Battle of Neighbourhoods

Finding out the best possible location for buying or renting home in UttarPradesh for Immigrants from other cities

Problem

The purpose of this Project is to help people (new immigrants) in exploring better facilities around their neighbourhood. It will help people making smart and efficient decision on selecting great neighbourhood out of numbers of other neighbourhoods in Uttar Pradesh, India based on amenities & budget.

Lots of people come to Uttar Pradesh for job, education & business purpose and needed lots of research for good housing prices and reputed schools for their children. This project is for those people who are looking for better neighbourhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.
It will help people to get awareness of the area and neighbourhood before moving to a new city, state, country or place for their work or to start a new fresh life.

Data Source & Data Frame

Fetching the data of districts in Uttar Pradesh from

• https://www.indianmirror.com/india-post/indianpincode.html

	Neighborhood	Latitude	Longitude
0	Agra	27.166667	78.083333
1	Allahabad	25.466667	81.900000
2	Aligarh	27.500000	79.666667
3	Ambedkar Nagar	26.433333	52.550000
4	Auraiya	26.466667	79.516667
...	...	...	...
65	Sant Ravidas Nagar	25.350000	82.466667
66	Sultanpur	26.266667	82.066667
67	Shravasti	27.916667	82.616667
68	Unnao	26.800000	80.716667
69	Varanasi	25.333333	83.000000

70 rows × 3 columns

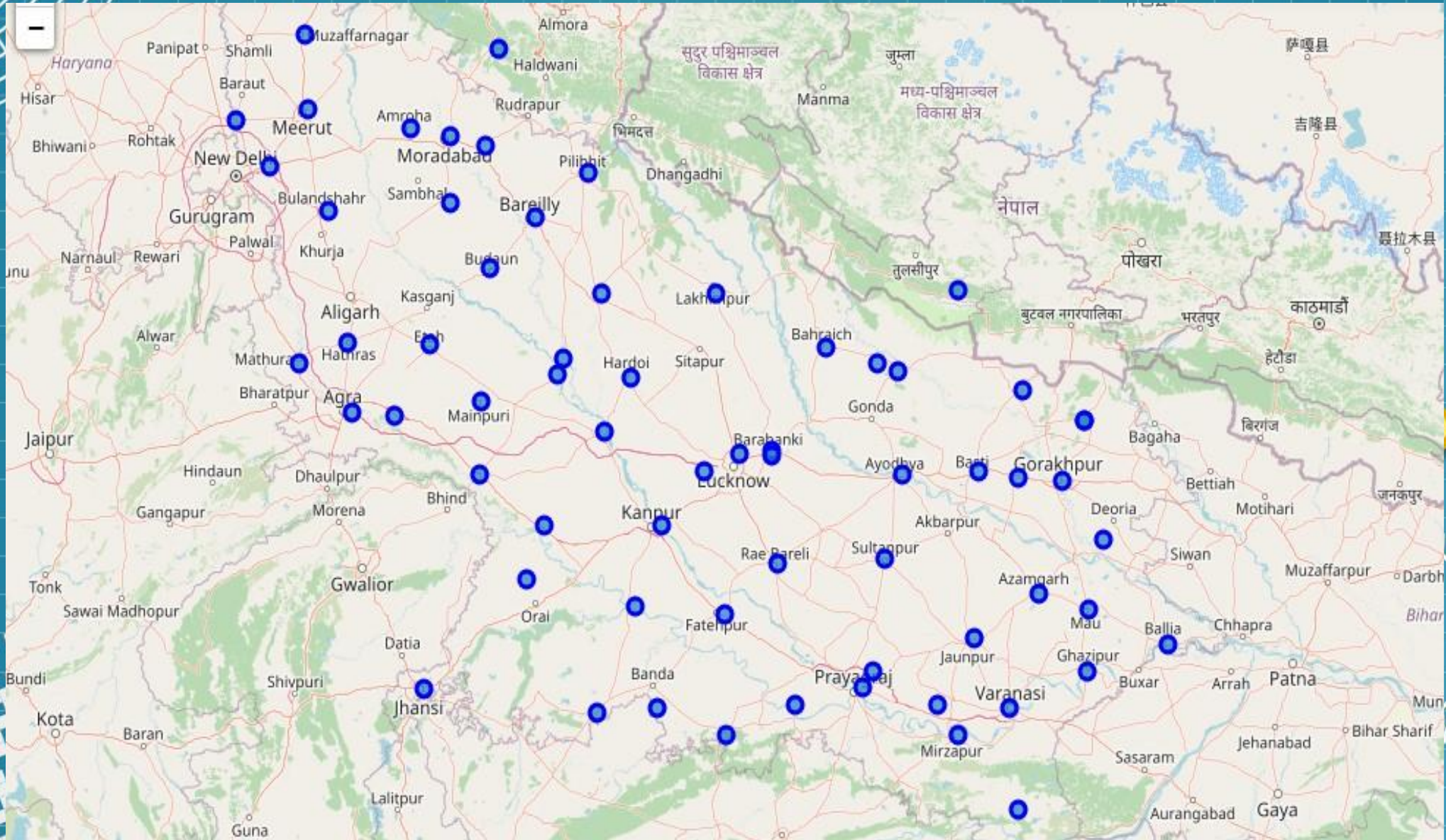# Methodology

- Using Geopy library for getting geographical coordinates of Uttar Pradesh

- The geograpical coordinate of Uttar Pradesh are 27.1303344, 80.859666.

