NEIGHBORHOOD CLUSTERING OF HYDERABAD CITY

Contents

Introduction

Description and discussion of the background

Hyderabad City is capital of Telangana state in India. The economy of the capital is based on traditional manufacturing, Information Technology and tourism. Starting in 1990s, economic landscape of the city changed from service hub to a diversified economy. Hyderabad is the largest producer of Telangana State's GDP. It is among the global centers of Information Technology and is also known as 'Cyberabad'. It exports have increased over the last two decades and major multinational firms located their offices in Hyderabad.

Many IT campuses have been developed in areas such as Madhapur, Gachibowli, Kondapur and Uppal. As more and more Information Technology Investment Region (ITIR) are coming up and offices are setup, there is a constant need amenities for people who work and live in that area.

It is significantly necessary to understand landscape of amenities/venues currently existing in the neighborhoods of Hyderabad so that any new venue can be opened by entrepreneurs to serve right necessity of the new people who started working or some of them could have moved to these neighborhoods.

Wiki webpage provides all the neighborhoods of Greater Hyderabad Municipal Council. This data can be analyzed, and each neighborhood, also known as Wards, can be clustered according to the venue density. This visualization of the clusters can be provide data insights to entrepreneurs to know what kind of new venue will successfully meet the need of people and also make a good business.

Data

Neighborhoods/Wards data

Greater Hyderabad Municipal Council data is obtained and scraped from Wiki webpage.

Geocoding

Using Google Maps Geocoding API, the latitude and longitude of the neighborhoods are retrieved and then the locations are stored in initial dataframe. This information will be used later to retrieve venues from Foursquare website using Foursquare APIs.

Venues

Once the latitude and longitude information is obtained from Geocoding APIs for each of the neighborhoods, this information is used to retrieve venues from Foursquare website using Foursquare APIs with a limit of 30 venues and within the radius of 1000 meters of the mentioned location attributes.

Methodology

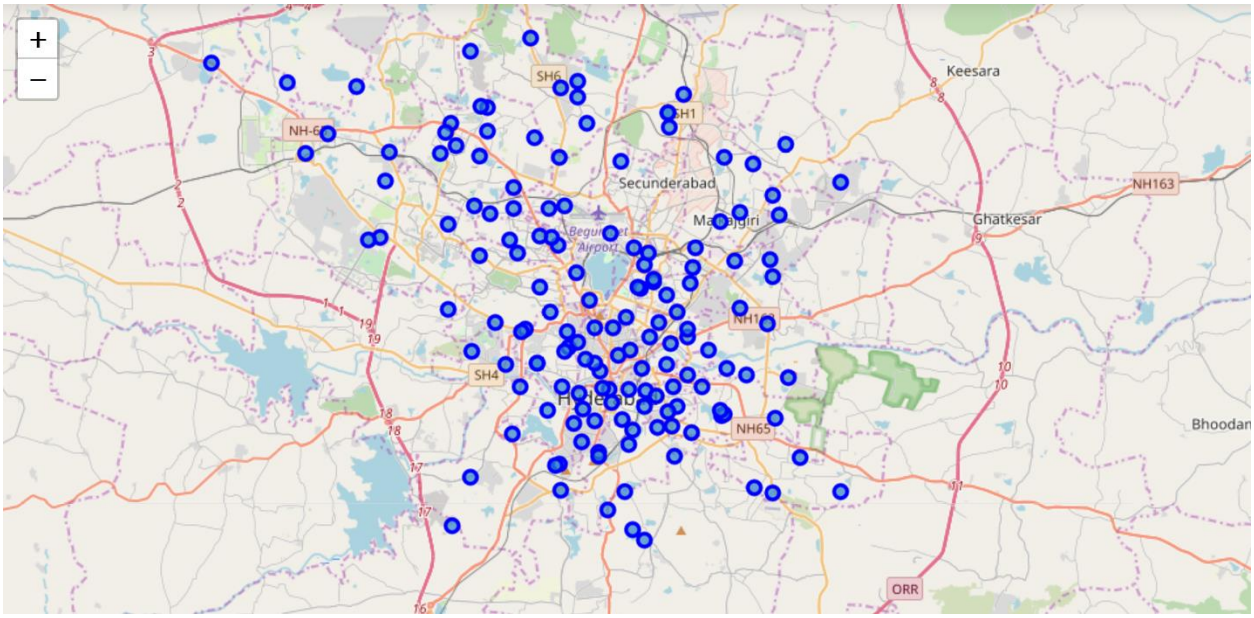
Geocoding APIs

Location information obtained from Geocoding APIs is tested for least amount of errors/collisions and it turned out to have only 1 collision from the data retrieved for over 150 neighborhoods.

Folium

Folium is used for data visualization on Hyderabad city map. Folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the leaflet.js library. Data is manipulated in python and then visualized on a leaflet map via folium.

