IBM Applied Data Science Capstone

Proposed Location for a New Shopping Mall in Bangkok, Thailand

By: Audrey Cheang

August 2021

Introduction

For many people, visiting shopping malls is one way to unwind during the weekends and vacations. There are a multitude of activities one can do at shopping malls, such as watching a movie, visiting fashion outlets, dining in at restaurants, purchasing groceries or just simply to window shop and look at the latest promotions and discounts that are available. Shopping malls are also a great hang out location to catch up with friends and family. Therefore, the location of a shopping mall is extremely vital to attract people to patronise it. Considering the perspective of retail shop and business owners, it is likely they will rent a place at a shopping mall to market their products if they envisage a large crowd and high population density in the area. On the same note, property developers also take advantage of the large crowds that shopping malls pull to set up hotels and residential properties in the area. Therefore, the location of shopping malls plays large part in deciding whether it would make constructing one profitable.

Business Problem

This capstone project will identify the optimal locations in Bangkok, Thailand to construct a new shopping mall. Leveraging on data science and analytics techniques as well as machine learning algorithms, this project aims to find the best solution to tackle the identified business problem: Where in the city of Bangkok will you recommend a property developer set up a new shopping mall?

Scope and Target Audience of Project

The target audience of this project is that of property developers, investors and retailers looking to invest in a new shopping mall in Bangkok, Thailand. Bangkok is one of the world's top tourist destinations with more than 22.7 million arrivals per year. Bangkok also ranks 4th in cross-border spending and has a diverse range of attractions to cater to both international and domestic tourism in the country. As such, it is a prime city for developers to build shopping malls as well as retail owners looking to market their brands by leveraging on the high volume of people in the area.

Data Science Methodology

To tackle this problem, we will need to obtain the following data:

- List of districts in Bangkok, along with the latitude and longitude coordinates of these districts

- Data on the existing shopping malls

Data Sources

The data on different district information can be obtained from Wikipedia (https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok). From this, we will web scrape the page to obtain the relevant information we need.

Thereafter, the Foursquare API will be used to get data on the existing shopping malls in these districts. Upon collection of data, machine learning techniques (k-means clustering) will be employed to determine the optimal location to build the new shopping mall.

Data Science and Analytics Skills

The data science skills employed in this project include web scrapping, utilising an API, cleaning, and wrangling of data, machine learning algorithms (k-means clustering) and data visualisation skills.

Methodology

Firstly, data regarding the districts in Bangkok were obtained from Wikipedia. Fortunately, the latitude and longitude data were available on the same page and hence, we scraped the page to obtain the data frame containing all the districts. After which, the columns irrelevant to our analysis were dropped and thereafter, the Folium package was used to visualise the map of Bangkok and its districts. As such, we could check to ensure that the coordinates provided were accurately plotted in the map.