- Creating a map of Uttar Pradesh with neaby cities superimposed on top

- Using Foursquare API to explore points of interest in each city within a radius of 40km

- There are 123 unique venue category in Uttar Pradesh

# Uttar Pradesh with cities

Methodology

- Analysing each city by converting each venue category into dummy values

- For K-Means Clustering approach group rows by neighbourhood city & taking the mean of the frequency of occurrence of each category in each city

Neighborhood	Vegetarian / Vegan Restaurant	ATM	Accessories Store	Airport	Airport Service	Airport Terminal	American Restaurant	Art Gallery	Art Museum	...	South Indian Restaurant	Spa	Speakeasy	Tea Room	Temple
Agra	0.0	0.000000	0.0	0.022222	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.00	0.000000
Aligarh	0.0	0.666667	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.00	0.000000
Allahabad	0.0	0.000000	0.0	0.055556	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.00	0.000000
Ambedkar Nagar	0.0	0.000000	0.0	0.333333	0.0	0.0	0.333333	0.0	0.0	...	0.0	0.0	0.0	0.00	0.000000
Auraiya	0.0	0.375000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.00	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Sitapur	0.0	0.500000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.00	0.000000
Sonbhadra	0.0	0.666667	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.00	0.000000
Sultanpur	0.0	0.300000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.00	0.000000
Unnao	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.02	0.000000
Varanasi	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.00	0.051282

Methodology

.For getting optimum number of clusters into which cities can be categorised based on amenities & budget, Elbow method is used first for getting an idea of number of clusters
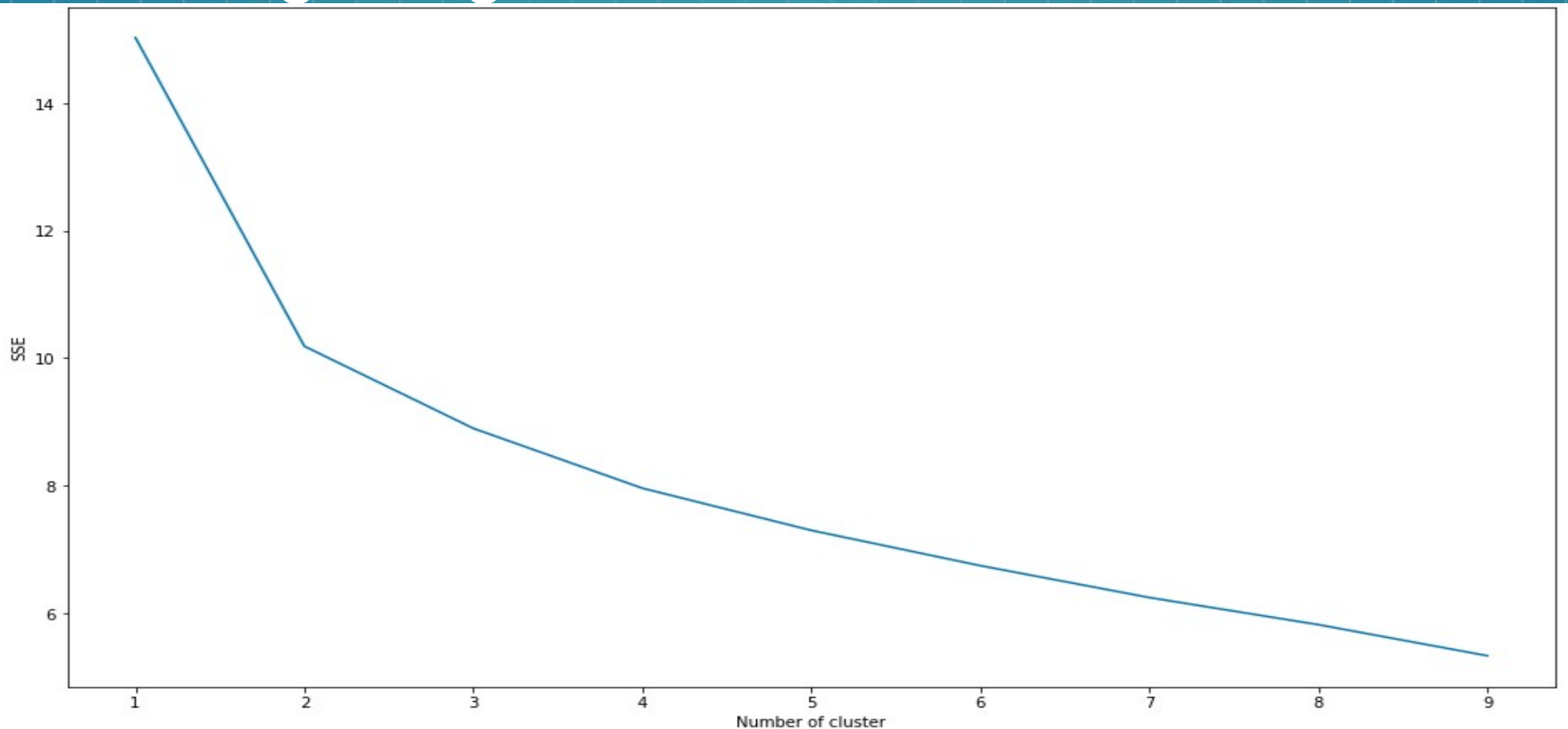
Methodology

