Introduction

Business Problem

The objective of this capstone project is to analyse and select the best location in Singapore to open a new bubble tea shop. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Singapore, if a franchise owner is looking to open a new bubble tea shop, where would you recommend that they open it?

Target Audience

This project is particularly useful to franchise owners and investors looking to open and invest in new bubble tea shops in Singapore. Despite the large supply of outlets around the country, there is still a demand for the sugary beverage. As a precautionary measure due to the COVID-19 outbreak, the Ministry of Trade and Industry announced that all standalone food and beverage outlets have to be temporarily shut. Following the sudden announcement, long queues were spotted at bubble tea shops across many parts of the island as Singaporeans tried to get their last bubble tea fix.

A queue at the Koi outlet in Seletar Mall on April 21, 2020. ST PHOTO: ALOYSIUS CHOW

As of June 2020, bubble tea shops in Singapore can only operate from a limited number of central kitchens. Consequently, this reduced supply has led to an even higher demand for the drink. GrabFood - a delivery service provider popular in South East Asia, offers a subscription plan exclusively for bubble tea lovers. Despite these unprecedented times, the amount of love Singaporeans have for the popular Taiwanese drink remains unwavered, indicating a continuous demand for bubble tea even after we tide over the pandemic.

Data

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

- List of MRT (mass rapid transit) stations in Singapore. Each MRT station is associated with its surrounding area. This defines the scope of the project which is confined to the city of Singapore.
- Latitude and longitude coordinates of those MRT stations. This is required to plot the map and also to get the venue data.
- Venue data, particularly data related to bubble tea shops. We will use this data to perform clustering on the MRT stations.

Sources of data and methods to extract them

We will read a CSV file which contains a list of 122 MRT stations in Singapore using the pandas library. Then we will get the geographical coordinates of the train stations using the geocoder package which will give us the latitude and longitude coordinates of the surrounding area. After that, we will use the Foursquare API to get the venue data for those train stations. The API provides many categories of the venue data, we are particularly interested in the 'Bubble Tea Shop' category in order to solve the business problem. This is a project that will make use of many data science skills, from extracting information from a CSV file, working with Foursquare API, data cleaning, data wrangling, to machine learning (k-means clustering) and map visualization (Folium).