# The Battle of Neighbourhoods | Business Proposal Opening a restaraunt in Hyderabad, India:

## Introduction :

Restaraunts are a great place to spend a quality time with family and friends while having delicious food together.There are different restaraunts having various cuisines.But the cuisine selection and what to offer to the customers depends on the locality of where the investor want to place the restaurant.Business investors contribute in the growth of the economy and also the restaurant business is a great way for their growth."Who dosen't invest in good food,right?".Here we suggest business investors to establish a restaurant at the desired location by investigating the in-sights of restaraunts in the city of Hyderabad,India

The objective of this capstone project is to analyze and select the best locations in the city of Hyderabad, India to open a new restaraunt. 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 business investor is to open a restaurant, where would you recommend that they open it?

## 2. Data Section:

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, 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 Restaraunts. We will use this data to perform clustering on the neighbourhoods.

### Foursquare API

We will need data about different venues in different neighborhoods of that specific city. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data

provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 2000 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Venue Latitude
6. Venue Longitude
7. Venue Category(Restaraunt)


# 3. Methodology Section:

Firstly, we need to get the list of neighbourhoods in the city of Hyderabad. Fortunately, the list is available in the page (https://commons.wikimedia.org/wiki/Category:Suburbs_of_Hyderabad,_India). I 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. I 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 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 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 "Restarauntsl" data, we will filter the "Restaraunts" as 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 "3" clusters based on their frequency of occurrence for "Restarauntsl". The results will allow us to identify which neighbourhoods have higher concentration of restaraunts while which neighbourhoods have fewer number of restaraunts. Based on the occurrence of restaraunts in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new Restaraunts.

# 4. Results Section:

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for "Restaurants":

• Cluster 0: Neighbourhoods with moderate number of Restaurants

• Cluster 1: Neighbourhoods with low number or no existence of Restaraunts

• Cluster 2:Neighbourhoods with high concentration of Restaurants

# 5. Discussion Section:

The aim of this capstone project is to analyze 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 is particularly useful to business investors looking to open or invest in new Restaurant in the Hyderabad city, India.

# 6. Conclusion Section:

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 business investors regarding the best locations to open a new Restaurantl. To answer the business

question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 0 are the most preferred locations to open a new Restaurant. The findings of this project will help the relevant investors to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a Restaurant.

# 7. References:

Details of Suburbs in hyderabad retrieved from https://commons.wikimedia.org/wiki/Category:Suburbs_of_Hyderabad,_India