

IBM Data Science Capstone

The Battle of Neighborhoods

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 Capital** 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 dine-outs 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 national newspaper The Times of India also reported in March last year that the true occupancy rates 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.

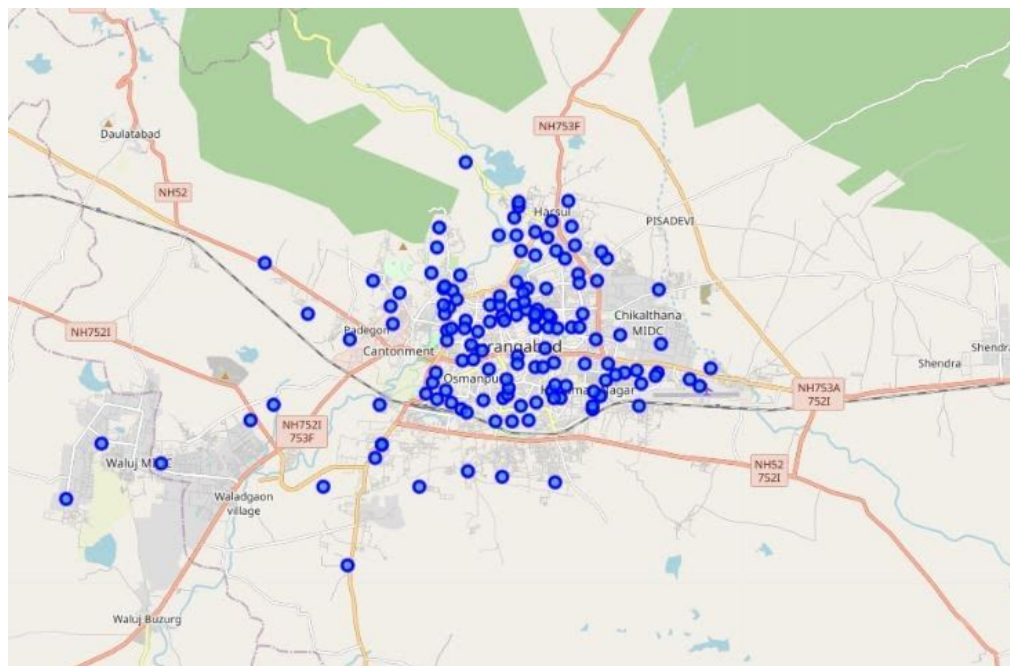
The latitude and longitude data was obtained via batch geocoding of the neighbourhoods from an open-source map site. The acquired data was consolidated using a spreadsheet tool (Google Sheets) and data cleaning, data wrangling to machine learning (K-means clustering) and map visualization (Folium) was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

Methodology

We have neighbourhood data of Aurangabad city processed in a [python notebook](#). This data consists of neighbourhood names, their Ward number, borough they follow in and their census data. The property rates data of each neighbourhood is also augmented in the same dataset. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Aurangabad.



A plot of Neighbourhoods of Aurangabad using the folium map

Next, we will use the Foursquare API to get the top 100 venues that are within a radius of 5 km. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the “Restaurant” data, we will filter the “Restaurant” as the venue category for the neighbourhoods.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 5 clusters based on their frequency of occurrence for “Restaurant”. The results will allow us to identify which neighbourhoods have a higher concentration of Restaurants while which neighbourhoods have a fewer number of Restaurants. Based on the occurrence of Restaurants in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new Restaurants.

Results

The final data frame of the neighbourhoods data of Aurangabad city can be observed as below:

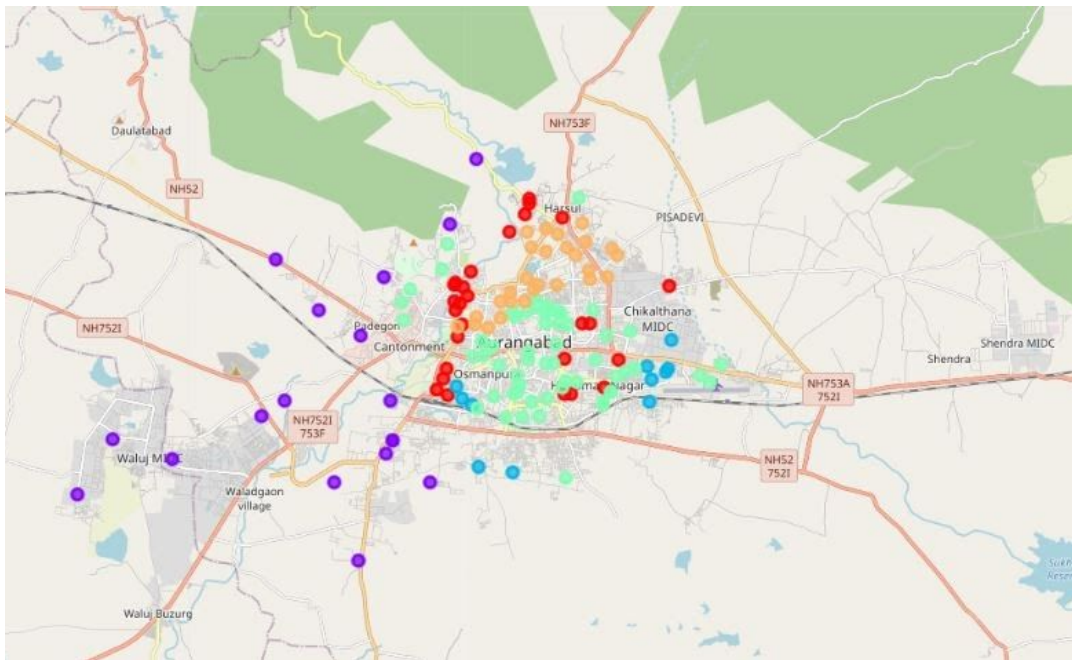
	Neighborhood	Ward	Latitude	Longitude	Borough	Total Population	Population Male	Population Female	Sex Ratio	No.of Houses	Area Hectare	Population Density
0	Harsul	1	19.917838	75.340586	N-A	11304	6548	4756	726	1257	2894.80	3
1	Bhagatsingh Nagar	2	19.919386	75.355821	N-A	16339	8015	8324	1039	2786	232.38	70
2	Mhasoba Nagar	2	19.913759	75.350888	N-A	15245	8658	6587	761	4275	365.50	41
3	Radhaswami Colony	3	19.918989	75.340617	N-A	10226	5374	4852	903	2238	295.12	34
4	Ambar Hill	4	19.930736	75.324129	N-A	21406	11586	9820	848	4002	458.70	46

After performing K-Means clustering on the dataset we get the result for each neighbourhood in the cluster, by taking mean we can have the following overall result

Mean Result	
Cluster	
0	0.0604
1	0.0000
2	0.1140
3	0.0792
4	0.0437

The above results from the k-means clustering show that we can categorize the neighbourhoods into clusters based on the frequency of occurrence for “Restaurant”:

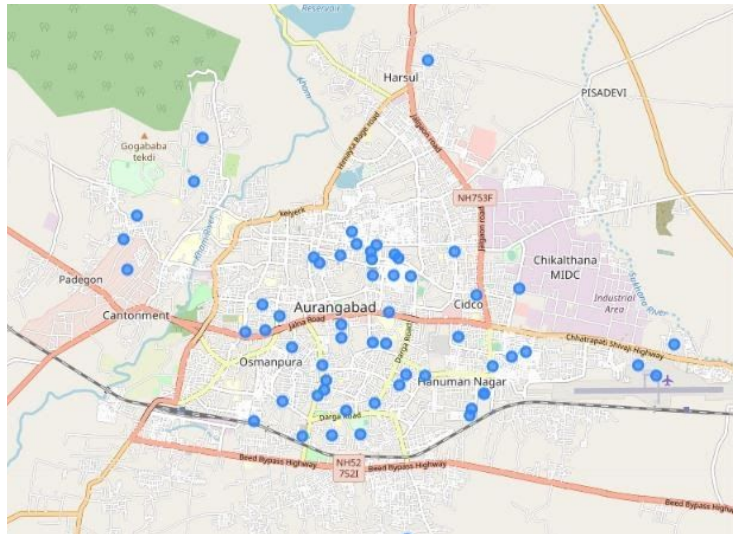
- Neighbourhoods with a moderate number of Restaurants: Cluster 0,3,4
- Neighbourhoods with a low number to no existence of Restaurants: Cluster 1
- Neighbourhoods with a high concentration of Restaurants: Cluster 2