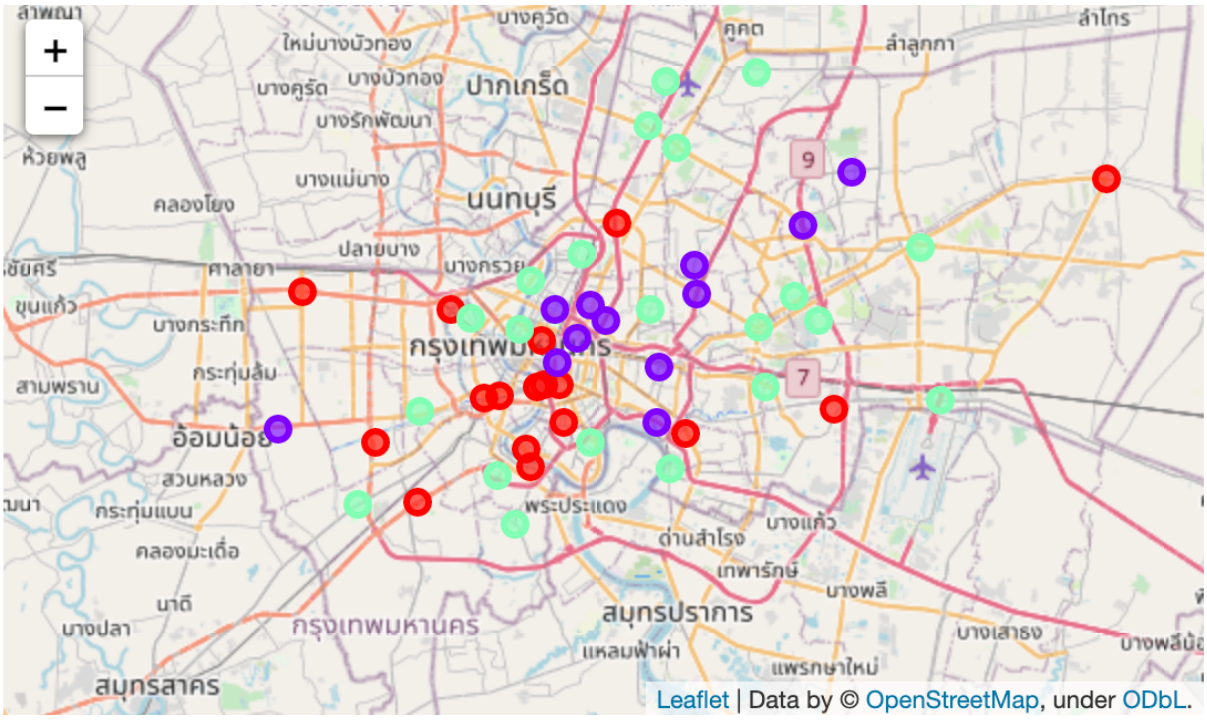
After cleaning and obtaining a visualisation of our location data, the Foursquare API was used to get different venues in each district, with a radius of 10,000. This was done by making an API call to Foursquare by passing in the coordinate values, and this process was automated by using a loop. From our collected data, we could determine the different venues and the categories they fall into. In addition, the number of unique categories was also analysed. For this project, as we were considering the shopping malls in Bangkok, we filtered the districts and the shopping malls in each of this district. Next, we took the mean to find out the concentration of shopping malls in each district.

Lastly, the data was clustered using k-means clustering. K=3 clusters were used, and the k-means clustering algorithm then allocates every data point to the nearest cluster. In our case, the districts were allocated to 3 different clusters based on the frequency of occurrence of ‘Shopping Mall’ in our data set. As a result, the concentration of shopping malls in each cluster was identified and plotted on a map using the Folium package.

Results

The map obtained from the clustering algorithm is shown below, whereby:

- Cluster 0 (red): Moderate concentration of shopping malls in district
- Cluster 1 (purple): High consternation of shopping malls in district
- Cluster 2 (green): Low concentration of shopping malls in district



Discussions

From the map of the different districts and their identified clusters, the green cluster represent the districts with the lowest concentration of shopping malls while the purple cluster represent the districts with the highest concentration of shopping malls.

This could possibly imply that the green areas are potential areas to open new shopping malls given the low levels of competition from existing shopping malls. On the other hand, purple areas are likely to be more populated with shopping malls and constructing a new shopping mall would imply that there would be intense competition in the area.

By zooming in on to the map, it was inferred that the purple districts are areas that are more highly populated, with many residential buildings observed, along with several essential services like hospitals and public transport stations. Ostensibly, the population density in these areas would be higher, and as expected, retailers would anticipate high volumes of people, resulting in the correspondingly higher concentration of shopping malls. On the flipside, the green districts have residential buildings more spread out, which would imply a lower shopping mall concentration. Therefore, this presents an opportunity to enter the

market as the low concentration of shopping malls coupled with the lack of options for consumers in the area would make a new shopping mall complex in these areas very attractive.

Limitations of Study

This project only considers the concentration of shopping malls in each district to identify potential areas to build a new shopping mall. However, there are other factors that influence whether a shopping mall should be built, and these include the accessibility to shoppers, the average wealth and income levels of a particular district, whether there are places of attraction in the district that would draw tourists. Further research should delve into these areas and identify the reasons that could have contributed to the existing high and low concentration in these districts.

Conclusion

In this project, we have undergone the process of solving a business problem by first identifying the data that we needed, extracting that preparing the data, analysing, and visualising the data and finally, utilising machine learning to cluster the data, in order to provide a solution to the business problem 'Where in the city of Bangkok will you recommend a property developer set up a new shopping mall?'.

Upon analysis, it was concluded that that low concentration districts, as identified from our study, are potential areas where developers can consider building a new shopping mall complex. However, it was also acknowledged that this study did not cover certain factors that could have possibly affected the status of the data that was obtained and analysed, and deeper analysis into these areas are required to come to a more comprehensive conclusion.