Applied Data Science Project | IBM Data Science Professional Certificate

Title: Where to setup more pharmacy outlets?

Table of Content:

Introduction: Problem Statement

Pharmacy outlets have become growingly important as one stop centers for drugs, supplement, medical and surgical supplies and consultation. This is induced by the population increase within the areas of inhabitant.

Despite the need of setting up more outlets, study on the demographic from the point of interest(POI) and existing outlets is required to understand the strategic location in order for the outlets to be accessible.

The POIs will be the key influece to the success of pharmacy outlets. The POIs in vicinity can be, but not limited to Hospital/Medical Centers, Department Store, Shopping Mall, Residential Building, Tourist Information Center, Market, Resorts etc.

The project aim to provide location recommendation based the analysis conducted on the point of interest and neighboring pharmacy outlets, believing that these places can easily attract consumers who are enroute to the specific POIs or those who need drugs/medical supply or prescription which near to the infirmary.

Data Definition

This project is to address the areas in Tanjung Tokong, Penang which are lacking pharmacy outlets and to propose new locations for outlets to be set up.

Data collection is based on the radius of the target neighborhood grouped as Neighborhood/Venue, Latitude and Longitude. The landmarks from the geographic coordinates will be mapped using Folium library.

The factors to be considered for decision on the recommended locations are:

- Number of point of interest that influence setting up of pharmacy outlets
- Number of existing pharmacy outlets in vicinity
- Residential or city demographics

I will focus on the areas where the center of where the POIs are located.
The approach of retrieving the pharmacy location and related information is by means of the Four-square API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Tanjung Tokong	5.450990	100.304527	Edo Ichi Japanese Cuisine 江戸一日本料理	5.450613	100.305675	Japanese Restaurant
1	Tanjung Tokong	5.450990	100.304527	Meraki Coffee	5.450268	100.305727	Café
2	Tanjung Tokong	5.450990	100.304527	Kan Nichi Kan Yakiniku Japanese BBQ restaurant (韓日館 KNK)	5.451530	100.306009	Japanese Restaurant
3	Tanjung Tokong	5.450990	100.304527	DAORAE Korean BBQ Restaurant	5.452626	100.306130	Korean Restaurant
4	Tanjung Tokong	5.450990	100.304527	Hairstory International	5.450269	100.305613	Health & Beauty Service
5	Tanjung Tokong	5.450990	100.304527	beÃngels Specialty Coffee	5.450464	100.305669	Café
6	Tanjung Tokong	5.450990	100.304527	K•POT	5.450738	100.304907	Korean Restaurant
7	Tanjung Tokong	5.450990	100.304527	The Coffee Bean & Tea Leaf	5.450521	100.305876	Coffee Shop
8	Tanjung Tokong	5.450990	100.304527	味坊包子	5.449004	100.302226	Dongbei Restaurant

Fig 1: Data retrieved with Foursquare API.

POIs are spelt out from the Venue Category list based on the data above. Among them are:

- Cosmetics Shop
- Department Store
- Miscellaneous Shop
- Residential Building (Apartment / Condo)
- Grocery Store
- Health & Beauty Service
- Shopping Mall
- Tourist Information Center
- Hospital
- Hotel Bar
- Gas Station
- Convenience Store
- Resort
- Market

Methodology

To obtain the geographic coordinates of the target borough, I will first search and retrieve the neighborhood coordinates from the area of Tanjung Tokong. I will then explore block radius to discover the landmarks based on the data with the use of Foursquare API.

Next, designated POIs will be segregated out from the landmarks together with the existing pharmacy outlets for visualiation to better understand the mapping of respective locations.