Neighborhoods of Hyderabad – visual view



One-hot encoding

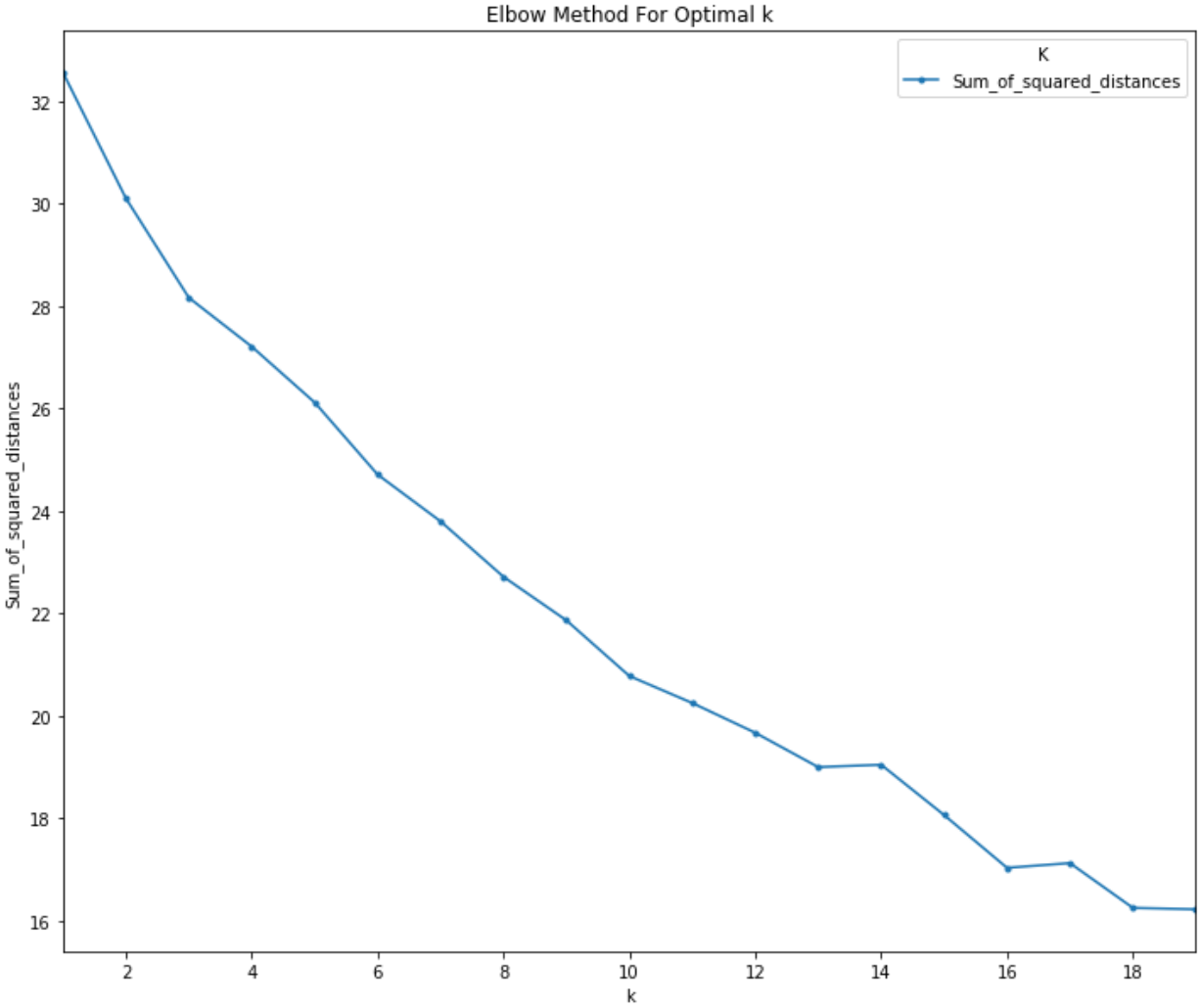
One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction. All unique items under 'Venue Category' are one-hot encoded for K-Means clustering machine learning algorithm.

As there are high number of venues, only the top 10 most common venues are selected and this data is used in K-Means clustering learning algorithm to train the model to identify right clusters of the venues.

Optimal number of clusters

Elbow method is used for determining optimal number of clusters. A diagram is plotted with number of clusters (K) on X-axis and sum of squared distances on Y-axis. Based on the plot, optimal number of cluster is 5 as seen in the below.

Number of Clusters vs Sum of Squared distance diagram



K-Means clustering

As number of 'Venue Categories' data is huge, K-Means algorithm is used. Venue data is trained using K-Means Clustering Algorithm and get desired number of clusters of the venues.

Cluster dominating venue/venues

Here is merged table with cluster labels for each neighborhood.

