Selection of best location to open a new restaurant in India

Minhaj Ahmed Ansari

1. Business Problem

1.1. Introduction

The India business landscape is prepared to grow in multiple areas due to several factors such as international trade, government stimulus, and an overall strong developing country. With a young population that is rising to leadership and technology driving growth and innovation, there are several business opportunities in multiple sectors that we are going to explore today.

Many small business start-ups, in order to reduce expenditure usually settle for a cheap location. Others believe that location doesn't matter provided the product is right. However, getting a good location is very critical to the success of your business but this can prove quite difficult because one of the challenges of starting a business successfully is getting a good business site.

Now, for a food joint or restaurant, the location in itself becomes one of the major contributing factor towards success. Because for people to like your food, delicacies and environment you offer, they first need to reach it and distinguish it among the competition. Assuming that a Multi-national fast food chain wants to open its new chain of restaurant in India, which location should they choose? Which type of cuisine should they serve? And most of all how to ensure people are easily able to reach the location.

1.2. Problem

Now India is a large country, with a population of more than 1.3 billion people. It will become impossible to simply zero in on perfect location by starting randomly choosing a place or city. So, it's better to first identify most viable target locations and then probably decide among them based on various virtues and features. For this reason alone, I restricted my analysis to 5 major cities of India, which in a way are Economic, Technological and demographic centres of India. They are,

- Delhi
- Mumbai
- Chennai
- Kolkata
- Bengaluru

Now, we need to just compare these cities and identify the best of them based on some common parameters. These parameters will eventually help us to select the best features to recommend best location.

- Optimal distance between city centre and the proposed location.
- Selection of cuisine.
- Density of local businesses at the proposed location.

- Traffic Conditions (also instrumental in case of online deliveries)

However, it is important to note that, these are not the only parameters that contribute in the selection of parameters other factors such as rent, condition of building etc. which also impact the success of business.

However, this analysis can give a head start to businesses which they can utilise to do further deeper analysis of other features.

1.3. Interests

This analysis would be really helpful for the small businesses and start-ups who are looking to locate their stores/ restaurants at optimal location to get a jump start.

2. Data Collection and wrangling

2.1. Data Sources

Now there are many location services which provide various details about and around a particular location in a city. However, Foursquare is one of the freely available services which return a list of the recommended venues around a location based on various factors which include distance, type of business and likes or dislikes.

An easy way to get this data is through Foursquare Places API's **explore** endpoint. The documentation on how to create foursquare developer account, rate limits and calling methods is available on the link here.

Another data source that we will be using in this project will be **Traffic API** provided by Bing Maps. This API can be used to get traffic information on a particular set of location. A complete detailed documentation on this REST API is available here.

2.2. Data cleaning

For calling Foursquare API's explore endpoint, first we need is the spatial details of the location around which you want to explore the venues. We can choose this location as the city centre as explained in section 1.2. Another important parameter that you need to specify is 'radius'. This will determine the radius within which the venues are returned. For the purpose of this analysis let us fix it to 10 KM as beyond that, a. customers might not travel in peak times b. online delivery might not be possible.

After initial cleaning of the returned json data file, we populate all the responses in separate data frames (one for each city) using Pandas library. Each data frame consists of almost 100 rows and 22 columns.

Next we will rename some of the columns so that it is easier to manipulate and understand them. For example - 'venue.categories' can be renamed as only category.

Another cleaning exercise that we will do would be to get the exact category name, since the column returned contains a list.

Next we will drop unnecessary columns which do not contribute in any meaningful way to our analysis. These include - **['venue.photos.count', 'venue.photos.groups', 'venue.venuePage.id', 'venue.location.neighborhood', 'referralId', 'reasons.count', 'reasons.items', 'venue.location.cc', 'venue.location.postalCode', 'venue.location.labeledLatLngs']**

After above all operations now we are left with following columns in our data frame - **['id', 'name', 'address', 'crossStreet', 'lat', 'lng', 'distance', 'city', 'state', 'country', 'formattedAddress', 'categories', 'category']**

2.3. Data Binning & One Hot Encoding

This step will especially help us in performing some Exploratory data analysis and Data Visualisation in upcoming Sections.

First, one hot encoding is possible for 'Category' column, this will help us in identifying the most popular category and their distribution across the cities.

Next, the 'distance' feature is actually a continuous value which will be difficult to plot for every venue. To achieve more appropriate analysis, let us bin them in four categories.

- Very Near = 0 to 2 KM
- Near = 2 to 4 KM
- Far = 4 to 6 KM
- Very Far = 6 to 8 KM

Note: No/ Negligible Venues are actually available beyond 8 KM radius of city centre of all cities

One hot encoding for these binned distances is also possible and it can be done for further EDA during our analysis.

2.4. Feature Selection

Above data wrangling and cleaning leaves us with all the features, which are required for our further Exploratory data analysis and Modelling. Further features may be dropped once we have completed our EDA.