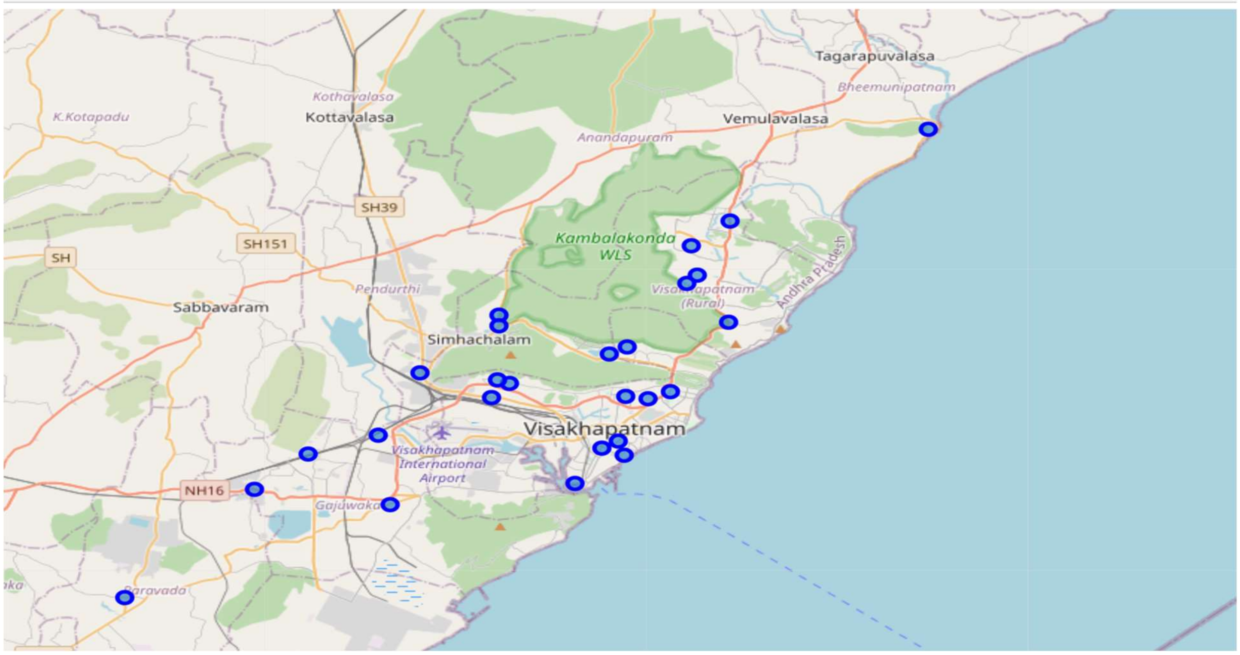
The Battle of Neighborhoods – Vizag



Applied Data Science Capstone by IBM on Coursera (Suresh Malla)

Introduction: Business Problem

This project deals with major venue categories in the neighborhoods of Vizag, the proposed executive capital of the Indian state of Andhra Pradesh. It is most populated and one of the largest cities in state of Andhra Pradesh. As the city drawing an attention in recent times, this Project would specifically help investors to choose right business to start.

The Foursquare API is used to access the venues in the neighborhoods. Since, it returns less venues in the neighborhoods, we would be analyzing areas for which countable number of venues are obtained.

Then they are clustered based on their venues using Data Science Techniques. Here the k-means clustering algorithm is used to achieve the task.

Folium visualization library can be used to visualize the clusters superimposed on the map of Chennai city. These clusters can be analyzed to help investors to select a suitable location for their business ideas.

Data Requirements:

Vizag has multiple localities to explore. Extracted all the localities by web scrapping Wikipedia page which has list of localities in Vizag. Used geocoder to get coordinates of longitude and latitude for each locality.

In order to obtain the venue details in each neighborhood Foursquare API is used.

Wiki page: https://commons.wikimedia.org/wiki/Category:Suburbs_of_Visakhapatnam"

Foursquare API:

https://api.foursquare.com/v2/venues/explore?

There is total 26 neighborhoods. Latitudes and Longitudes are obtained for each neighborhood and explored the venues from foursquare Api. A total of 62 venues are returned by foursquare API.