• Based on Elbow plot, I decided to cluster the cities into 2, 3, 4, 5, 6 clusters. Now, for getting accuracy, I used Silhouette score

```
For n_clusters = 2 The average silhouette_score is : 0.42916292288748314
For n_clusters = 3 The average silhouette_score is : 0.6251876926750025
For n_clusters = 4 The average silhouette_score is : 0.826294079517989
For n_clusters = 5 The average silhouette_score is : 0.6214774768554949
For n_clusters = 6 The average silhouette_score is : 0.4166696513773977
```
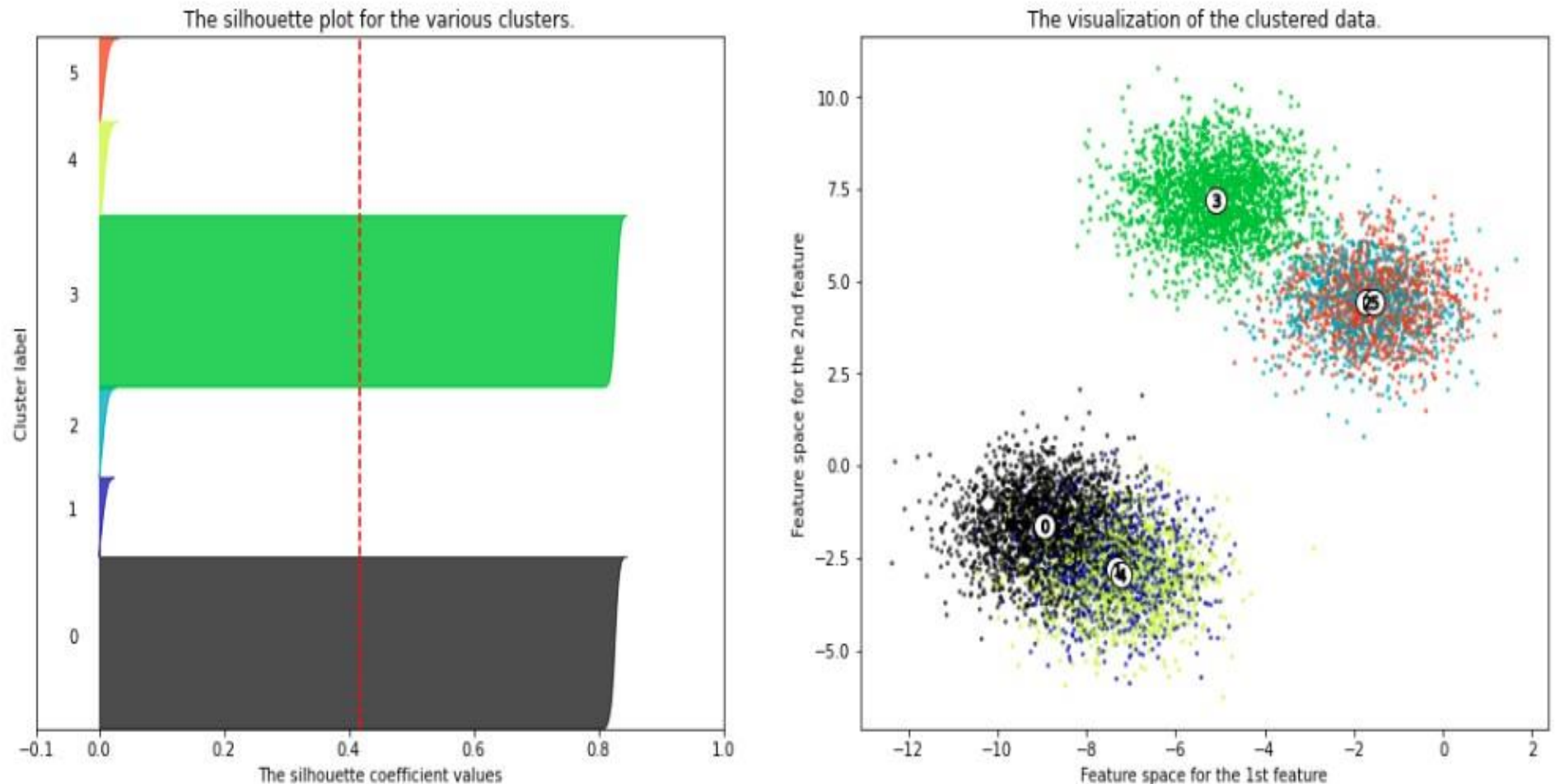
Methodology

