Coursera Capstone

IBM Applied Data Science Capstone

Opening a café near train stations (MRT) and shopping malls in Singapore

By: Lee Jun Yi

February 2020

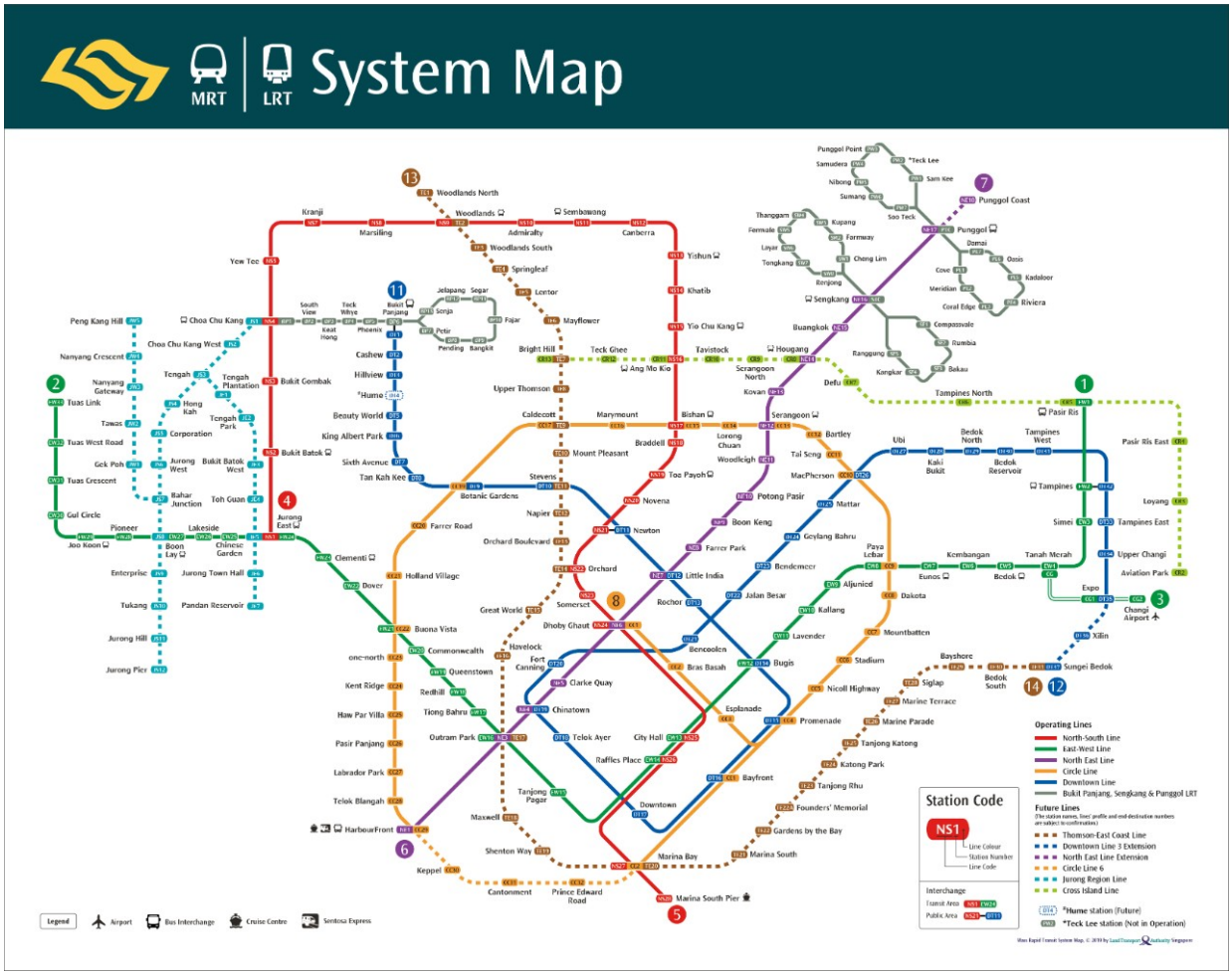

Image source : https://www.lta.gov.sg/content/ltagov/en/getting_around/public_transport/rail_network.html

Introduction

For many business owners, opening a shop is a costly decision and many of them depend on Real Estate salesperson to facilitate the search for a suitable location. However, this introduces a biased selection based on the limited number of properties that the salesperson can sell or rent.

Selecting Singapore for this analysis, it is a densely populated city with most area accessible by train and a shopping mall is located at a stone's throw every few stations away. While shopping malls provide the towns and neighborhoods a centralized location for shopping, dining, and entertainment, it naturally attracts the crowd to there to get products or services. The heavy traffic near shopping malls and stations will benefit any business daily with the large number of potential customers.

To make an informed decision by first understanding the competitors and businesses around shopping malls close to every station, business owners can save the time of researching each area and focus on the locations that satisfy their criteria. While this solution is designed to find out the businesses situated at the most convenient locations, it is highly adaptable for other uses e.g. finding eateries in all neighborhoods, concentration of cafes in every town, schools available at each area, etc.

Business Problem

The objective of this Capstone project is to enable users to select the best locations in the city of Singapore for opening a shop in the vicinity of a shopping mall beside the train station. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to the business need of finding the most suitable location with high human traffic in Singapore. In essence, the question is : where should a business owner in Singapore open a shop?

Target Audience of this project

This project is particularly useful for business owners or investors who intends to understand the business landscape and concentration around shopping malls conveniently located beside train stations. While there is no lack of stations or shopping malls, businesses struggle to demonstrate a Unique Selling Point to consumers since shopping malls can already provide a one-stop location for a wide variety of products and services. Hence, business owners or investors can utilize this project to assess whether a shop is strategically located near to (or away from) the similar businesses.

Data

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

- List of train stations in Singapore — This defines the scope of this project and the area is confined to the country Singapore in South East Asia.
- Latitude and longitude coordinates of every station — This is required in order to plot the map and also get the venue data.
- Venue data, especially data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_Singapore_MRT_stations) contains a list of train stations in Singapore, with a total of 226 stations. I will be using web scraping techniques to extract the data table from the Wikipedia page, with the help of Python's "Request" and "BeautifulSoup" packages. Then we will get the geographical coordinates of the stations using Python Geocoder package which will give us the latitude and longitude coordinates of the stations.

After that, I will be using Foursquare API to get the venue data for those stations. As we have learnt from this course, Foursquare has one of the largest database of places and is used by most developers.

Foursquare API will provide many categories of the venue data, we are particularly interested in the "Shopping Mall" category in order to help us to solve the business problem listed. This is a project that will make use of many data science skills, from web scraping (using BeautifulSoup on Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning modeling, and map visualization (Folium).