

Capstone Project Final Report
Classification of Neighbourhoods in Singapore
by Their Tastes

Yixuan Zhong

23 September 2019

I. Table of Contents

I.	Table of Contents	2
II.	Acknowledgement	2
III.	Executive Summary	2
IV.	Introduction	3
V.	Methodology	4
VI.	Results	6
VII.	Discussions	8
VIII.	Conclusion	8
IX.	References	9

II. Acknowledgement

The author would like to thank Coursera and IBM for providing this series of IBM Data Science courses, all the course coordinators for their effort in delivery of these courses, and peers in the same course for their help of grading my assignments.

III. Executive Summary

In this report, the author has identified the problem of finding the most popular food establishment in terms of their number in each neighbourhood and tried to categorise each neighbourhood into distinct clusters by the most common type of food establishment.

IV. Introduction

Singapore is a city-state in Southeast Asia. Modern Singapore is a global hub for education, finance, healthcare, innovation, manufacturing, technology, tourism, trade, and transport.

The city-state is home to 5.6 million residents, 39% of whom are foreign nationals, including permanent residents. (En.wikipedia.org, 2019)

Since Singapore situates between various Asian countries, there is a diversity in food and culture. Singapore is a heaven of food, even for the most picky epicure. Singapore's diversity of cuisine is touted as a reason to visit the country, one of the best locations when it comes to a combination of convenience, variety, quality and price. Singaporean food is a significant cultural attraction for tourists and visitors. Some Singaporean dishes have become internationally known. In 2011, four Singaporean dishes were included in the list of 'World's 50 Most Delicious Foods (Readers' Pick)' – a worldwide online poll by 35,000 people held by CNN International. They are [Hainanese chicken rice](#) (13th), [chili crab](#) (29th), [Katong Laksa](#) (44th) and [roti prata](#) (45th). (En.wikipedia.org, 2019)

Singaporeans also love international cuisines. There are a lot of French, Italian, Spanish, Indian, Middle Eastern, Thai and many other restaurants located across the entire island nation.

There is a total of 55 Planning Areas in Singapore, which are the main urban planning and census divisions of Singapore delineated by the Urban Redevelopment Authority. A Development Guide Plan is then drawn up for each planning area, providing for detailed planning guidelines for every individual plot of land throughout the country. (En.wikipedia.org, 2019)

So which kind of cuisine and restaurant is the most popular in each neighbourhood? Are there any connections among the neighbourhoods with similar tastes? By the popularities of places in each area, we will be surprised by the power of data.

V. Methodology

3.1. Analytic Approach

The analytic approach of this project will be diagnostic (statistical analysis).

3.2. Data Requirements

We will need the names of neighbourhoods, location of neighbourhoods, list of food venues in each neighbourhood.

3.3. Data Collection

Data sources used in this project include but not limited to:

- [Planning Areas of Singapore](#)

The name, region and population of each area was used by web scraping.

- [Foursquare APIs](#)

Foursquare APIs were used to get the restaurants in each area.

3.4. Data Understanding and Preparation

With BeautifulSoup4 library, this Wikipedia page [Planning Areas of Singapore](#) was scraped to get the relevant information.

The data is subsequently transferred into a pandas dataframe. We noticed in the 55 neighbourhoods 11 of them have population no more than 50 people. Those neighbourhoods are mostly industrial estates, national parks and remote islands. A small population size is likely to introduce errors to our study, and there are few food options in those neighbourhoods. Therefore, neighbourhoods with no more than 50 residents are ignored in our analysis. The result includes 44 neighbourhoods.

The names of the areas are then used to query their respective geographical coordinates with Geocoders.

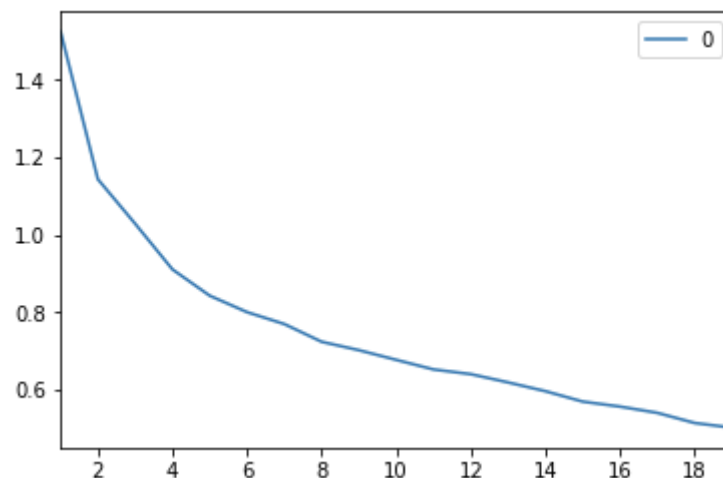
Since we are looking only at venue category of food, only those categories will be retained in our dataset. After filtering, there are 90 types of food options inside our data.

3.5. Modelling and Evaluation

The mean of the frequency of occurrence of each category is calculated for subsequent K-means clustering.

