# COURSERA CAPSTONE

## IBM Applied Data Science

## Neighborhood Recommendation to open a new Indian Restaurant in Ahmadabad City of India

**By: I S Kushagra**

**May 2020**

## INTRODUCTION

Food is the most essential part of any living beings life. Food provides us our daily fuel to work, walk, sleep , play and what not. Humans, being an advanced species has always been fond of food and we don’t miss to try a different cuisine from time to time. Restaurants are our one stop solution for food whenever we are outside, on vacation or just bored of home made food. The restaurants focused around a particular cuisine offers are always famous among people. Indian food is one such cuisine, which people love to have.

But, selecting a perfect location for a Restaurant is the most important and mind boggling decision. This analysis will help define the features of the city and provide a suggestion of neighborhoods where one might open an Indian Restaurant to stay profitable.

## BUSINESS PROBLEM

The main objective of this capstone project is to analyze and select the best location in the city of Ahmedabad, India to open a new restaurant which server Indian Cuisines. Using the methodology and tools of data science and machine learning technique like clustering, this project aims to provide solutions to answer the one major question which is, If in the city of Ahmedabad, a person is looking to open a new Restaurant, where would one recommend that they open it?

# DATA

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

- List of neighborhoods in Ahmedabad. This defines the scope of this project which is confined to the city of Ahmedabad, a city in the state of Gujarat in India
- Latitude and longitude coordinates of those neighborhoods. 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 this data to perform clustering on the neighborhoods.

## Sources of data and methods to extract them

This webpage (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Ahmedabad) contains a list of neighborhoods in Ahmedabad, with a total of 81 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful soup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

Following which, we will use Foursquare API to get the venue data for those neighborhoods. 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, we are particularly interested in the Indian Restaurants category in order to help us to solve the business problem put forward.

Upon receiving the data about the top venues such as museums, shops, restaurants, atms etc from foursquare for every neighbourhood defined in the city of Ahmedabad. Based on the details we can identify the areas where there is a high footfall, i.e. more number of people visit that neighborhood(eg: shopping malls, museums, parks etc), and would most probably be the best location to open the restaurant at.

For example :
Assume a neighborhood “A” and “B” respectively. Upon analysis of these neighbourhood we find that neighborhood A has 2 shopping malls, 10 restaurants and 1 museum. While neighborhood “B” has 2 shopping malls, 3 restaurants and no museums.

In this case, it might seem that the footfall at “A” would be higher than ‘B’, as there are 3 visiting locations, but there are already 10 restaurants in the vicinity, which will increase the competition by a lot. Hence, to stay profitable, opening a restaurant at neighborhood “B” would be more logical.
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). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.