COURSERA CAPSTONE
IBM APPLIED DATA SCIENCE
CAPSTONE PROJECT

By Jyothi Mayuri

❑ INTRODUCTION /BUSINESS PROBLEM

1. Hyderabad is the capital of Indian state of Telangana.
2. The City is the most populous city in Telangana. It is multicultural. It provides lot of business opportunities and business friendly environment. Hyderabad of today is buzzing with business activities. It is a global hub of business and commerce.
3. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance , fashion.
4. This also means that the market is highly competitive. Although you will find many restaurants in the city, still there is ample scope for new restaurants to make good money. Theme-based restaurants can be a profitable business venture in the city of Hyderabad.

Problem Description:

- A restaurant is a business which prepares and serves food and drink to customers in return for money, either paid before the meal, after the meal, or with an open account. The Hyderabad City is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.
- It is famous for not just Biryani, but also for fine desserts which satisfy your carvings some of these are Khubani Ka Meetha,Shashi Tukda, halwa etc
- So it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the Location such as :
- City Population
- City Demographics
- Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
- Are there any venues like Gyms, Entertainment zones, Parks etc nearby where floating population is high etc
- Who are the competitors in that location?
- Cuisine served / Menu of the competitors
- The list can go on...
- Even though well funded XYZ Company Ltd. need to choose the correct location to start its first venture.If this is successful they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

Target Audience:

- To recommend the correct location, XYZ Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighbourhood of New York city will be best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made.
- This would interest anyone who wants to start a new restaurant in New York city.

- **Success criteria:**

- The success criteria of the project will be a good recommendation of borough/Neighborhoods choice to XYZ Company Ltd based

Data

- One city will be analysed in this project : ***Hyderabad City***.
- We will be using the below datasets for analysing Hyderabad city

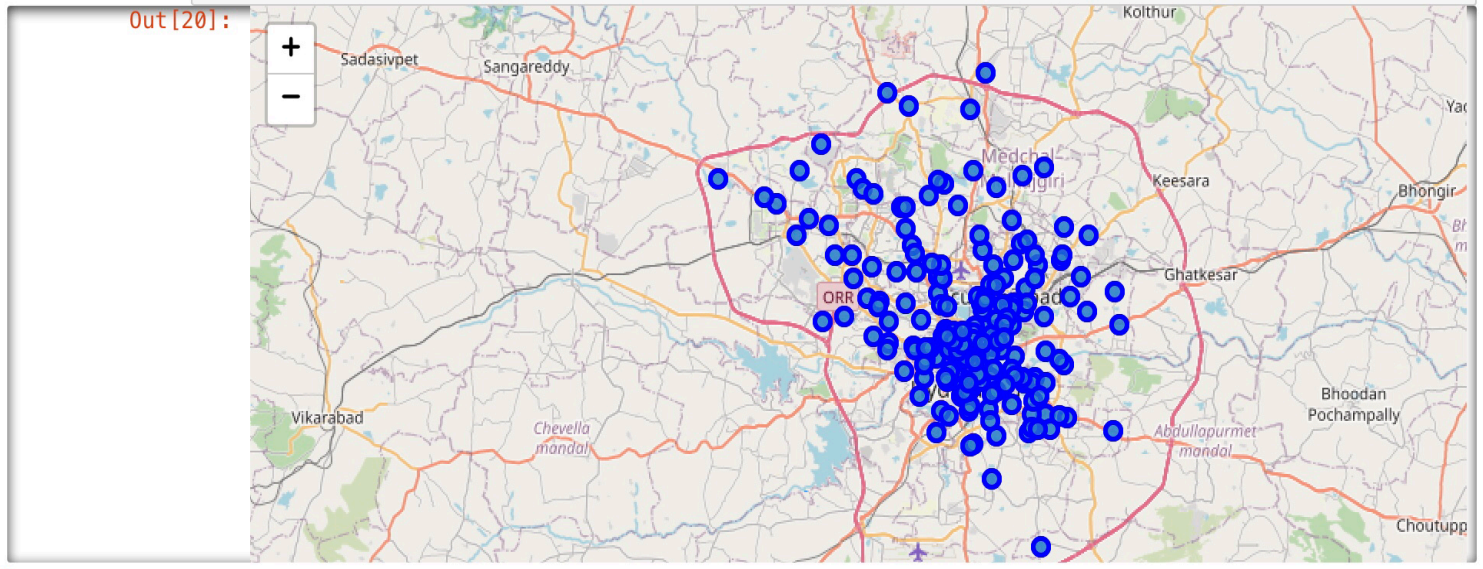
- ***Data 1 :*** In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the neighbourhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighbourhood.
- This dataset exists for free on the web. Link to the dataset is : https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Hyderabad

- ***Data 2:*** Hyderabad city geographical coordinates data will be utilised as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood. We will use the Foursquare API to explore neighbourhoods in Hyderabad City.

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.

Out[20]:



- 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.

