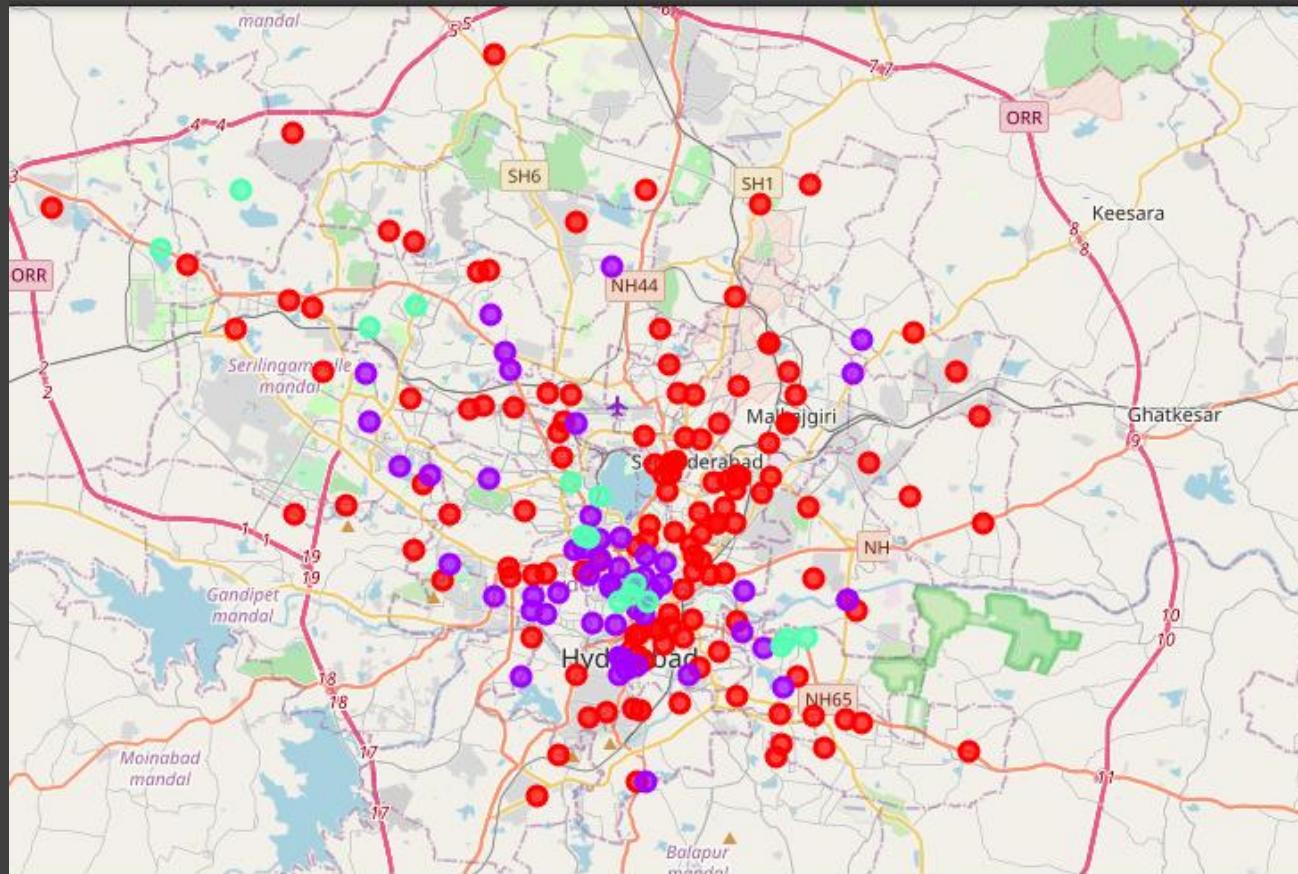


The Battle of Neighborhoods – Hyderabad, India



Coursera Capstone Project

Problem Statement: To identify lucrative neighborhoods in Hyderabad set up a restaurant

Techniques Used:

- Web scraping
- Data Preprocessing Cleansing
- K-means Clustering

Created By: Neha Madnekar

Introduction: Business Problem

Objective: The objective of this project is to find all the major neighborhoods in Hyderabad and accordingly cluster them to identify lucrative localities to set up a restaurant. This project will be helpful to individual businessmen, groups or conglomerates who want to identify optimal locations in Hyderabad to set up restaurants, hotels, food trucks etc.

