The Battle of Mumbai Restaurants

Selecting the perfect restaurant in Mumbai, India

By: **Ankit Mandal**

August 2021

A. Introduction

Whenever a user is visiting a city, they start looking for places to visit and restaurants to eat at during their stay. They primarily look for restaurants based on their ratings and average price, such that the venue fits in their budget.

The aim of the project is to identify restaurant venues in Mumbai, India based on their ratings and average prices. In this notebook, we will identify various venues in the city of **Mumbai, India**, using **Foursquare API** and **Zomato API**, to help visitors select the restaurants that suit them the best. Once we have a plot with the restaurant venues, any company can launch an application using the same data and suggest users such information.

B. Business Problem

The objective of this capstone project is to analyze and select the best restaurants in Mumbai, India based on their rating and average prices.. Using Data Science methodology and Machine Learning techniques like Clustering, this project aims to provide solutions to answer the following business question:

In the city of Mumbai, India, if a foodie is looking for a great restaurant, where would you recommend that they go?

C. Target Audience of this project

This project will be useful to everyone, especially foodies who are looking for an awesome restaurant in the capital city of India i.e., Mumbai. This project is of utmost importance as currently the demand for a good restaurant in Mumbai, with a unique theme and innovative cuisine, is enormous. Owing to this huge demand, new restaurants are coming up almost every day in Mumbai, each with its own specialty dishes and themes. However, with so many restaurants to choose from, people are often simply confused. This project aims to help these people to select a good restaurant, according to their preferences.

D. Data

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

- **Latitude and longitude coordinates of restaurant venues in Mumbai** - This is required in order toplot the map and also to get the venue data.
- **Venue data** - We will use this data to performclustering on the neighborhoods.

E. Sources of data and methods to extract them

This is a project that will make use of many data science skills, from working with APIs (Foursquare and Zomato), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

1. To begin with, we will take a look at **Mumbai on the Map** using the `Folium` library.
2. We will also fetch the required restaurant data from **Foursquare API** and **Zomato API**.

Foursquare API: We will use the Foursquare API to fetch important data like restaurant names, coordinates in Mumbai starting from the middle up to 4 Kilometers in each direction.

Zomato API: The Zomato API utilizes the restaurant coordinates received from the Foursquare API to provide information about various restaurant venues, including the complete address, user ratings, price for two people, price range and a lot more.

F. Methodology

This project aims at identifying the venues in Mumbai based on their rating and average costs. This would enable any visitor to identify the venues he/she wants to visit based on their rating and cost preference.

As a first step, we will retrieve the data from two APIs (Foursquare and Zomato). We extract venue information from the center of Mumbai, upto a distance of 4 Km. The latitude and longitude values are then used to fetch venue rating and price from Zomato.

Secondly, we explore the data retrieved from the two APIs on the map and identify the top category types. The data from the two sources is carefully combined based on the name, latitude and longitude values from the two sources. The final data set would include the rating and price values for each venue.

Next, we will analyze the data that we will create based on the ratings and price of each venue. We will identify places where many venues are located so that any visitor can go to one place and enjoy the option to choose among many venue options. We will also explore areas that are high rated and those that are low rated while also plotting the map of high and low priced venues. Lastly, we will cluster the venues based on the available information of each venue. We will achieve this with the help of K-means Clustering algorithm. This will allow us to clearly identify which venues can be recommended and with what characteristics.

Finally, we'll discuss and conclude which venues to be explored based on visitor requirement of rating and cost.