



The Battle of Neighborhoods

**Opening a Shopping Mall in
Hyderabad , Telangana, INDIA**

Capstone Project of IBM Data Science

-by Vamshi Krishna Gundu, July
2020



Business Problem

- Location of the Shopping malls is one of the most important decisions that will determine whether the mall will run in profits or not
- Objective: To analyze and select best locations in the city of Hyderabad to open a new shopping mall
- This project is timely as the city is suffering from undersupply of shopping malls
- Business question: “ In the city of Hyderabad , Telangana state of India, if someone is looking to open a Shopping Mall, where would you recommend that they open it? ”



Data

Required data:

- List of Neighborhoods in Hyderabad
- Latitude and Longitude coordinates of those neighborhoods
- Venue data, Shopping mall category in particular

Sources of above data:

- Wikipedia url:
"https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Hyderabad"
- Geocoder library to get coordinates
- Foursquare API for venue data



Methodology

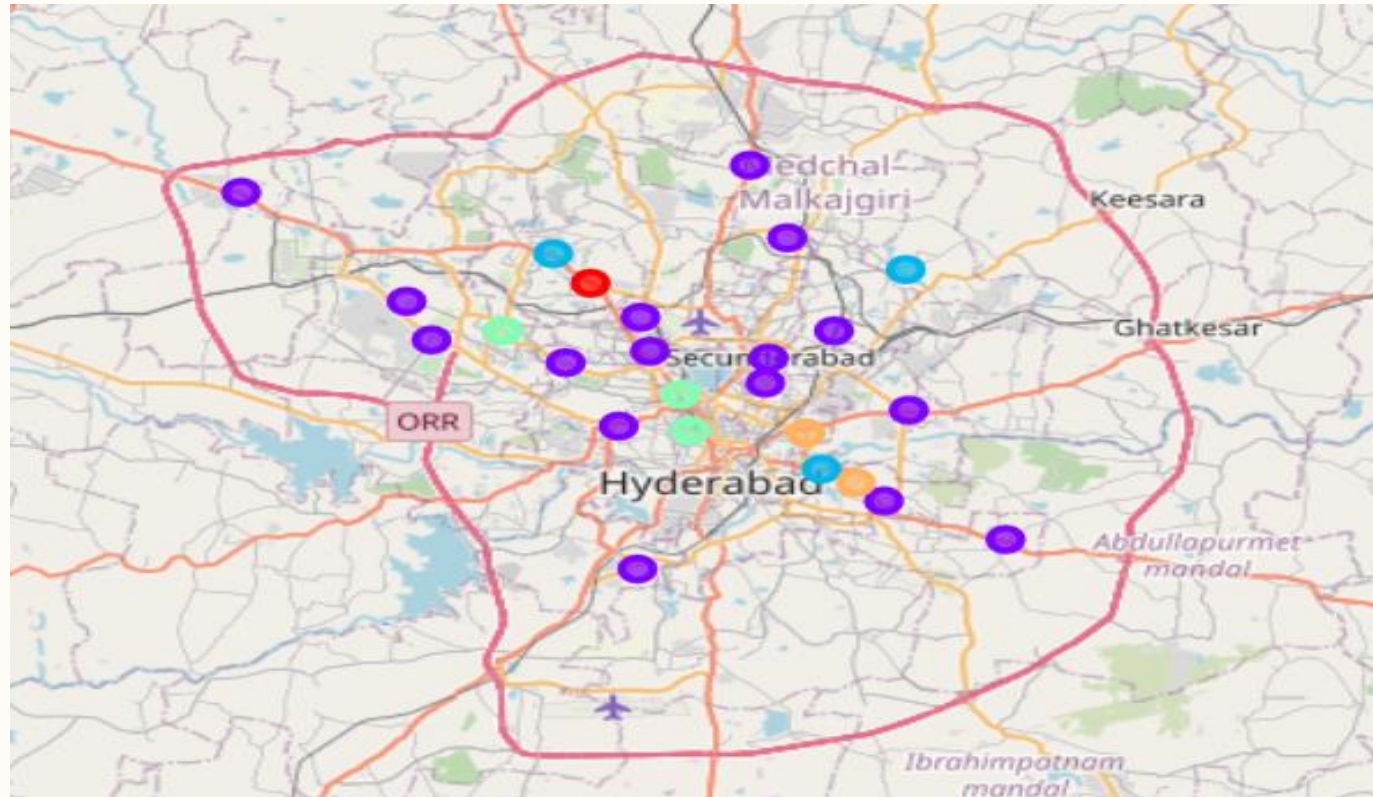
- Web scraping wiki page to get neighborhoods list
- Get coordinates using geocoder library
- Use FourSquare API to get venue data
- Group by Neighborhood and taking the mean of the frequency occurrence of the each venue category
- Filter venue category by “Shopping Mall”
- Form clusters using K-means
- Create Map using cluster and Neighborhood data



Results

- Cluster 0 (Red colour) : With only 1 Neighborhood with mean value more than all the clusters as it is an industrial area with small shopping malls around it
- Cluster 1 (Purple colour) : With almost no shopping malls present in these neighborhoods, some of the areas might be having one or two shopping malls but our model did not detect it because of the radius(2000m) that we have chosen.
- Cluster 2,3,4 (Blue, Green, Orange) : These clusters 2,3 and 4 has enough number of shopping malls in these neighborhoods with different frequencies of mean. Neighborhoods in clusters 2 and 3 comes under main city or popular areas of the city Hyderabad.

Clusters formed using K-means





Conclusion

- Answer to business question: The neighborhoods in cluster 1 are the most preferred locations to open a new shopping mall
- Visualization and the Foursquare API for community data processing, it is possible to cluster city shopping mall data on the basis machine learning techniques – K-Means Algorithm.
- This study can be reproduced easily for any other cities across the world