- Figure below showing Silhouette plot for n = 2 i.e. number of clusters = 2

Silhouette analysis for KMeans clustering on sample data with n_clusters = 2
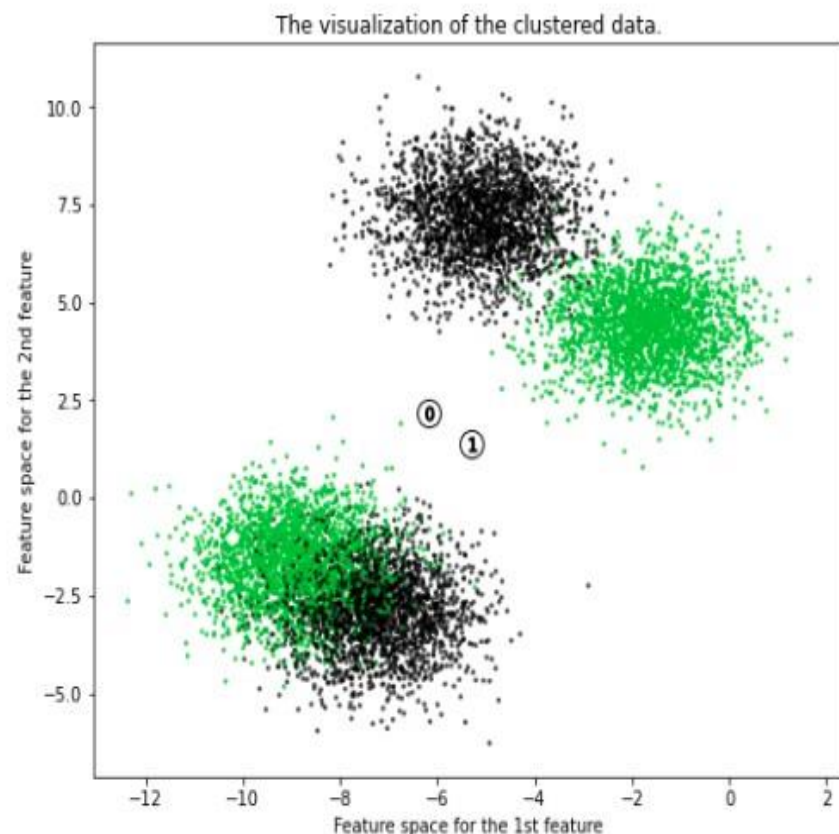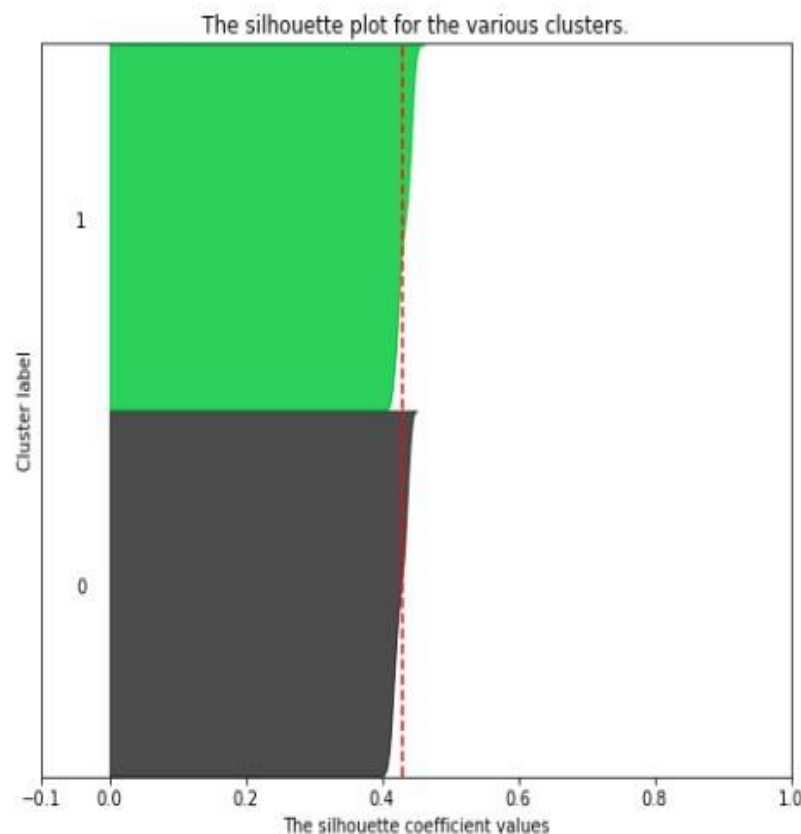
Methodology
Figure below showing Silhouette plot for n = 3 i.e. number of Clusters = 3

