Data Science Capstone Project 2020

The Battle of Neighborhoods (Week 1 Report)

By: Dennis Lam

INTRODUCTION

Sandakan is located in state of Sabah, East Malaysia. It is the second largest town. The town was founded in 1878 and today has approximate 160000 residents staying in there.

This Capstone project will try to find out where and which suburbs neighbourhood is suitable for setting up business.

BUSINESS PROBLEM

As a town, the people here are mainly ranges from mid-income to low income earners, hence small businesses are more suitable instead of building large shopping malls.

We shall do an analysis of Sandakan neighborhoods and present the report which will consists of methodology, results and conclusions.

The stakeholders are those who wish to setup any new business can refer to this report.

DATA

The neighborhoods data is found in Sandakan Municipal website: http://www.mps.sabah.gov.my/isandakan.cfm#penduduk

The focus will be on these neighbourhoods (they are also called housing estates) around Sandakan. Potential for business setups will be within housing estates and areas surrounding it.

Python libraries like pandas, geopy, folium, scikit-learn will be used to explore and cluster these neighbourhoods to find out which one is suitable.

Foursquare APIs will be used to find out any interesting venues like food, offices, entertainment etc.

Google maps and Geopy library will be used to find latitude and longitude for each neighbourhood as geolocation data is not available online.