Methodology:

Now we have the neighborhoods data of Vizag (26 neighborhoods). We also have the most popular venues in each neighborhood obtained by using Foursqaure API. A total of 62 venues have been obtained in whole city and 37 unique categories. Observation here is data with venues for each neighborhood is very less which varies from 1-12 venues.

Perform one-hot encoding on the dataset and use it finding the 10 most common venue category in each neighborhood.

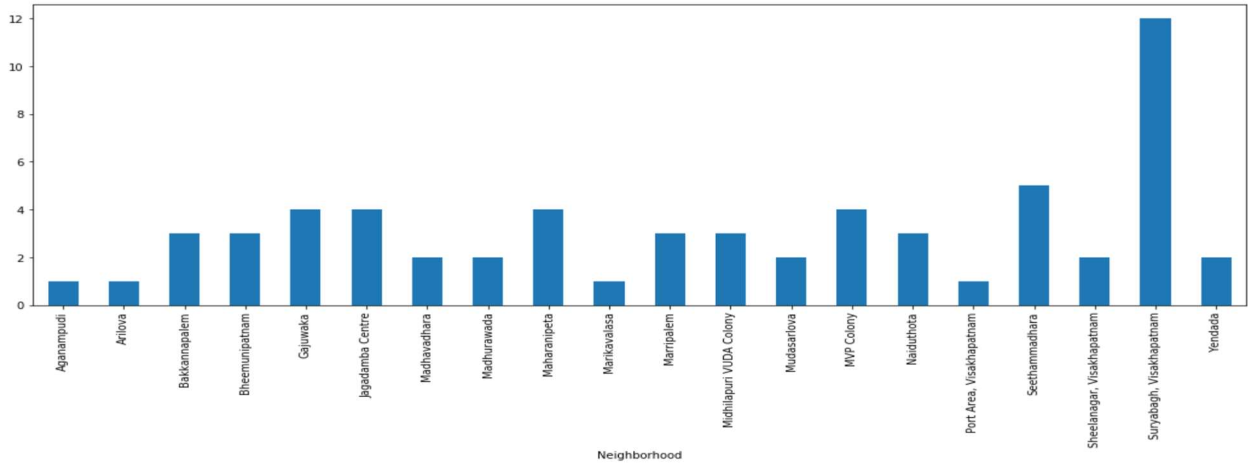
K- Nearest Neighbor clustering technique have been used to find optimum number of clusters.

Each cluster is analyzed to find major type of venue categories in each cluster.

This data can be used to suggest investors to start suitable business and location based on the category.

ANALYSIS:

Looking into the dataset, venues returned are varying from 1- 12 range. We can see there is not enough data that foursquare API return which could lead to in accurate results. But for this capstone, I am proceeding with the same data.



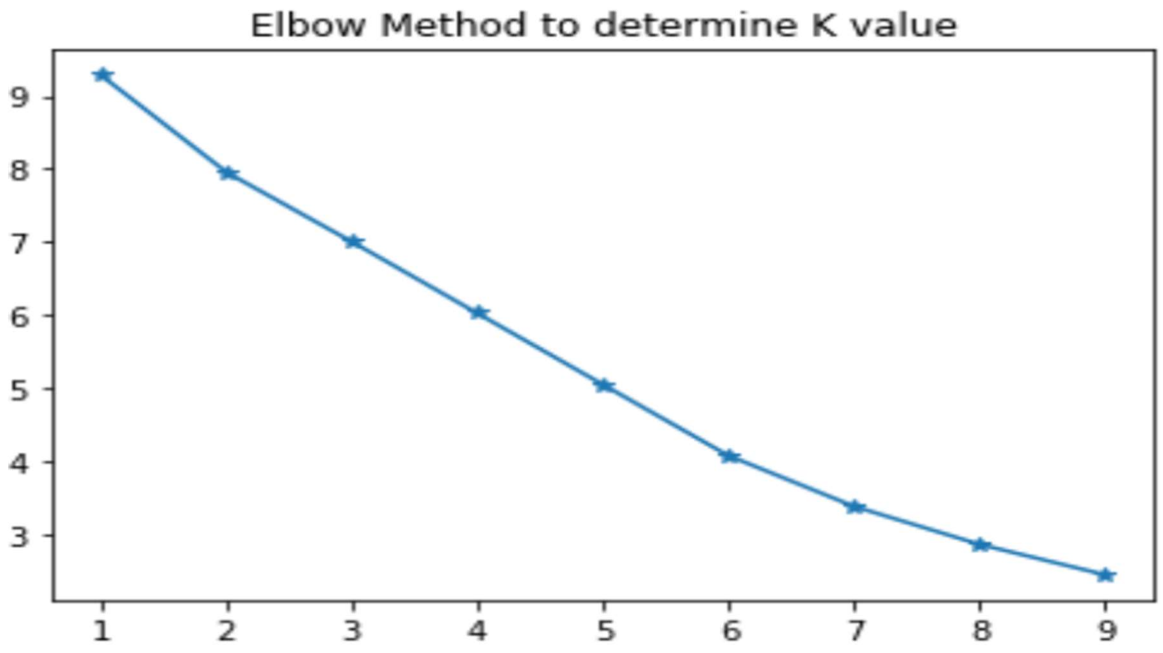
Perform one-hot encoding on the dataset and use it finding the 10 most common venue category in each neighborhood.

