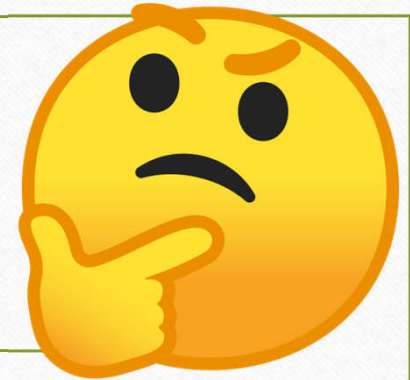# Battle of Neighborhoods

Building a Pizza Place in Hyderabad

Business Problem

- To analyse and select the best locations in the city of Hyderabad, India to build a new Pizza outlet. Using data science methodology and machine learning techniques like clustering, this project aims to solve this and get an answer to the business question:

  "In the city of Hyderabad, India, if a business person or property developer is looking to build a new pizza outlet, where would you recommend that they do so?"

Data Analysis
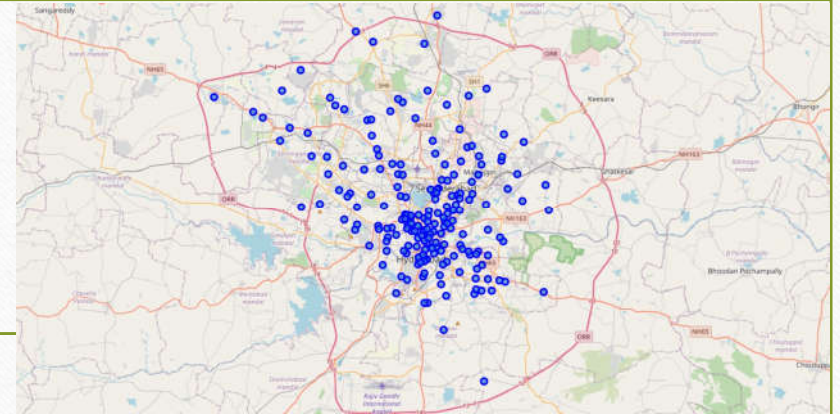
- This Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India) contains a list of neighbourhoods in Hyderabad, with a total of 200 neighbourhoods. We will use web scraping techniques for extracting the data from the Wikipedia page, with the help of Python requests and beautiful soup packages.

- Then we will use Python Geocoder package to get the geographical coordinates of the neighbourhoods which will give us the latitude and longitude coordinates of the neighbourhoods.

- After that, we will be using Foursquare API to get the venue data for those neighbourhoods.

Methodology

- We will do web scraping using Python requests and beautiful soup package to extract the list of neighbourhood's data. This is only a list of names

- We will use the Geocoder package to convert address into geographical coordinates in the form of latitude and longitude

- Now, we will be populating the data into a pandas Data Frame and then visualize the neighbourhoods in a map using Folium package. Next, we will use Foursquare API to get the venues.

- 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 extract the venue name, venue category, venue latitude and longitude

|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	A. C. Guards	17.395015	78.459812	Hotel Marjaan	17.395373	78.455115	Café
1	A. C. Guards	17.395015	78.459812	Milan Juice Center	17.392266	78.458102	Snack Place
2	A. S. Rao Nagar	17.411200	78.508240	Srinivasa Textiles	17.407338	78.509908	Clothing Store
3	A. S. Rao Nagar	17.411200	78.508240	Reign Gym & Spa	17.407340	78.509920	Gym
4	A. S. Rao Nagar	17.411200	78.508240	turnhalle gym	17.407347	78.509941	Gym

Continued...

- Since we are analysing the "Pizza Place" data, we will filter the same as venue category for the neighbourhoods.

- Lastly, will perform clustering on the data using k-means clustering.