Silhouette analysis for KMeans clustering on sample data with n_clusters = 3
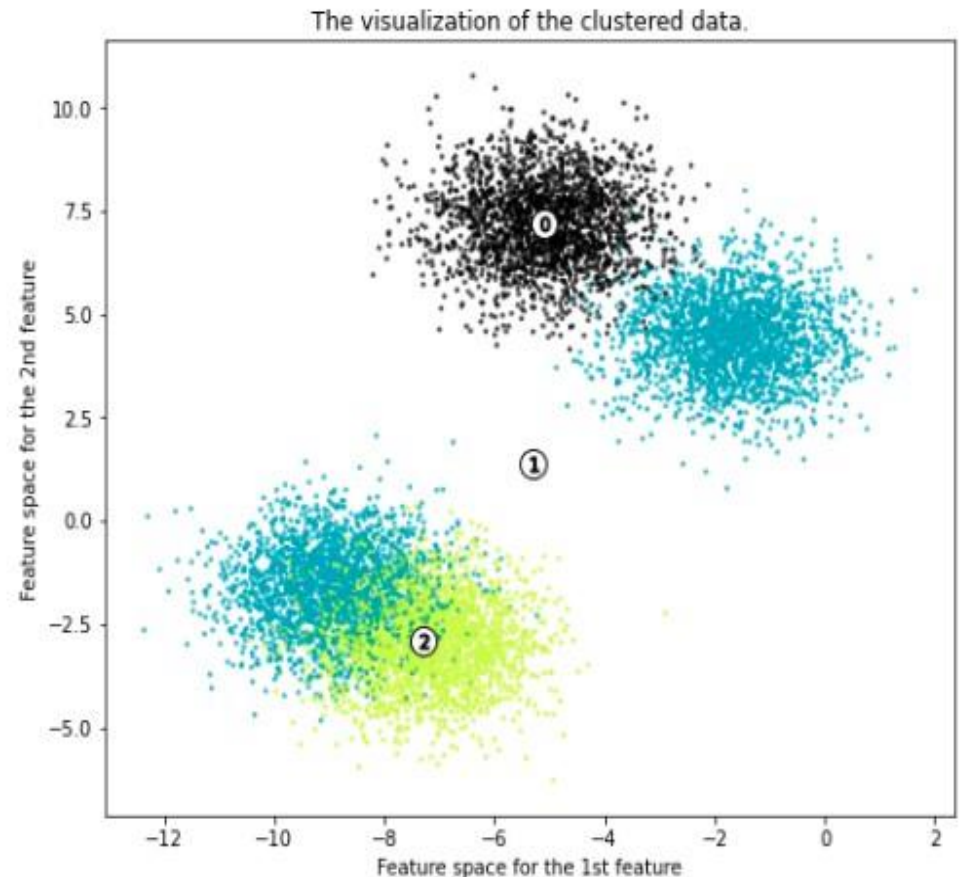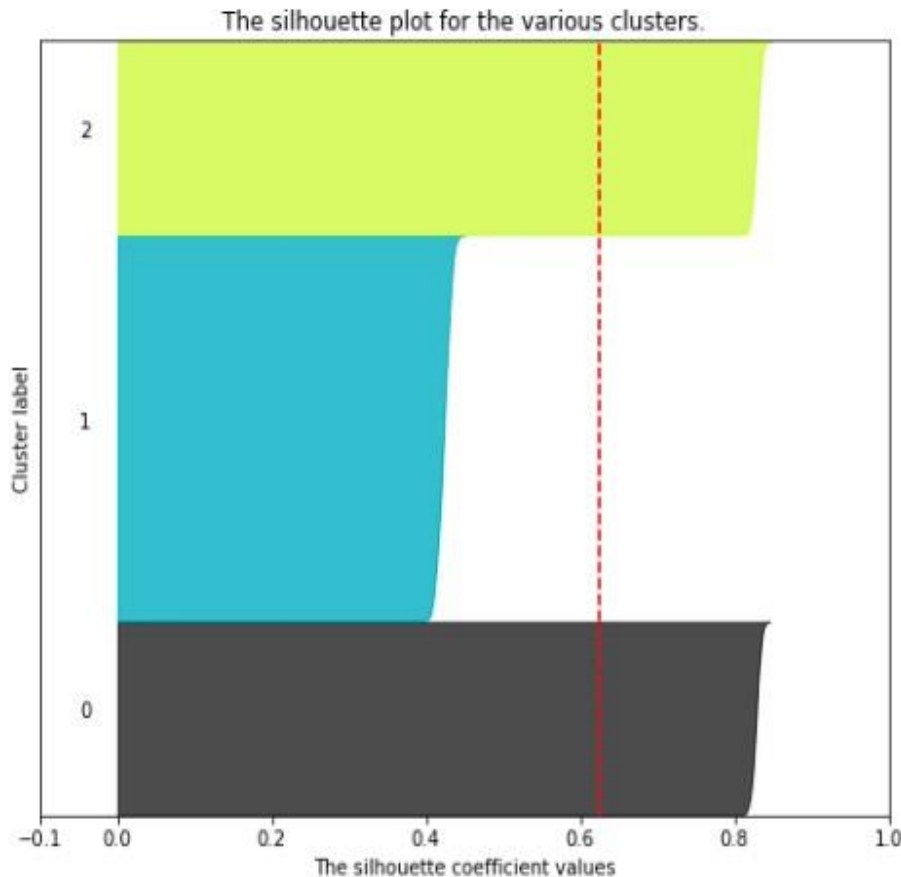
Methodology
Figure below showing Silhouette plot for n = 4 i.e. number of Clusters = 4

