

CAPSTONE PROJECT REPORT

IBM Applied Data Science Capstone

Opening a new RESTAURANT in MUMBAI, INDIA

By

Likhith Harish

INTRODUCTION:

For many citizens, dining at restaurants is a major part of their lives. So, this has been a huge market for investors and business people to invest into many restaurants. So, there is a high competition into this market. Hence opening a new restaurant is a difficult task. The first problem comes with the location. A new market player can not open his restaurant in a location where many restaurants are already triumphing over competition. So, they should find a place which has low concentration of restaurants and with more customers. If a new comer opens a restaurant in a location with high competition, the business wont be feasible and they finally have to shut down .

BUSINESS PROBLEM:

The objective of this capstone project is to find an appropriate location in city of Mumbai, INDIA. So, we use data science methodologies and machine learning techniques like clustering to solve a very important question for this business. That is, what will be the most feasible location to open a new restaurant in city of Mumbai, India?

TARGET AUDIENCE:

This data and output of the project is most useful to new business entries that are thinking of opening a new restaurant in the Mumbai city. This project is most important to people because of oversupply of restaurants in the city.

DATA

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

- List of neighbourhoods in Mumbai. This defines the scope of this project which is confined to the city of Mumbai, the most populated metropolitan city of INDIA.
- 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 various restaurants. We will use this data to perform clustering on the neighbourhoods. Venue data will include the latitude and longitude also.

SOURCES OF DATA & METHODS TO EXTRACT THEM:

The Wikipedia page https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Mumbai contains the list of neighbourhoods in Mumbai. There are about 135 neighbourhoods in Mumbai. With the help of python requests and Beautiful soup package, we will scrape data from the Wikipedia page and store it in a data frame for further evaluation. Next, we will use Geocoder package in python to get the geographical coordinates of all the neighbourhoods and add them as latitude and longitude columns to the data frame.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest databases 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 Restaurant category in order to help us to solve the business problem put forward. So using Foursquare API , we will obtain various venue names , their latitude and longitude and along with the venue categories and add them to our dataset. We will then utilise the dataset and choose those neighbourhoods with a venue category of Restaurant. But we can anticipate various categories with Restaurant being included in them as Afghan Restaurant , Indian Restaurant etc. As the location is specified to INDIA and reduce complexity venue category has been limited to Restaurant & Indian Restaurant.

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).