Coursera Capstone

IBM Applied Data Science Capstone

Starting new Multiplex in Hyderabad, India

Presented By:

Harish Reddy Kunduru

March 2019

Business Problem

- The objective of this capstone project is to analyze and select best locations in the city of Hyderabad for starting new Multiplex.

- Using Data science methodologies and Machine Learning techniques like Clustering this project aims to provide solutions to answer the business question:

"In the city of Hyderabad, India, if a property developer is looking to open a new Multiplex, 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, particularly data related to Multiplex.

Source of Data

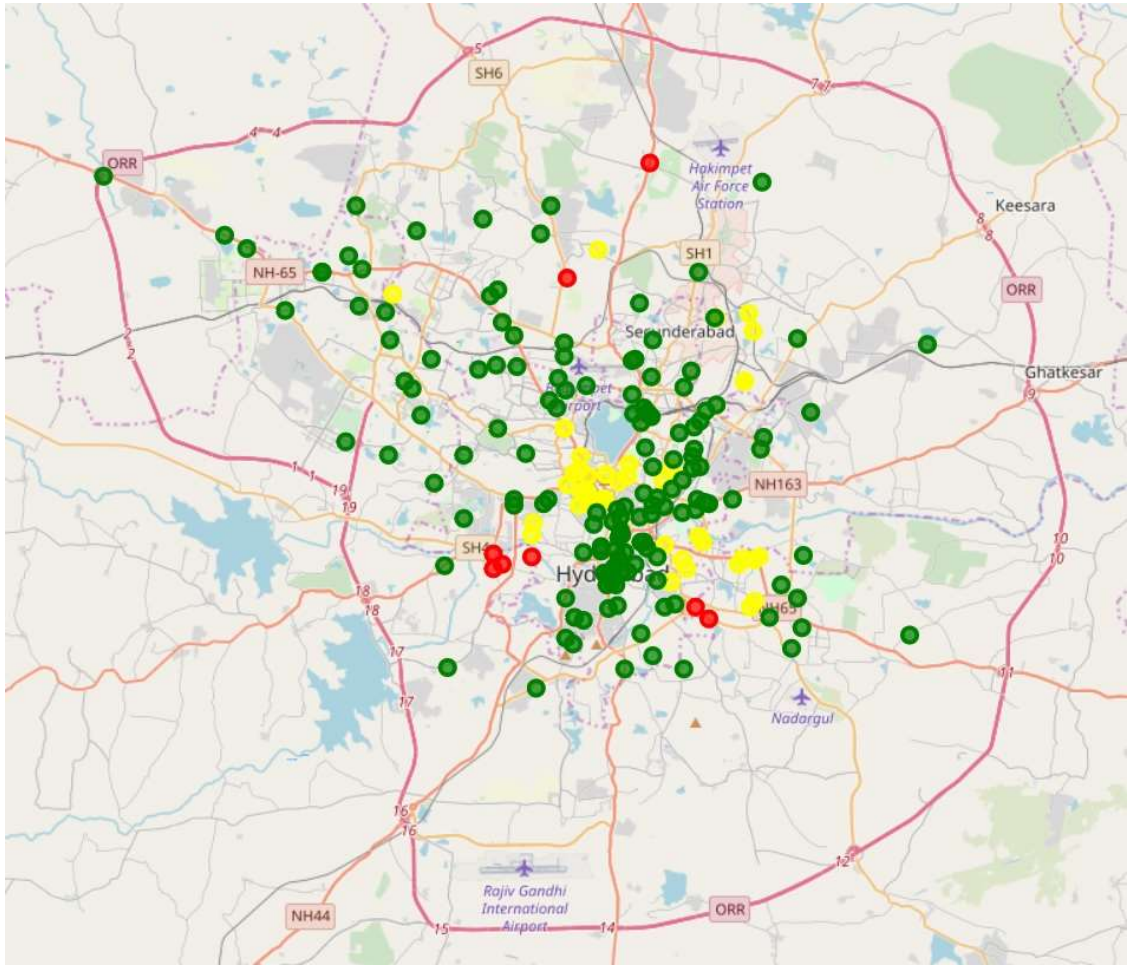
- The Wikipedia Page which contain data of neighborhoods in Hyderabad
- Foursquare API to get latitude and longitude of neighborhoods
- Geocoder to get the geographical data of neighborhoods

Methodology

- Web scraping Wikipedia page for neighborhood list
- Get latitude and longitude of neighborhoods using geocoder
- Use Foursquare API to get venue data
- Group data by neighborhood and taking the mean of frequency of occurrence of each venue category
- Filter venue category by Multiplex
- Perform clustering on data by using K-Means Clustering
- Visualize the clusters in map using Folium

Results

- Cluster 0: neighborhoods with moderate number of Multiplexes (Yellow)
- Cluster 1: neighborhoods with low number to no existence of Multiplexes (Green)
- Cluster 2: neighborhoods with high concentration of Multiplexes (Red)



Discussion

Most of Multiplexes in Hyderabad city are found in areas of cluster 2.It is not good to start a new multiplex in neighborhoods of cluster 2 because of very high competition. Neighborhoods in cluster 0 have moderate multiplex numbers but still there will be tough competition. But Neighborhoods of cluster 1 has less number of multiplex when compared to other clusters. There are many Neighborhoods without any multiplex, It is good to start in neighborhoods of cluster 1 as there will be little or no 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 Multiplex. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighborhoods in cluster 1 are the most preferred locations to start a new Multiplex. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new Multiplex.