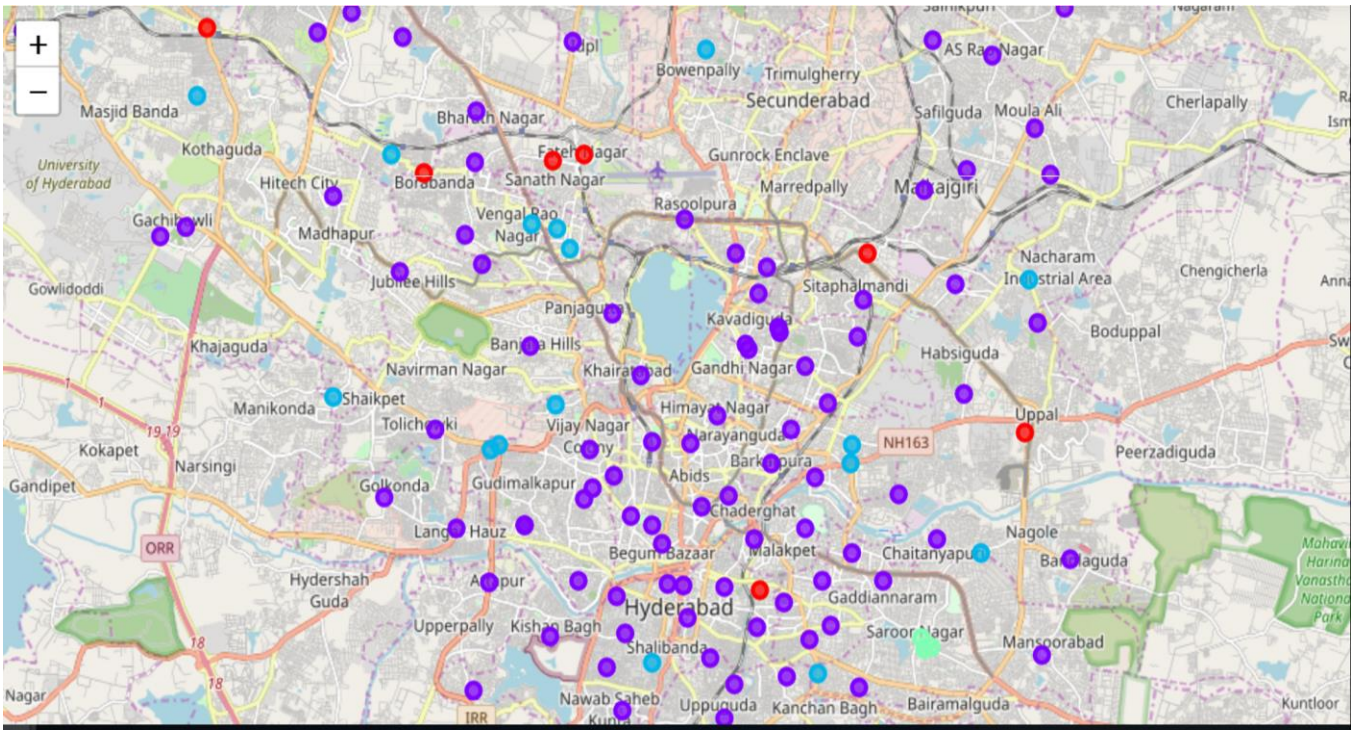
Ward	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
Kapra	17.488525	78.568172	1.0	Bakery	ATM	Pharmacy	Bar	Spanish Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit
Dr AS Rao Nagar	17.478478	78.550637	1.0	Café	Department Store	Clothing Store	Indian Restaurant	Fast Food Restaurant	Pizza Place	Bank	Farmers Market	Falafel Restaurant
Meerpet HB Colony	17.453117	78.564488	1.0	Neighborhood	Movie Theater	Garden	Mountain	Department Store	Falafel Restaurant	Exhibit	Electronics Store	Donut Shop
Nacharam	17.430809	78.559548	2.0	Indian Restaurant	Movie Theater	Electronics Store	Zoo	Diner	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit
Chilukanagar	17.421812	78.561472	1.0	ATM	Playground	Department Store	Farmers Market	Falafel Restaurant	Exhibit	Electronics Store	Donut Shop	Diner
Habsiguda	17.406623	78.543774	1.0	Indian Restaurant	Food	Cricket Ground	Park	Gas Station	Café	Athletics & Sports	Fast Food Restaurant	Stadium

Based on the data, number of '1St Most common Venue' and number of '1st and 2nd most common venue' are estimated. Below chart gives an idea on which venue dominates the cluster.

	Cluster - 5	
Cluster #	**2nd Most Common Venue**	**1st Most Common Venue**
0	Train Station	Train Station
1	Bakery + Café + Indian Restaurant	Café +Indian Restaurant
2	Indian Restaurants + Others	Indian Restaurants
3	Pharmacy	Pharmacy
4	Health & Beauty Service	Health & Beauty Service

Visualization of Clusters

Each neighborhood will have clusters found using optimal approach.



Discussion

Based on the clusters chart, it is clear that Cluster 0 clearly indicates that it has train stations nearby. Cluster 1 and 2 are Indian Restaurants and Cafés dominated can Cluster 3 is clearly a Pharmacy cluster whereas Cluster 4 is for Health and Beauty Services.

Conclusion

- Cluster 0 (Red), indicate that there is a potential to have more Indian restaurants near Train stations/Light rail stations to serve the food needs of passengers. Light rail stations are new to the city and not many restaurants might be opened yet.
- Neighborhoods such as Gachibowli, Vinayaknagar, Madhapur have Cluster 1(Purple) dots, these are the areas where multinational companies have opened offices in last decade and there might be a need of Pharmacies and ATMs in those neighborhoods