Capstone Project - The Battle of the Neighborhoods

Finding Optimal Locations to Open an Indian Restaurant in Singapore

Prepared by: Akshaya Suresh

Introduction/Business Problem

- ➔ Singapore is home to a wide range of cultures, ethnicities and religions, making it a very attractive hub for restaurateurs who are contemplating to open a restaurant serving ethnic cuisine.
- ➔ In this project, we aim to find the **most optimal location to recommend to stakeholders who are planning to open an Indian restaurant in Singapore.**
- ➔ We will define an optimal restaurant location based on the following criteria:
  - ◆ Competition - Being close to established competition may help with business marketing, but if the new restaurant is too close to its competition, it may have a tough time gaining a foothold in the community. Hence, **we prefer neighborhoods that don't already have many Indian restaurants but at the same time don’t have too few restaurants.**
  - ◆ Accessibility - We need to consider the amount of accessibility to make it as easy as possible for customers to visit the establishment. We should also keep in mind that tourists tend to visit eateries that are in or around the city center. Hence, **we prefer more centrally located neighborhoods.**

Data

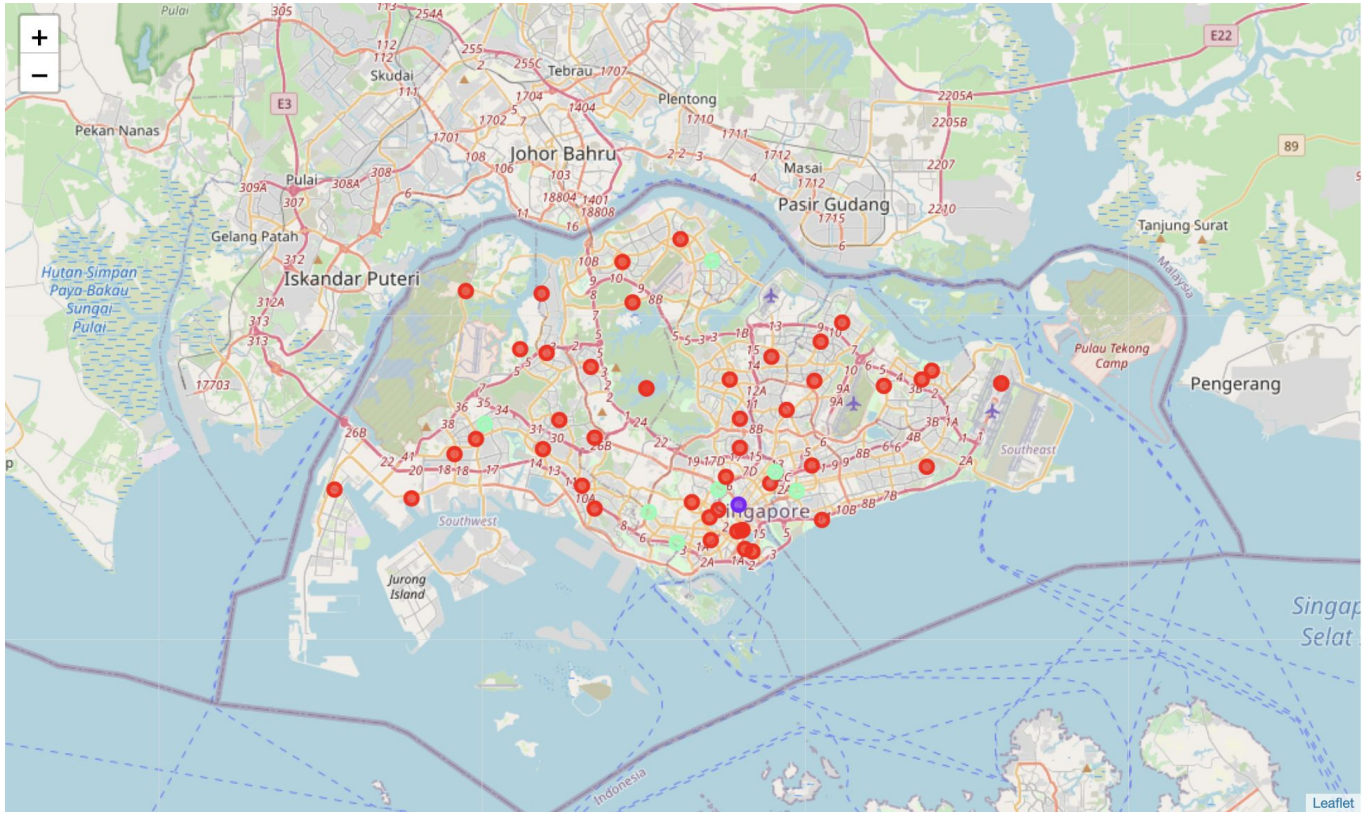
- ➔ Wikipedia page titled "Planning Areas of Singapore"
  - ◆ We will scrape this Wikipedia page, wrangle the data and clean it to prepare the dataset of **neighborhoods and their corresponding regions** in Singapore.
- ➔ Geocoder Python package
  - ◆ We then proceed to obtain the **latitude and the longitude coordinates** of each neighborhood.
- ➔ Foursquare API
  - ◆ Finally, we will **analyze the venue data, cluster the neighborhoods with respect to the number of Indian restaurants, and finally examine the results** to make recommendations based on the data.

