Coursera Capstone Project

IBM APPLIED DATA SCIENCE CAPSTONE

Opening a new restaurant in Dhaka

By: S. M. Asif Mahfuz

October 2019

Introduction:

Restaurants are the hottest business everywhere in the world. Day by day people are turning to takeout foods because of their increase in working time. For many owners this is a great chance to earn more on this situation. For this reason more and more people are turning on to open their own restaurants. But this is not just opening a restaurant anywhere. Placement of a location is more important than people think. Most of the times it can be the most influential factor for whether a restaurant business will succeed or fail.

Problem Statement:

The target of this capstone project is to analyze the best location for a person to open a new restaurant in the city of Dhaka, Bangladesh with the help of Data Analysis and Unsupervised Machine Learning technique called K-Means Clustering.

Target Audience:

This project is particularly targeted for anyone who want to open a new restaurant. In the city of Dhaka, Bangladesh restaurant business has gone up so much in the past few years. But many faces failure as a result of choosing a poor place. And as 18% tax is for restaurants set by the Government of Bangladesh it is really crucial to choose the place for the restaurant.

Data:

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

1. List of neighborhoods in Dhaka, the capital city of Bangladesh.
2. Latitude and Longitude of these neighborhoods.
3. Venue data for these neighborhoods.

The list of neighborhood will be obtained from https://en.wikipedia.org/wiki/Dhaka using python beautiful soup package. Latitude and longitude will be obtained from python geocoder package. And finally the venue data will be obtained from foursquare.com with a developer API.

Methodology:

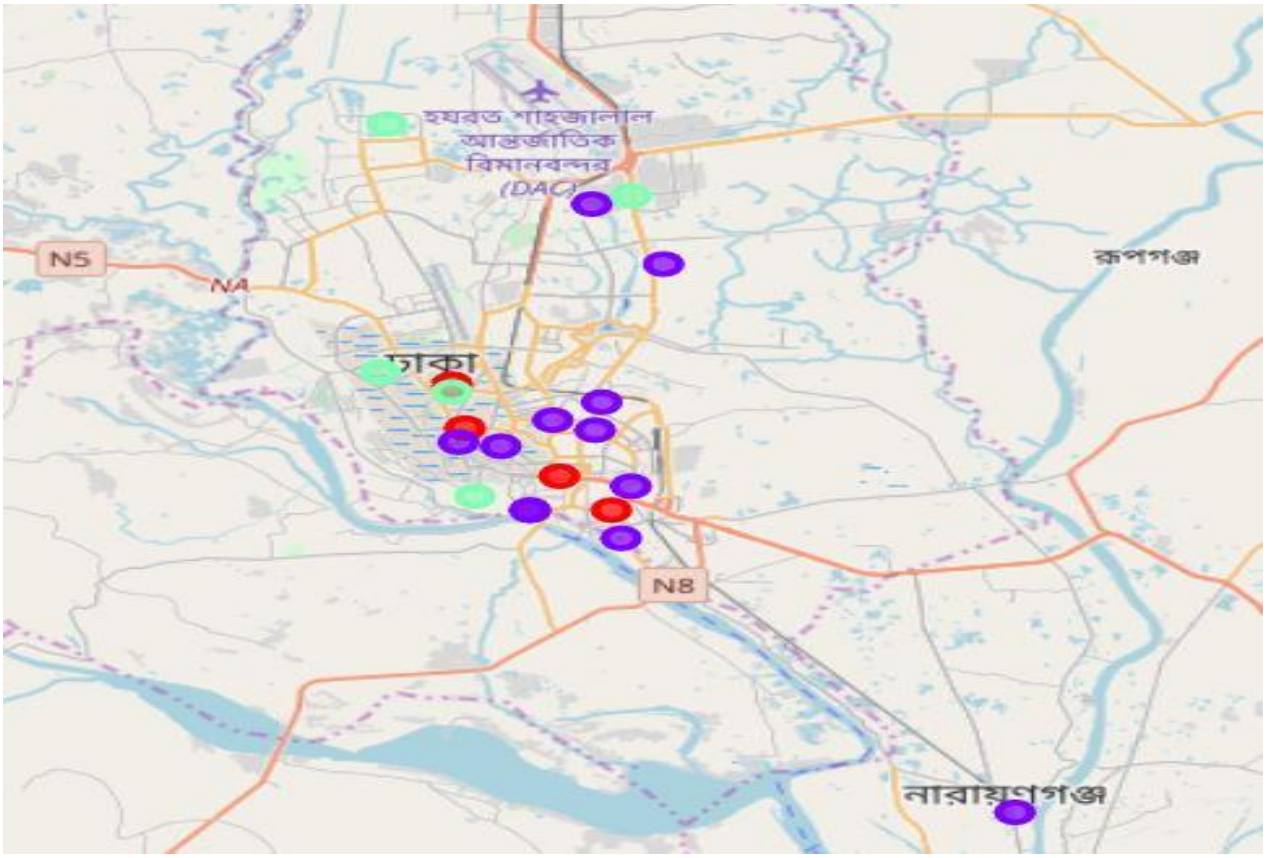
At first we obtained the neighborhoods of Dhaka, Bangladesh from Wikipedia (https://en.wikipedia.org/wiki/Dhaka) using request package from Python. Then we used Beautiful Soup package to obtain the neighborhood names as a python data frame. Then using geocoder package we obtained the latitudes and longitudes of the neighborhoods as an array. Then we added those latitudes and longitudes to the neighborhood data frame. Then we used folium package to visualize the neighborhoods.

After that we obtained the venues in those neighborhoods and added them to the data frame. We sorted out the restaurants in the data frame by one hot encoding. Then we found the means for our K-means clustering. We set the value of K to be three and performed the clustering.

Result:

The result shows that we can divide the neighborhoods in three categories based on frequency.

1. Cluster 0: Neighborhoods with moderate no. of restaurants denoted by red color.
2. Cluster 1: Neighborhoods with low no. of restaurants denoted by purple color.
3. Cluster 2: Neighborhoods with high no. of restaurants denoted by mint color.



Discussion:

From the map shown in the results it can be shown that most of the restaurants are located in the top part of Dhaka city. A few are located in the bottom part. But it is clear that the cluster 2 locations, which have the most number of restaurant frequencies, are on the top part. The moderate

frequency of restaurants are found to be evenly distributed which is shown by cluster 0, denoted by red. Other areas are not that popular for restaurants.

Conclusion:

In conclusion, it can be said that one can open a restaurant and get by in cluster 1. But if you are confident enough to take up the competition, you can consider opening your business at cluster 2. Cluster 1 is not suggested for the business. This is only based on frequency of distributions of restaurants. Further studies can be done using other parameters.