Predicting Location in Delhi for Mall

Satyendra

Applied Data Science Capstone Project for

IBM Data Science Professional Certificate

Problem

- **Finding a perfect location in the city of Delhi for opening a shopping mall.**
- **Decision making on right choices keeping some factors in mind.**
- **Extracting useful insights using dataset available.**

Data Required

- To solve the problem, we will need the following data:
- List of neighbourhoods in New Delhi. This will elaborate the scope of data.
- For plotting the locations we need longitudes and latitudes.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.
- Sources of data and methods to extract them .
- This Wikipedia page (https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi) contains a list of neighbourhoods in New Delhi.

Methodology

- **Web scraping-getting data from wikipedia.**

- **Geocoder - getting longitudes and latitudes.**

- **Foursquare Api- exploration of data.**

- **Performing k means clustering.**

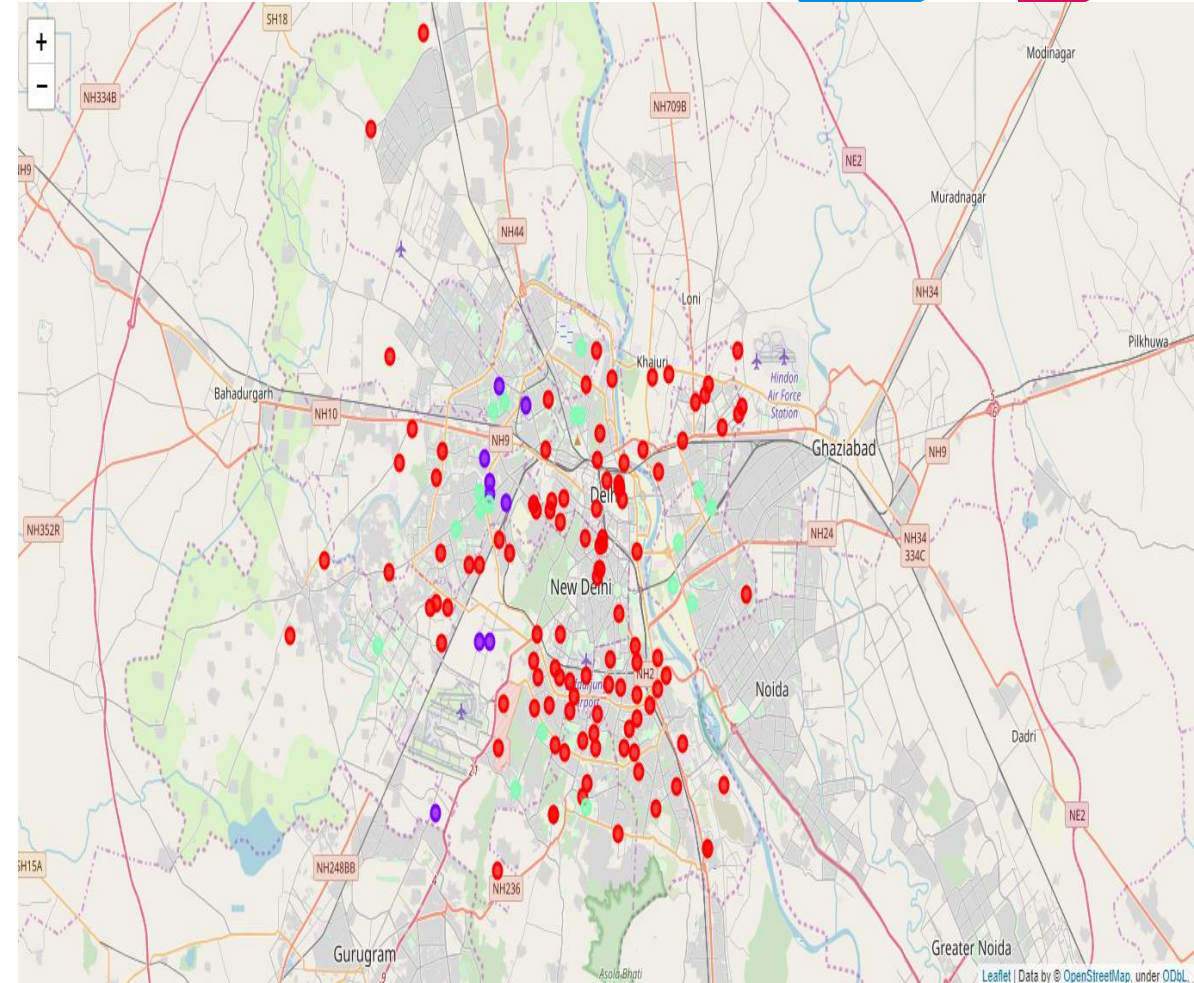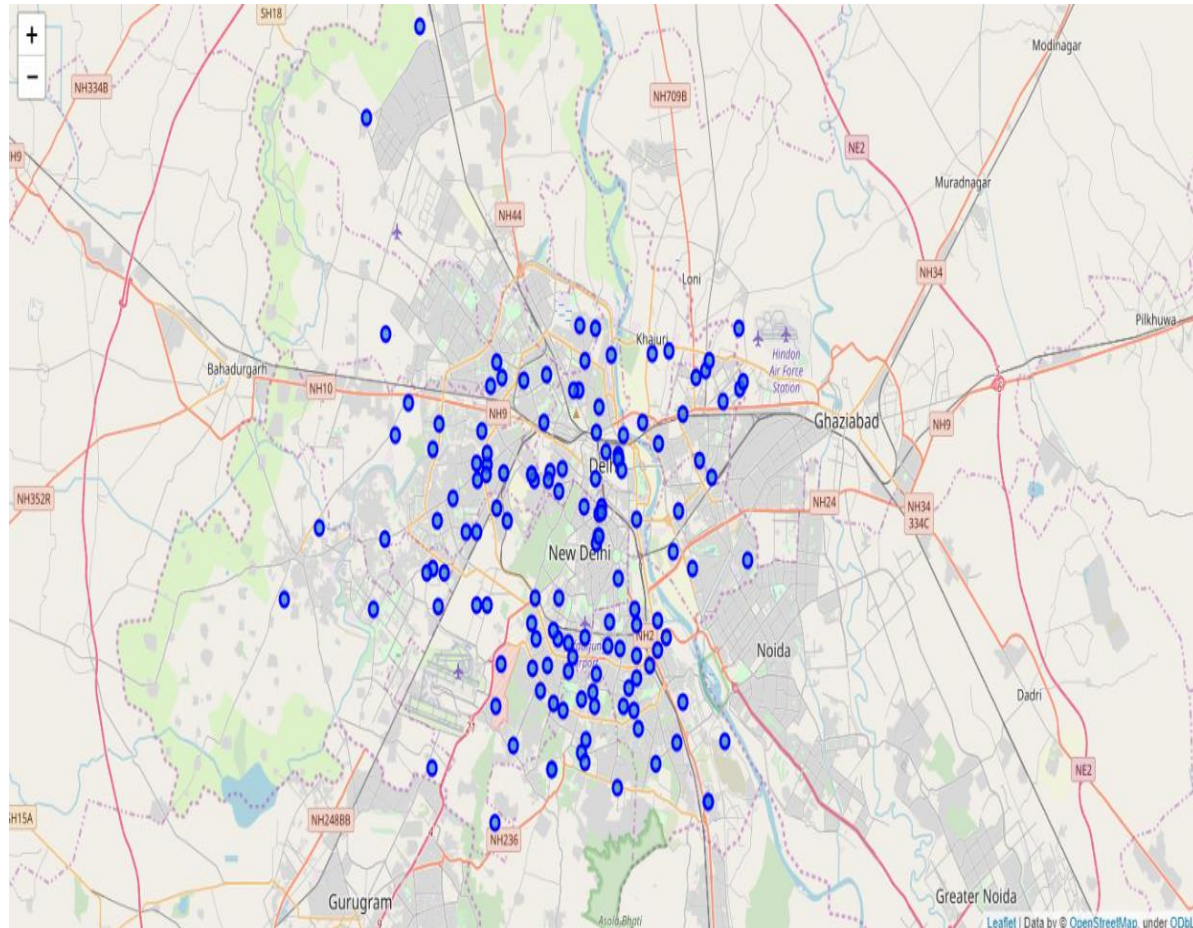- **Visualising using folium library of python.**

Result

The data will be divide in three clusters which represents different intensities of the shopping malls situated in the area

- Cluster 0: Moderate density of the shopping complex It is represented by red colour.
- Cluster 1: Low density of shopping complexes. It is represented by purple colour.
- Cluster 2:High density of shopping complexes. It is represented by mint green colour.

Area visualization with clusters

Discussion

We can deduce the fact from the map that the most of the malls are located in south Delhi and central Delhi. These areas have moderate concentration but localities with moderate concentration is quite high.

- Cluster 0:South and Central Delhi neighbourhoods are mostly in cluster 0. West and east Delhi have less number of mall concentration in cluster0.

- Cluster 1: Neighbourhoods are mostly located in western and Southern part of Delhi. East Delhi have very few neighbourhoods in cluster 1. It is also noticeable that North Delhi do not have areas in cluster 1.

- Cluster 2: West Delhi is the only region where the most no of cluster 2 neighbourhoods are present.

Recommendation

South-West and part of the capital has very less frequency of malls and investors can invest after judging some local factors. Also investors must take the note that they should not try to invest near southern part as there is very high concentration of malls in these areas. Some part of western areas very highly concentrated with shopping malls some regions in central part are also less concentrated.

Conclusions

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. 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 shopping mall.