Methodology

1. We scrape the Wikipedia page using *Beautiful Soup*, wrangle and clean it to obtain a dataset of all the neighborhoods in Singapore and their corresponding regions.
2. We use the *Geocoder Python package* to get the geographical coordinates of each neighborhood in our list.
3. We utilize *Foursquare API* to explore the neighborhoods. We first get the top 100 venues that are in each neighborhood within a radius of 500 meters.
4. We make use of *one hot encoding* to convert the categorical venue data to integer data. After which, we group the rows by neighborhood and by taking the mean of the frequency of occurrence of each category. We can finally create a dataframe to view only the mean of the frequency of Indian restaurants in each neighborhood.
5. We use *k-means clustering* to cluster the neighborhoods in Singapore into three clusters.
6. We can visualize the resulting clusters by creating a map of Singapore with the various clusters superimposed on top, using the *Folium package*.

Results

- The three clusters are:
  - Cluster 0 - Neighborhoods with a **low number** of Indian restaurants
  - Cluster 1 - Neighborhoods with a **high number** of Indian restaurants
  - Cluster 2 - Neighborhoods with a **medium number** of Indian restaurants



Legend

- Cluster 0 (Low number of Indian Restaurants)
- Cluster 1 (High number of Indian Restaurants)
- Cluster 2 (Medium number of Indian Restaurants)

Discussion

Competition

- Cluster 1 (Neighborhoods with a **high number** of Indian restaurants)
  - This cluster only consists of the neighborhood Rochor.
  - This makes sense as Little India, a buzzing historic area that shows off the best of Singapore's Indian community, is situated in Rochor.
  - This cluster is likely already suffering from high competition and would not be a good choice for restaurateurs to pick as a location for a new Indian restaurant.

- Cluster 0 (Neighborhoods with a **low number** of Indian restaurants)
  - These neighborhoods may be considered by stakeholders.
  - But more research should be done in order to examine these neighborhoods further
  - There might be a range of reasons as to why they have been unpopular venues for Indian restaurants.

- Cluster 2 (Neighborhoods with a **medium number** of Indian restaurants)
  - An indication that these neighborhoods have been viable options for restaurateurs to conduct business in.
  - More detailed analysis needs to be done to find an optimal location that has taken all factors into account.

Discussion

Accessibility

➔ Since we prefer more centrally located neighborhoods in order to cater to tourists, we give preference to the neighborhoods in cluster 2 (which has the most ideal amount of competition as discussed above) that are situated in the central location. These neighborhoods are **Bukit Merah, Geylang, Marine Parade, Newton and Queenstown.**

	Neighborhood	Region	Latitude	Longitude	Cluster Labels	Indian Restaurant
3	Boon Lay	West	1.34805	103.71216	2	0.083333
5	Bukit Merah	Central	1.28307	103.81667	2	0.076923
14	Geylang	Central	1.31147	103.88218	2	0.075000
23	Marine Parade	Central	1.32144	103.87004	2	0.125000
25	Newton	Central	1.31218	103.83912	2	0.041667
34	Queenstown	Central	1.29966	103.80172	2	0.043478
54	Yishun	North	1.43621	103.83582	2	0.058824

Conclusion

Limitation

- ➔ All 55 neighborhoods in our dataset are of different areas.
- ➔ Some neighborhoods are each less than 1 square kilometer in area, while some are each more than 20 square kilometers in area.
- ➔ Hence, using a 500 meter radius to gather venue data from each of the 55 neighborhoods is problematic as they are all of unequal areas.
- ➔ This could be mitigated by creating a hexagonal grid of area candidates, equally spaced and centered around the central region of Singapore (instead of using the list of neighborhoods from Wikipedia as our area candidates).

The final decision on an optimal location for a new Indian restaurant should be made by stakeholders based on specific characteristics of every recommended neighborhood in this analysis, taking into consideration additional factors like the demographics of the neighborhood, proximity to parking lots, surrounding traffic patterns, real estate availability and price.