K- Nearest Neighbor clustering technique have been used to find optimum number of clusters.

Each cluster is analyzed to find major type of venue categories in each cluster.

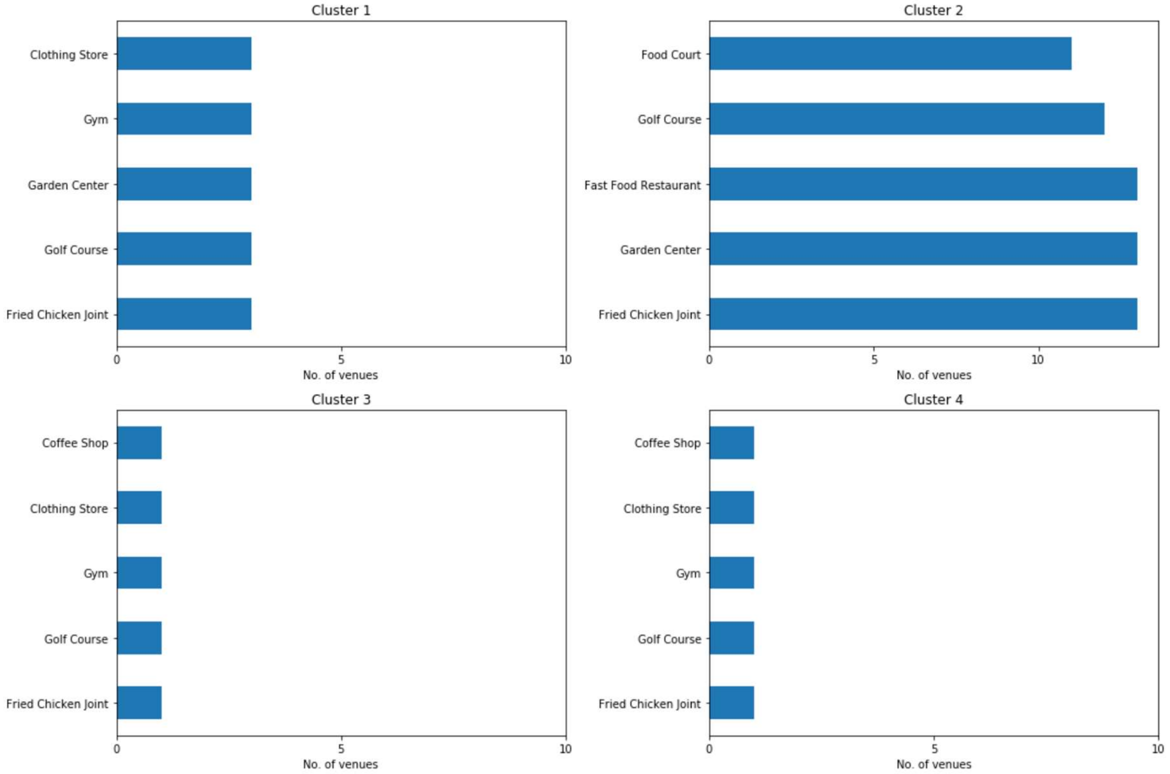
This data can be used to suggest investors to start suitable business and location based on the category.

