Analyzing best location for opening a restaurant in Mumbai, India

Capstone Project
IBM Applied Data Science Certificate Program

Project Documentation

Kumar Lokesh Singh
April 2020

Contents

1 Introduction

Mumbai (formerly called Bombay) is a densely populated capital city of state Maharashtra on India's west coast. It is the country's financial and commercial center and its principal port on the Arabian Sea. Mumbai is India's most-populous city, and it is one of the largest and most densely populated urban areas in the world with an estimated city proper population of 12.5 million living under Municipal Corporation of Greater Mumbai. The metropolis has a roadside fast food consisting of Maharashtrian Pav Bhaji, Vada pavs, Dabeli, Panipuri, Bhelpuri, etc. South Indian and Chinese food is also popular in the city. Lebanese, Korean, Thai, Italian, Mexican, Mughlai, Punjabi, Mālvani, and Continental cuisine are also popular in Mumbai.

Mumbai has some of the oldest restaurants in India. Delhi Darbar, Sindhudurg, Highway Gomantak, Samrat, Vitthal Bhelwala, Mahesh Lunch Home, Kailas Parbat, and Adarsh are some of the oldest restaurants in the city.

Mumbai, being the financial capital, has a large migrant population. Members of the migrating families are working with no other family support, hence ordering food from close by restaurants or calling for pre-packaged meals (dabba as called locally) is an accepted thing. These 'Dabbas' are the usually delivered by a special local delivery organization known as the Dabbawallas. Owing to these recent trends, several online food-ordering services have cropped up in the city.

Mumbai has the largest of the organized food service markets pegged at nearly Rs 41,000 crore. Mumbai has the highest share in the organized market -which mainly accounts for standalone restaurants. Interestingly, in Mumbai, among international cuisines, 33% prefer Italian compared to 29% who like Chinese. People in the island city also prefers south Indian food to north Indian cuisines.
The analysis will mainly help new restaurants in examining the factors affecting their restaurant location.

2 Purpose of study

As Mumbai and its sub-urban area is growing, the number of restaurants is increasing day by day. Currently which stands at approximately 13,790 restaurants. The key issues that continue to pose a challenge to them include high real estate costs, rising food costs, shortage of quality work force, fragmented supply chain and over-licensing.
With each day new restaurants, opening the industry is yet to saturate and the demand is increasing regularly. Despite increasing demand, it however has become difficult for new restaurants to compete with established restaurants. Most of them serving similar food and most of the people are dependent on the restaurant food. With such an overwhelming demand of restaurants, it has therefore become important to study the demography of a location. What kind of a food is more popular in a locality? Do the entire locality loves vegetarian food etc.? This kind of analysis can be done using the data, by studying the factors such as

- Approx. price of food
- Location of the restaurant
- Theme based restaurant or not?
- Which locality of that city serves those cuisines with maximum number of restaurants?
- Customers who are striving to get the best cuisine of the neighbourhood
- Is a neighbourhood famous for its own kind of food?

3 Properties

3.1 File Format

The data is in csv format. For best results read the data using Python. You can refer to the notebook for Data Cleaning in this Week5_Final_Project Git folder get a glimpse of how the data was cleaned.

3.2 Content

The dataset contains 16 variables all of which were obtained from Kaggle website. The dataset contains details of more than 6000 restaurants in Mumbai in each of its neighborhood. Detailed explanation about the variables are also available in section 5.

3.3 Size

The total size of the dataset is approximately 3MB. The dataset examined has the following dimensions:

Feature	Result
Number of Observations	6000 plus
Number of Variables	16

4 Source

The data was of Mumbai restaurants of Zomato website was obtained from Kaggle website. Remaining data for clustering was pulled using Foursquare API.

5 Variable names and description

Variable	Type	Description
Cost_for_two	int	contains the approximate cost for meal for two people
Features	object	What are the exclusive features of restaurant and features for which it is known?
Home_Delivery	bool	whether home delivery is available in the restaurant or not
Operational_hours	object	operational timing
Restaurant_Location	object	contains the neighborhood name in which the restaurant is located in Mumbai
Restaurant_Name	object	contains the name of the restaurant
Restaurant_Type	object	restaurant type
View_Menu	bool	contains the data whether restaurant menu can be viewed
Rating	float	contains total number of rating for the restaurant
Votes	int	User votes given to restaurant
Rating_Category	object	What is the category of rating ranging from excellent, good or fair?
Operational_after_Midnight	bool	Whether restaurant is operational after mid-night or not?
Cuisines	object	Food types separated by commas
Cuisine_count	int	Number of cuisines of a restaurant
Competitors_in_Location	float	Contains number of competitors in that location
Score	float	Overall score of the restaurant

6 Codes and Scripts

The code of final project can be found here (final project data). The code to clean the data can be found from this link - data cleaning.

7 Procedure

7.1 Phase I

Phase I involves cleaning of dataset in CSV file and checking each field for different data types. The dataset was optimised to fit as per our project requirement.

7.2 Phase II

For each neighborhood, Geopy module to convert an address into latitude and longitude values. For each neighborhood's coordinate, we will call Foursquare API to get the trending venues in that location.

8 Software used

For implementation of this Python Programming language with the help of Foursquare API.

Software	Version
Python	3.7.2
Selenium	3.14.0
Pandas	0.24.2
Numpy	1.16.2