Capstone Final project report

Advise restaurant owners on neighbourhoods of Colombo, SriLanka to open new restaurants

Vishal Kulkarni

August 2019

Capstone Final project report

Introduction

Restaurants are places of congregation and communication. They're public yet very private rooms in which many people might feel more at ease than at home when in certain company. They're equalizers - the executive with six houses and the employee with a mortgage under water get the same chairs, the same plates, and the same food. Restaurants provide shelter and entertainment. For many many restaurantgoers, visiting restaurants is a great way to relax and enjoy themselves during anytime of the day particularly weekends and holidays. For restaurant business owners, the location and the customers visiting the restaurants provides a great distribution channel to expand their business and services as restaurant chains based on their unique selling propositions. Business owners are taking advantage of this trend to build more restaurants to cater to the demand. As a result, there are many restaurants in the city of Colombo, SriLanka and many more are being built. Opening restaurants allows business owners to earn consistent revenue and also increase employment. Of course, as with any business decision, opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure.

* Competitiona and other business
* Corrlation of restaurant location and rent

* Size
*Accessibility

* Local Zoning regulations

Capstone Final project report

Business Problem

The objective of this capstone project is to analyse different neighbourhoods in Colombo, SriLanka to advise restaurant business owners on best locations to open new restaurants to expand business. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to resolve the business question: If a restaurant owner wants to open new restaurants in Colombo, SriLanka to expand the business, where would you advise them to open?

Stakeholders for this project

This project is particularly useful to Restaurant Business owners and investors looking to open or invest in new restaurants in the capital city of Colombo, SriLanka. This project is timely as the city is currently going through oversupply of restaurants to cater to increase tourist footfall. Data from the Sri Lanka tourism report released last year showed that an additional 12 per cent growth is projected to existing restaurants.

Capstone Final project report

Data

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

- Neighbourhood details in Colombo, SriLanka. This defines the scope of this project that is confined to the city of Colombo, the capital city of the country of SriLanka in Asia.
- Geographical coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to restaurants in Colombo. This data is used to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Colombo) contains a list of neighbourhoods in Colombo, with a total of 63 neighbourhoods. Webscraping techniques are used to to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then geographical coordinates of the neighbourhoods are extracted using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, Foursquare API is used to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers.

Foursquare API will provide many categories of the venue data, for this project particular interest is in the restaurants category in order to help to solve the business problem put forward. This is a project that will make use of many 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). The next section will detail the Methodology where the steps taken in this project will be discussed, the data analysis performed and the machine learning technique that was used.

Capstone Final project report

Methodology

To perform the data analysis, following methodology is used:

- Webscrape Colombo neighbourhoods wikipedia page for neighbourhoods list - To start with, need to get the list of neighbourhoods in the city of Colombo, Sri Lanka. Fortunately, the list is available in the Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Colombo). Webscraping techniques using Python requests and beautifulsoup packages are used to extract the list of neighbourhoods data. The result is a list of neighbourhood names
- Extract geographical coordinates using Geocoder - get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, Geocoder package is used that will allow to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, data is populated into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package. This allows to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Colombo
- Explore neighbourhoods using Foursquare API - Foursquare AP is used to get the top 100 venues that are within a radius of 2000 meters. Foursquare Developer Account is registered in order to obtain the Foursquare ID and Foursquare secret key. API calls are made to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format, venue name, venue category, venue latitude and longitude.
- Group data by neighbourhood by taking mean of the frequency of each venue category - With the data, venue count returned for each neighbourhood is checked to examine how many unique categories can be curated from all the returned venues. Analysis is performed on each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category that helps in preparing the data for further restaurants category analysis and clustering
- Analyse each neighbourhood as per the venue category and filter venue category by restaurants – Analyse the restaurants data by filtering “restaurants” as venue category for the neighbourhoods.
- Perform clustering by using k-means clustering - Clustering on the data is performed 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. Neighbourhoods are clustered into 3 clusters based on their frequency of occurrence for “restaurants”

