



APPLIED DATA SCIENCE CAPSTONE (WEEK 4)

suitable location in Singapore for a new Food Court

1. Introduction

For many people living in Singapore, there is a wealth of local food that can satisfy the palate of even the most fastidious. Food courts and coffee shops are the most popular locations for people to have their three meals as unlike restaurants, they are more affordable. In addition, food courts and coffee shops offer a variety of dishes, both local and regional cuisines under one roof.

Food courts and coffee shops offers great business opportunities in Singapore due to its metropolitan environment where most people prefer to eat out rather than cook. My client is interested to setup a food court as he believes that food court provides a cosier and more comfortable eating place as it is larger and air-conditioned, in comparison to a coffee shop which is smaller and open-air, albeit at a higher capital outlay. This to him is especially important, given Singapore's hot and humid climate.

My client has considered the food Court's prevalence, but he believes that as more and more people are eating out because of convenience, there is still considerable demand in this market segment. As with any other business decision, opening a food court requires serious consideration and is a lot more complicated than it seems. In particular, the location of the food court is one of the most important factors that will determine the profitability and sustainability of the business.

2. Business Problem

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

In Singapore, which is small, densely populated and with a metropolitan vibe, what is the best location to set up a food court, amidst the prevalence of eating places in the republic?

3. Data

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

- List of neighbourhoods in Singapore. For the listing, I have decided to look at the district codes in Singapore and use the listed neighbourhoods of the respective district codes to provide a holistic coverage of all the residential and commercial areas of Singapore.
- Latitude and longitude coordinates of these neighbourhoods. This is required in order to plot the map and to acquire the venue data.
- Venue data, which will be used to perform K-means clustering on the neighbourhoods.

Sources of data and methods to extract them:

This web page (<https://www.mingproperty.sg/singapore-district-code/>) contains a list of locations that are tied to the district codes of Singapore with a total of 68 neighbourhoods. I will use web scraping techniques to extract the data from web page with the help of Python Pandas requests. The geographical coordinates of the neighbourhoods will then be acquired either through Python Geocoder package or manually through the website <https://www.maps.ie/coordinates.html>.

After that, I will use Foursquare API to get the venue data for these neighbourhoods. Foursquare API will provide many categories of the venue data: I am particularly interested in the following categories:

- 'Food Court' and 'Coffee Shop', which are competing categories that will be assigned negative scores; and
- 'Shopping Mall' and 'Metro Station' which are complementing categories because of the crowd that they generate, in addition to being potential locations for setting up a food court; positive scores will be assigned to these categories.

Data science skills used for the project as follow:

- web scraping (Wikipedia),
- working with API (Foursquare),
- data cleaning and data wrangling (Pandas),
- machine learning (K-means clustering) and
- map visualization (Folium)

In the next section, I will present the methodology used and steps taken to arrive at my recommendation.