Capstone of Coursera

IBM Professional Data Science Certificate

New shopping mall in Kuala Lumpur in Malaysia

Deepak Mittal(@yesdeepakmittal)

May, 2020

Github code here

Blog here

Introduction

A shopping mall is one of the hot places in any city. It is an awesome place to do the shopping and make yourself relax and enjoy it. The shoppers here do shopping, look at various new outlets, and familiar with new fashions. Marketers here do marketing experiments and evaluate ROI. Retailers here find most of their items and find it a central place. Companies launch their products to monitor the market reach from here. But if someone wants to open up a new mall, it is beneficial to analyze the competition at various places so that after opening a new mall, sales will not affect due to competition.

Business Problem

The business problem is to select the location for the shopping mall in the city of Kuala Lumpur, Malaysia having the least competition using the data available on Wikipedia page, and using foursquare API to evaluate the hot venues near to it, and using KMeans Clustering Unsupervised Machine Learning Algorithm.

Target Audience

The target audience is shopping mall builders and those who want to open up a new shopping mall in the Kuala Lumpur of Malaysia.

Data

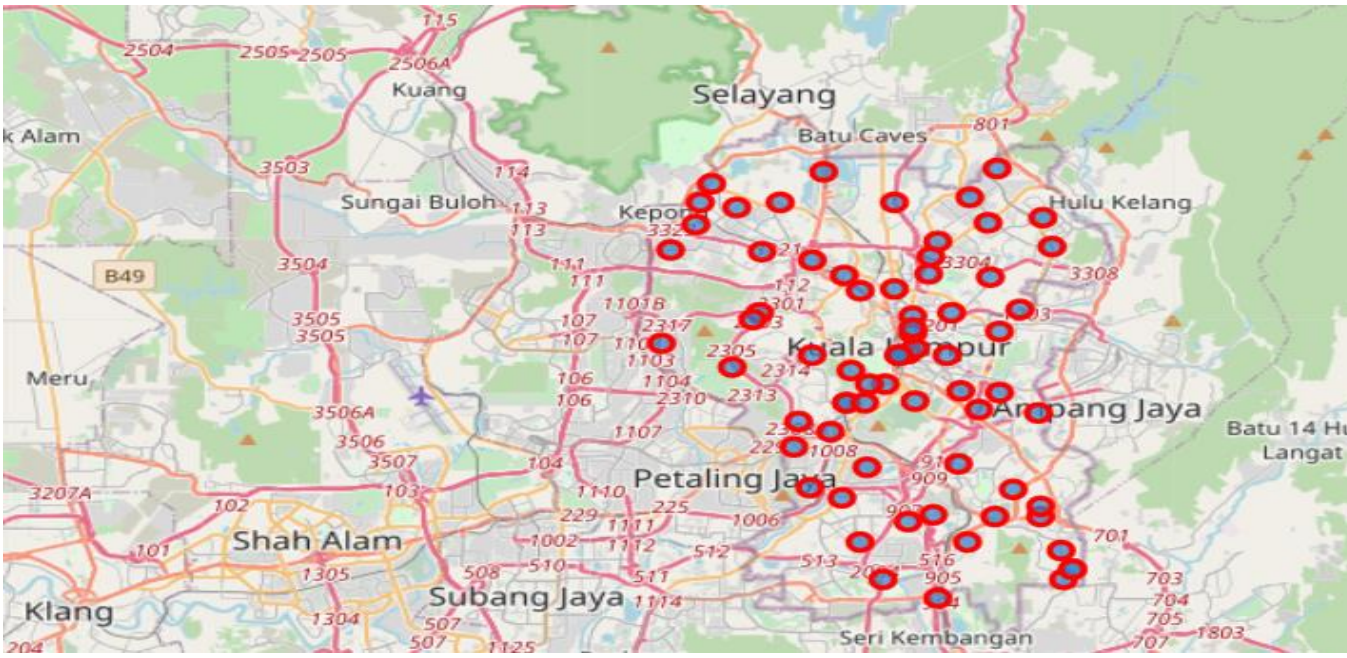
The following data is required for the analysis:

1. List of neighbors of the city.
2. Coordinates of it.
3. Venue data
4. Foursquare API and its client credentials.

Source:
https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur

Methodology

- Get the list of neighbours of Kuala Lumpur of Malaysia from https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur using python library bs4.
- Get coordinates of each location using Geocoder library.
- Visualize the data using Folium library of python on geographical map.

