2020

The Battle of Neighbourhoods: Where to Open a Restaurant in Singapore?

IBM Applied Data Science Capstone

Sng Yu Xin

4/2/2020

1. Introduction

1.1. Background

Singapore, being a multi-cultured country, is famous for having various type of cuisines. There are dozens of famous food bloggers in Singapore who would try out different food outlets in Singapore and post their reviews. Each of these food bloggers could have up to hundred thousands of followers, who would not mind travelling a little further than usual to try out the tasty food.

Tourists would also travel to Singapore with the intention to try out the various type of cuisines available to complete their travel experience to Singapore.

1.2. Business Problem

Based on the statistics shared in 2016, there are a total of 7,679 food and beverage establishments, where 2,712 of them were restaurants. The operating receipts of Singapore's food and beverage establishments totalled up to $9,023 million of which $3,380 million were contributed by the restaurants.

With over 2,000 restaurants available in Singapore, it would be a tough decision for new restaurant owner to decided where to open a new restaurant that would yield profit. Therefore, the objective of this project is to analyse and recommend the best location in Singapore to open a Japanese restaurant.

Few points of consideration include, but not limited to:

- Number of restaurants in the vicinity

1.3. Target Audience

This project would be useful for entrepreneurs who have the intention to open a food establishment in Singapore, helping entrepreneurs to make informed decision when deciding the location of the restaurant.

2. Data

The following data sources would be used for this project.

2.1. Names of Train Station

Data would be retrieved from Singapore's Public Data via the following link, https://data.gov.sg/dataset/train-station-chinese-names?resource_id=65c1093b-0c34-41cf-8f0e-f11318766298. This table would be required to establish the possible estates for a food establishment to be opened in Singapore.

2.2. Coordinates of locations

The geographical coordinates, latitude and longitude, of each station would then be obtained using Python Geocoder package.

2.3. Venue Data nearby locations

Foursquare API would subsequently be used to search for food establishments nearby the locations. This data would allow target audience to better understand the potential competitors should the restaurant be decided to established in the area.

3. Methodology

3.1. Methodology

Firstly, the csv file that contained the names of train stations was obtained from the Singapore's Public Data and loaded with Pandas. Stations not connected to major rail lines were dropped with the intent to focus only on areas with relatively higher human traffic. Subsequently, duplicated station names were removed. In addition, only column that contained the names of train station was retained to serve as the foundation of the dataframe.

Secondly, the geographical coordinates of each location were obtained using the Geocoder package which allowed the conversion of address into geographical coordinates. The geographical coordinates were input into the dataframe that contained the names of train

station. The geographical coordinates were essential to aid with the query from Foursquare API required to be performed.

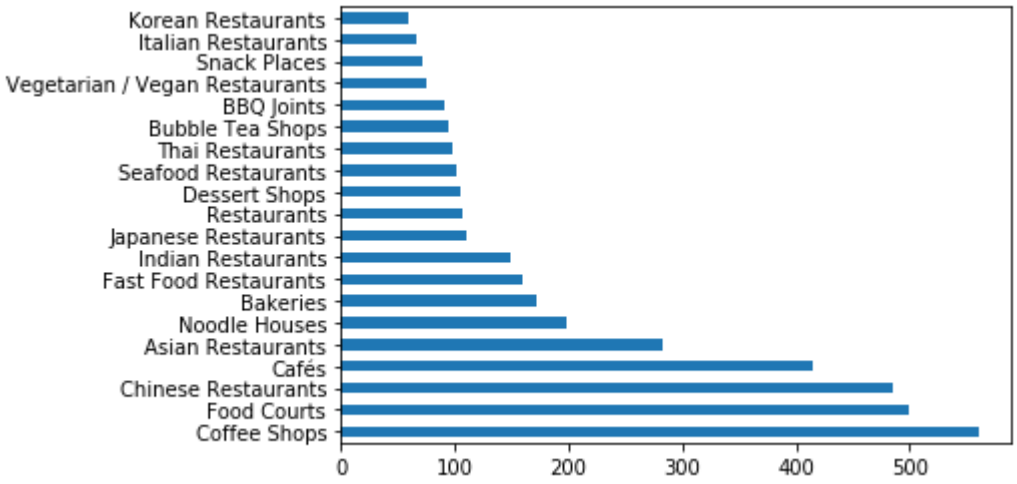
Thirdly, the locations were plotted on a map with the help of folium package. This step aided in the visual verification of the geographical coordinates obtained in the second step.

