

Coursera IBM

Data Science Professional Certificate Capstone Project

Topic: Investigate the optimal location for setting up a new café business along the Ma On Shan Rail in Hong Kong

Prepared by: Lawrence Lee

Date: 22 May, 2020

1. Introduction

The Ma On Shan Rail in Hong Kong is a rapid transit line that serving the new towns of Shatin and Ma On Shan in the northeastern N.T. The ridership is 153,100 weekday average in 2014 research figure. Due to the extension of Ma On Shan Rail to Kai Tak station on 14 Feb 2020, it is predicted that the ridership in Ma On Shan Rail will be sharply increased. As a result, venues near the railway stations in Ma On Shan lines are suitable for small business to setup as more people will pass by those area.

2. Business Problem

In order to find the suitable location and type of business to start with, it is important to grab the venues data for the places nearby for further analysis. Recently, many Hongkongers like to go for café for meal or chatting with friends. The cafés are not only treated as “café” and they always offer entertainment services like karaoke, boardgames and books for reading. It is world to invest a café as a start up business as the capital involved is low. The main objective of this project is to find out the ideal location near the Ma On Shan Line for starting a new café business. The locations with more restaurants but less café will be selected as business location.

3. Data

The data for this project had been processed with different sources and ensured the method of analysis is accurate.

3.1 MTR Station

The MTR Station data of Ma On Shan line will be scraped from a Wikipedia webpage. AS there are only 9 stations in Ma On Shan line, the data will be directly saved and written to a list without using BeautifulSoup library in Python.

3.2 Geocoding

For the geometric data like the latitude and longitude of each of the MTR stations, Google Maps Geocoding API will be used to extract the data and those dataframe will be placed into the station dataframe.

3.3 Venue Data

After the location data is achieved from geocoding and Wikipedia, the FourSquare API will be used to find out the nearby venues and create a new dataframe to involve all the venue data nearby the MTR station. It is assumed that maximum 30 nearby venues to be explored for each MTR Stations and the radius of exploration is 1000m.