South	0.00	2	Chinese Restaurant	Park	Café	Japanese Restaurant	Seahood Restaurant	Scenic Lookout	Trail	Boutique	Resort	Cocktail Bar	Coffee Shop	Performing Arts Venue	Food Court	Bar	Modern European Restaurant	Italian Restaurant	Asian Restaurant	Gym / Fitness Center	Greek Restaurant	Gourmet Shop
4	5	Southern Village	0.00	2	Bakery	French Restaurant	Ice Cream Shop	Garden	Chinese Restaurant	Pizza Place	Italian Restaurant	Coffee Shop	Park	Performing Arts Venue	Café	Japanese Restaurant	Multiplex	Food Court	Seahood Restaurant	Bar	Indian Restaurant	Gourmet Shop	Asian Restaurant	Gym / Fitness Center
20	21	Central West	0.00	2	Bakery	Japanese Restaurant	Indian Restaurant	Park	Chinese Restaurant	Italian Restaurant	Dessert Shop	French Restaurant	Food Court	Pizza Place	Café	Ice Cream Shop	Trail	Coffee Shop	Multiplex	Performing Arts Venue	Playground	Thai Restaurant	Kitchen Supply Store	Korean Restaurant
22	23	North West	0.02	2	Chinese Restaurant	Bakery	Coffee Shop	Japanese Restaurant	Food Court	Park	Supermarket	Nature Preserve	Indian Restaurant	Italian Restaurant	Asian Restaurant	Bar	Scenic Lookout	Vegetarian / Vegan Restaurant	Ice Cream Shop	Trail	Department Store	Dessert Shop	Bowling Alley	BBQ Joint

Cluster 3 (central business district area of singapore surrounded by hotel, cafe, and different types of restaurants including Vegetarian / Vegan)

</

Cluster 4 (far east area of singapore near airport)

```
In [316]: # Cluster 4 (far east area of singapore near airport)
sg_merged.loc[sg_merged['Cluster Labels'] == 4, sg_merged.columns[0] + [1] + list(range(5, sg_merged.shape[1]))]]

Out[316]:
```

District	Area	Vegetarian / Vegan Restaurant	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	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue	16th Most Common Venue	17th Most Common Venue	18th Most Common Venue	19th Most Common Venue	20th Most Common Venue	
16	17	Far East	0.0	4	Airport Lounge	Café	Golf Course	Boutique	General Entertainment	Asian Restaurant	Coffee Shop	Candy Store	Canal	Supermarket	Electronics Store	Hotel	Airport	Housing Development	Hobby Shop	Smoke Shop	Burger Joint	Garden	Ice Cream Shop	Shoe Store

Cluster 5 (far north west area + north east of singapore surrounded by coffee shop, cafe, food court, and different kind of restaurants excluding Vegetarian / Vegan)

```
In [317]: # Cluster 5 (far north west area + north east of singapore surrounded by coffee shop, cafe, food court, and different kind of restaurants excluding Vegetarian / Vegan)
sg_merged.loc[sg_merged['Cluster Labels'] == 5, sg_merged.columns[0] + [1] + list(range(5, sg_merged.shape[1]))]
```

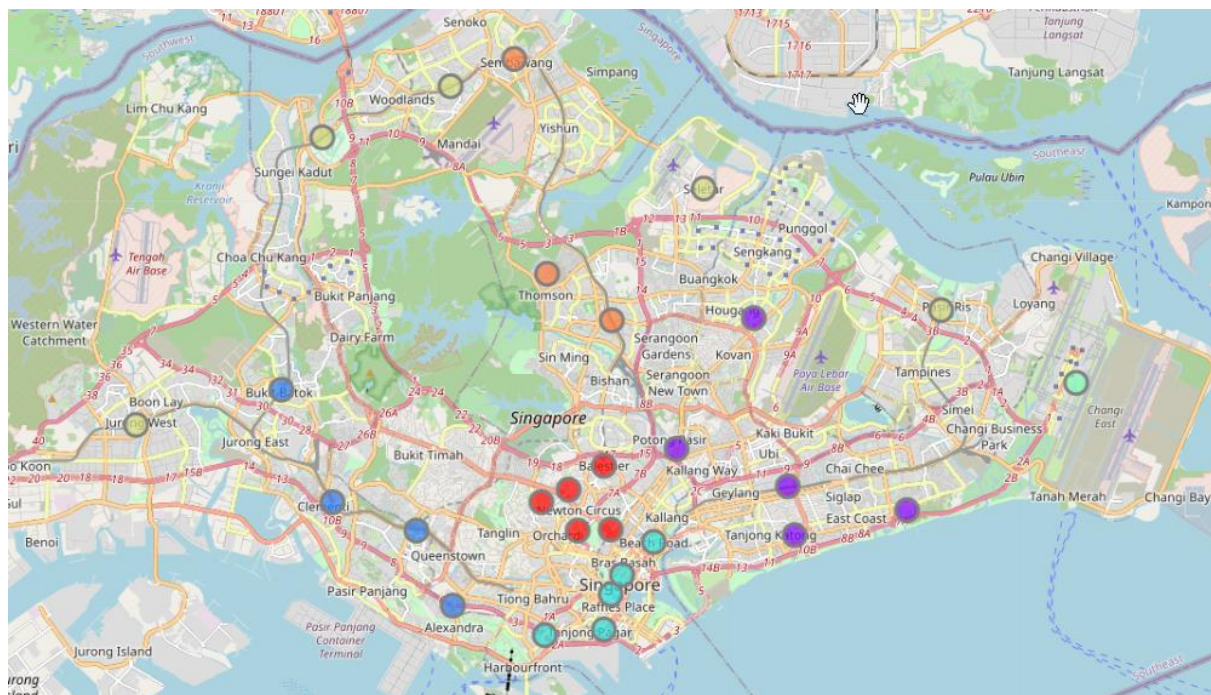
Out[317]:

