



# IBM Applied Data Science Capstone Opening a new Food Court in Singapore

By: Lee Kwong Yee May 2020

# Business Problem

- My client has considered the food Court's prevalence, but believes that as more people are eating out, there is still untapped potential in this market segment.
- Location of the food court is an important factor that will determine the profitability and sustainability of the business.
- The objective of this project is to analyse and assist my client to select the best locations in the Singapore to open a new food court using Data Science.

# Data

## Data required

- List of neighbourhoods in Singapore
- Latitude and longitude coordinates of these neighbourhoods
- Venue data

## Sources of data

- <https://www.mingproperty.sg/singapore-district-code/> for list of neighbourhoods
- <https://www.maps.ie/coordinates.htm> for coordinates data
- Foursquare API for venue data

# Methodology

- Web-scrape web page for list of neighborhoods
- Get latitude and longitude coordinates
- Use Foursquare API to get venue data
- Group data by neighborhood and take the mean of the frequency of occurrence for each venue category
- Retain only the following categories of interest:
  - a. *'Food Court' and 'Coffee Shop', which are competing categories will be assigned negative scores; and*
  - b. *'Shopping Mall' and 'Metro Station' which are complementing categories will be assigned positive scores*

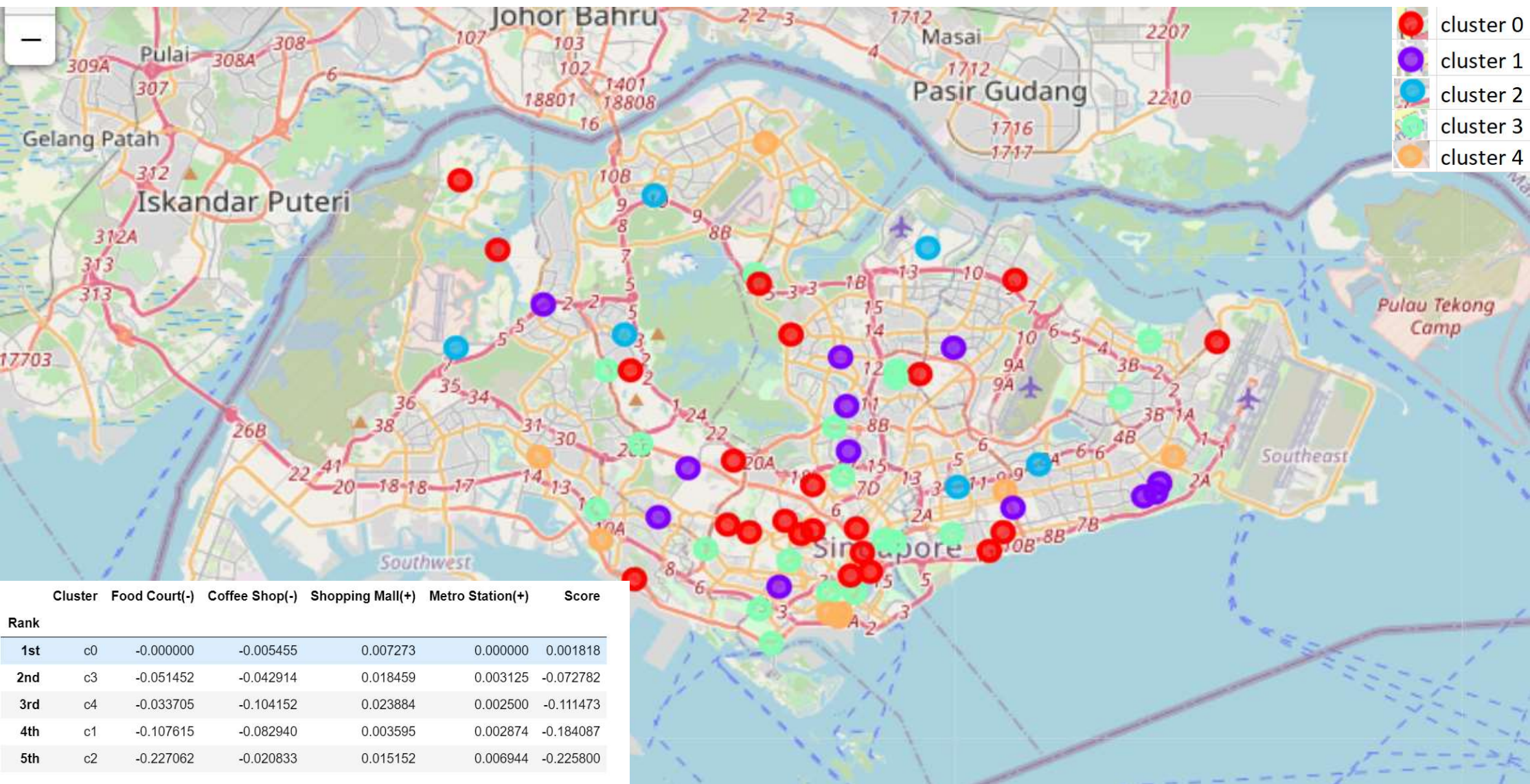
## Methodology (Cont....)

- Find optimal value for number of cluster (k) and perform k-means clustering
- Visualize the clusters in a map using Folium
- Take the mean value of the 4 categories for each cluster and compute cluster score based on the following formula:

$$\text{Cluster score} = ('Food Court' + 'Coffee Shop') * (-1) + ('Shopping Mall' + 'Metro Station')$$



# Results



# Discussion

	Cluster	Food Court(-)	Coffee Shop(-)	Shopping Mall(+)	Metro Station(+)	Score
Rank						
1st	c0	-0.000000	-0.005455	0.007273	0.000000	0.001818
2nd	c3	-0.051452	-0.042914	0.018459	0.003125	-0.072782
3rd	c4	-0.033705	-0.104152	0.023884	0.002500	-0.111473
4th	c1	-0.107615	-0.082940	0.003595	0.002874	-0.184087
5th	c2	-0.227062	-0.020833	0.015152	0.006944	-0.225800

- Cluster 0 is ranked the highest based on the scoring. However, this is mainly due to the lack of Food Court and Coffee Shop in the vicinity. Although less competitive, mean value of the Shopping Mall is comparatively low, and Cluster 0 has also a low mean value for Metro Station
- All these suggest crowd volume might not be ideal to derive the necessary demand.
- Recommend cluster 3 instead (ranked second) as the cluster is moderate for Food Court and Coffee Shop, and at the same time dense for Shopping Mall and Metro Station.

# Conclusion

- Based on holistic analysis of the cluster results, it is deduced that cluster 3 would be an ideal location to set up a Food Court for my client.
- However, other factors such as population and income of residents could influence the location decision of the Food Court and such data are not available at neighbourhood level required by this project.
- Nevertheless, site survey would be able to bridge this gap to affirm and select the most ideal location amongst all neighbourhoods listed within cluster 3.