Silhouette analysis for KMeans clustering on sample data with n_clusters = 4
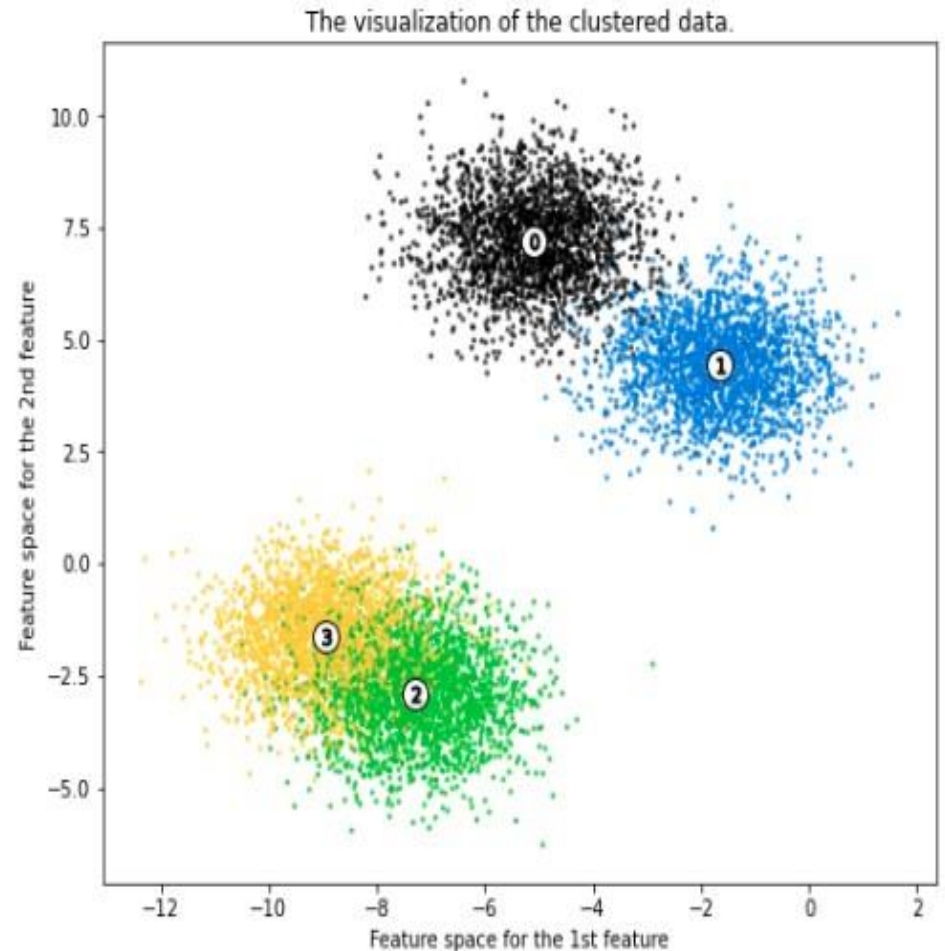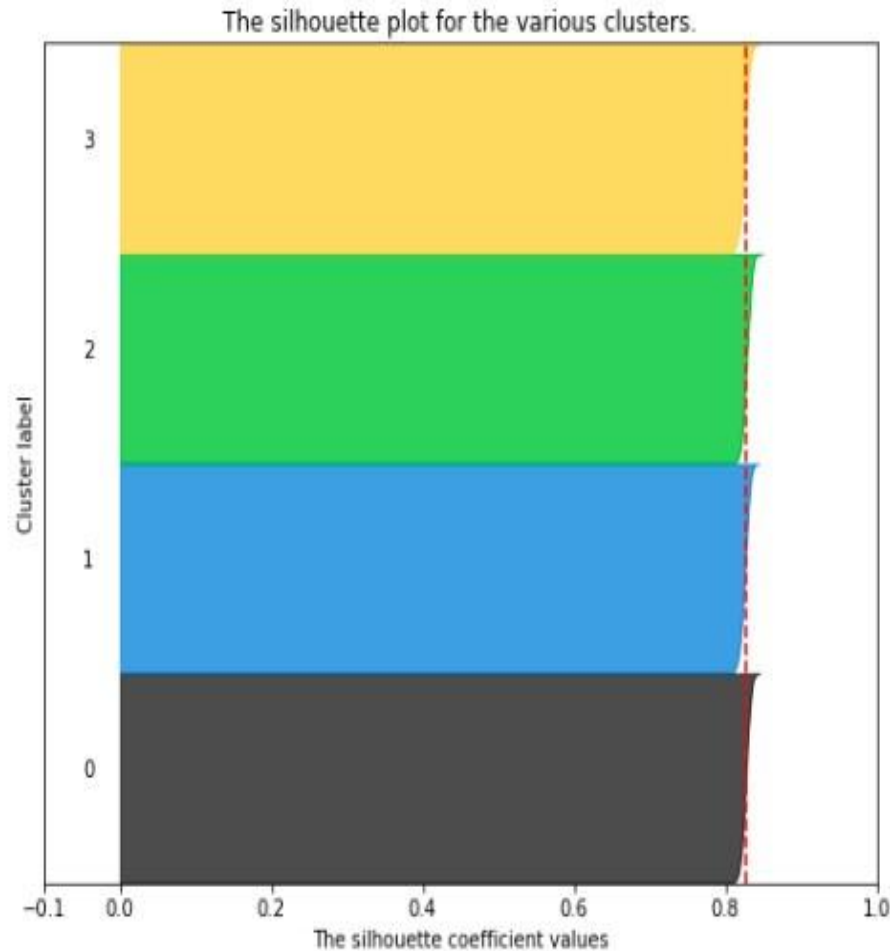
Methodology

