

IBM Applied Data Science Capstone

Opening a new Restaurant in Colombo, Sri Lanka



Project Created by: Tisal Nanayakkara

Date: July, 2020

Introduction

Colombo is the commercial capital and the largest city in Sri Lanka by population. It is the financial center of the island and a tourist destination.

Majority of the Sri Lankan cooperation's head offices are situated in Colombo, hence the workforce in the country focuses on traveling or staying in Colombo on a daily basis. The Colombo Metropolitan area is the most important industrial, commercial and administrative center in Sri Lanka.

When we talk about Colombo as a tourist destination, it boasts a rich colonial heritage as a mix of different cultures, religions and races. Colombo's multi-ethnic legacy can be observed with churches, mosques and temples congregating on the same streets, and one could also delve into iconic sites including Fort Station, the Red Mosque, and the Khan clock tower. A shopper's paradise, exquisite boutiques feature international brands, and shops offer colorful local art and textured fabrics. Featuring a fascinating heritage and enticing culture, Sri Lanka's capital is one of Asia's favored travel destinations. Offering a vast selection of accommodation options, including rustic beachside shacks to lavish colonial mansions and stunning hotels, tourists can satisfy their every taste and desire.

The rapid development that can be seen in Sri Lanka during the post war era of the country has encouraged investors to set up shop specially around the Colombo region and there has been a sharp increase in the local crowd who explore Colombo and spend their time leisurely or to meet up with people in restaurants and bars. Colombo has become a lively city full of opportunity and growth.

From an Investor's point of view, Sri Lanka is one of the best places to invest or start a business due to the following reasons:

- Government Stability
- Annual Tourism records which continue to improve every year
- Internal Infrastructure
- Upcoming investments, ports and airports being built in the country

Due to the above reasons, investing in opening a new restaurant in Colombo could become a highly profitable venture for both local as well as foreign investors.

Business Problem

Since we have established the fact that Colombo is a suitable area to invest in Sri Lanka, the objective of this project is to analyze and find the best locations to open a new restaurant in Colombo. Data will be obtained through websites such as Wikipedia and also from the Foursquare API for this project and machine learning algorithms such as clustering algorithm will be used to come up with the answers for the questions an investor might have when investing in a new venture.

Target Audience

- Investors who would want to invest or open new restaurant in Colombo
- Any local or foreign individual who would want to know about the existing restaurants and other interesting places around Colombo so that they could get an idea about it.
- Any other Data Analyst or Data Scientist who would find the observations we find through completing this project to be important for their own analysis.

Data required for our analysis

The Data required to understand our problem so as to select the best area to open a new restaurant included:

- The list of cities or neighborhoods which fell under the jurisdiction of the district of Colombo – Web scraping was done using Pandas and the data required for this purpose was obtained through Wikipedia page ‘Postal Codes of Sri Lanka’ under the following link
https://en.wikipedia.org/wiki/Postal_codes_in_Sri_Lanka
- The coordinates of those neighborhoods which were the Longitudes and the Latitudes of those cities – Geocoding, which is the process of converting physical addresses to a location on the Earth’s surface was used to obtain the coordinates of the cities. Geopy and geocoders was installed to extract the required information.
- Data about the different shops and venues in those area, specially data about current restaurants so as to build a model which would best fit our requirements – Foursquare API was used to obtain the relevant Venue data and special emphasis was given to restaurants.
- The data that we obtained from the above methods were transformed and converted to a form which was useable through data cleaning and data wrangling so that only the relevant information was obtained and the best model to predict a place to open a new restaurant could be created.

Methodology

The first step in analyzing this problem was to obtain the required data of the Colombo district and its cities. To extract this data, we used the 'Postal Codes of Sri Lanka' Wikipedia page. The cities which fell under the jurisdiction of Colombo, along with their postal codes was extracted using pandas and was stored in a data frame for analysis. Next we required the geographical coordinates of the cities we obtained. Geopy and Geocoders libraries were used to get the relevant coordinates and the data was appended to the data frame consisting of the cities we extracted before. After the data was gathered, Folium was used to visualize the Colombo district, along with the cities superimposed on it to check if the coordinates we obtained from the earlier steps was correct. The result confirmed that all coordinates were correct.