The results of the clustering are visualized in the folium map

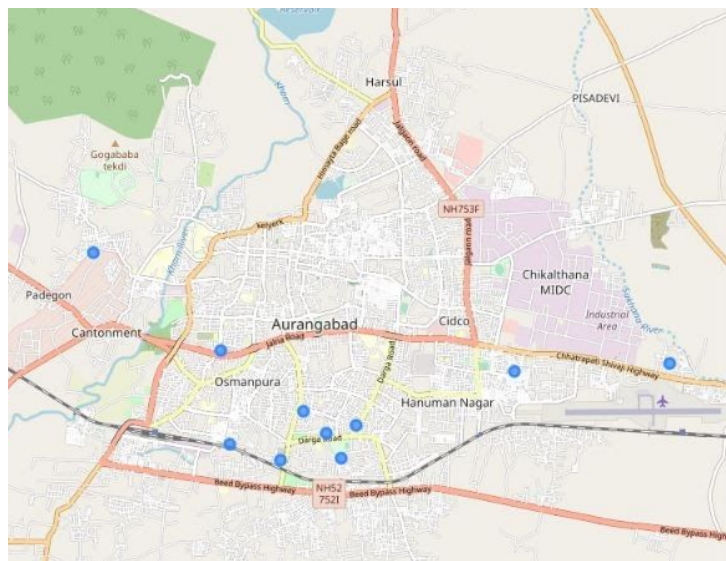
Most of the restaurants are concentrated in the central area of Aurangabad city, with the highest number in cluster 2, moderate number in cluster 0, cluster 3 and cluster 4. On the other hand, cluster 1 has a very low number to totally no restaurants in the neighbourhoods.

This represents a great opportunity and high potential areas to open a new Restaurant as there is moderate competition from existing restaurants. Meanwhile, restaurants in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of restaurants. From another perspective, this also shows that the oversupply of restaurants mostly happened in the central area of the city, with the suburb area still have very few Restaurants



Advised cluster to open a new restaurant in Aurangabad City

Therefore, this project recommends capitalizing on these findings to open a new restaurant in neighbourhoods in cluster 3 with moderate competition. New franchise with unique service propositions can stand out from the competition can also open new restaurants in neighbourhoods in cluster 3 with moderate competition.



Top -10 neighbourhoods in the cluster to open a new restaurant in Aurangabad City

Lastly, newcomers to this industry are advised to avoid neighbourhoods in cluster 2 which already have a high concentration of restaurants and suffering from intense competition.

Discussion

As observations noted from the map in the Results section, most of the restaurants are concentrated in the central area of Aurangabad city, with the highest number in cluster 2 and moderate number in cluster 0, 3 and 4. On the other hand, cluster 1 has a very low number of restaurant in the neighbourhoods. This represents a great opportunity and high potential areas to open a new restaurant as it has very little to no competition from restaurants. Meanwhile, restaurants in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of restaurants.

From another perspective, the results also show that the oversupply of restaurants mostly happened in the central area of the city, with the suburb area still have very few restaurants. Therefore, this project recommends a new franchise, investors to capitalize on these findings to open new restaurants in neighbourhoods in cluster 3 with moderate competition. Franchise owners with unique service propositions to stand out from the competition can also open new restaurants in neighbourhoods in cluster 0 and 4 with even less competition.

Lastly, property developers are advised to avoid neighbourhoods in cluster 2 which already have a high concentration of restaurants and suffering from intense competition.

Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of restaurants, there are other factors such as population and income of residents that could influence the location decision of a new restaurant. However, to the best knowledge of this researcher, such data are not available to the neighbourhood level required by this project.

Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new restaurant. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of a paid account to bypass these limitations and obtain more results.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new restaurant.

To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 1 are the most preferred locations to open a new restaurant. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new restaurant.

References

Suburbs in Aurangabad. Retrieved from Aurangabad Municipal Corporation site:

<http://www.aurangabadmahapalika.org/RtsPortal/CitizenHome.html>

Foursquare Developers Documentation. Foursquare. Retrieved from

<https://developer.foursquare.com/docs>

Aurangabad demographical overview, census:

http://censusindia.gov.in/2011census/dchb/2719_PART_B_DCHB_AURANGABAD.pdf

Real estate overview, property rates derived from

<https://www.makaan.com/price-trends/property-rates-for-buy-in-aurangabad>