Coursera Capstone

IBM Data Science Capstone

Opening of New Multiplex in Hyderabad, India

By
Vinay K

Introduction:

A multiplex is a movie theater complex with multiple screens within a single complex. They are usually housed in a specially designed building. There are so many movie lovers increasing day by day. So day by day so many Multiplexes were getting built in major cities like Hyderabad. But most of the Multiplexes are built in the same place which will result in less profits. But due to the unavailability of Multiplexes near to there area and also the traffic in the city makes it difficult to watch movies in Multiplexes. Particularly the location of Multiplexes is one of the most important factors which decides the success of the Multiplexes.

Business problem:

Day by day so many Multiplexes were getting built in major cities like Hyderabad. But most of the Multiplexes are built in the same place which will result in less profits. So by using data science to find the perfect place to build a new Multiplex will increase their profits.

Target Audience:

This project will be useful for the developers and investors who are interested in building Multiplexes in Hyderabad.

Data:

To solve this data we want three types of data:

- List of popular suburbs in Hyderabad. We can get this data fro wikipedia page "https://commons.wikimedia.org/wiki/Category:Suburbs_of_Hyderabad,_India"
- Latitude and longitude coordinates of these suburbs. We can get this data from python geocoder package.
- Venue data to find the shopping malls. We can get this data from Foursquare API

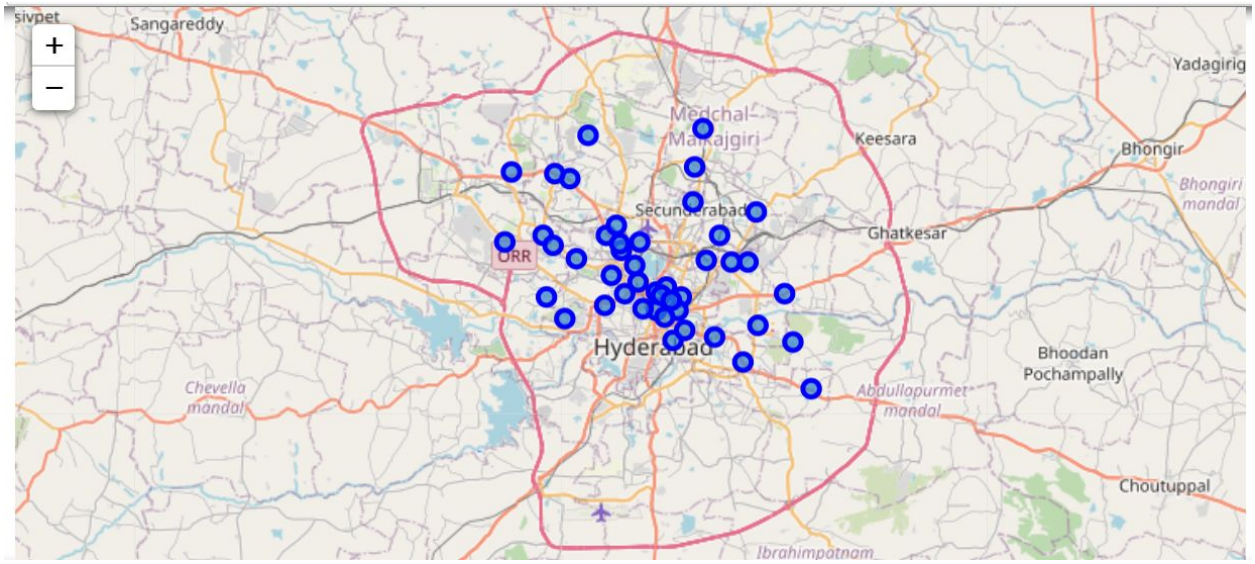
Methodology:

I utilized the python Geocoder package to find the latitude and longitude of all the locations as shown in picture below:

	neighborhood	latitude	longitude
0	Abids	17.389478	78.477182
1	Alwal	17.502229	78.508858
2	Ameerpet	17.437501	78.448251
3	Bandlaguda, Rangareddy	17.364495	78.589584
4	Banjara Hills	17.417746	78.439901

I used a python folium library to visualize geographic details of Hyderabad and its main suburbs and I created a map of Hyderabad

with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:



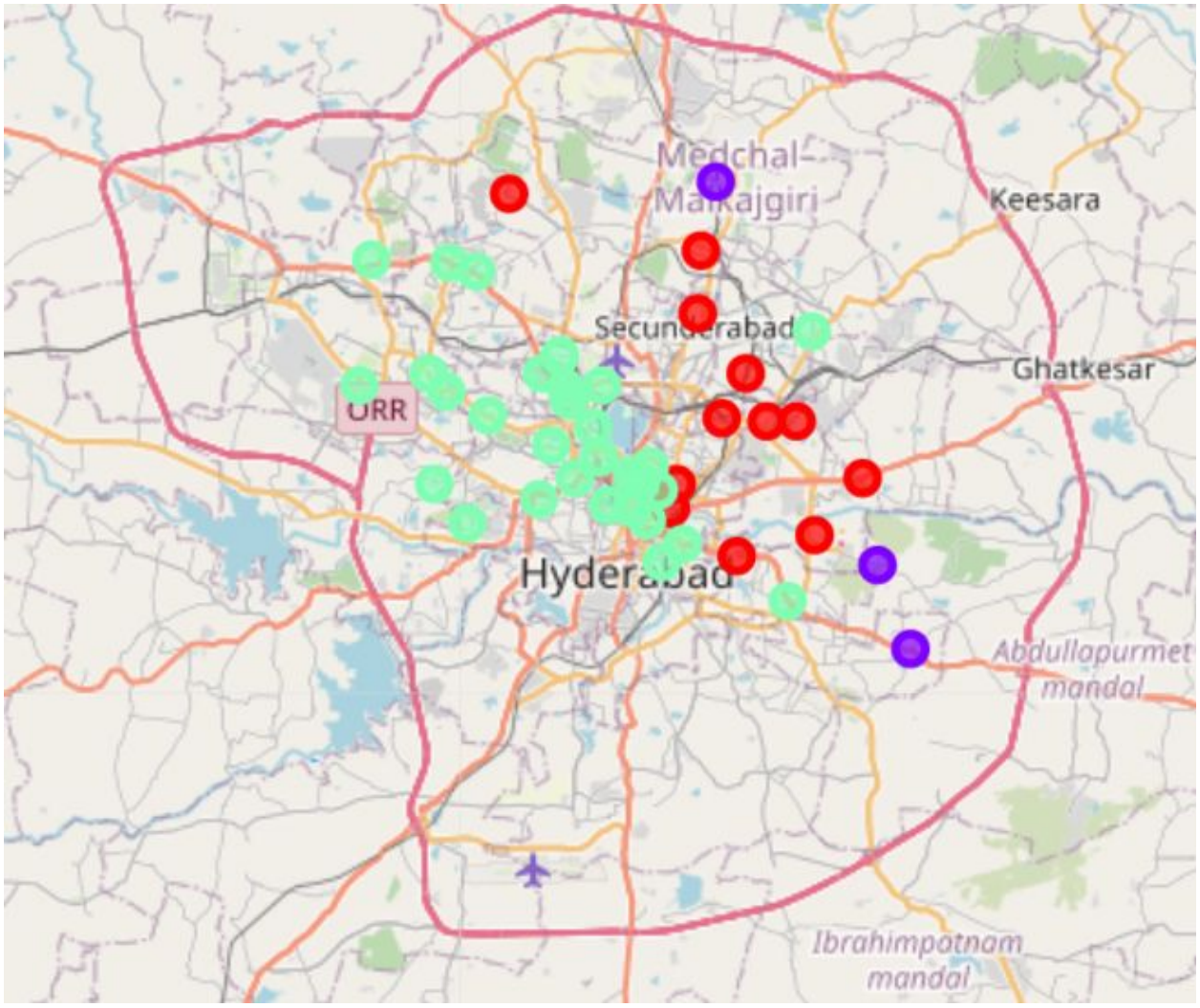
I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venues and the radius 10000 meter for each borough from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Abids	17.389478	78.477182	Pragati	17.388088	78.481134	South Indian Restaurant
1	Abids	17.389478	78.477182	Mayur Pan Shop	17.388894	78.480578	Juice Bar
2	Abids	17.389478	78.477182	Cafe Bahar	17.399595	78.478566	Indian Restaurant
3	Abids	17.389478	78.477182	Subhan Bakery	17.392412	78.464712	Bakery
4	Abids	17.389478	78.477182	King & Cardinal	17.400678	78.488575	Burger Joint

We can filter the Multiplexes near each place and we can cluster them

Results:

We can see the clusrered multiplexes location in hyderabad in below figure:



Conclusion:

As we can see more Multiplexes are concentrated in green clusters compared to remaining clusters. So there is more chance of getting profits when we build in remaining clusters as there is less competition in those areas.