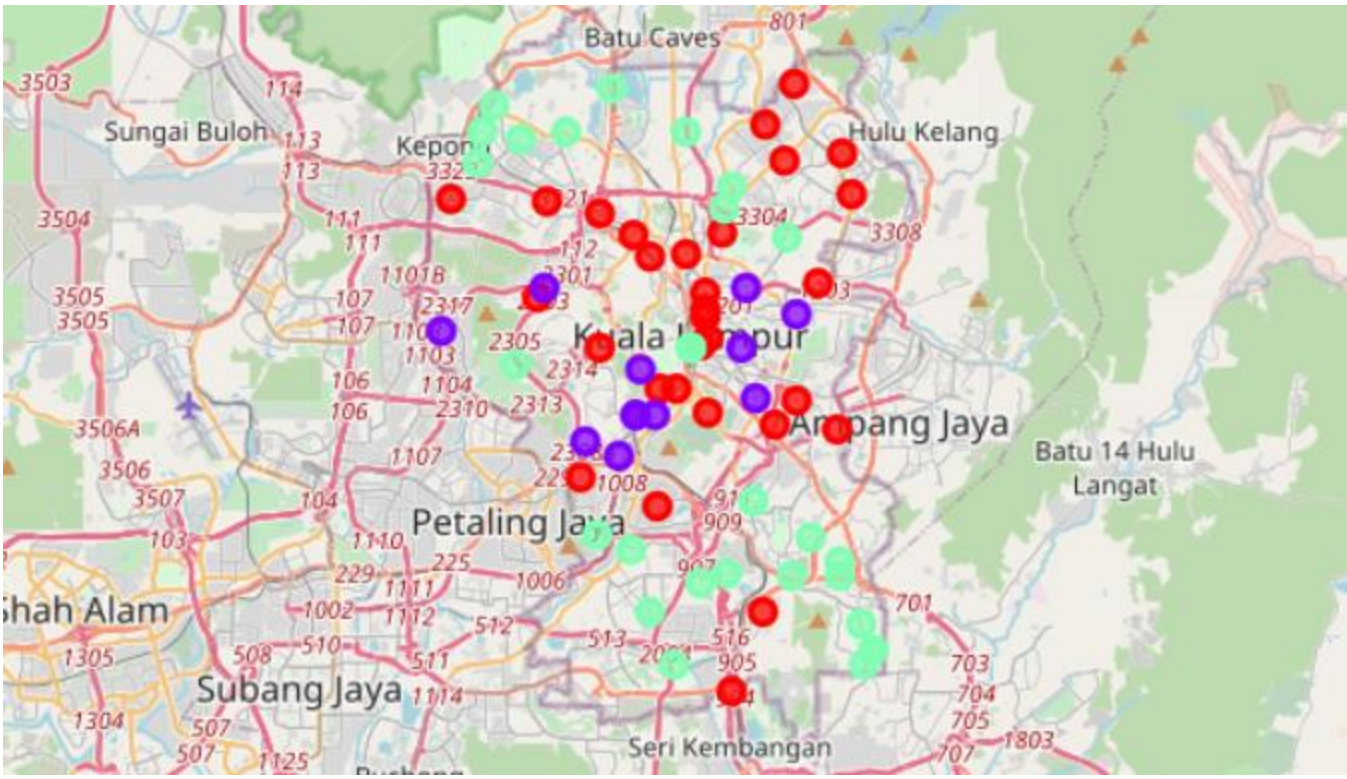


- Get 100 top venues within the radii of 2000 meters using Foursquare Developer Account client credentials.
- Extract the data from json formatted data returned by the Foursquare API.
- Take mean of occurrence of each venue category.
- Filter the shopping mall data from it.
- Cluster the data into specified centroids (in this case 3).
- Examine each cluster.

Result

After examining all three clusters, it is found that:

- Cluster 0: Moderate number of shopping malls and hence moderate competition.
- Cluster 1: Highest number of shopping malls ranging from 3 to 5 and hence highest competition.
- Cluster 2: No malls in this region and hence no competition.



Discussion

As it is observed that cluster 2 has no malls while cluster 0 has a moderate number of malls and cluster 1 has the highest number of malls, a marketing manager should precisely monitor the competition and allocation of marketing budget in all these regions. In the high competition region, he should allocate more budget for the marketing as compare to the region where only 1 or 2 malls are there in which both are selling his products.

Conclusion

It is concluded that the business problem can be significantly solved using the results of experimentation on data. A marketing manager, a builder, a businessman, and other stakeholders should at least do data analysis for the next experimentation in the market.