Techniques Used:

1. Web Scraping: A list of all neighborhoods in Hyderabad is fetched from Wikipedia using several python requests along with their latitude and longitude details.. Foursquare API is then used to access venues around these neighborhoods
2. Data Preprocessing and Cleansing: A data frame of all these neighborhoods mapped against their geo location details as well as venues is created indicating the presence/ absence of the venue category in that area
3. Clustering: K-means clustering technique is then used to cluster all the neighborhoods based on the presence of all the venues except for the venue category – restaurants. Folium library is imported to help visualize the map of Hyderabad and superimpose the clustered data points onto it.

Data Requirements

Neighborhood Data:

A list of all the neighborhoods of Hyderabad can be found on Wikipedia using the link below

https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India

Geolocation Data:

Python Package – Geocoder is used to fetch the latitude and longitude details and map them against the respective neighborhoods

Venue Data:

All the venue details associated with the particular neighborhoods are then fetched using the foursquare API

Data Visualization:

Python's Folium Package is used to superimposed the neighborhood location datasets onto the map of hyderabad

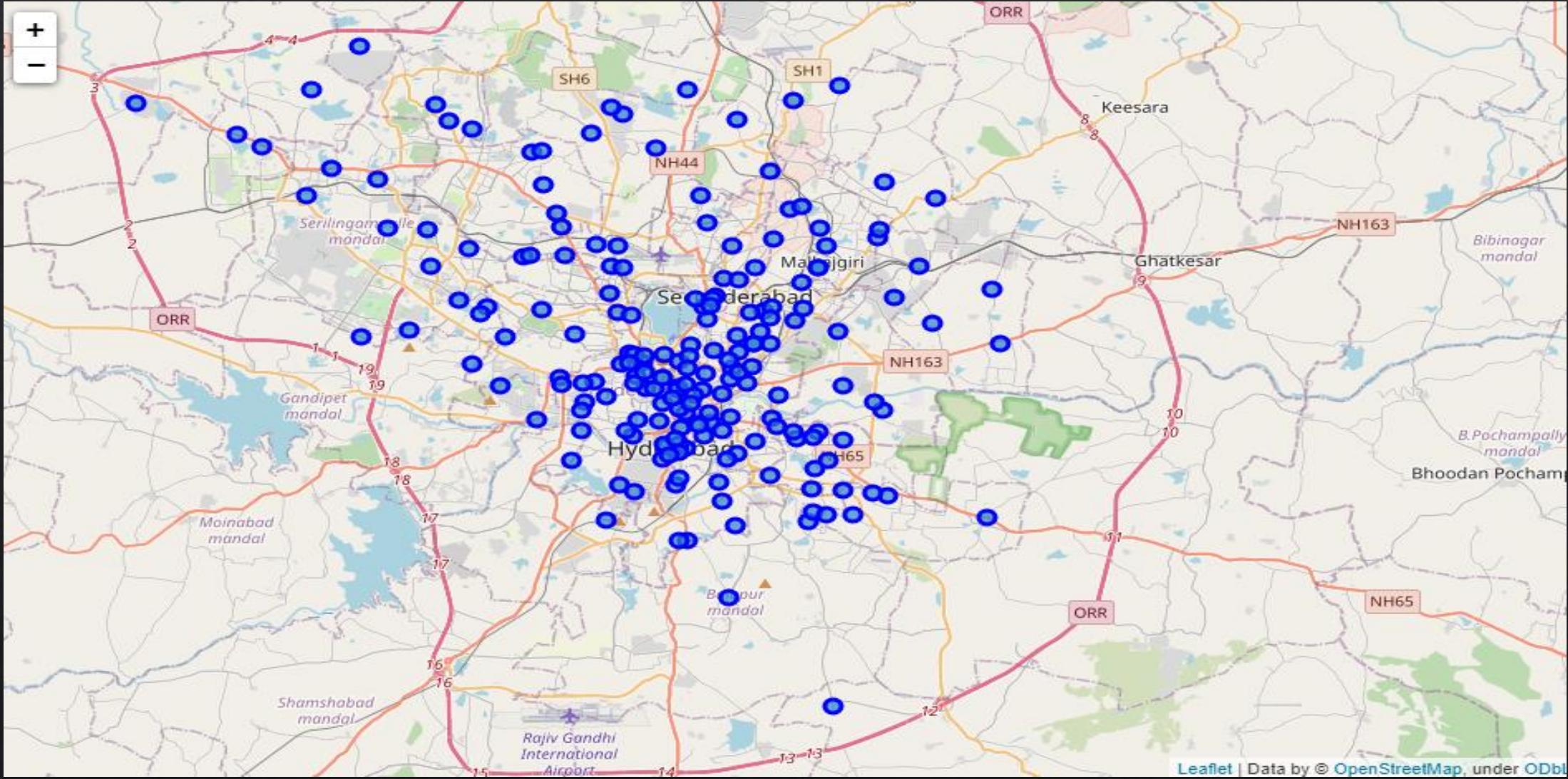
Methodology

The following is the step-by-step methodology that have been used to get the results:

1. Web Scraping: Perform this step using the information provided on the Wikipedia page and use several python requests and beautiful soup packages to retrieve this data in a Data Frame Format
2. Geolocations: Use the Geocoder package to map the location details to latitude and longitude values to be able to fetch the venue details using Foursquare API