Out[27]:

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	A. C. Guards	17.395015	78.459812	Cafe Niloufer & Bakers	17.399715	78.462881	Café
1	A. C. Guards	17.395015	78.459812	Subhan Bakery	17.392412	78.464712	Bakery
2	A. C. Guards	17.395015	78.459812	Taiba Bakers & Confectioners	17.402530	78.456823	Bakery
3	A. C. Guards	17.395015	78.459812	Chicha's	17.403255	78.460152	Hyderabadi Restaurant
4	A. C. Guards	17.395015	78.459812	Nizam club	17.403221	78.468729	Lounge
5	A. C. Guards	17.395015	78.459812	Rayalaseema Ruchulu	17.403084	78.463012	South Indian Restaurant
6	A. C. Guards	17.395015	78.459812	Prince Hotel	17.394736	78.442410	Indian Restaurant
7	A. C. Guards	17.395015	78.459812	DineHill	17.405256	78.451674	Indian Restaurant
8	A. C. Guards	17.395015	78.459812	Spice 6	17.409007	78.450559	Bistro
9	A. C. Guards	17.395015	78.459812	Laxman Ki Bandi	17.378895	78.463973	South Indian Restaurant

getting the total venues for each neighborhood

```
In [28]: venues_df.groupby(["Neighborhood"]).count()
```

Out[28]:

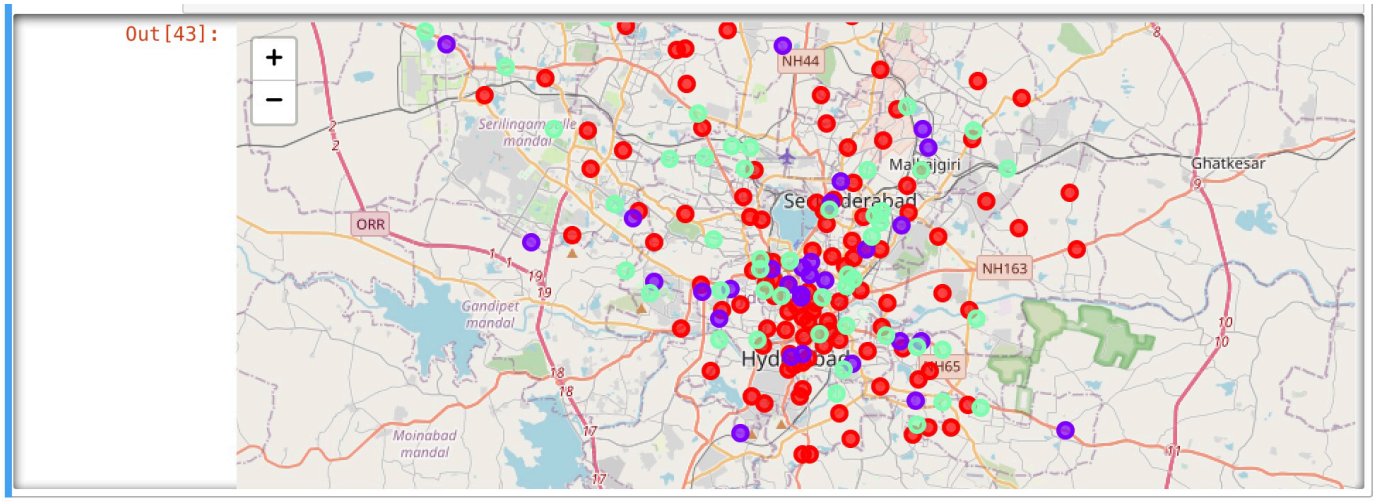
Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
A. C. Guards	69	69	69	69	69	69
A. S. Rao Nagar	28	28	28	28	28	28
Abhyudaya Nagar	9	9	9	9	9	9
Abids	77	77	77	77	77	77
Adikmet	23	23	23	23	23	23
Afzal Gunj	42	42	42	42	42	42
Aghapura	58	58	58	58	58	58
Aliabad, Hyderabad	9	9	9	9	9	9
Alijah Kotla	13	13	13	13	13	13
Allwyn Colony	16	16	16	16	16	16

Analyzing the neighborhoods

- 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 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 “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 Restaurant .
- Therefore, this project recommends property developers to capitalise on these findings to open new Restaurant in neighbourhoods in cluster 0 with little to no competition.

Visualization of clusters

Out[43]:



In [44]:
```
df_merged.loc[df_merged['Cluster Labels'] == 0]
```

Out[44]:

	Neighborhood	Restaurant	Cluster Labels	Latitude	Longitude
0	A. C. Guards	0	0	17.395015	78.459812
102	Kukatpally	0	0	17.385940	78.483380
103	Kushaiguda	0	0	17.427540	78.420630
104	L. B. Nagar	0	0	17.487350	78.420870
105	Laad Bazaar	0	0	17.481130	78.583700
106	Lab quarters	0	0	17.512650	78.441290
108	Lal Darwaza	0	0	17.335170	78.495370
99	Kothapet, Hyderabad	0	0	17.533180	78.481020
109	Lallaguda	0	0	17.405050	78.462890
111	Laxminagar Colony, Mehdipatnam	0	0	17.440300	78.527910

Examine the clusters

- As observations noted from the map , most of the Restaurants are concentrated in the central area of Hyderabad city, with the highest number in cluster 1 and moderate number in cluster 2. On the other hand, cluster 0 has a very low number.
- This represents a great opportunity and high potential areas to open new Restaurant as there is very little to no competition . Meanwhile, Restaurant in cluster 1 are likely suffering from intense competition .
- Therefore, this project recommends property developers to capitalise on these findings to open new Restaurant 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 shopping malls in neighbourhoods in cluster 2 with moderate competition. Lastly, property developers are advised to avoid neighbourhoods in cluster 0 which already have a high concentration of Restaurants and suffering from intense competition.

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 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 Restaurant .
- To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 2 are the most preferred locations to open a new Restaurant.
- The findings of this project will help the relevant stakeholders to capitalise on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new Restaurant .