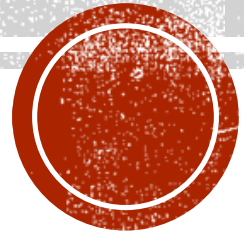
IBM APPLIED DATA SCIENCE

Capstone Project

Opening a new restaurant in Colombo, Sri Lanka



by
Ashane Fernando

INTRODUCTION

- Colombo is a booming commercial hub, with a thriving nightlife and a rapidly developing restaurant culture.
- Due to the rapid boom in tourism, there is a large influx of foreign tourists to the city who are eager to patronize the local establishments.
- Although there are already many restaurants within Colombo, there is still ample opportunity for a new restaurant to thrive, provided that they position themselves well; both in terms of location as well as business strategy.
- The presence of other types of venues, such as hotels and entertainment venues in the immediate area of a restaurant can play a major role in its success.

BUSINESS PROBLEM

- Location is a key factor which can determine whether a restaurant succeeds or fails.

- **The objective of this project is to determine the best neighborhoods in Colombo to open a new restaurant.**

- We would use the presence of other restaurants as well as other types of venues as a metric to determine the suitability of a neighborhood for a new restaurant.

- **<u>Stakeholders who would be interested in this project</u>**
  - Local and foreign investors and restauranteurs who would like to open a restaurant.
  - Current restauranteurs who may wish to expand further.

DATA

- **Data Requirement**
  - A list of neighborhoods within the Colombo city limits
  - Geographical coordinates for each of the neighborhoods.
  - Venue data for the neighborhoods, with particular emphasis on restaurants, hotels and entertainment venues.

- **Data Sources**
  - The Wikipedia page of neighborhoods in Sri Lanka
    (https://en.wikipedia.org/wiki/Postal_codes_in_Sri_Lanka)
  - Geocoder package to obtain location data for the neighborhoods.
  - Foursquare API to obtain data on venues.

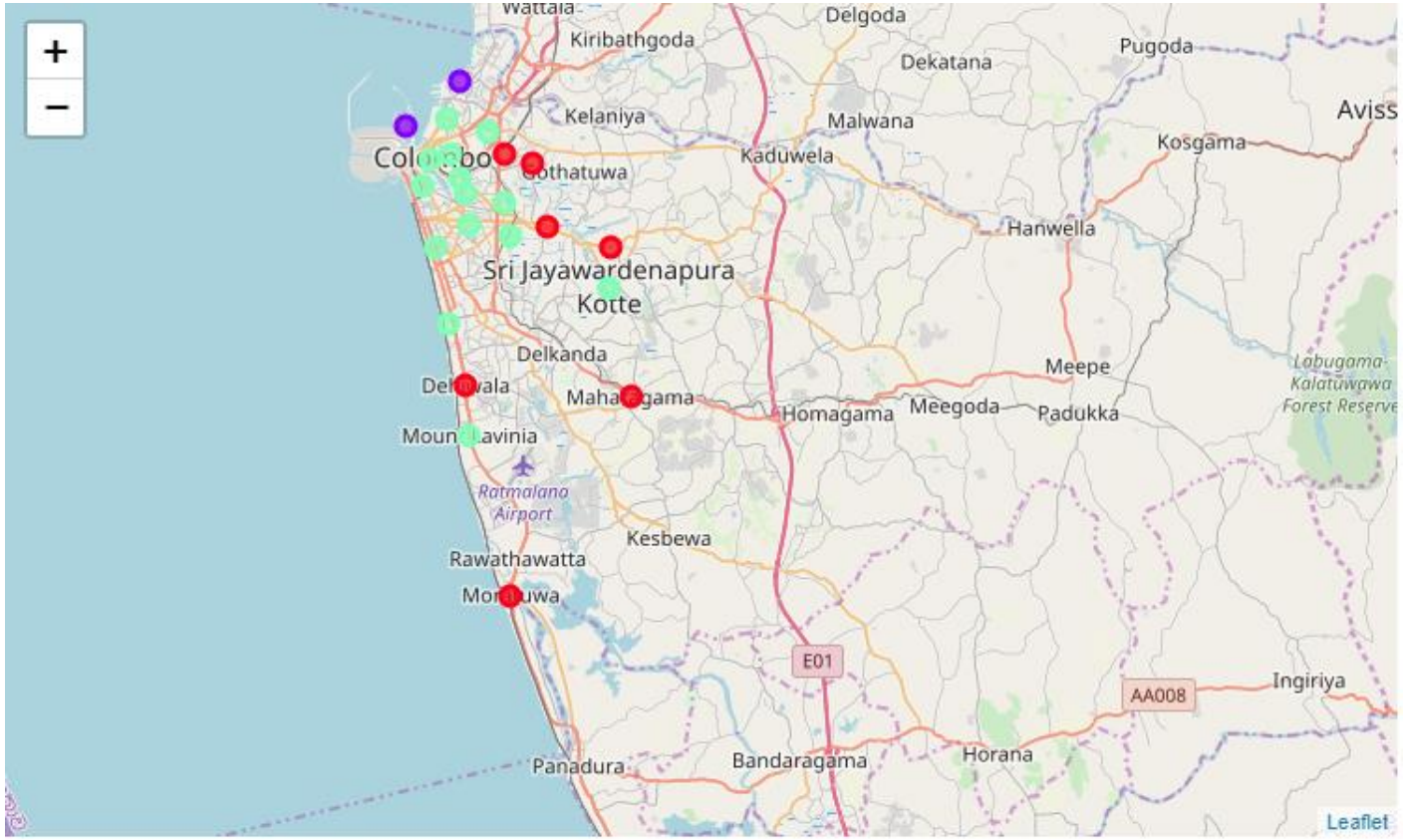
METHODOLOGY

- Scrape the Wikipedia page to obtain the list of neighborhoods.
- Obtain the latitude and longitude coordinates for the neighborhoods using Geocoder.
- Use the Foursquare API to extract the venue data for each of the neighborhoods.
- Group the venues by neighborhood and then determine the average proportion of hotels, restaurants and entertainment venues in the area.
- Cluster the neighborhoods according to the above proportion, using K-mean clustering.
- Visualize the clusters on a map generated using Folium.

RESULTS

The neighborhoods were categories into three clusters.

Cluster 0: Red
No hotels or entertainment venues, restaurants present
Cluster 1: Purple
No restaurants, hotels and entertainment venues present.
Cluster 3: Lime green
High proportion of restaurants, but hotels and entertainment venues present.



DISCUSSION

- Cluster 0 has a high concentration of restaurants, and thus a new restaurant would face heavy competition. This cluster also has no hotels or entertainment venues, therefore the number of prospective clients in the area would be expected to be low.
- At first glance, the neighborhoods in Cluster 1 appear to be the ideal place for a restaurant since they contain hotels and entertainment areas, but no restaurants. However on closer look we find that these locations are not as developed as we hoped.
- Cluster 2 has a rather high concentration of restaurants as well, but also has a fair number of hotels and entertainment venues, which means that we can expect a fair number of prospective clients in the area.

RECOMMENDATIONS

- Ideally the restaurant should be opened in a neighborhood in Cluster 2.
- Neighborhoods in Cluster 0 can also be considered since although they don't have much in the way of other venues, the number of existing restaurants in the area imply that there is still some business opportunity there.

CONCLUSION

- Thus our business question is answered.
  - The neighborhoods in Cluster 2 would be the preferred locations for a new restaurant.

- The findings of this project would assist any investor or restauranteur in selecting a good neighborhood to start a restaurant