

CAPSTONE PROJECT

IBM APPLIED DATA SCIENCE

Setting up a restaurant in Hyderabad, India

By Rithik Kotha

INTRODUCTION

Restaurants have always played an essential role in the business, social, intellectual and artistic life of a thriving society. The major events of life, personal and professional, are celebrated in restaurants. Acquaintances become friends around a table in the safe and controlled environment of a restaurant. Restaurants are more important than ever.

Beyond the basic purpose of restaurants to provide food and drink, restaurants have fulfilled a human need for connection and shaped social relations. Restaurants occupy an increasingly important place in shaping our overall economy and the nature and makeup of our cities.

Restaurants have always been the leading edge of reshaping neighbourhoods. In certain places, people dine more outside the home. People do crave the social experience of dining. Economy and Employment depend on restaurants in a certain way.

BUSINESS PROBLEM

The objective of this capstone project is to analyse and select the best locations in the city of Hyderabad, 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 city of Hyderabad, India if a property developer is looking to open a restaurant, where would you recommend that they open it?**

Although the restaurant industry is very competitive and as the number of people have less time, resources, and ability to cook for them it is important that the restaurant is well positioned for the current interest and people get healthier foods at moderate to low prices.

TARGET AUDIENCE OF THE PROJECT

This project is particularly useful to property developers and investors looking to open or invest in restaurants in the capital city of Telangana i.e. Hyderabad. This should be carefully planned and executed so that it will be beneficial to everyone. People can come and enjoy the variety of cuisines and delicious food.

DATA

Build a data frame of neighbourhoods in Hyderabad, India by web scraping the data from Wikipedia page Get the geographical coordinates of the neighbourhoods by Python Geocoder package. Obtain the venue data for the neighbourhoods from Foursquare API Explore and cluster the neighbourhoods Select the best cluster to open a new restaurant.

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 capital city of Telangana which is in South India

- 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 shopping malls. We will use this data to perform clustering on the neighbourhoods.

SOURCES AND METHODS TO EXTRACT DATA

We will use **Foursquare API** to get the venue data for those neighbourhoods. Foursquare has one of the largest databases of 105+ million places and is used by many developers across the world. Foursquare API will provide many categories of the venue data, we are particularly interested in the restaurant category in order to help us to solve the business problem put forward. This is a project that will make use of many data sciences kills, from **web scraping** (Wikipedia), working with **API (Foursquare)**, **data cleaning**, **data wrangling**, to machine learning (**K-means clustering**) and map visualization (**Folium**).

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 **beautiful soup 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.

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 beautiful soup 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 Data Frame 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

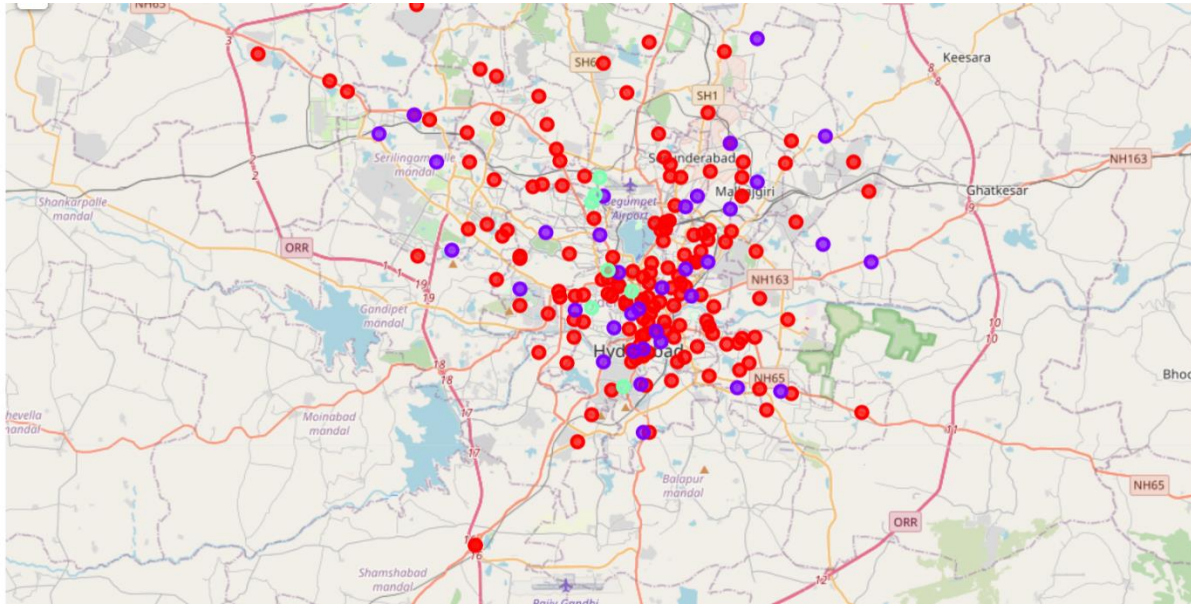
examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the data, we will filter the restaurant as venue 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 3 clusters based on their frequency of occurrence . 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 these in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new restaurants. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighbourhoods in cluster 2 with little to no competition.

RESULTS :

After detailed analysis of the data and critical thinking of the data the neighbourhoods have been divided into clusters depending upon the frequency in which people go to these restaurants. To set up a new restaurant we will be looking at a place where people go more often or in other words where more population density is there.

They have been divided into clusters and the restaurant we will be looking to set up will be based on these clusters .

Clusters with high , moderate ,no restaurants are plotted and depicted in the below picture.



DISCUSSION

As observations noted from the map in the Results section, most of the suitable places are concentrated in the central area of Hyderabad city, with the highest number in cluster 1 and moderate number in cluster 0. On the other hand, cluster 2 has a very low number in the neighbourhoods. This represents a great opportunity and high potential areas to open new restaurants as there is very little to no competition from existing competitors. Meanwhile, 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 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 2 with little to no competition. Property developers with unique selling propositions to stand out

from the competition can also open new places in neighbourhoods in cluster 0 with moderate competition. Lastly, they are advised to avoid neighbourhoods in cluster 1 which are facing intense competition from fellow competitors.

CONCLUSION :

We have narrowed down the possibilities so as to where we can set up a new restaurants so that it can be successful enough in its domain.

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 3 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 shopping mall. 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 setting up their restaurants.

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/>

HYDERABAD MAP WITH ALL PLACES WHERE WE CAN DINE