- Visualize the clusters details in a map generated - The results will allow to identify which neighbourhoods have higher concentration of restaurants while which neighbourhoods have fewer number. Based on the occurrence of restaurants in different neighbourhoods, it will help to answer the question as to which neighbourhoods are most suitable to open new restaurants

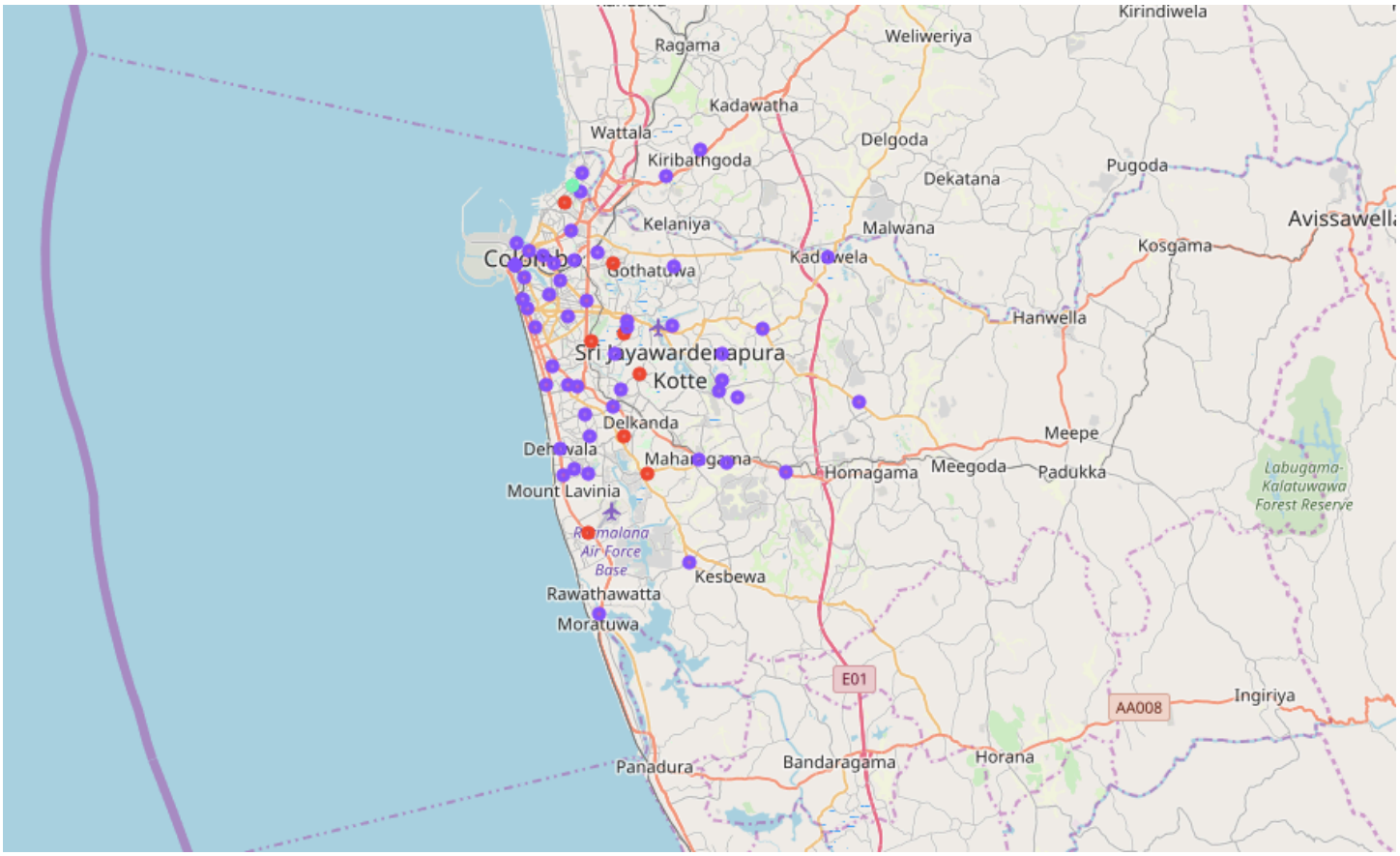
Capstone Final project report

Results

The results from the K-means clustering show that neighbourhoods are clustered into 3 clusters based on the frequency of occurrence for “restaurants”:

- Cluster 1: Neighbourhoods with approx 13% (moderate number) of restaurants
- Cluster 2: Neighbourhoods with approx 85% (high concentration) of restaurants
- Cluster 3: Neighbourhoods with only one restaurant

The results of the clustering are visualized in the map below with cluster 1 in red colour, cluster 2 in purple colour, and cluster 3 in mint green colour.



Results

As observations noted from the map in the Results section, most of the restaurants (approx. 85%) are concentrated in a single neighbourhood of Colombo city, with the highest number in cluster 2 and moderate number (approx. 13%) in cluster 1. On the other hand, cluster 2 has very only 1 restaurant in the neighbourhoods. This represents a great opportunity and high potential areas to open new restaurants as there is very little to no competition from existing restaurants. Also, restaurants in cluster 2 are likely going through intense competition due to surplus and high concentration of restaurants. From another perspective, the results also show that the surplus of restaurants mostly happened in a neighbourhood areas of single cluster of the city, with other cluster neighbourhood areas still have very few restaurants. Therefore, this project recommends restaurant business owners to take advantage of these findings to open new restaurants in neighbourhoods in cluster 3 with little to no competition. Restaurants with exclusive cuisines who stand out from the competition can also open new restaurants in neighbourhoods in cluster 1 with moderate competition. Lastly, restaurant business owners are advised to shun neighbourhoods in cluster 2 which already has high concentration of restaurants and going through intense competition.

Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of restaurants, there are other factors such as population and income of residents that could influence the location decision of a new restaurants. However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new restaurant. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

Capstone Final project report

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. restaurants business owners and investors regarding the best locations to open a new restaurant. To answer the business query that was raised in the introduction section, the response proposed by this project is: The neighbourhoods in cluster 3 are the most preferred locations to open a new restaurant. The findings of this project will help the relevant stakeholders to take advantage of the opportunities on high potential locations while avoiding intense competition areas in their decisions to open a new restaurant.

References

Category:Suburbs of Colombo, SriLanka. *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Category:Suburbs_of_colombo

Foursquare Developers Documentation. *Foursquare*. Retrieved from https://developer.foursquare.com/docs

Appendix

Cluster1:

Sri Jayawardenepura Kotte
Kotahena
Batuwatta
Boralesgamuwa
Koswatte
Kolonnawa
Narahenpita
Ratmalana

Cluster2:

Nawala	Kottawa	Kotikawatta
Mount-Lavinia	Madampitiya	Wellawatte
Moratuwa	Bambalapitiya	Hulftsdorp
Athurugiriya	Battaramulla	Wickramasinhapura
Nugegoda	Borella	
Maradana	Cinnamon Gardens	
Maligawatta	Colombo	
Mattakkuliya	Dalugama	
Pamankada	Dehiwala	
Pelawatte	Dehiwala-Mount Lavinia	
Pannipitiya	Dematagoda	
Malabe	Fort (Colombo)	
Pettah, Sri Lanka	Grandpass	
Piliyandala	Havelock Town	
Ragama	Maha Nuge Gardens	
Rajagiriya	Hokandara	
Slave Island	Ja-Ela	
Template:Suburbs of Colombo	Kadawatha	
Thalawathugoda	Kaduwela, Western Province	
Union Place	Kalubowila	
Wattala	Kandana	
Welikada	Kiribathgoda	
Panchikawatte	Kirulapana	
Maharagama	Kohuwala	
	Kollupitiya	

Cluster 3:

Modara