Predicting a suitable residential location for new immigrants in Uttar Pradesh

Introduction:

Uttar Pradesh is a state in northern India. With roughly 200 million inhabitants, it is the most populous state in India as well as the most populous country subdivision in the world. It was created on 1 April 1937 as the United Provinces of Agra and Oudh during British rule, and was renamed *Uttar Pradesh* in 1950, giving it the acronym *UP*. The state is divided into 18 divisions and 75 districts, with the capital being Lucknow.

The state is bordered by Rajasthan to the west, Haryana, Himachal Pradesh and Delhi to the northwest, Uttarakhand and an international border with Nepal to the north, Bihar to the east, Madhya Pradesh to the south, and touches the states of Jharkhand and Chhattisgarh to the southeast. It covers 240,928 km2 (93,023 sq mi), equal to 7.34% of the total area of India, and is the fourth-largest Indian state by area.

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.

Interest:

- New immigrants, who are ready to relocate from other states or country for the purpose of job, business or education.
- Families, who are looking for a residential location based on budget & amenities.

Data acquisition & cleaning:

For getting latitude & longitude values of different cities in Uttar Pradesh, I scraped the data from https://www.indianmirror.com/india-post/indianpincode.html website. This website consists of latitude, longitude, pin-code, population, area & population density of various cities/districts of different states of India.

Using requests & Beautiful Soup libraries I scraped the data of Uttar Pradesh. Regarding feature selection, I selected 2 columns as 'Districts', 'Latitude & Longitude'. I split the Latitude & Longitude column into 2 columns namely 'Latitude' & 'Longitude'. Then, I converted the latitude & longitude values into geological coordinates. The data-frame has 3 columns & 70 rows.

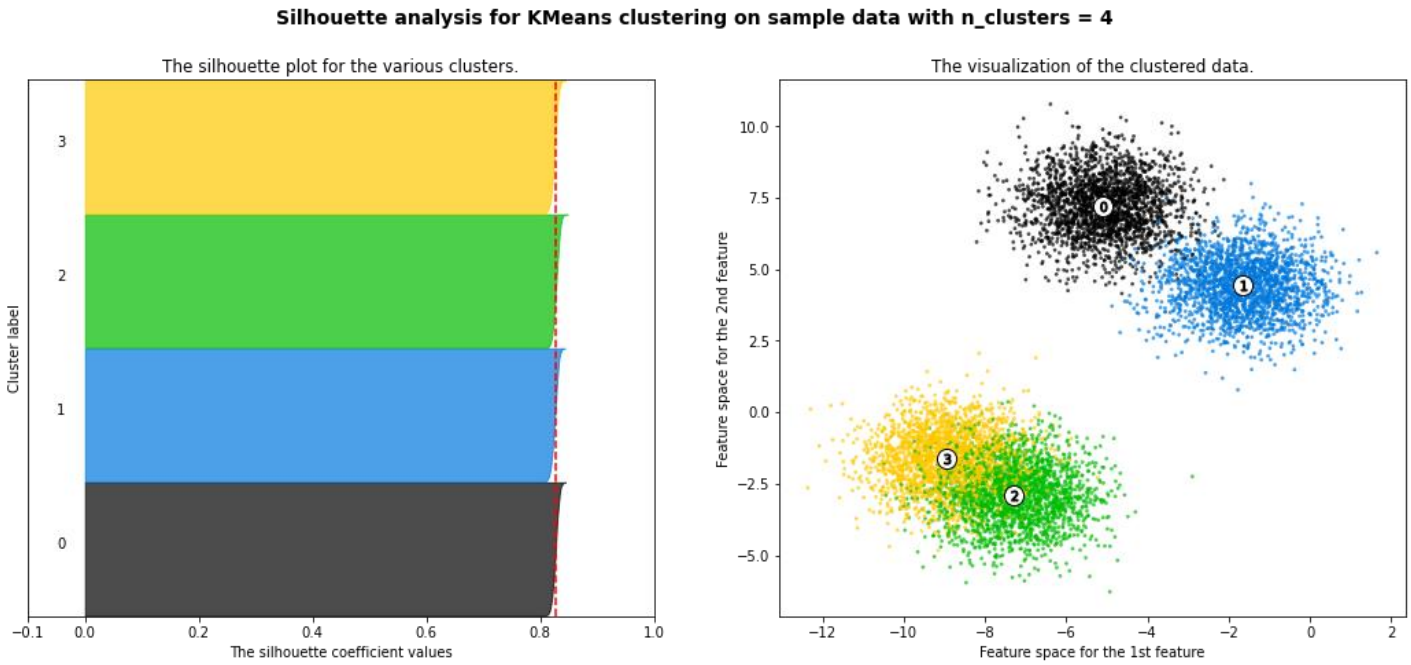
Methodology:

I used Geopy library to get the geological coordinates of Uttar Pradesh. Then, using Foursquare API & setting the LIMIT = 100 points of interest as well as a radius of 40 km from each city center, I extracted the venue name, venue latitude, venue longitude & venue category for

each city in Uttar Pradesh. I found that there are 123 unique venue categories in Uttar Pradesh.

I decided to use clustering approach to distribute the cities into clusters based on amenities & financial status (budget). For that, I assigned dummy variable to each venue category & then by counting the mean of the frequency of occurrence of each venue category in every city, I got top 10 most common venues for each city in Uttar Pradesh.