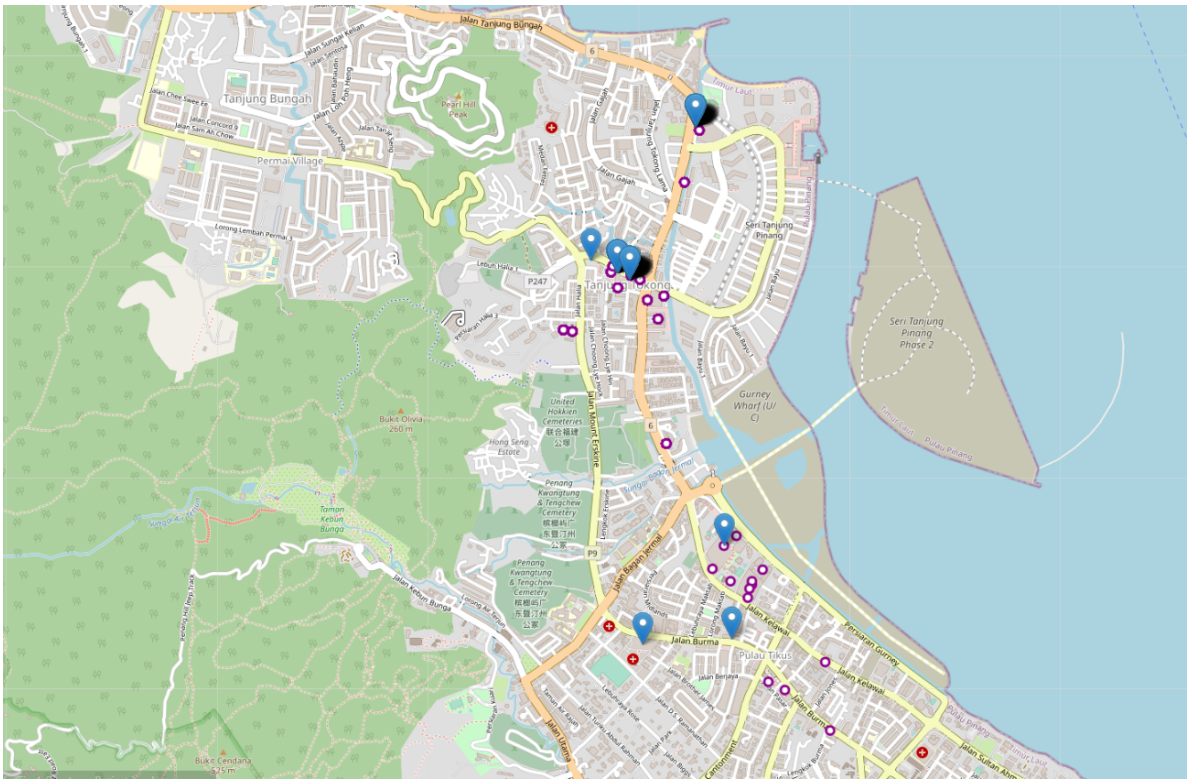

Fig 2: Map out the 7 existing pharmacy outlets (blue marker) and the POIs(violet circle marker)

Data Analysis

Data cleansing is done before exisiting pharmacy outlet and POI locations are segregated for visualization later. In the process, I have already replaced some Neighborhood names which are in the same region with similar name to better group the region.

From this exercise, I am able to see how the POIs affect the decision of the existing outlets setups. For example, the outlets are usually built near or at the center of the POIs.

Prior to the cluster the POIs, segregated POIs and pharmacy outlets data will be tested for the optimum k-cluster count with elbow method.

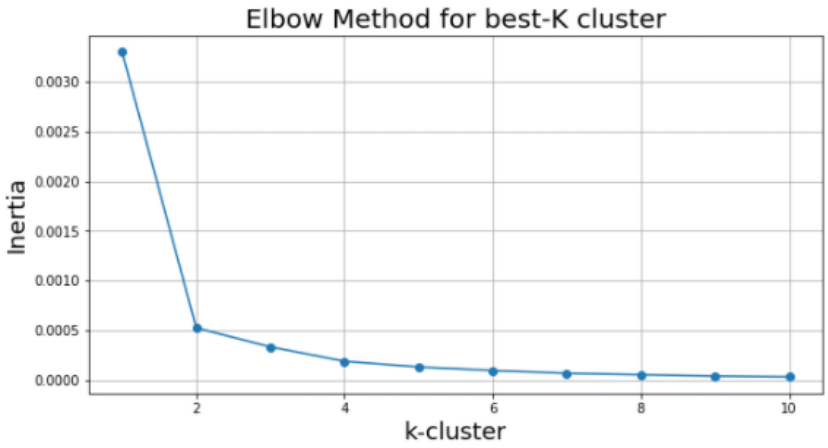

Fig 3: Inertia starts to converge at k=3.

