COURSERA CAPSTONE

IBM Applied Data Science Capstone

Opening a New Restaurant Chain in Ahmedabad, India

By: **SUPAN SHAH**

February, 2020

INTRODUCTION

Ahmedabad is the largest city in the state of Gujarat in India. It is also known as the “Manchester of India”. Ahmedabad has emerged as an important economic and industrial hub in India.

Gujarati people are known for their love of food (after the fact that they also love travelling, of course!). The people here are extremely fond of food, so much that on average there are more number of foodies in a household than there are households - well, obvious, isn’t it?

As a result, the food industry in the pertaining area has never been out of business. Opening a restaurant is one of the most profitable businesses in this part of the world. You can find restaurants that have been managed by single families, since ages! Hence, opening restaurants allows firms and entrepreneurs to earn consistent income. Of course, 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 the restaurant will be a success or a failure.

BUSINESS PROBLEM

The objective of this capstone project is to analyse and select the best locations in the city of Ahmedabad, Gujarat, India 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 the city of Ahmedabad, if a company / an entrepreneur is looking to open a new restaurant (chain), where would you recommend that they open it?***

TARGET AUDIENCE

This project will be particularly useful to companies as well as entrepreneurs and investors looking to open or invest in new restaurants in the city of Ahmedabad.

DATA

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

- List of neighbourhoods in Ahmedabad. This defines the scope of the project which is confined to the city of Ahmedabad, Gujarat in India.
- Latitude and Longitude coordinates of those neighbourhoods. This is required in order to plot the map and also get the venue data.
- Venue data, particularly data relating to restaurants. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods employed:

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

After that, we will use the 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 Restaurant category in order to help us to solve the business problem put forward. This is a project that will make use of many skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling to machine learning (K-means clustering) as well as map visualization (Folium).