Next, food establishments within 500 meters of each location were searched using Foursquare API using the food category ID and passing the geographical coordinates of the locations in a loop and storing each venue into a list and subsequently a dataframe. Subsequently, the focus was shifted to only Japanese Restaurants.

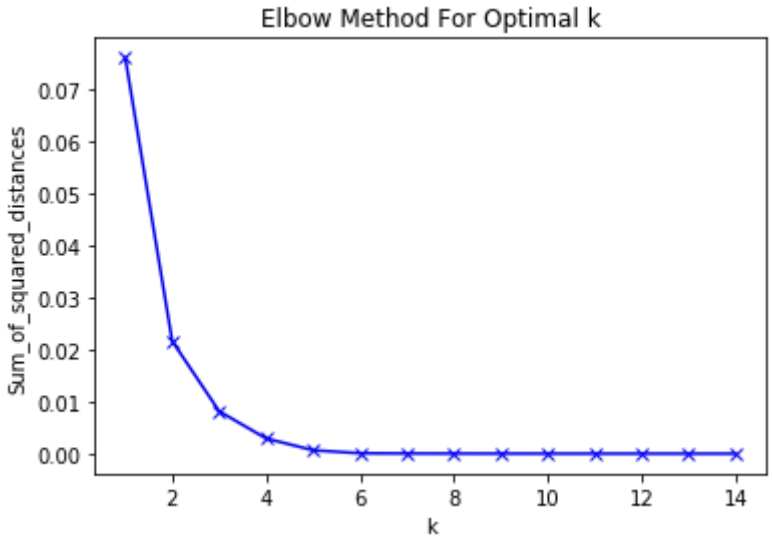
Lastly, k-Means clustering was used to cluster the locations. K-Means clustering was used as it would help to group unsupervised data based on the similarity of the objects which was aligned with the objective of this project.

3.2. Analysis

All food establishments within 500 meters of each location were searched using Foursquare API and the following chart illustrated the top 20 categories of food establishments obtained.