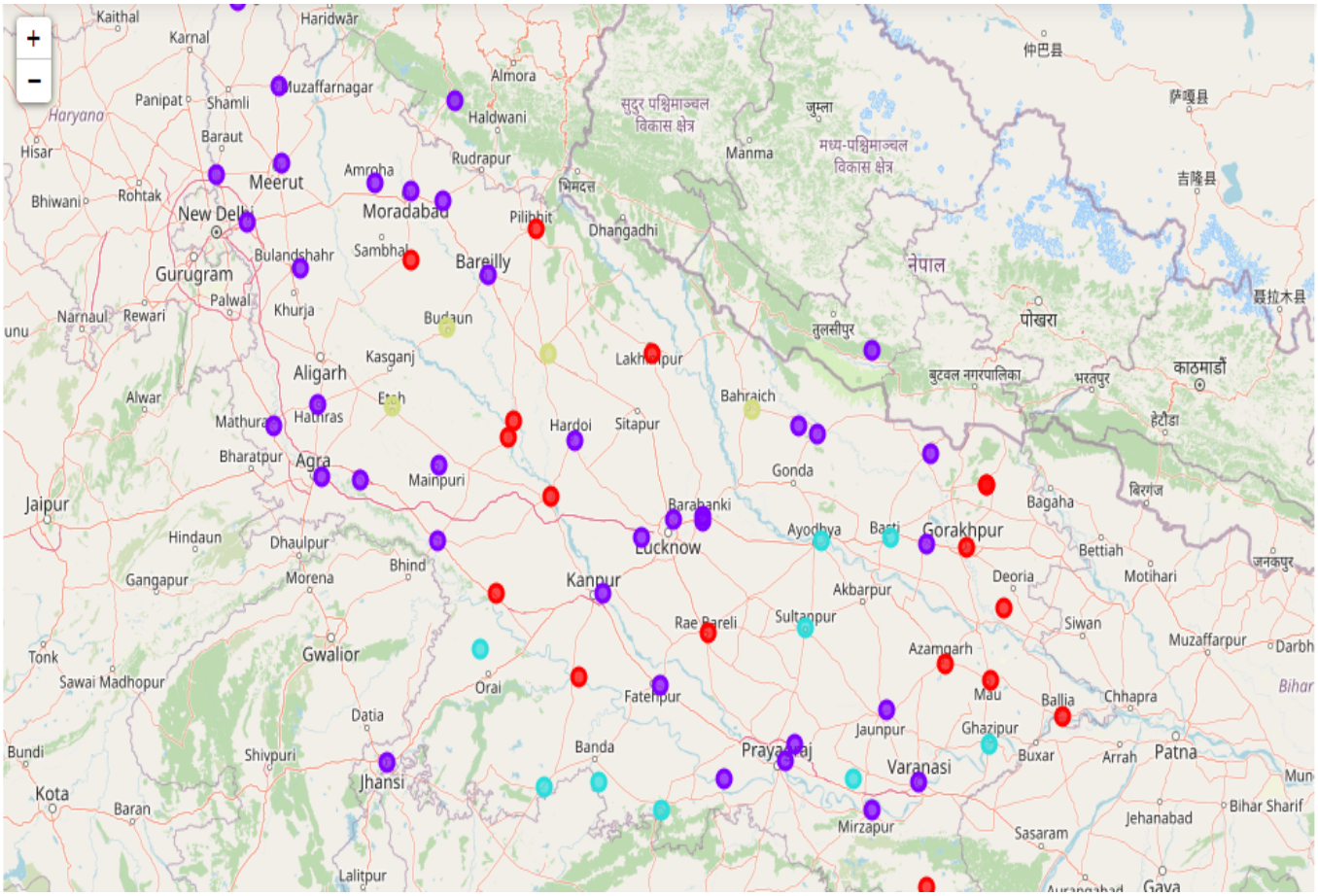
To find the optimum number of clusters into which the cities can be distributed, I used elbow method. But, considering the data & for getting more accuracy, I used Silhouette score.



Considering the thickness of each label & Silhouette coefficient of > 0.8 in the figure above, k-value = 4 seems to be optimum i.e., we can distribute the cities into 4 clusters based on amenities.

The figure below shows clustered map of Uttar Pradesh.

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



Results & Discussions:

Uttar Pradesh is a popular destination for new immigrants in north India to reside. Being home to various religious groups and places of worship. Cluster 1 & Cluster 2 are amenities rich as they fulfil almost all day-to-day life needs of people of business class. Cluster 4 consists of only 4 cities & is good for students & service class persons. Cluster 3 is ideal for families who want to live in some localities or in gated society.

Conclusions & Future directions:

The major purpose of this project, is to suggest a better neighbourhood/place in a new city for the person who are shifting there. Social presence in society in terms of like-minded people. Connectivity to the airport, railway station, city center, markets and other daily needs

things nearby. In this project, using k-means cluster algorithm I separated the neighbourhood into 4(four) different clusters and for 70 different latitude and longitude from dataset, which have very-similar neighbourhoods around them. Using the figure above results presented to a particular neighbourhood based on amenities have been made.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.

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 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.