Opening a new shopping mall in Delhi, India

Introduction:

The purpose of this project is to find the best place for shopping mall in Delhi, India. It will help property developer to make a smart and efficient decision on choosing the area in Delhi.

This project aims to find the best place by initially finding the location of the neighbourhoods and then finding the venue using the API's and then doing the analysis of shopping malls in each neighbourhood.

Data Description:

Data is collected through web scraping.

Wikipedia : https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi

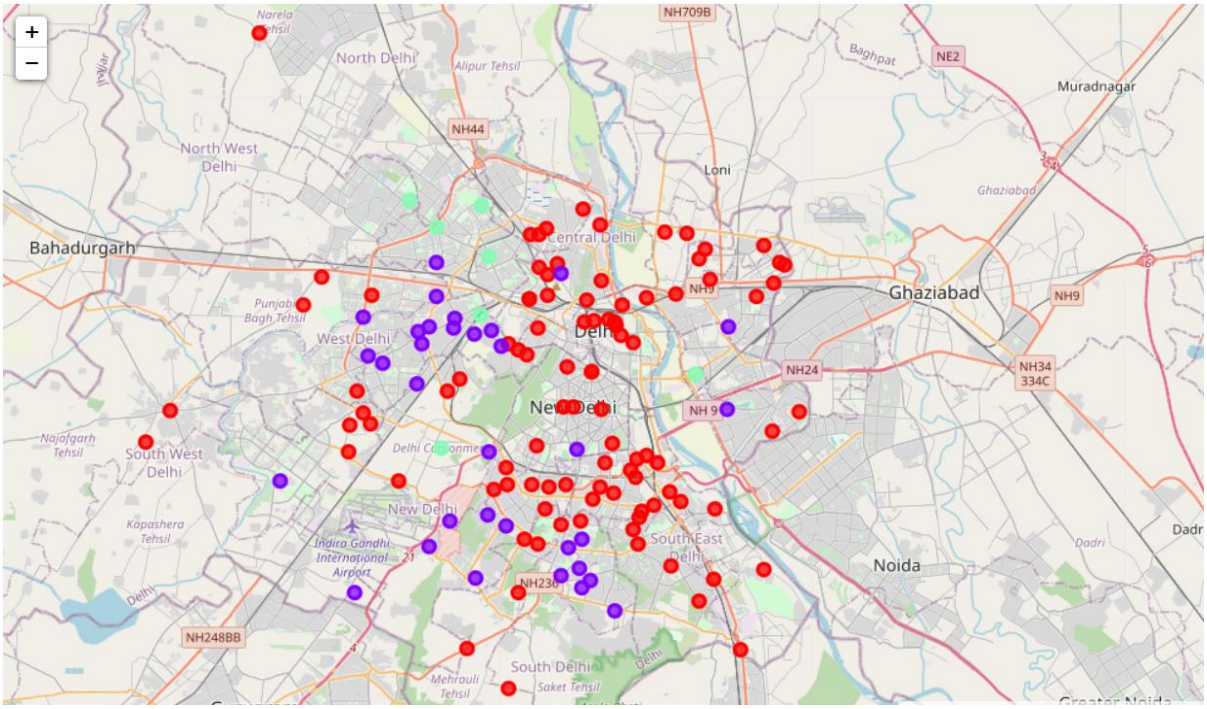
The location of each neighbourhood is determined by using geopy library in python.

Using Foursquare API the data of venue around 2.5 km of each neighbourhood is collected.

Methodology:

1. Web Scraping: Lists of neighbourhood is collected through web scraping.
2. Determining the location of each neighbourhood to find the latitude and longitude with the help of geopy library.
3. Determining the venue information which is around 2.5 km radius of each neighbourhood with the help of foursquare API
4. Grouping the venues categories by neighbourhood and taking the mean by the frequency of neighbourhood.
5. Filtering it to only shopping mall.
6. Clustering the data using K means.

Result:



Using the K means clustering and the plotting the neighbourhoods in the map we found:

- Cluster 0 which is red mark in the map has very less number of shopping mall around 2.5 km of radius
- Cluster 1 which is purple mark in the map has moderate number of shopping mall
- Cluster 2 which is light green mark has higher number of shopping mall

Recommendation:

- It can be seen that cluster 2 is not recommended due to intense competition.
- Cluster 1 is recommended only if shopping mall teams are quite confident to sell products or have unique products to sell
- Cluster 0 is highly recommended as there is almost no competition.

Future Works:

This project can be continued by making it better on deciding place by not only seeing the no. of shopping malls but also the price of the land on that neighbourhood , different venues which will make decision much better and efficient.

Libraries Used:

Pandas: Best library for handling the dataframe.

BeautifulSoup: Best library for web scrapping in python.

Matplotlib: plotting and visualizing the data in graph.

Folium: For visualizing the data on the map.

Geopy: To determine the location of the places.

Sklearn: For using k means clustering algorithm.

Json: to handle json files.