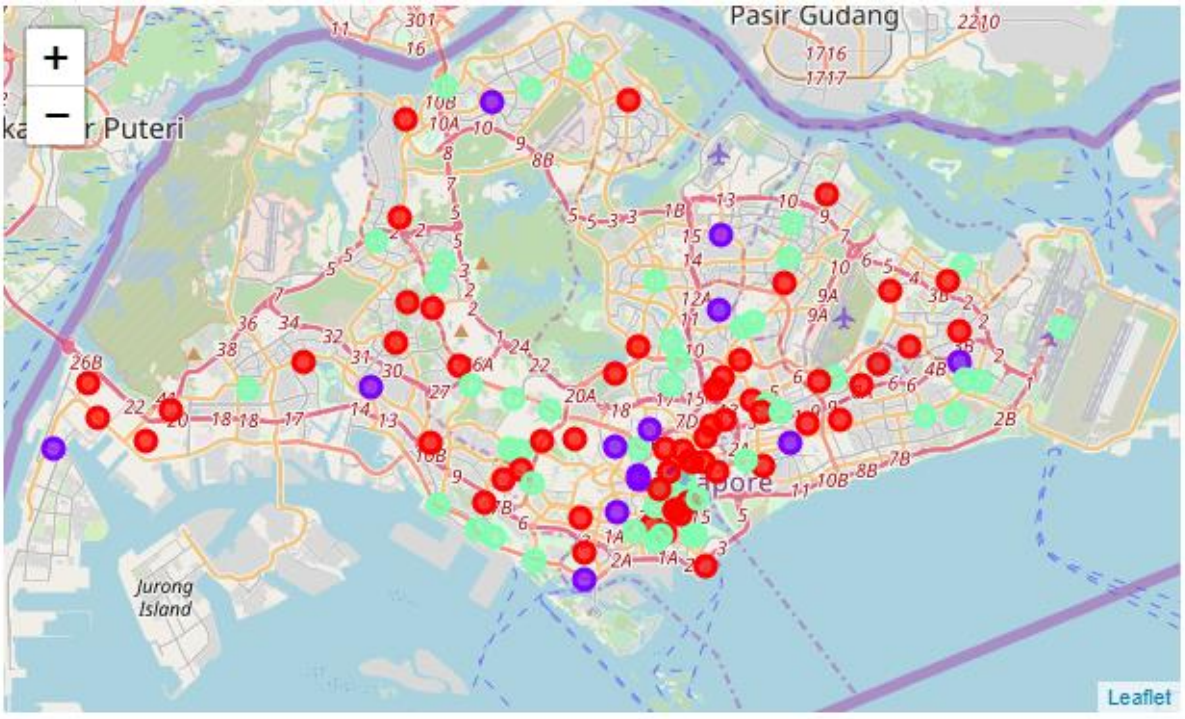


K-Means clustering was performed twice, first on all food establishments identified and secondly, focusing only on venue category named Japanese Restaurants. In this section, only the k-Means clustering performed on Japanese Restaurants would be discussed as it was the main focus of this study.



The optimal value for k was obtained using the elbow method. The value for k was determined to be 3 as the error would not decrease much after this point.

4. Results



The geographical visualisation of the clusters could be seen above and the 3 clusters were inferred to be:

- Cluster 0 (Red): No Japanese Restaurants within the vicinity

- Cluster 1 (Purple): Most number of Japanese Restaurants, in relative to cluster 0 and cluster 2, within the vicinity
- Cluster 2 (Green): Lesser number of Japanese Restaurants, in relative to cluster 1, within the vicinity

5. Discussion

Based on this study, Japanese Restaurants were the tenth most common food establishments within the vicinity of the locations studied. This showed that entrepreneurs planning to set up a Japanese Restaurant would expect to face competition from other established brands.

With the three clusters listed above, it could help entrepreneurs determine the location of restaurant to be set up taking into consideration the presence of other Japanese Restaurants.

Based on the above analysis, opening a Japanese Restaurant in cluster 0 would be the most appropriate option as it would most likely face the least amount of competition and it would be fresh option for the population in cluster 0 where no Japanese Restaurants had been established.

One of the limitations of the study was the number of venue recommendations obtained from Foursquare was limited, hence, it may not necessarily be the most representative of the available venues. Therefore, the recommendation and analysis provided in this study should serve only as a reference.

One of the possible improvements for future study would be to also take the price of rental into consideration to provide a more holistic view of the problem. It would also be beneficial if the rating of the nearby food establishments and the human traffic of each location could be taken into consideration.

6. Conclusion

In this study, the most viable location for a Japanese Restaurant to be established in Singapore was studied using data science. Data obtained from Singapore's public data, combined with geographical coordinates obtained using Geocoder and data obtained from Foursquare API helped with the analysis, where k-Means clustering was used. It was recommended for entrepreneurs to consider setting up a Japanese Restaurant in cluster 0. It was also shared that as the study was based on limited data, therefore, the recommendation and analysis provided in this study should serve only as a reference.

7. References:

Image: https://www.forbes.com/sites/alexcapri/2018/09/21/5-reasons-why-the-worlds-tech-firms-are-moving-to-singapore/#5c5ce1202aa0

Train Station Chinese Names: https://data.gov.sg/

Hawkers and Food Establishments statistics: https://www.singstat.gov.sg/find-data/search-by-theme/industry/hawkers-and-food-establishments/latest-data