Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Restaurant in Aurangabad, India

Geo-Economical analysis of Restaurant Industry
in Aurangabad Neighbourhoods.

By
Chinmay Gaikwad
June 2020

Bibi Ka Maqbara, Aurangabad, India

Introduction

Aurangabad known as "City of Gates" is a historic city in the Indian state of Maharashtra. It is the **largest city** in the Marathwada region. Aurangabad is the **fourth-most populous** urban area in Maharashtra with a population of 2,175,116. The city is Tourism Captial of Maharashtra and a popular tourist hub, with tourist destinations like the Ajanta and Ellora caves lying on its outskirts, both of which have been designated as **UNESCO World Heritage** Sites since 1983. Other tourist attractions include the Aurangabad Caves, Daulatabad Fort, Grishneshwar Temple, Jama Mosque, Himayat Bagh, Panchakki and Salim Ali Lake.

Outline of Aurangabad Metropolitan region

Along with tourism, Aurangabad is one of the **fastest-growing industrial, educational and trading business** cities in Asia. In 2019, the Aurangabad Industrial City (AURIC) became the **first greenfield industrial smart city** of India under the country's flagship Smart Cities Mission.

Restaurants and dining out are an important part of the tourism industry and are a major business in their own right. Restaurants not only cater to food and drinking needs but also alleviate the city’s development and maintain coordination between other businesses.

For restaurants, the central location and the large crowd at tourist and industrial places provide a great distribution channel to market their services. Restaurant chains and franchise developers are also taking advantage of this trend to build more outlets to cater to demand. As a result, there are many small-big restaurants and hotels in the city of Aurangabad and many more are being built. Opening restaurants, food outlets allow property developers to earn consistent rental income. Of course, as with any business decision, opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location 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 sort out the best locations in the neighbourhoods of the city of Aurangabad to open a new restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to come up with a solution to answer the business question:

In the city of Aurangabad, MH, India, **‘ If a franchise owner is looking to start a new restaurant, where would you recommend them to open it? ‘**

Target Audience of this project

This project is particularly useful to franchise owners, entrepreneurs, property developers and investors looking to open or invest in new restaurant or chain of outlets in the tourism capital city of Maharashtra i.e. Aurangabad. This project is would be helpful for local authorities to bring coordination among businesses as the city is currently suffering from unorganised and oversupply of restaurants, small food outlets and food zones.

Data from the Ministry of Tourism (MH) released last year showed that an additional 12 per cent will be added to existing space, and the government body predicted that total occupancy may dip below 86 per cent. The local newspaper The Times of India also reported in March last year that the true occupancy rates in malls may be as low as 40 per cent in some areas.

Data

In order to find a solution to the problem, we will require the following mentioned data:

1. **List of neighbourhoods** in Aurangabad. This defines the scope of this project which is confined to the city of Aurangabad, the tourism capital city and industrial hub of the state of Maharashtra in the western state of India.
2. **Data of latitude and longitude** coordinates of these neighbourhoods. This data is extremely important for plotting the neighbourhoods on the map and also to retrieve the venue data.
3. **Venue data**, particularly data related to the **category 'Restaurant'**. We will employ this data to carry out clustering on the neighbourhoods.

Sources of data

The outcome of any project heavily depends on the quality and quantity of the data that has been incorporated to perform data analytics on it. While quality refers to the authenticity and closeness to accuracy, quantity helps to verify consistency among various sources.

Unfortunately, the data required for this project is not readily available over the internet and is present in fragments. Thus to create a dataset of all the required information requires human intervention such as manual picking of the data through articles, sites and web portals. It also requires manual pre-processing, filtering of the data and finally consolidation of all the fields to bring out a single, satisfying and suitable dataset.

The latitude and longitude data was obtained via batch geocoding of the neighbourhoods from an open-source map site. The acquired data was consolidated using a spreadsheet tool (Google Sheets) and data cleaning, data wrangling to machine learning (K-means clustering) and map visualization (Folium) was performed using a python notebook.

List of sources:

1. Neighbourhood Data:
    a. Aurangabad Municipal Corporation (AMC) governing site : http://rts.aurangabadmahapalika.org/RtsPortal/CitizenHome.html
    b. Government of India, Census: https://censusindia.gov.in
2. Latitude and Longitude data: Batch geocoding: https://www.geocod.io
3. Pricing data (Property rates) of Aurangabad Neighbourhoods:
    a. Makan: https://www.makaan.com
    b. Housing: https://housing.com/in/buy/aurangabad_maharashtra