- We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for "Pizza Place". The results will allow us to identify which neighbourhoods have higher concentration of housing development and IT offices

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
A. C. Guards	2	2	2	2	2	2
A. S. Rao Nagar	4	4	4	4	4	4
Abhyudaya Nagar	6	6	6	6	6	6
Abids	19	19	19	19	19	19
Adikmet	3	3	3	3	3	3
...	...	...	...	...	...	...
Saroornagar	6	6	6	6	6	6
Secunderabad	11	11	11	11	11	11
Shah-Ali-Banda	6	6	6	6	6	6
Shahran Market	8	8	8	8	8	8
Shanker Mutt	4	4	4	4	4	4

Results

➤ Cluster 0: Neighbourhoods with high number of Pizza places (red colour)

➤ Cluster 1: Neighbourhoods with moderate number of Pizza places (purple colour)

➤ Cluster 2: Neighbourhoods with low or no concentration of Pizza places (green colour)
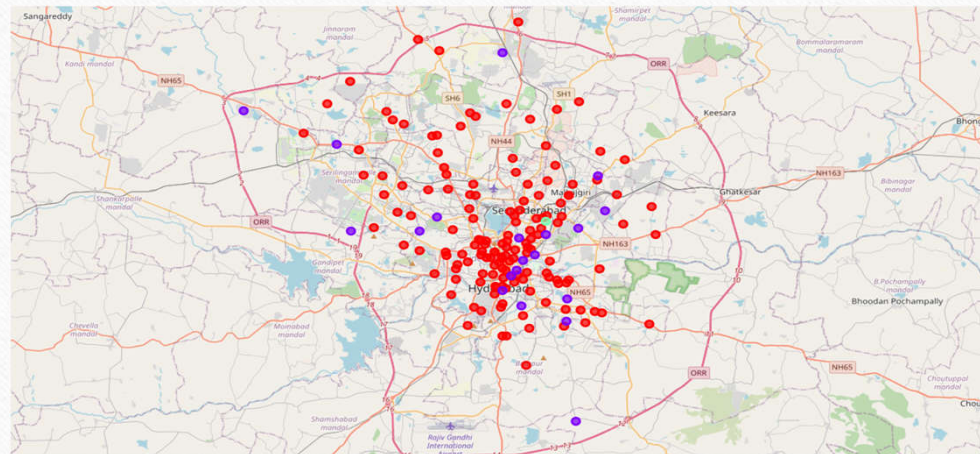
Discussion

- Most of the Pizza places are concentrated in the central area of Hyderabad city, with the highest in cluster 0 and moderate in cluster 1. Cluster 2 has no Pizza places in the neighbourhoods. This represents a great opportunity and high potential areas to build a new pizza outlet

- Pizza places in cluster 0 are facing tough competition, causing some inconvenience to the business. From another perspective, the results also show that the oversupply of pizza places mostly happened in the central area of the city.

- Therefore, this project recommends property developers to capitalize on these findings to build pizza outlet in neighbourhood of cluster 2. Lastly, property developers are advised to avoid neighbourhoods in cluster 0 which have a very high concentration of pizza places.

	Neighborhood	Pizza Place	Cluster Labels	Latitude	Longitude
0	A. C. Guards	0	0	17.395015	78.459812
104	Mahatma Gandhi Road (Secunderabad)	0	0	17.487350	78.420870
105	Malakpet	0	0	17.481130	78.583700
106	Malkajgiri	0	0	17.512650	78.441290
107	Malkajgiri mandal	0	0	17.361450	78.472090
...	...	...	...	...	...
114	Mehdipatnam	1	1	17.350670	78.534040
40	Chilkalguda	1	1	17.377330	78.489740
66	Izzat Nagar	1	1	17.391920	78.505640
145	Punjagutta	1	1	17.433510	78.566730
150	Raj Bhavan Road	2	2	17.423410	78.515340

Conclusion

- In this project, we started the process by identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters depending on similarities, and then providing recommendations to the relevant stakeholders i.e., property developers and investors regarding the best locations to build a new pizza outlet.

- 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 build new pizza outlet. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding areas in their decisions to build a new pizza outlet.