The Battle of Neighbourhoods

SETTING UP A RESTAURANT IN HYDERABAD, INDIA

Coursera Applied Data Science Capstone Project
Week: 5 Submission – Final Report

By

Kurapati Venkatesh

21 August 2021

Introduction:

The purpose of this Project is to help clients in exploring better opportunities around the City. It will help people making smart and efficient decisions on selecting great places for their investments in Restaurant business.

Restaurants is a service business and it is a place where people are come and pay for sitting and eating meal, India's extensive cuisine is an amalgamation of influences from various cultures and traditions. Its popularity put the country on the global culinary map with some chefs earning prestigious awards for their restaurants. Furthermore, restaurants and hotels accounted for the highest consumption among food services market across the country in financial year 2020. Increasing disposable incomes, urban lifestyles and an expanding market allowed forgoing home cooked meals occasionally. A symbiotic growth between the food industry and the growing middle class rendered the success of Indian restaurants.
A location should be a prime concern. It is not that people do not travel for food. There are instances where people plan to travel 25 km for a restaurant and end up not attending due to traffic and might get late. There is a need to be very careful while choosing a location keeping an eye on the footfall available in that area - is it a business estate or a residential. So, the location is the 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 select the best locations in the Hyderabad city of Telangana, India to open a new Restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the Hyderabad city of Telangana, India, if a property developer is looking to open a new restaurant, where would you recommend that they open it?

Target Audience of this project:

This project is particularly useful to Start-ups and investors looking to open or invest in new restaurants in the capital city of Telangana, India i.e. Hyderabad. This project is timely as the city is currently suffering from oversupply of restaurants. The Indian Restaurant Market is one of the fastest-growing in the world. It is expected to reach INR 5.99 lakh crore by 2022-23, growing at a

compounded annual growth rate of 9 percent, as per the National Restaurant Association of India (NRAI) report. Despite the hardships and challenges faced by restaurant owners in recent years, opening a restaurant in the near future would bring a host of new opportunities for budding restaurant owners.

The restaurant business is quickly boost upping and growing fast nowadays mainly two factions are involved one is customer that means buyer and other is seller. Customers focused on service, quality, price and so many other factors while seller only wants to earn money. Nowadays restaurant business producing a lot with very little waste and various type of appropriate catering and restaurant services includes celebration of every event or a party. Even if it is a birthday party celebration, farewell parties for universities and colleges ,happy new -year party, social and culture related parties, traditional, wedding catering related to big business or political event - quality and delicious food serving is a must to complete the event.

Data:

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

- List of neighbourhoods in Hyderabad. This defines the scope of this project which is confined to the city of Hyderabad, the metropolitan city of the country of India in South Asia.
- 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 restaurants. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them:

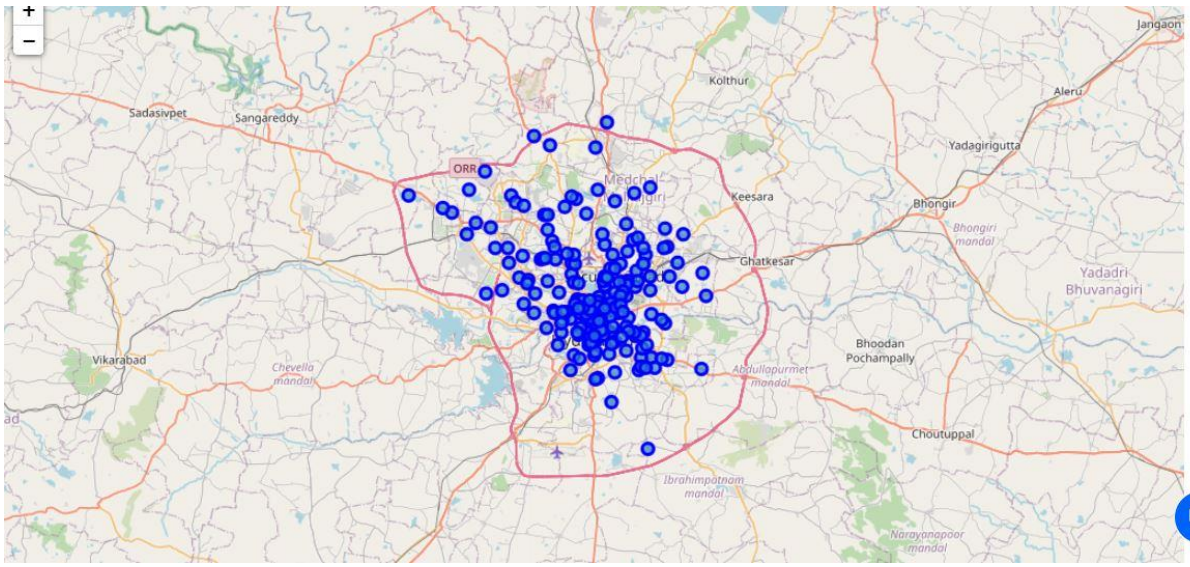
This Wikipedia page – (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India) contains a list of neighbourhoods in Hyderabad, with a total of 200+ neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database 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 Restauarant category in order to help us to solve the business problem put forward. 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). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

HYDERABAD MAP:

This map represents all the 200 neighbourhoods of the city Hyderabad. We then try to make clusters by using the Machine learning technique (K-means clustering) and group all the neighbourhoods into different clusters.



