APRIL 19, 2020

How about Opening a New Shopping Complex in Malaysia?

IBM Applied Data Science Capstone

SREEJITH R

Introduction

For some, customers, visiting shopping centers is an incredible method to unwind and have a good time during ends of the week and occasions. They can do shopping for food, eat at cafés, shop at different style outlets, watch motion pictures and perform a lot more exercises. Shopping centers resemble a one-stop goal for a wide range of customers. For retailers, the focal area and the huge group at the shopping centers give an incredible dispersion channel to advertise their items and administrations. Property designers are likewise exploiting this pattern to construct additionally shopping centers to take into account request. Accordingly, there are many shopping centers in the city of Kuala Lumpur and a lot more are being fabricated. Opening shopping centers permit property designers to win predictable rental salary. Obviously, similarly as with any business choice, opening another shopping center requires genuine thought and is significantly quite confounded. Especially, the area of the shopping center is one of the most significant choices that will decide if the shopping center will be a triumph or a disappointment.

Business Problem

The target of this capstone venture is to dissect and choose the best areas in the city of Kuala Lumpur, Malaysia to open another shopping center. Utilizing information science procedure and AI strategies like grouping, this undertaking plans to give answers for answer the business question: In the city of Kuala Lumpur, Malaysia, if a property designer is hoping to open another shopping center, where might you suggest that they open it?

Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the capital city of Malaysia i.e. Kuala Lumpur. This project is timely as the city is currently suffering from an oversupply of shopping malls. Data from the National Property Information Centre (NAPIC) released last year showed that an additional 15 per cent will be added to existing mall space, and the agency predicted that total occupancy may dip below 86 per cent. The local newspaper The Malay Mail also reported in March last year that the true occupancy rates in malls may be as low as 40 per cent in some areas, quoting a Financial Times (FT) article cataloguing the country's continued obsession with building more shopping space despite chronic oversupply.

Data

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

1. List of neighbourhoods in Kuala Lumpur. This defines the scope of this project which is confined to the city of Kuala Lumpur, the capital city of the country of Malaysia in South East Asia.
2. Latitude and longitude coordinates of those neighbourhoods. This is required to plot the map and also to get the venue data.
3. Venue data, particularly 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/Category:Suburbs_in_Kuala_Lumpur) contains a rundown of neighborhoods in Kuala Lumpur, with a sum of 70 neighborhoods. We will utilize web scratching strategies to separate the information from the Wikipedia page, with the assistance of Python demands and wonderful soup bundles. At that point we will get the land directions of the areas utilizing Python Geocoder bundle which will give us the scope and longitude directions of the areas.

From that point forward, we will utilize Foursquare API to get the scene information for those areas. Foursquare has probably the biggest database of 105+ million places and is utilized by more than 125,000 engineers. Foursquare API will give numerous classifications of the scene information, we are especially keen on the Shopping Mall class to assist us with solving the business issue set forward. This is a task that will utilize numerous information science aptitudes, from web scratching (Wikipedia), working with API (Foursquare), information cleaning, information wrangling, to AI (K-implies bunching) and map representation (Folium). In the following area, we will introduce the Methodology segment where we will talk about the means taken in this undertaking, the information examination that we did and the AI procedure that was utilized.

Methodology

Firstly, we need to get the list of neighbourhoods in the city of Kuala Lumpur. Fortunately, the list is available in the Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur). We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Kuala Lumpur.

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the "Shopping Mall" data, we will filter the "Shopping Mall" as venue category for the neighbourhoods.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for "Shopping Mall". The results will allow us to identify which neighbourhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

Results

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Shopping Mall”:

- Cluster 0: Neighbourhoods with moderate number of shopping malls
- Cluster 1: Neighbourhoods with low number to no existence of shopping malls
- Cluster 2: Neighbourhoods with high concentration of shopping malls

The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour.

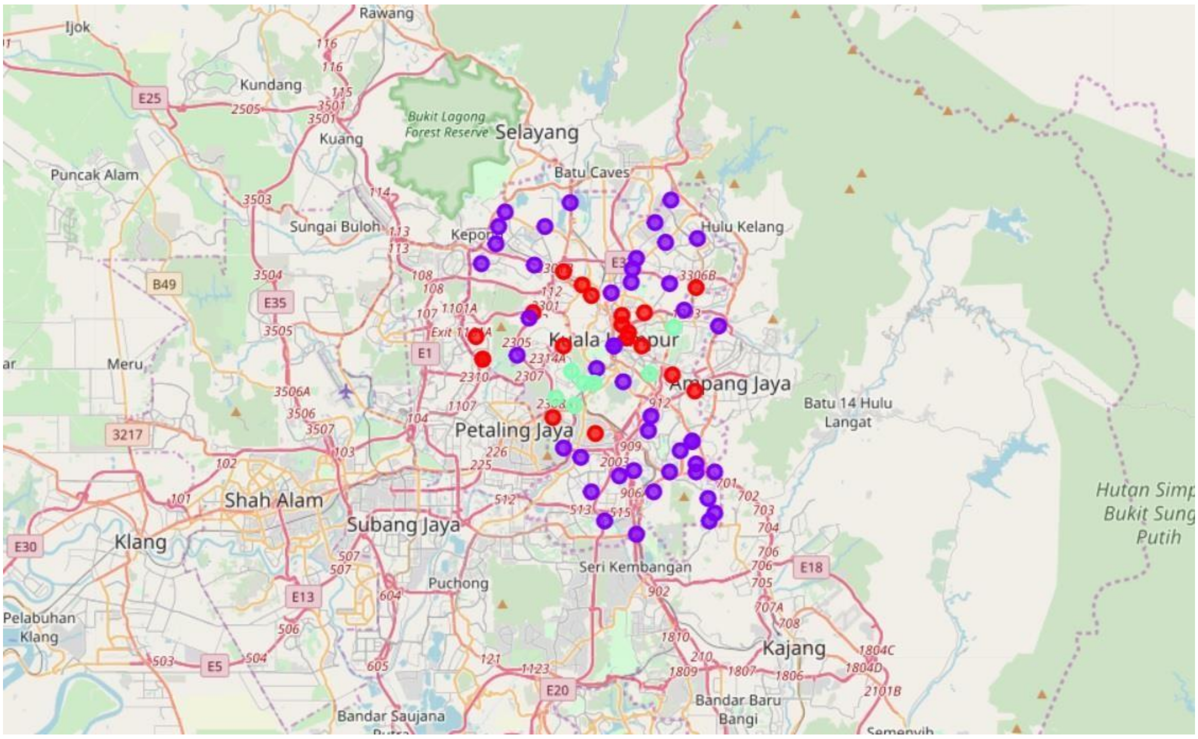

Fig: Map Visualisation for clusters

Discussion

As perceptions noted from the guide in the Results segment, the vast majority of the shopping centers are amassed in the focal zone of Kuala Lumpur city, with the most noteworthy number in cluster 2 and moderate number in cluster 0. Then again, cluster 1 has low number to no shopping center in the areas. This speaks to an incredible chance and high potential regions to open new shopping centers as there is almost no to no opposition from existing shopping centers. In the interim, shopping centers in cluster 2 are likely experiencing extraordinary rivalry because of oversupply and high centralization of shopping centers. From another point of view, the outcomes additionally show that the oversupply of shopping centers for the most part occurred in the focal region of the city, with the suburb zone despite everything have not many shopping centers. Thusly, this venture prescribes property designers to capitalize on these discoveries to open new shopping centers in neighborhoods in cluster 1 with almost no opposition. Property engineers with interesting offering recommendations to stand apart from the opposition can likewise open new shopping centers in neighborhoods in cluster 0 with moderate rivalry. Ultimately, property designers are encouraged to maintain a strategic distance from neighborhoods in cluster 2 which as of now have high convergence of shopping centers and experiencing serious rivalry.

Limitations and Suggestions for Future Research

In this undertaking, we just consider one factor for example frequency of event of shopping centers, there are different factors, for example, populace and pay of inhabitants that could impact the area choice of another shopping center. Be that as it may, to the best information on this researcher such information are not accessible to the local level required by this task. Future research could devise a methodology to gauge such information to be utilized in the bunching calculation to decide the favored areas to open another shopping center. Furthermore, this undertaking utilized the free Sandbox Tier Account of Foursquare API that accompanied constraints with regards to the quantity of API calls and results returned. Future research could utilize paid account to sidestep these restrictions and get more outcomes.

Conclusion

In this venture, we have experienced the way toward distinguishing the business issue, determining the information required, extricating and setting up the information, performing AI by bunching the information into 3 groups dependent on their likenesses, and ultimately giving proposals to the important stakeholders for example property engineers and financial specialists with respect to the best areas to open another shopping center. To respond to the business question that was brought up in the presentation segment, the appropriate response proposed by this task is: The areas in group 1 are the most favored areas to open another shopping center. The discoveries of this venture will assist the pertinent stakeholders with capitalizing on the open doors on high potential areas while staying away from packed zones in their choices to open another shopping center.

References

- Category:Suburbs in Kuala Lumpur. *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur
- Foursquare Developers Documentation. *Foursquare*. Retrieved from https://developer.foursquare.com/docs
- Malay Mail. (2018, March 14). Malls facing meltdown as glut continues. *Malay Mail*. Retrieved from https://www.malaymail.com/s/1597735/malls-facing-meltdown-as-glut-continues
- Tan, H. H. (2018, April 5). An oversupply of retail space in Malaysia. *StarProperty.my*. Retrieved from http://www.starproperty.my/index.php/articles/property-news/an-oversupply-of-retail-space-in- malaysia/

Appendix

Cluster 0

- Bangsar South
- Bukit Bintang
- Bukit Nanas
- Bukit Tunku
- Chow Kit
- Damansara Heights
- Damansara Town Centre
- Damansara, Kuala Lumpur
- Dang Wangi
- Jalan Cochrane, Kuala Lumpur
- Jalan Duta
- Kampung Baru, Kuala Lumpur
- Medan Tuanku
- Mont Kiara
- Segambut
- Setiawangsa
- Shamelin
- Taman Desa
- Taman Tun Dr Ismail

Cluster 1

- Alam Damai
- Ampang, Kuala Lumpur
- Bandar Menjalara
- Bandar Sri Permaisuri
- Bandar Tasik Selatan
- Bandar Tun Razak
- Batu 11 Cheras
- Batu, Kuala Lumpur
- Bukit Jalil
- Bukit Kiara
- Bukit Petaling
- Cheras, Kuala Lumpur
- Desa Petaling
- Federal Hill, Kuala Lumpur
- Happy Garden
- Jinjang
- Kampung Datuk Keramat
- Kepong
- Kuchai Lama
- Maluri
- Miharja
- Pantai Dalam
- Putrajaya
- Salak South
- Semarak
- Sentul Raya
- Setapak
- Sri Hartamas
- Sri Petaling
- Sungai Besi
- Taman Bukit Maluri
- Taman Cheras Hartamas
- Taman Connaught
- Taman Ibukota
- Taman Len Seng
- Taman Melati
- Taman Midah
- Taman OUG
- Taman P. Ramlee
- Taman Sri Sinar
- Taman Taynton View
- Taman Wahyu
- Titiwangsa
- Wangsa Maju

Cluster 2

- Bangsar
- Bangsar Park
- Brickfields
- KL Eco City
- Lembah Pantai
- Pudu, Kuala Lumpur
- Taman U-Thant