3. Create Neighborhood Data Frame: Populate all details (location, latitude and longitude) in a Pandas Data Frame
4. Visualize Maps: Use the folium package to visualize the map of all the data sets in the data frame
5. Venues: Use the foursquare API to explore all the neighborhoods and get the venue details nearby the locations
6. Create Venues Data Frame: Merge the venues details against each of the neighborhoods create a larger D. In this DF, each venue is analyzed against the presence of venues and the most popular locations with more venues are chosen to be ideal
7. Clustering: Cluster the neighborhoods using K-means clustering technique and identify localities that already have higher number of restaurants to help answer the question as to which neighborhoods are most suitable to open new restaurants
8. Visualize: Use Python's Package folium to visualize the clusters by color coding 3 clusters created in the earlier step

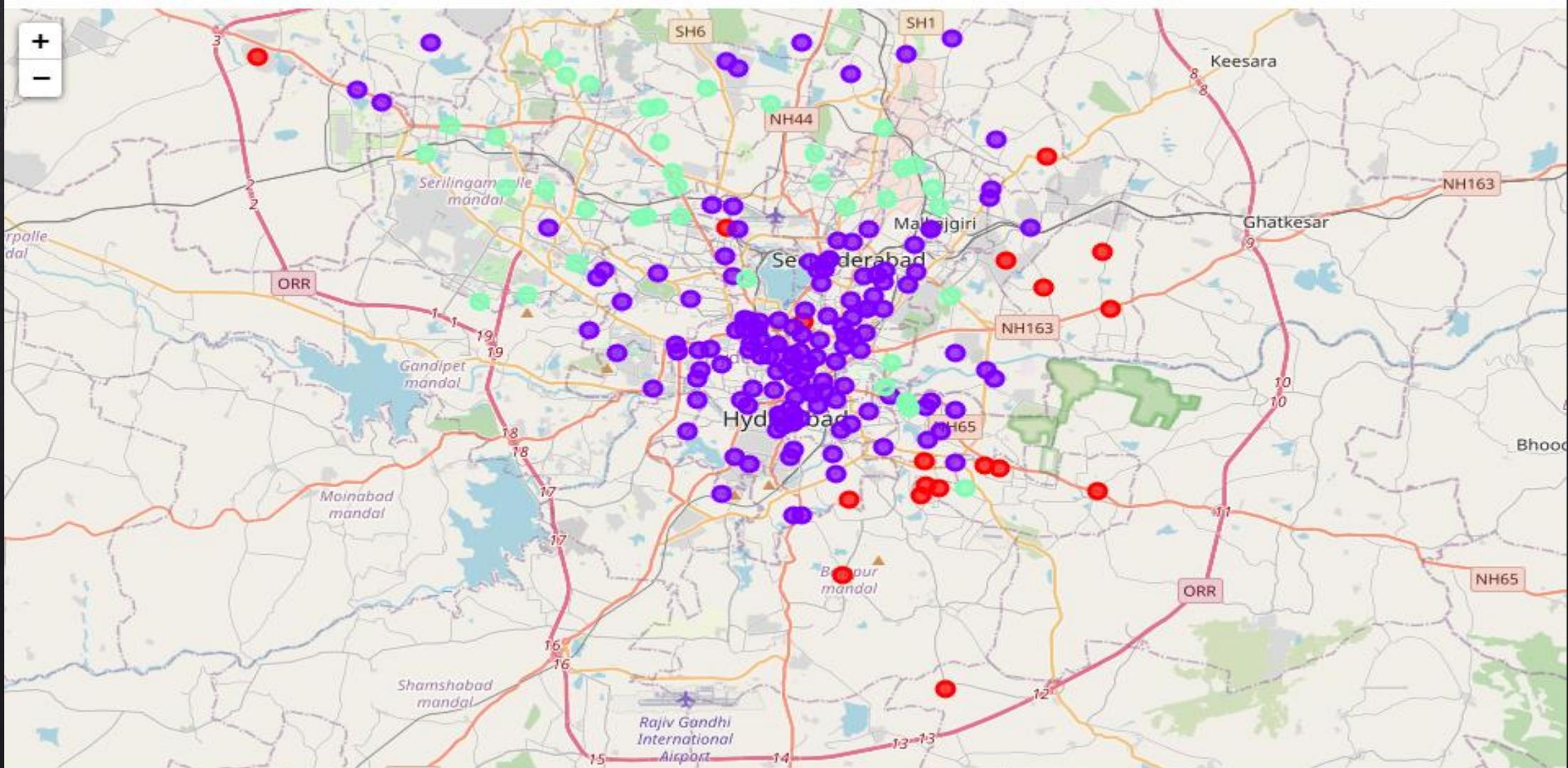
Analysis (Discussion)