METHODOLOGY

Firstly, we need to get the list of neighbourhoods in the city of Hyderabad. Fortunately, the list is available in the Wikipedia page - (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India)

We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert the address into geographical coordinates in the form of latitude and longitude. 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 Hyderabad. Next, we will use the Foursquare API to get the top 100 venues that are within a radius of 2000 meters.

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.

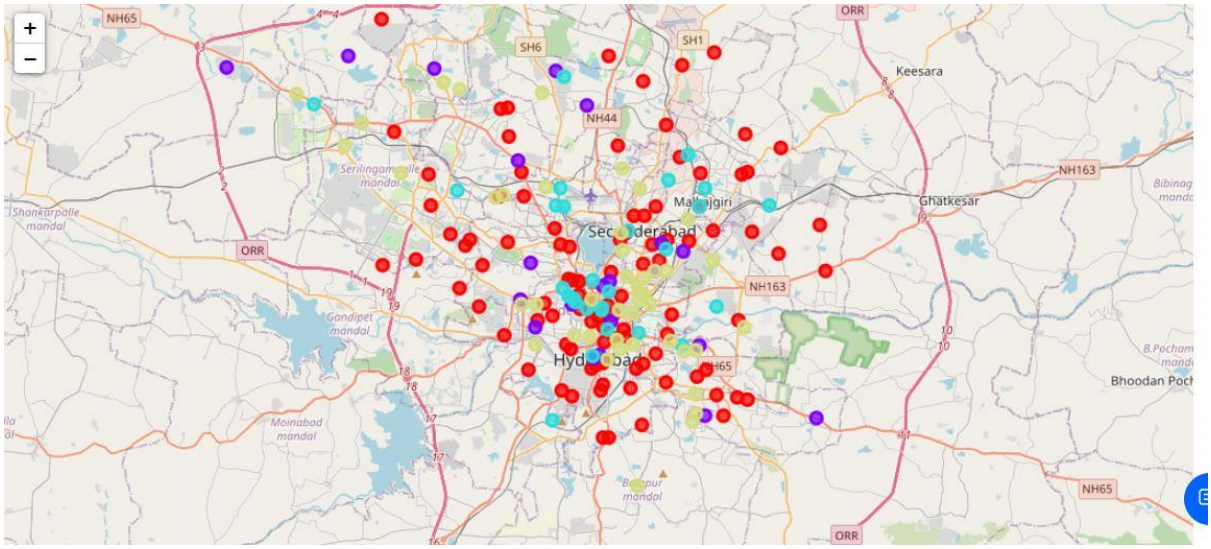
Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the "Restaurant" data, we will filter the "Restaurant" as venue allocates every data point to the nearest cluster, while keeping the centroids as restaurant 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 3 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 5 which neighbourhoods are most suitable to open new Restaurants. Therefore, this project recommends property developers to capitalize on these findings to open new Restaurants in neighbourhoods in cluster 2 with little to no competition

RESULTS:

The results from the k-means clustering show that we can categorize the neighbourhoods into 4 clusters based on the frequency of occurrence for "Restaurant":

Major number of restaurants are concentrated in the central area of Hyderabad city, with the highest number in cluster 1 & 2 and moderate number in cluster 3. This represents a great opportunity and high potential areas to open new restaurants as there is very little to no competition from existing restaurants. Meanwhile, restaurants in cluster 2 are likely suffering from intense

competition due to oversupply and high concentration of restaurants. Therefore, this project recommends property developers to capitalize on these findings to open new restaurants in neighbourhoods in cluster 0 with relatively less competition. Property developers with unique selling propositions to stand out from the competition can also open new restaurants in neighbourhoods in cluster 3 with moderate competition. Lastly, property developers are advised to prefer central area of Hyderabad city with the recommended cluster because lot of demand can be seen and with best business model can yield a best restaurant.



DISCUSSIONS:

As observations noted from the map in the Results section, most of the Restaurants are concentrated in the central area of Hyderabad city, with the highest number in cluster 0 and moderate number in cluster 1. This represents a great opportunity and high potential areas to open new Restaurants as there is very little to no competition from existing restaurants. Meanwhile, Restaurants in cluster 1 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 property developers to capitalize on these findings to open new Restaurants in neighbourhoods in cluster 0 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new Restaurants in neighbourhoods in cluster 1 with moderate 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 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 4 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:

- Category: Neighbourhoods in Hyderabad, India. Wikipedia Retrieved from https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India
- Foursquare Developers Documentation. Foursquare. Retrieved from https://developer.foursquare.com/docs/