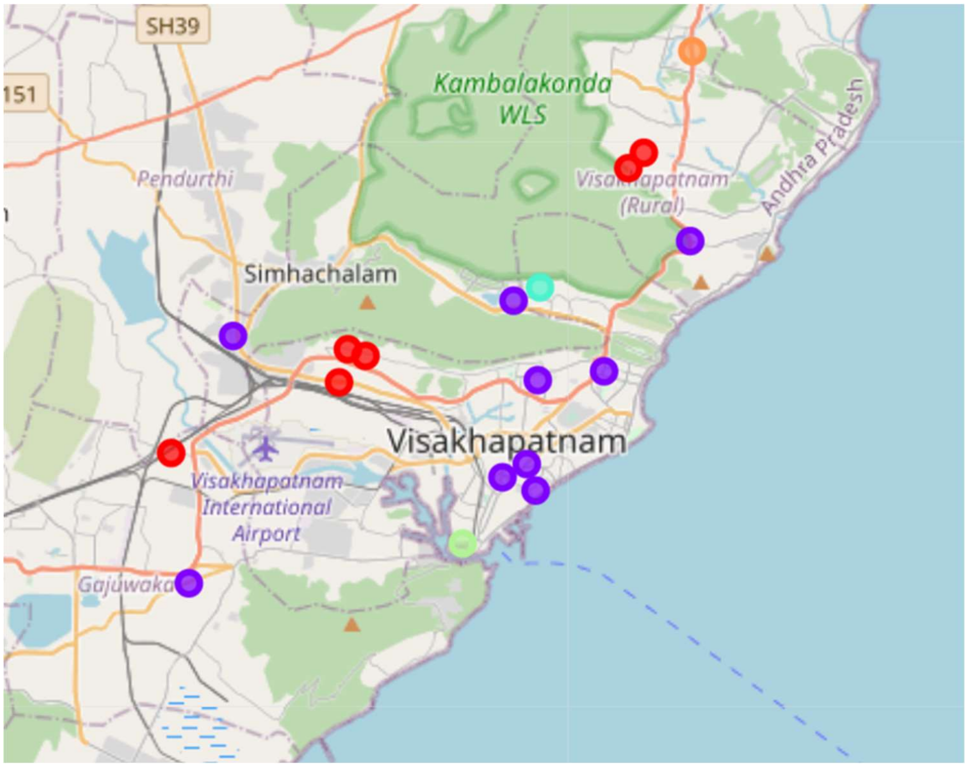
A range of values from 1 to 10 was considered, KNN clustering was performed on the dataset and plotted a elbow plot. From the elbow plot we can see that a value of k-value of 6 provides the best score. This k-value is used for K-means clustering technique-means labels obtained were included in the top neighborhoods dataset for examining the characteristics of each cluster.



Results and Discussion

Using the clusters and top venue categories lets visualize the top 5 venue category in each cluster.


Cluster 1
Clothing Store
Gym
Garden Center
Golf Course
Fried Chicken Joint
0
5
10
No. of venues
Cluster 2
Food Court
Golf Course
Fast Food Restaurant
Garden Center
Fried Chicken Joint
0
5
10
No. of venues
Cluster 3
Coffee Shop
Clothing Store
Gym
Golf Course
Fried Chicken Joint
0
5
10
No. of venues
Cluster 4
Coffee Shop
Clothing Store
Gym
Golf Course
Fried Chicken Joint
0
5
10
No. of venues


SH39
151
Kambalakonda
WLS
Pendurthi
Visakhapatnam
(Rural)
Andhra Pradesh
Simhachalam
Visakhapatnam
Visakhapatnam
International
Airport
Gajuwaka

After going through the neighborhoods of Visakhapatnam, India and looking the cluster information, cluster1 is already occupied with good number of restaurants, grocery, food and women store but see an opportunity for someone to start off with a gym which is missing in the cluster. Whereas starting with restaurant, clothing business as an opportunity to start in cluser2,3,4 considering the less competition and uniqueness of the business.

The main challenge here with the analysis is Foursquare API has returned very less data points and this can be improved by trying with other API's which gives better data for our analysis to recommend business ideas in Vizag location.