All the neighborhoods in Hyderabad



Analysis (Discussion)

Hyderabad neighborhood divided into 3 clusters based on the presence of restaurants



Results

Attribute	Result
Number of neighborhood locations fetched	200
Total number of unique venue categories	122
Examples of venue categories	Café, Bakery, Lounge, Indian Restaurant, Ice Cream Shop, South Indian Restaurant, Hyderabadi Restaurant, Juice Bar, Scenic Lookout, Bistro etc.
Total number of localities with restaurants around:	187
Number of clusters:	2
Clustering based on	Existing presence of restaurants

Cluster	Restaurants	Number of Neighborhoods
0	Max 1	23
1	Max 3	139
2	Max 6	38

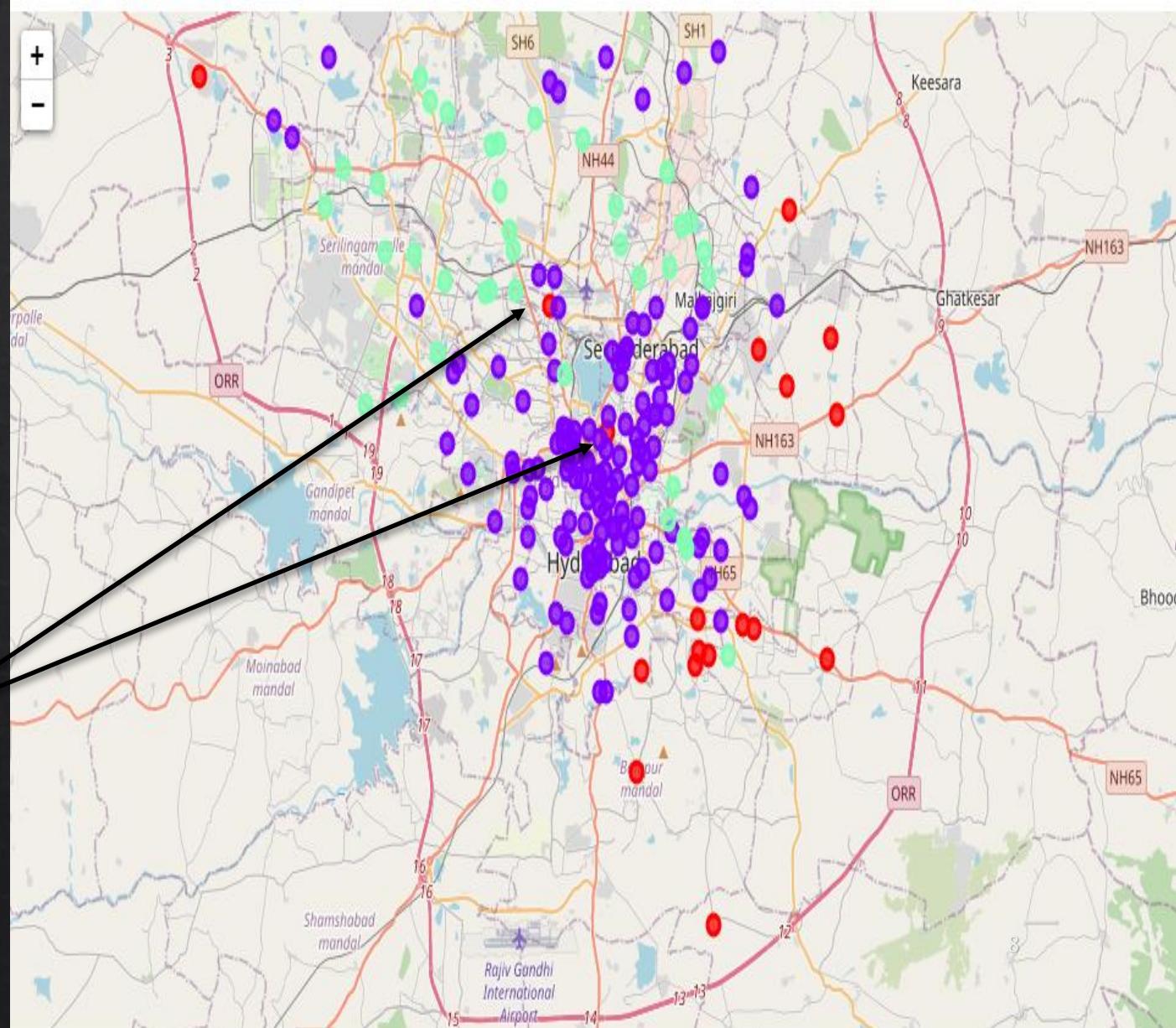
Conclusion(1/3)

As you may observe, all the data points belonging to the first cluster are on the borders of the city where they might not be any restaurants organically due to lower amount of population in that area.

Therefore, the first most lucrative areas to set up a restaurants would be the data points belonging to the cluster 0 that are well within the city limits to ensure population density and footfall.

In our results, we found two such data points:

1. Hayathnagar
2. Balapur

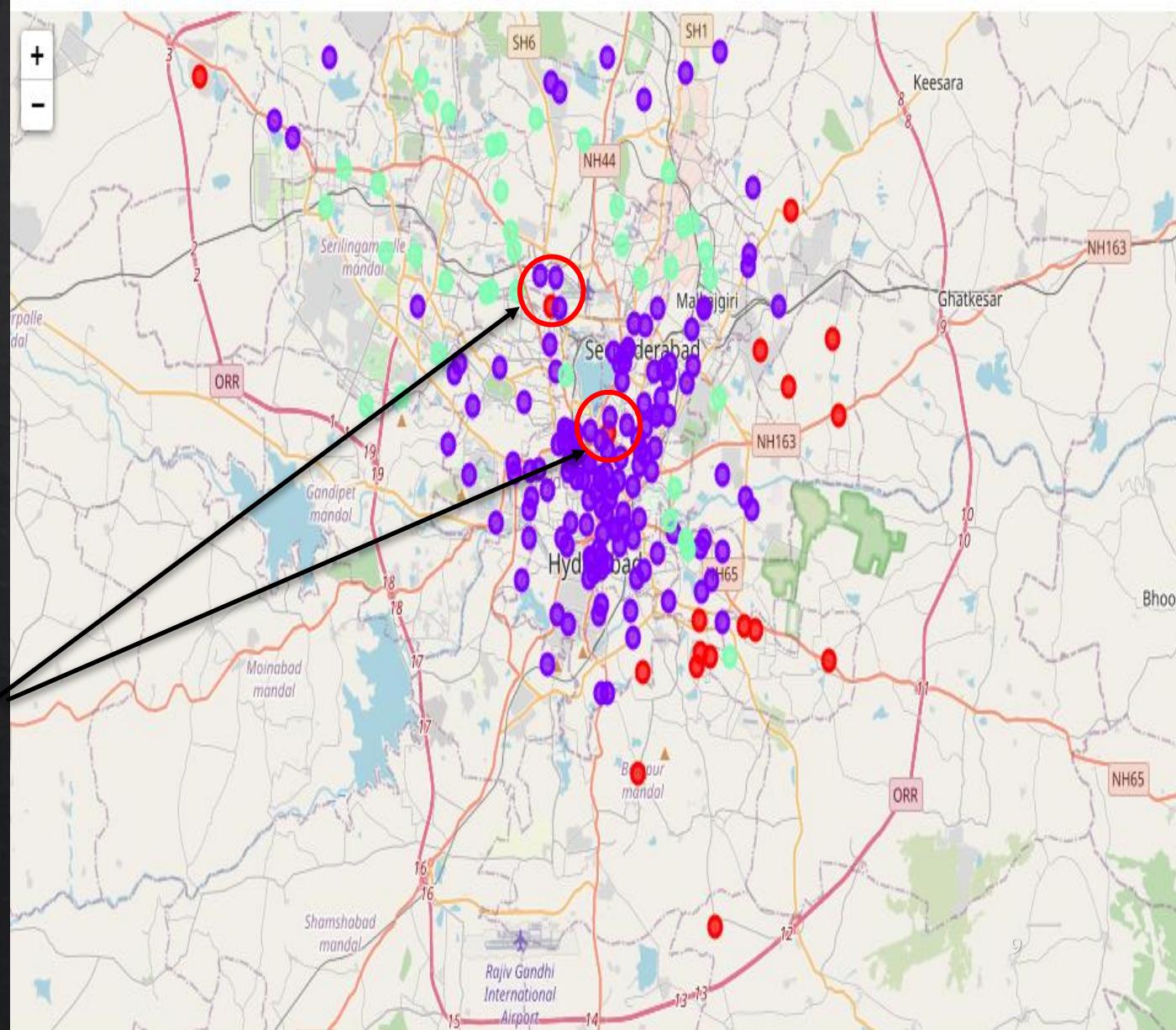


Conclusion(2/3)

The second most lucrative list of possible areas for setting up a restaurant would be areas within the city limits nearby the data sets from the cluster 0 belonging to cluster 1. The reason behind this would be the population living in the neighborhoods of the first cluster might consider areas that might be far more popular near to them

Even though, the neighborhoods in cluster 2 have certain number of restaurants present, its always safer to set up shop in an area that already attracts a bit of crowd. Some of the areas in this bucket would be:

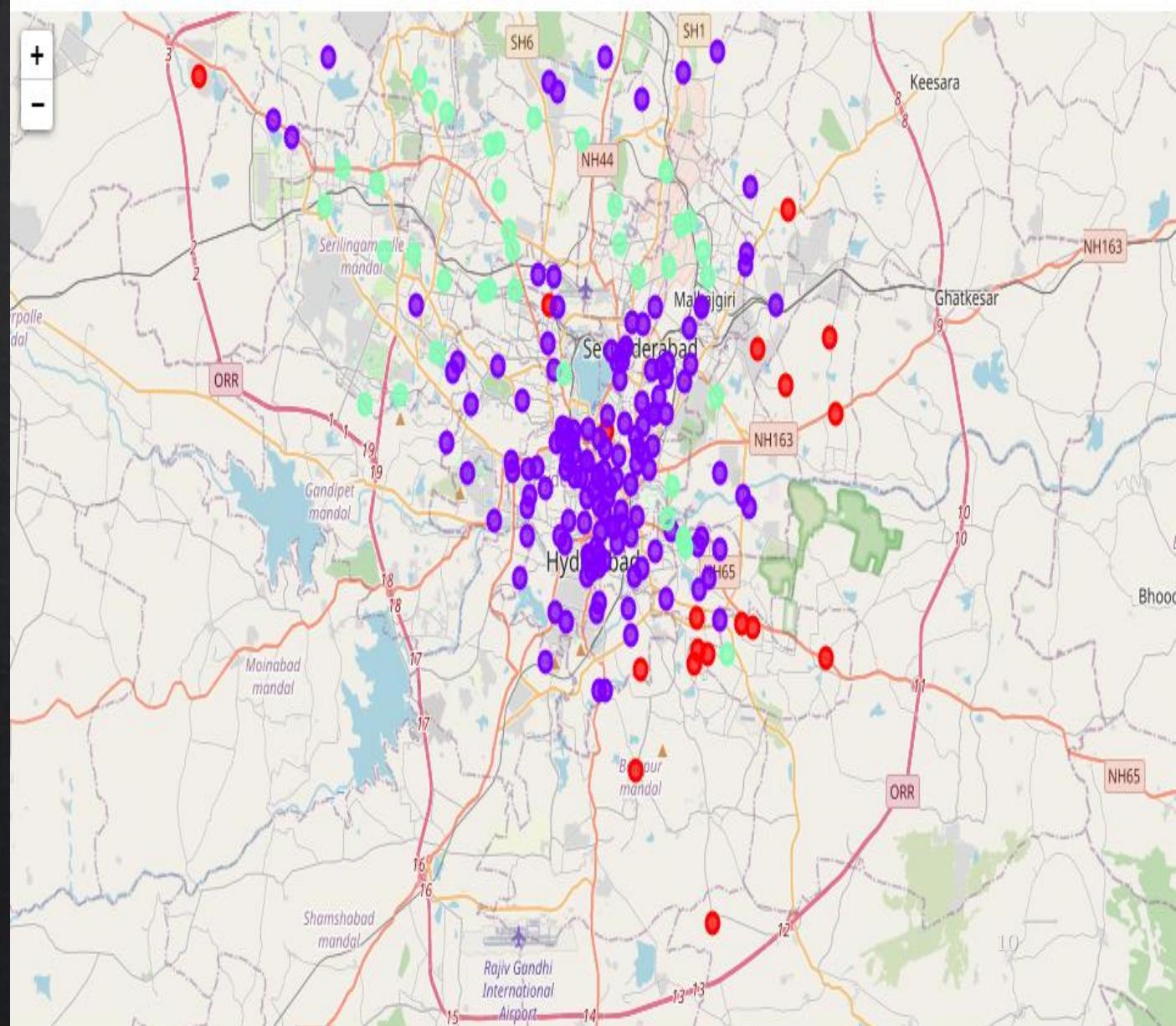
1. Domalguda
2. Hyderguda
3. Barkatpura
4. Ameerpet



Conclusion(3/3)

The neighborhoods of cluster 2 should not be considered as there are already high number of restaurants present in the area.

Restaurants in these clusters are likely already suffering with immense amount of competition because of oversupply and high concentration of existing restaurants



Drawbacks

1. Major issue with this project is that Foursquare API returns very few number of venues
2. This project was a demonstration of the skills developed during the course of this certification. For setting up a restaurant there are several other factors that need to be considered like cost of rent, the surroundings around the shopping mall, the kind of people in the locality-if it's a luxurious area many people prefer going out, their lifestyle will be different from others and therefore spend a lot.
3. The demographic of the neighborhood should also be considered. That is, if the people living around the neighborhood prefer going out or ordering in etc.

Useful Links

- ❖ Github Link
- ❖ <https://github.com/nehamadnekar/Capstone-Final-Project/blob/fd9426d795ce92ca024079361f32640bb5a1ca30/Capstone%20Final%20Project.ipynb>
- ❖ NBViewer Link
- ❖ <https://nbviewer.jupyter.org/github/nehamadnekar/Capstone-Final-Project/blob/fd9426d795ce92ca024079361f32640bb5a1ca30/Capstone%20Final%20Project.ipynb>
- ❖ Jupyter Notebook Link
- ❖ https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/8fafab4c-1d6c-463b-8160-2441852973b2/view?access_token=36cb94890a6a6f1bd2f95c3fbadd7eb3c9b19d0f6488a6e5af013ea1c5d3316