District	Area	Vegetarian / Vegan Restaurant	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	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue	16th Most Common Venue	17th Most Common Venue	18th Most Common Venue	19th Most Common Venue	20th Most Common Venue		
17	18	Tampines	0.0	5	Coffee Shop	Park	Cafe	Bakery	Thai Restaurant	Supermarket	Furniture / Home Store	Gym	Japanese Restaurant	Fast Food Restaurant	Food Court	Steakhouse	Dessert Shop	Fried Chicken Joint	Clothing Store	Asian Restaurant	Grocery Store	Seafood Restaurant	Scandinavian Restaurant	Sandwich Place	
21	22	Far West	0.0	5	Fast Food Restaurant	Chinese Restaurant	Food Court	Coffee Shop	Asian Restaurant	Sandwich Place	Zoo Exhibit	Exhibit	Cafe	Supermarket	Garden	Japanese Restaurant	Noodle House	Park	Pool	Seafood Restaurant	Shopping Mall	Indian Restaurant	Bookstore	Snack Place	
23	24	Far North West	0.0	5	Cafe	Coffee Shop	Chinese Restaurant	Sandwich Place	Park	Malay Restaurant	Clothing Store	Ice Cream Shop	Thai Restaurant	Zoo	Food Court	Frozen Yogurt Shop	Japanese Restaurant	Juice Bar	Indian Restaurant	Boutique	Snack Place	Bakery	Hainan Restaurant	Monument / Landmark	Movie Theater
24	25	Far North	0.0	5	Chinese Restaurant	Coffee Shop	Park	Food Court	Bar	Indian Restaurant	Fast Food Restaurant	Thai Restaurant	Asian Restaurant	Supermarket	Nightclub	Clothing Store	Cafe	Sandwich Place	Malay Restaurant	Frozen Yogurt Shop	Seafood Restaurant	Exhibit	Bistro	American Restaurant	
27	28	North East	0.0	5	Food Court	Coffee Shop	Cafe	Fast Food Restaurant	Supermarket	Sandwich Place	Park	Asian Restaurant	Noodle House	Chinese Restaurant	Japanese Restaurant	Bar	Pool	South Restaurant	Restaurant	Seafood Restaurant	Shopping Mall	Diner	Bridge	Club House	

Cluster 6 (north area of singapore surrounded by coffee shop, cafe, food court, park and different kind of restaurants including Vegetarian / Vegan)

</

The results of the clustering are visualized in the map below with cluster 0 in red, cluster 1 in purple, cluster 2 in blue, cluster 3 in light blue, cluster 4 in aquamarine, cluster 5 in lime, and cluster 6 in orange colour.



Discussion and recommendation

Based on the clustering result above. We will recommend entrepreneur to focus on cluster 5 which are suburban areas without much presence of vegetarian / vegan restaurants.

1) District 18 (Pasir Ris, Simei, Tampines)

District 18 consists of Simei, Tampines and Pasir Ris – residential areas with numerous amenities and facilities within the vicinity. A great and convenient place to live, it is located in the eastern part of Singapore.

2) District 22 (Boon Lay, Jurong, Tuas)

Located to the West of Singapore, District 22 is mainly an industrial and residential area with landscaped parks. In this District you can also find a golf course and Singapore's only ice-skating rink. District 22 comprises of Boon Lay, Jurong and Tuas. It is densely populated with many industrial buildings, contributing to Singapore's booming economy.