In this project, we use K-means clustering to find out the relationships and connections among those 44 neighbourhoods by their favourite type of food.

To determine the optimal size of clusters k , elbow method is applied for k ranges from 1 to 19. The result is shown in the graph below.

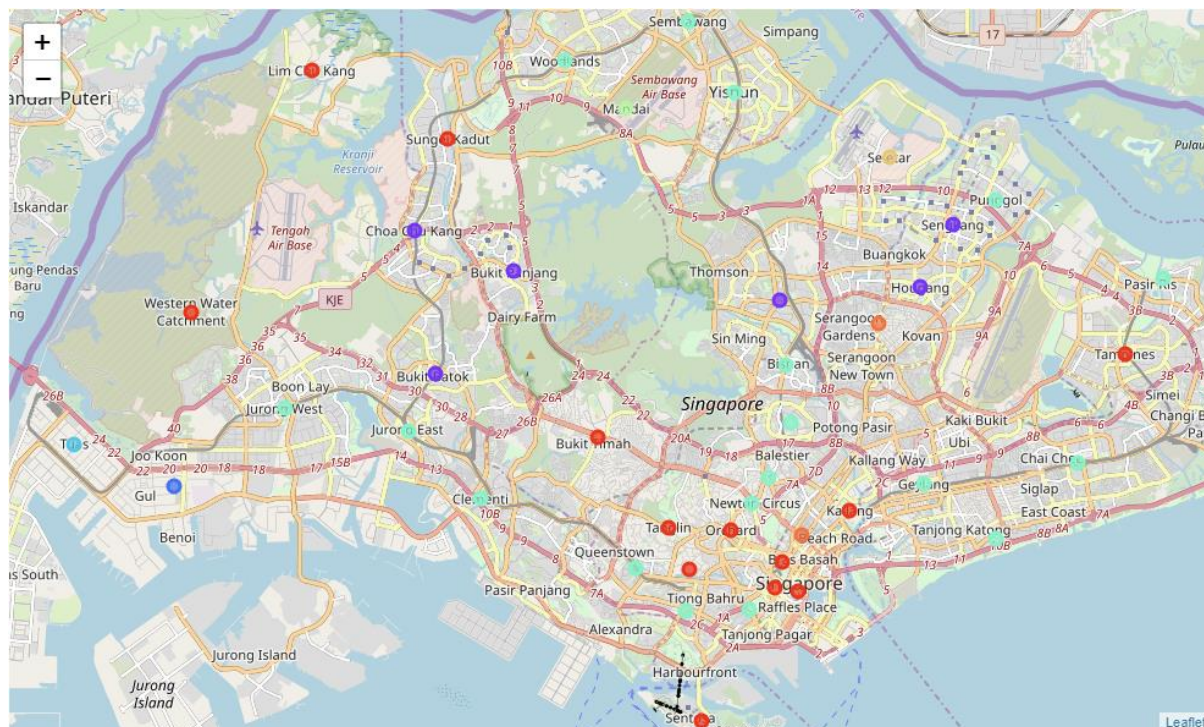
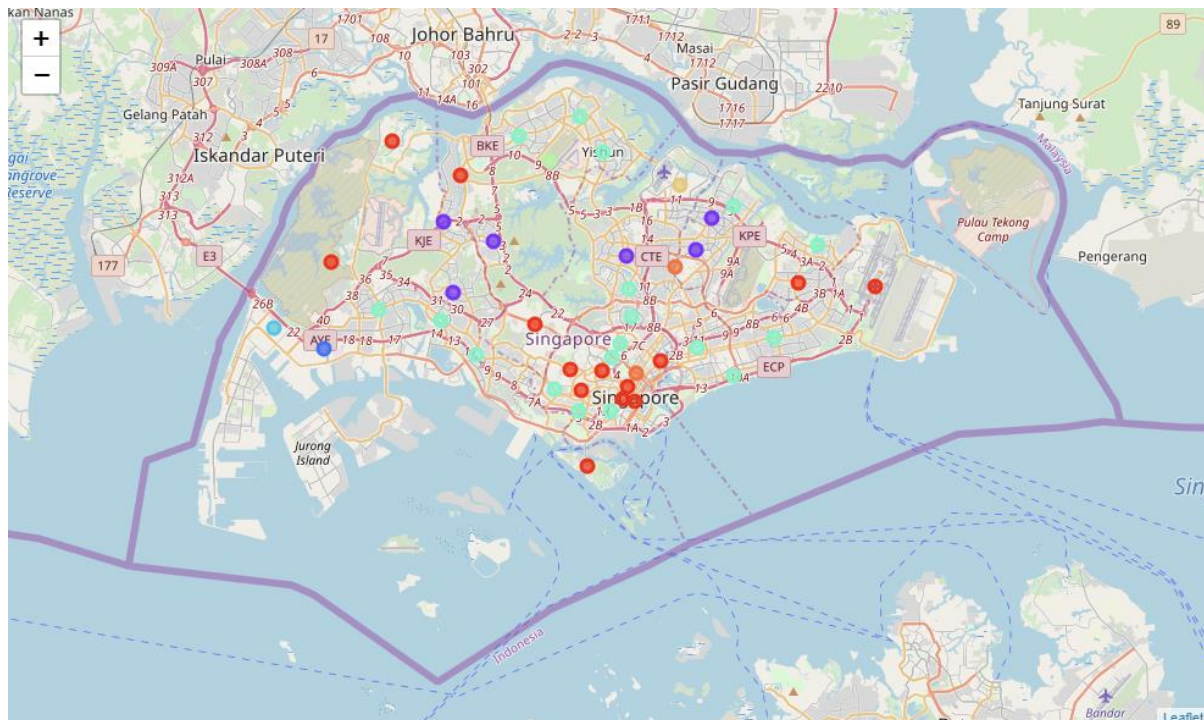