Figure below showing Silhouette plot for n = 5 i.e. number of Clusters = 5

Silhouette analysis for KMeans clustering on sample data with n_clusters = 5
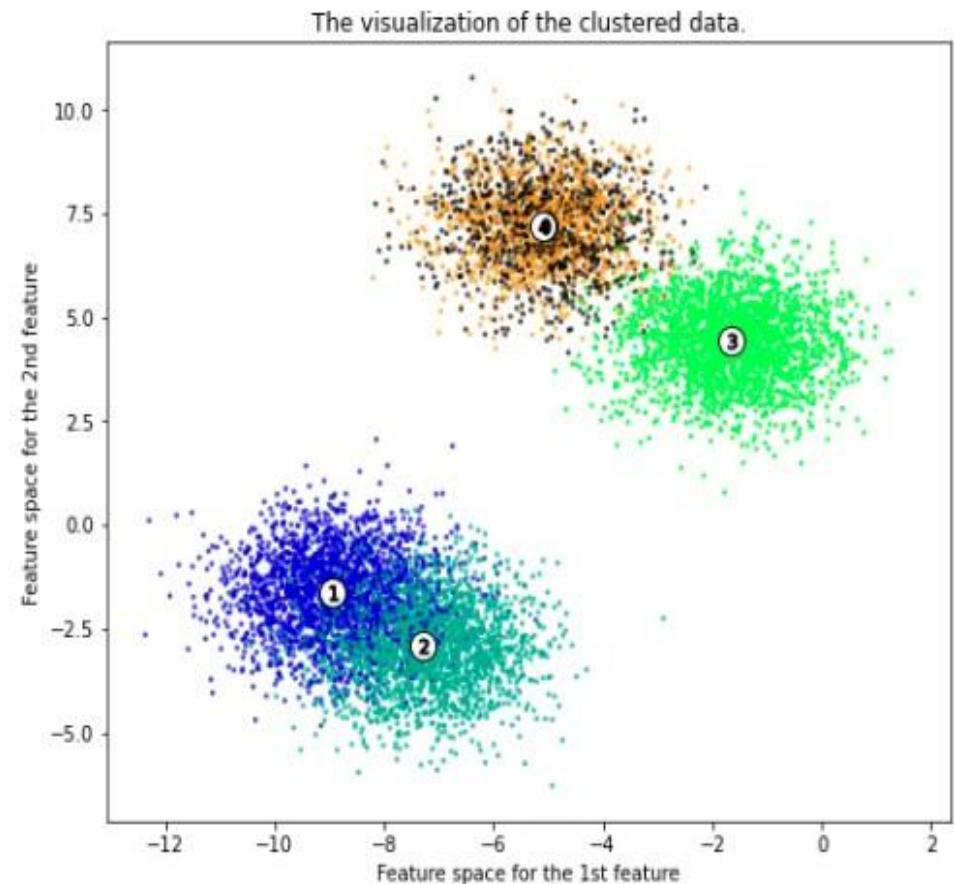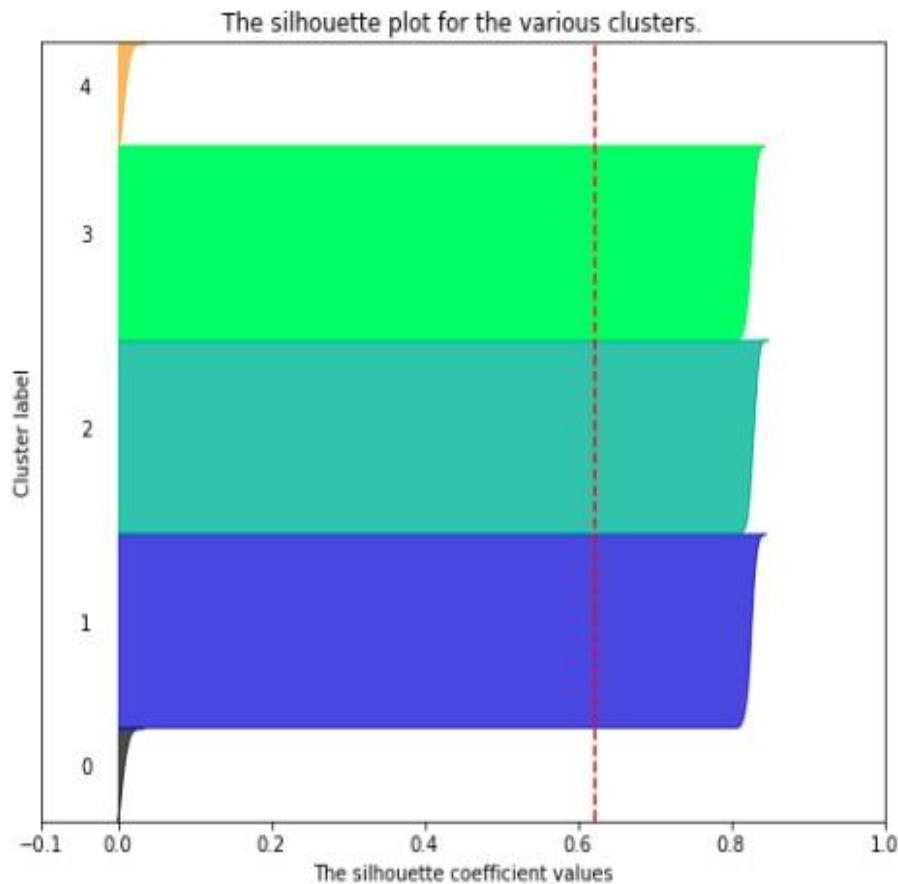
Methodology

Figure below showing the Silhouette plot for n = 6 i.e. number of clusters = 6

Silhouette analysis for KMeans clustering on sample data with n_clusters = 6

Methodology

As can be seen from the above plots, distributing cities into four clusters gives optimum results with Silhouette score value > 0.8 & also the thickness of each cluster label is more or less the same.

Finally, creating our Uttar Pradesh map with clustered data (see figure in the next slide)

Clustered Map of Uttar Pradesh

Result & Observations

Cluster 1 (red dots), Cluster 2 (purple dots), Cluster 3(sky-blue dots) & Cluster 4 (Yellow dots).

Cluster 1 & Cluster 2 are amenities rich as they fulfil almost all the day-to-day life needs for people of business class. Cluster 4 consists of only four cities & is good for students & service class persons. Cluster 3 is ideal for families who want to live in some locality or gated society.

Future directions

This project can be continued for making it more precise in terms to find best residential location in every city of Uttar Pradesh. Best means on the basis of all required things(daily needs or things we need to live a better life around and also in terms of cost effective. But there is not enough data about schools and hospitals in Foursquare API for the location Uttar Pradesh. So thinking to use google maps API for getting the best results