

# *Opening a New Restaurant in Indore, India*

Sakshi Mahajan

16 Jan 2020

## **1.Introduction:**

### **1.1. Background:**

A restaurant or an eatery, is a business that prepares and serves food and drinks to customers. Restaurants vary greatly in appearance and offerings, including a wide variety of cuisines and services models ranging from inexpensive fast food restaurants and cafeterias, to mid-priced family restaurants, to high-priced luxury establishments.

There are many restaurants in Indore city and many more are being built. Opening Restaurants allows property developers to earn consistent rental income or a big one-time investment. 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 it will be a success or a failure.

### **1.2. Problem:**

The objective of this capstone project is to analyse and select the best locations in Indore city to open a new Restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Indore city of Madhya Pradesh , India ,if a property developer is looking to open a new Restaurant, where would you recommend that they open it?

### 1.3. Interest:

This project is particularly useful to property developers and investors looking to open or invest in new Restaurant in the food capital of Madhya Pradesh i.e Indore.

This project is timely as there is no other city in India that has such a level of food crazed people, as you would find here in Indore.

The locals are also interested to visit up new places nearby their localities.

## 2. Data acquisition and cleaning

### 2.1 Data sources:

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

- List of neighbourhoods in Indore from:-  
[https://en.wikipedia.org/wiki/Category:Neighbourhoods\\_in\\_Indore](https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Indore). This defines the scope of this project which is confined to Indore, the food capital city of Madhya Pradesh ,India in Asia.
- Latitude and longitude 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. we will use Foursquare API to get the venue data for those neighbourhoods We will use this data to perform clustering on the neighbourhoods.

### 2.2. Data cleaning:

We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare API will provide many categories

of the venue data, we are particularly interested in the Restaurants category in order to help us to solve the business problem put forward.

## 2.3. Methodology:

- We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data. However, this is just a list of names.
- We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude.
- After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted.
- We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude.
- Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering.
- Lastly, we will perform clustering on the data 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.
- The results will allow us to identify which neighbourhoods have higher concentration of Restaurant while which neighbourhoods have fewer number of Restaurants. Based on the occurrence of Restaurants in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new Restaurants.