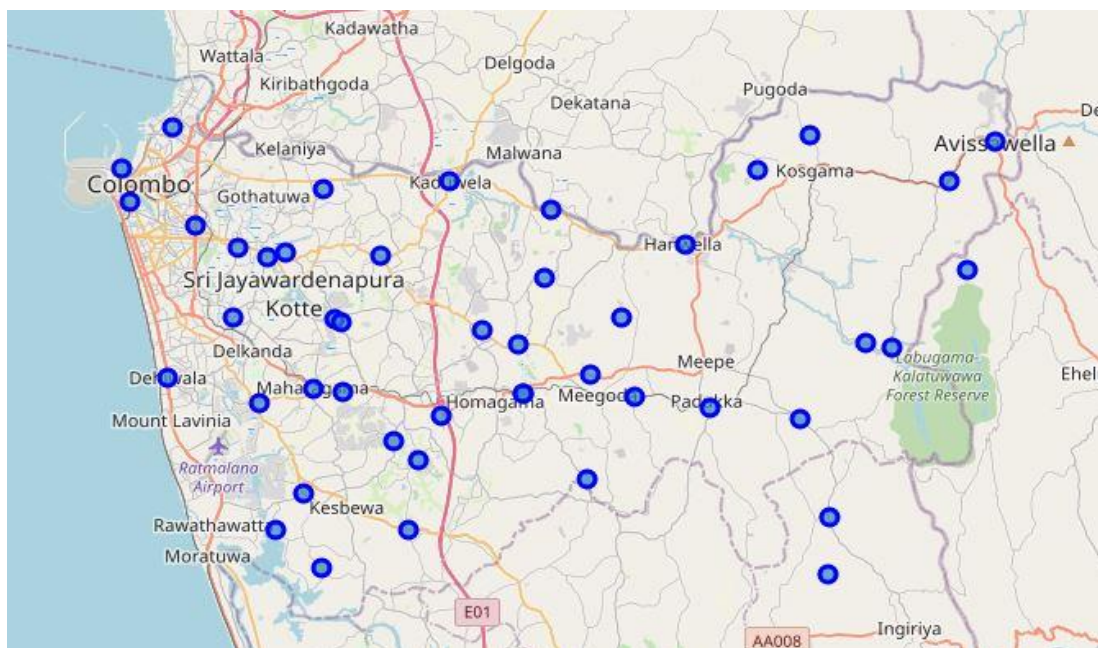


Figure 1 – Map of Colombo

Next, Foursquare API was used to retrieve the venues around the cities we obtained. Venues which were within the range of 1000m were selected and the number of venues was limited to 100. The data we obtained was saved into a data frame and the names of the venues, type of the venues and their coordinates were examined. The data of the venues were grouped according to venue count in each city and since we are interested in opening a restaurant, the different types of restaurants were grouped together to analyze the number of restaurants in each city.

as well as the type of restaurant in each city. It was understood that the location to be picked to open the restaurant should be busy with high foot traffic as well as a place which is highly developed. Hence, the 10 cities with the highest number of venues was extracted and used for further analysis.

Each city was analyzed and the cities were grouped by taking the mean of the frequency of occurrence of each category. Using this data, cities along with their top 10 most visited venues was saved into a data frame for exploratory analysis. K-means algorithm, which is an unsupervised machine learning algorithm was used to cluster the cities based on the frequency of their venues. Once the clustering was complete, the results were viewed to understand which location would be the best to open up a new restaurant.

Results

- The data was clustered into 6 clusters and it could be seen that when examining these clusters, cities like Fort, Hokandara, Battaramulla and Maharagama were cities where restaurants were not the most frequently found venues.
- In contrast to those clusters, Dehiwala, Nugegoda, Rajagiriya and Sri Jayawardenapura were cities where most of their common venues were restaurants
- Cluster 3 contained the busiest cities in the Colombo district, which are Borella and Slave Island, which consisted of the most venues in a 1000m radius.

Discussion and Recommendations

- If we examine the data from a competition stand point, Hokandara is the city with the least competition for an upcoming restaurant since the frequency of restaurants as a common venue is low compared to other cities. So if an investor wants to select a city where its comparatively safe to open a restaurant, he can select a city such as Hokandara
- But, according to entrepreneurs and other experts in this field, what makes a restaurant locations great are the facts that they are accessible and visible. Opening up a restaurant in a high traffic area, urbanized and developed area gives restaurants much more visibility. Also a location in a busy night time city will also attract a lot of customers. Hence Cities like Borella and Slave Island which are urbanized and having a busy night life will be a perfect city

for an investor who would like to invest in a high risk but high reward venture.

- The data that was obtained about the venues was obtained from Foursquare only. With bigger datasets and more variables such as the land prices, parking availability etc., a much more comprehensive model could have been built.
- When using the K-means algorithm, one of the most important factors to consider is the number of clusters, or the k value. To obtain the best k value, an iterative function was written to check the squared error of the model against the number of clusters and the 'Elbow method' was used to pick the k value. The result didn't show a clear elbow hence the value of 6 was chosen as the number of clusters.
- There are other clustering methods such as DBMS clustering methods which can cluster shapes other than circles hence might give better intuition to our data.

Conclusion

The purpose of this project was to come up with good locations for anyone interested in opening up a new restaurant. Foursquare API and Wikipedia was used as the sources for our data. Based on the results of our project, a person can come up with the sort of restaurant they would like to open and the location they choose to be the best. This data could also be used to build more comprehensive models for other data scientists who might be interested in this research. This project gave me the insights of the work a data scientist has to perform to analyze as well as represent the data to its target audience.