By fitting the coordinates of POI and pharmacy outlets to the k-Mean clustering model with k=3 clusters, the model gives us the final recommend region as below:

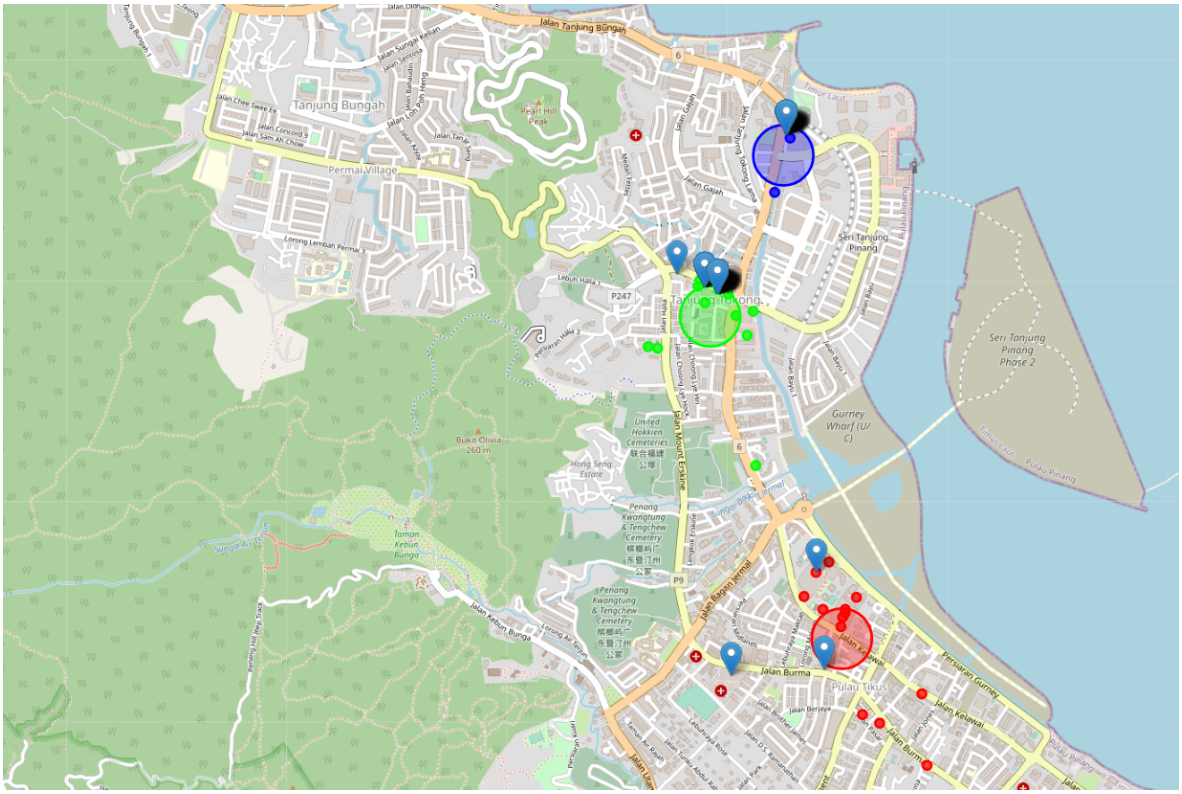

Fig 4: Recommended region for new outlets in large circle markers.

Results/Discussion

Based on the marker, there are only **7 existing outlets** available in 3 different neighborhoods, 'Tanjung Bungah','Tanjung Tokong' and 'Georgetown', having Tanjung Tokong with the most outlets at one close area and another 2 outlets nearer to the bay area. Georgetown area shares the same number of outlets, however the outlet locations are equally spaced.

The POIs on the Tanjung Tokong area are the most densed (displayed in small green circle markers) which see the setup of 3 outlets. This is followed by Georgetown area with 13 POIs.

The analysis shows the larger circle markers which recommend the area where new outlets can be setup. Judging from the 3, the **violet region** is the best as it will not get so much competition from the existing outlets and at the same time, the area is not very far from the POIs displayed in the **green region**.

Conclusion

The objective of this project is to provide a strategic location for new pharmacy outlets based on the POIs, the likely competition from the existing outlets and how near the locations are to the city and residential areas.

With the venue and neighborhood information obtained by Foursquare API, I am able to have clearer view on the density of existing outlets and POIs to decide whether the regions suggested are potentially suitable for new oulets.

The recommended areas from the 3 large circle markers will allow potential business owners to fathom the demographic and density of the landmarks, prior to the considerationof other variable factors such as traffic conditions, accessibility, prices and business models among others, to land on the decision on whether to start an outlet at the specified location.