As we can see the results presented in the below graph, after $k = 8$ the decrease in errors with the increase of cluster size almost become linear.

Hence, we will choose to use $k = 8$ for clustering.

VI. Results

The result is presented in a graphical form – map plotted with dots with different colour standing for neighbourhoods of different clusters by the folium library.



The name of each neighbourhood, the region, the cluster label, the cluster colour and the most common venue of each neighbourhood are show in the table.

Neighbourhood	Region	Cluster Labels	Cluster Colour	1st Most Common Venue
Bukit Timah	Central	0		Food Court
Changi	East	0		Coffee Shop
Downtown Core	Central	0		Japanese Restaurant
Kallang	Central	0		Coffee Shop
Lim Chu Kang	North	0		Vegetarian / Vegan Restaurant
Museum	Central	0		Japanese Restaurant
Orchard	Central	0		Japanese Restaurant
River Valley	Central	0		Italian Restaurant
Singapore River	Central	0		Japanese Restaurant
Southern Islands	Central	0		Coffee Shop
Sungei Kadut	North	0		Ramen Restaurant
Tampines	East	0		Café
Tanglin	Central	0		Café
Western Water Catchment	West	0		Wings Joint
Ang Mo Kio	North-East	1		Food Court
Bukit Batok	West	1		Food Court
Bukit Panjang	West	1		Food Court
Choa Chu Kang	West	1		Fast Food Restaurant
Hougang	North-East	1		Food Court
Sengkang	North-East	1		Food Court
Pioneer	West	2		Coffee Shop
Tuas	West	3		Coffee Shop
Bedok	East	4		Coffee Shop
Bishan	Central	4		Coffee Shop
Bukit Merah	Central	4		Chinese Restaurant
Clementi	West	4		Coffee Shop
Geylang	Central	4		Asian Restaurant
Jurong East	West	4		Chinese Restaurant
Jurong West	West	4		Fast Food Restaurant
Marine Parade	Central	4		Chinese Restaurant
Newton	Central	4		Chinese Restaurant
Novena	Central	4		Chinese Restaurant
Outram	Central	4		Café
Pasir Ris	East	4		Coffee Shop
Punggol	North-East	4		Fast Food Restaurant
Queenstown	Central	4		Chinese Restaurant
Sembawang	North	4		Asian Restaurant
Toa Payoh	Central	4		Chinese Restaurant
Woodlands	North	4		Food Court
Yishun	North	4		Food Court
Mandai	North	5		Asian Restaurant
Seletar	North-East	6		Asian Restaurant
Rochor	Central	7		Indian Restaurant
Serangoon	North-East	7		Asian Restaurant

VII. Discussions

In cluster 0, most neighbourhoods are in the central region and the most common food establishment is Japanese restaurants, followed by Italian restaurants, Café, Vegan restaurants, etc. It comes with no surprise as those are CBD and wealthy neighbourhoods with private housing.

In cluster 1 we see most food court and fast food. Those are residential areas with most public housing.

There is only 1 venue each in cluster 2 and 3. They are industrial estates but Tuas is also a custom checkpoint which brings some difference to Pioneer.

Cluster 4 includes most residential neighbourhoods. Chinese restaurant is most common as the majority of population in Singapore is Chinese ethnic.

There is only 1 venue each in cluster 5 and 6. For cluster 5, Mandai is near the natural reserves hence not many residential buildings are there. Seletar is a small neighbourhood with a small airport there.

For cluster 7, Rochor and Serangoon are famous neighbourhoods with mostly South Asian ethnic.

VIII. Conclusion

We can conclude that the K-means clustering method has successfully distinguished among the 44 neighbourhoods in Singapore by their characteristics, in this case the most common type of food establishment.

IX. References

En.wikipedia.org. (2019). *Singapore*. [online] Available at: <https://en.wikipedia.org/wiki/Singapore> [Accessed 23 Sep. 2019].

En.wikipedia.org. (2019). *Planning Areas of Singapore*. [online] Available at: https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore [Accessed 23 Sep. 2019].

En.wikipedia.org. (2019). *Singaporean cuisine*. [online] Available at: https://en.wikipedia.org/wiki/Singaporean_cuisine [Accessed 23 Sep. 2019].