**Introduction:**

In this project we will try to ﬁnd an optimal location for a zostel. Speciﬁcally, this report will be targeted to stakeholders interested in opening an zostel in DELHI, India

### Zostel is a place where backpackers can come and spend their time with other tourists and gain knowledge about cultures, different types of food and how to make friendship. This business was started with a vision to promote traveling as a way of life.

### Since the stakeholder wants to provide the immersive experience to the travelers/tourists. We want to get the cluster information about the delhi, based on the surrounding available environment so that we can analyze the cluster. Secondly,

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed and get the cluster character, so that best possible ﬁnal location and zostel type can be chosen by stakeholders.

**Data Acquisition And Cleaning**

### Based on deﬁnition of our problem, factors that will inﬂuence our decision are:

* number of existing Borough in the delhi region
* number of existing Neighborhood in each Borough in the delhi region
* Latitude and longitude of each neighborhood

Get the detailed information in each circle by using Foursquare API

Besides we use other packages, likes:

Pandas Numpy Json Geopy Matplotli

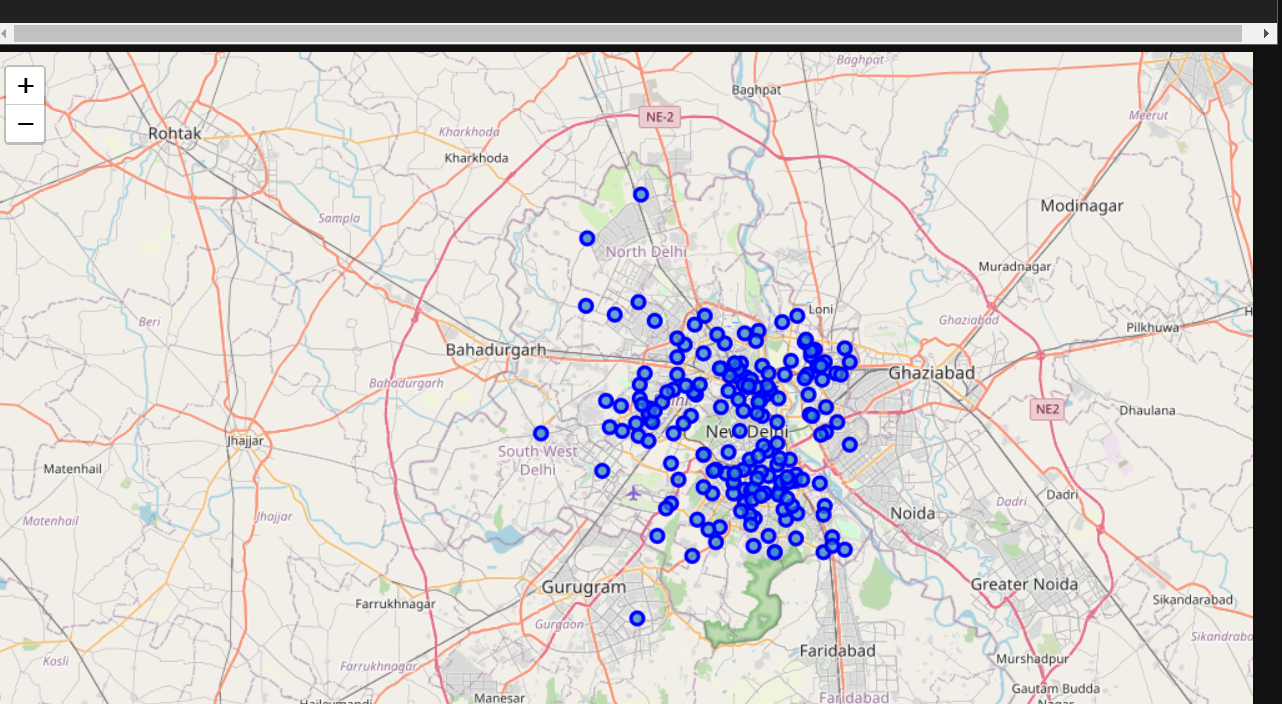
Sklearn Folium Requests

We get the basic information like:

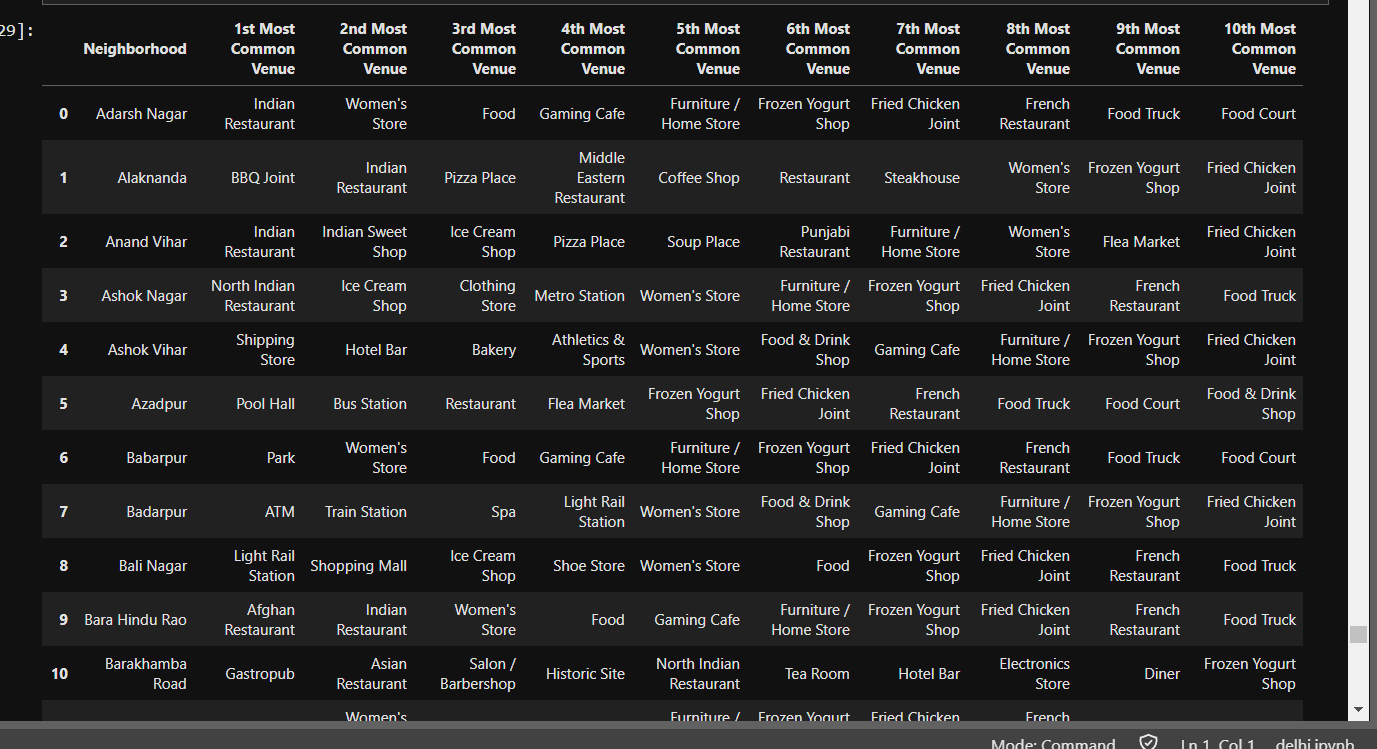


**Methodology**

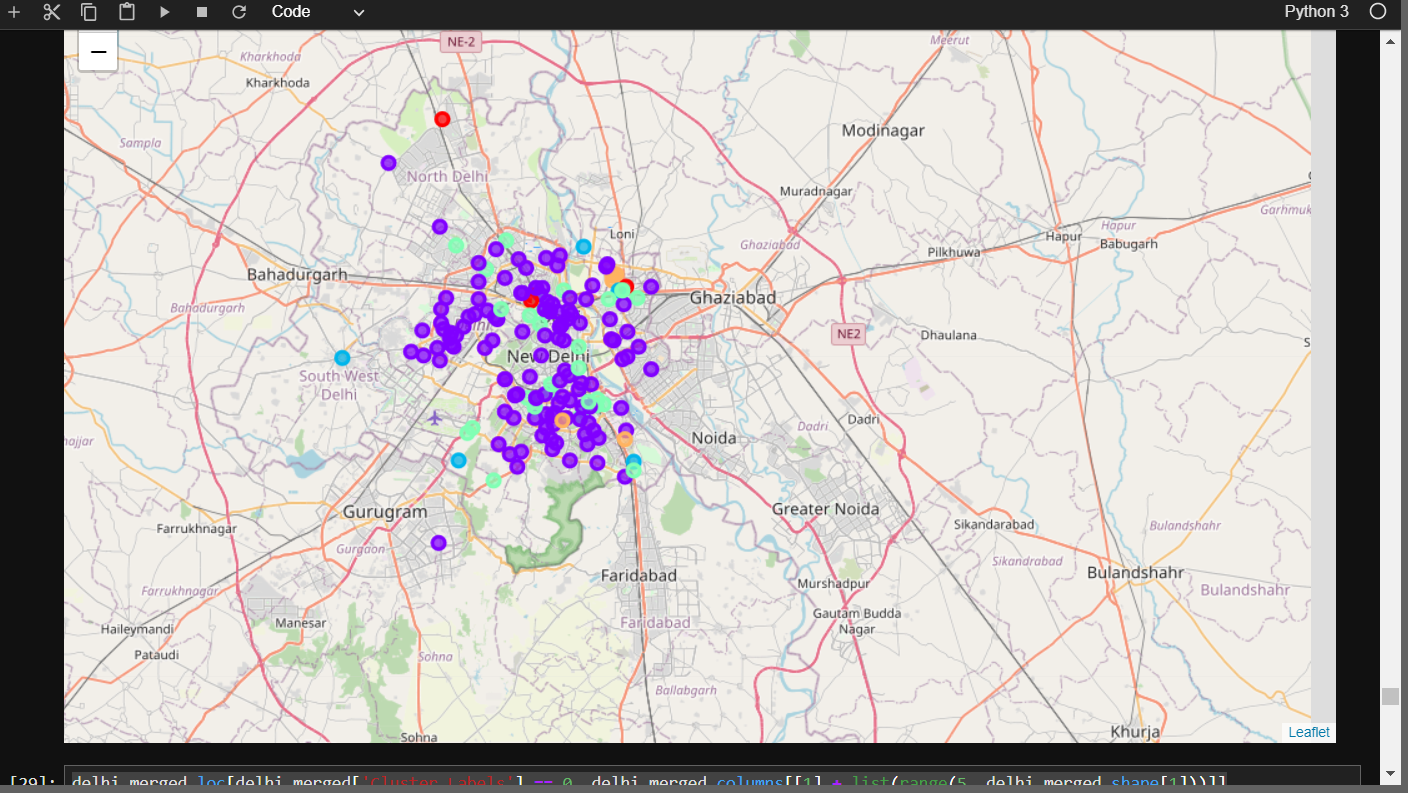
We create the data ﬂow to get the data by using Foursquare API, then extract information about the surrounding , we can visualize the distribution of the surrounding



Then create the new dataframe and display the top 10 venues for each neighborhood.



Finaly, we use the KMeans method.



**Results and Discussion**

In search of new place for zostel, explored different categories of venues. In this analysis, initially, a sum of 9 unique borough has been observed. Using geocoder of Geopy library for delhi, INDIA,

we split the data into longitude and latitude of the neighborhood. analysis of each neighborhood is carried out with the help of one-hot encoding. The result of one-hot encoding identified 173 types of places. Followed by one-hot encoding we explore the top 10 common venues in each of the neighborhoods.

After one hot encoding and venue exploration, we applied the k-means algorithm for k=5. It has identified 5 cluster centroids c0, c1, c2, c3, and c4. These five clusters identified in five different colors. Cluster 0 is having only four neighborhood, cluster 1 is having the maximum number of neighborhoods.

## Cluster 1:

## C:\Users\mkgtc\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Screenshot (112).png

## Cluster 2:

## C:\Users\mkgtc\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Screenshot (114).png

## Cluster 3:

## C:\Users\mkgtc\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Screenshot (115).png

## Cluster 4:



## Cluster 5:



**Conclusion**

The purpose of this project was to identify areas for zostel close to the center with a high number of different venues to aid stakeholders in narrowing down the search for the optimal location for a new zostel. By calculating different categories of venue distribution from Foursquare data we have first identified general boroughs that justify further analysis that the cluster number 2 has the highest number of different categories.

Final decisions on optimal zostel location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise/proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood.