Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Kuala Lumpur, Malaysia

By: Rishabh Singh

March 2020

Introduction

For some, customers, visiting shopping centers is an extraordinary method to unwind and live it up during ends of the week and occasions. They can do shopping for food, feast at cafés, shop at the different design 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 gives an incredible appropriation channel to showcase their items and administrations. Property designers are additionally exploiting this pattern to construct all the more shopping centers to take into account the interest. Thus, there are many shopping centers in the city of Kuala Lumpur and a lot more are being fabricated. Opening shopping centers permits property engineers to procure predictable rental pay. Obviously, similarly as with any business choice, opening another shopping center requires genuine thought and is much surprisingly entangled. 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 goal of this capstone venture is to break down and choose the best areas in the city of Kuala Lumpur, Malaysia to open another shopping center. Utilizing information science philosophy and AI methods like grouping, this venture intends 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 prescribe that they open it?

Target Audience of this project

This task is especially valuable to property designers and speculators hoping to open or put resources into new shopping centers in the capital city of Malaysia for example Kuala Lumpur. This venture is convenient as the city is right now experiencing oversupply of shopping centers. Information from the National Property Information Center (NAPIC) discharged a year ago indicated that an extra 15 percent will be added to existing shopping center space, and the organization anticipated that absolute inhabitance may plunge underneath 86 percent. The nearby paper The Malay Mail additionally detailed in March a year ago that the genuine inhabitance rates in shopping centers might be as low as 40 percent in certain zones, citing a Financial Times (FT) article inventoriing the nation's proceeded with fixation on building all the more shopping space in spite of incessant oversupply.

Data

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

- List of neighborhoods in Kuala Lumpur. This characterizes the extent of this venture which is bound to the city of Kuala Lumpur, the capital city of the nation of Malaysia in South East Asia.
- Latitude and longitude directions of those areas. This is required so as to plot the guide and furthermore to get the setting information.
- Venue information, especially information identified with shopping centers. We will utilize this information to perform bunching on the areas.

Wellsprings of information and strategies to separate them

This Wikipedia page (Category:Suburbs in Kuala Lumpur - Wikipedia) contains a rundown of neighborhoods in Kuala Lumpur, with a sum of 70 neighborhoods. We will utilize web scratching systems to separate the information from the Wikipedia page, with the assistance of Python requests and beautifulsoup packages. 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 venue data for those areas. Foursquare has one of the biggest database of 105+ million places and is utilized by more than 125,000 engineers.

Foursquare API will give numerous categories of venue data, we are especially keen on the Shopping Mall classification so as to assist us with solving the business issue set forward. This is a task that will utilize numerous data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the following segment, we will introduce the Methodology segment where we will talk about the means taken right now, the data analysis that we did and the machine learning technique that was utilized.

Methodology

Right off the bat, we have to get the rundown of neighborhoods in the city of Kuala Lumpur. Luckily, the rundown is accessible in the Wikipedia page

We will do web scraping

utilizing Python requests and beautifulsoup packages to remove the rundown of neighborhoods information. Notwithstanding, this is only a rundown of names. We have to get the geological facilitates as latitude and longitude so as to have the option to utilize Foursquare API. To do as such, we will utilize the awesome Geocoder package that will permit us to change over location into topographical facilitates as scope and longitude. In the wake of social occasion the information, we will populate the information into a pandas DataFrame and afterward visualize the neighbourhoods in a guide utilizing Folium package. This permits us to play out a once-over to make sure everything seems ok to ensure that the topographical directions information returned by Geocoder are effectively plotted in the city of Kuala Lumpur.

Next, we will utilize Foursquare API to get the best 100 venues that are inside a range of 2000 meters. We have to enroll a Foursquare Developer Account so as to get the Foursquare ID and Foursquare secret key. We at that point make API calls to Foursquare going in the geological directions of the areas in a Python loop. Foursquare will restore the venue data in JSON format and we will separate the venue name, venue category, venue latitude and longitude. With the data, we can check what number of venues were returned for every area and look at what number of special classifications can be curated from all the brought venues back. At that point, we will break down every area by gathering the lines by neighborhood and taking the mean of the recurrence of event of every venue category. Thusly, we are additionally setting up the data for use in bunching. Since we are breaking down the "Shopping Mall" data, we will filter the "Shopping Mall" as venue classification for the areas.

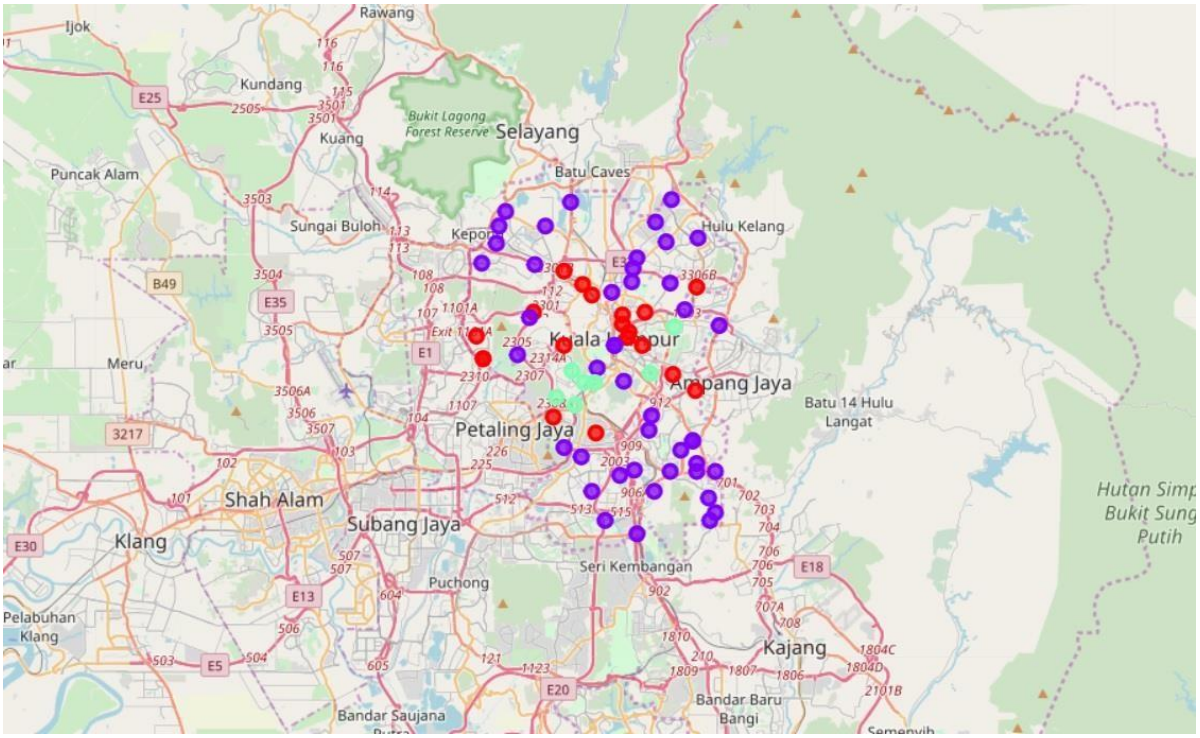
In conclusion, we will perform clustering on the information by utilizing k-means clustering. K-means clustering calculation recognizes k number of centroids, and afterward apportions each datum point to the closest bunch, while keeping the centroids as little as could reasonably be expected. It is one of the least complex and well known solo AI calculations and is especially fit to take care of the issue for this undertaking. We will group the areas into 3 clusters dependent on their recurrence of event for "Shopping Mall". The outcomes will permit us to distinguish which neighborhoods have higher centralization of shopping centers while which neighborhoods have less number of shopping centers. In view of the event of shopping centers in various neighborhoods, it will assist us with answering the inquiry about which neighborhoods are generally appropriate to open new shopping centers.

Results

The outcomes from the k-means clustering show that we can order the areas into 3 clusters dependent on the recurrence of event 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.



Discussion

As perceptions noted from the map in the Results segment, the vast majority of the shopping malls 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 extremely low number to no shopping mall in the areas. This speaks to an incredible chance and high potential zones to open new shopping malls as there is almost no to no opposition from existing malls. In the mean time, shopping malls in cluster 2 are likely experiencing extraordinary rivalry because of oversupply and high convergence of shopping malls. From another viewpoint, the outcomes additionally show that the oversupply of shopping malls for the most part occurred in the focal region of the city, with the suburb zone despite everything have not very many shopping malls. Along these lines, this task prescribes property engineers to exploit these discoveries to open new shopping malls in neighborhoods in cluster 1 with next to zero rivalry. Property designers with one of a kind offering suggestions to stand apart from the opposition can likewise open new shopping malls in neighborhoods in cluster 0 with moderate rivalry. Finally, property designers are encouraged to evade neighborhoods in cluster 2 which as of now have high centralization of shopping malls and experiencing extraordinary rivalry.

Limitations and Suggestions for Future Research

Right now, just consider one factor for example recurrence of event of shopping malls, there are different factors, for example, populace and pay of inhabitants that could impact the area choice of another shopping mall. Be that as it may, to the best knowledge of this specialist such information are not accessible to the local level required by this task. Future research could devise a strategy to gauge such information to be utilized in the clustering calculation to decide the favored areas to open another shopping mall. What's more, this undertaking utilized the free Sandbox Tier Account of Foursquare API that accompanied confinements with regards to the quantity of API calls and results returned. Future research could make utilization of paid record to sidestep these restrictions and get more outcomes.

Conclusion

Through this project, we have experienced the way toward recognizing the business issue, indicating the information required, separating and setting up the information, performing machine learning by clustering the information into 3 clusters dependent on their similarities, and in conclusion giving proposals to the pertinent stakeholders for example property designers and speculators with respect to the best areas to open another shopping mall. To respond to the business question that was brought up in the presentation segment, the appropriate response proposed by this undertaking is: The areas in bunch 1 are the most favored areas to open another shopping mall. The discoveries of this undertaking will assist the important stakeholders with capitalizing on the open doors on high potential areas while maintaining a strategic distance from packed regions in their choices to open another shopping mall.