3) District 24 - Kranji, Lim Chu Kang, Tengah

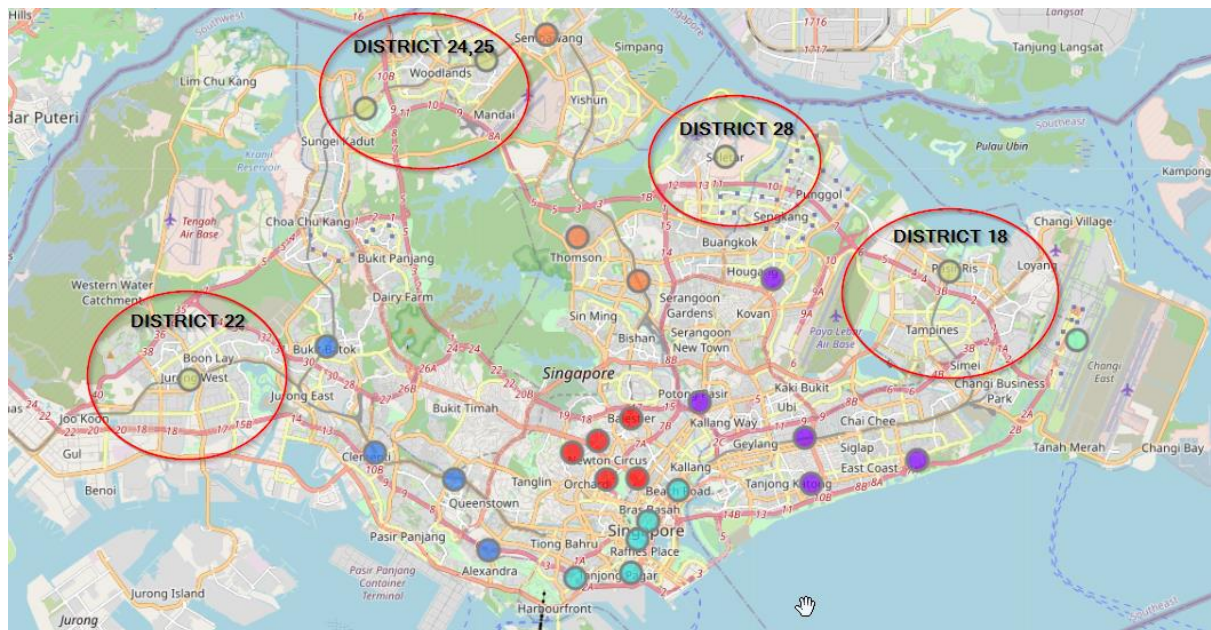
Located in the northwestern part of Singapore, District 24 is occupied fully by Lim Chu Kang. This area is characterized by its abundant foliage, the few remaining farms in Singapore, a jetty, a forest reserve, numerous campsites and cemeteries.

4) District 25 (Admiralty, Woodlands)

From District 25's Woodlands, you can actually see Malaysia. There is the Causeway bridge which links Singapore and Malaysia together. District 25's Kranji is an industrial area with a rich cultural history and the location of Singapore's only racecourse.

5) District 28 - Yio Chu Kang, Seletar

Comprising of Seletar and part of Yio Chu Kang of District 26, District 28 is a historical and residential area that holds many heritage and cultural stories. The British naval and air bases used to be located here. In the future, the Seletar Aerospace Park will be built to serve Singapore's aviation industries.



To verify our result, we go to HappyCow website to get their veggiemap using location=Singapore&filters=vegan-vegetarian

<https://www.happycow.net/veggimap/?location=singapore&filters=vegan-vegetarian&zoom=5&clat=1.3147307000000001&clng=103.84701910000001>



It is indeed showing similar result compare to our recommendation that District 18, 22, 24, 25 are having the lowest concentration of vegetarian / vegan restaurants.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 7 clusters based on their similarities, and lastly providing recommendations to entrepreneur regarding the best locations to open a new vegetarian / vegan restaurant. To answer the business question that was raised in the introduction section, the answer proposed by this project is the districts in cluster 5 are the most preferred locations to open a new vegetarian / vegan restaurant. The findings of this project will help the entrepreneur to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new vegetarian / vegan restaurant.

Limitation and Future Work

In this project, we only consider one factor i.e. frequency of occurrence of vegetarian / vegan restaurant, there are other factors such as population, race, religion and income of residents that could influence the location decision of a new vegetarian / vegan restaurant. However, to the best knowledge of this researcher such data are not available to the district level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new vegetarian / vegan restaurant. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

References

<https://thehoneycombers.com/singapore/vegan-scene-in-singapore/>

<https://www.todayonline.com/daily-focus/veganism-gaining-ground-singapore>

<https://www.asiaone.com/lifestyle/can-you-save-money-becoming-vegetarian-singapore>

<https://www.petaasia.com/living/food/top-10-vegan-friendly-cities-asia/>

<https://www.happycow.net/asia/singapore/?filters=vegan-vegetarian>

<https://www.happycow.net/veggieemap/?location=singapore&filters=vegan-vegetarian&zoom=5&clat=1.3147307000000001&clng=103.84701910000001>

https://www.iproperty.com.sg/resources/District_Guide.aspx

<https://keylocation.sg/singapore/districts-map>

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