

BATTLE OF NEIGHBORHOODS

Introduction

The world is big, but having multiple ways of transportation made it small. Today everyone is looking for new opportunities and to grab these one need to move out of the town and sometime to other cities. This is a problem for many, as they are not aware of the new city, and feels lost in the new city. One of the main problem is finding a place to stay. Most of the time people want to have same amenities near their place which they used to have at their old place, and rest want only some of the important facilities near their place. But both being equally difficult. To conquer this problem one need to compare the two cities and the place they used to live before and then make a choice whether or not the new place is suitable for them. TO compare the cities we'll be requiring list of venues near the Neighbourhood and to get done with this the foursquare API will be useful. And finally choosing the place that fits in criteria and best out of rest. Solving this problem will help the people living in the different cities but planning to move to other, or the person living in the same city is planning to move somewhere else within the same with same facilities. Using this will ease them to choose and pick their future place.

Data

In order to achieve this first we need to gather that data. Firstly we need to get the data regarding the location and name of the city we used to live previously. The data for the new city we are planning to move. In this case Delhi is considered to be current city whereas Mumbai will be the future city. After collecting the geospatial data regarding the two cities. We need to find out the places or venues near a place, and foursquare API will be taking care of this. Now we have the location data for the neighbourhood and the places near them. So now it'll will become easy to compare the two cities.

The Name of neighbour of Delhi and Mumbai will be taken from Wikipedia this will be in the html format, so we need to do some web scrapping and take out the required data from it and leave the rest of the data.

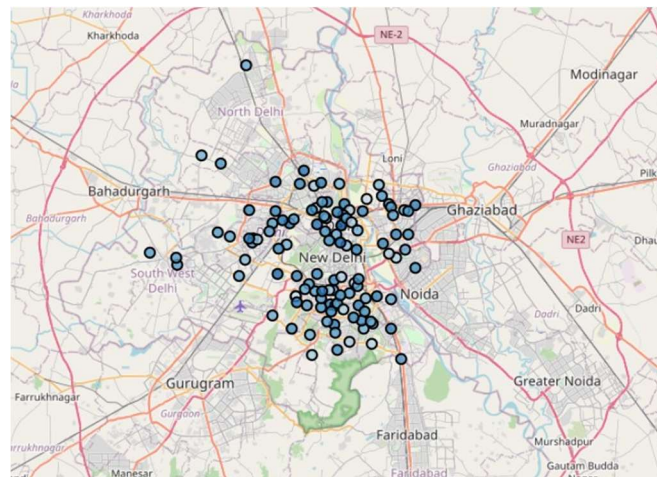
Location data is taken from OpenStreetMap Nominatim. This is require in order to use foursquare API because it needs the location of the place that is latitude and longitude to find the places nearby

And finally the list of venues near the neighbourhood is taken from Foursquare API. Using this we will get all the venues near that location with some other information like the category, location, tips etc. we will take the name of the venues, categories of venues and location of them

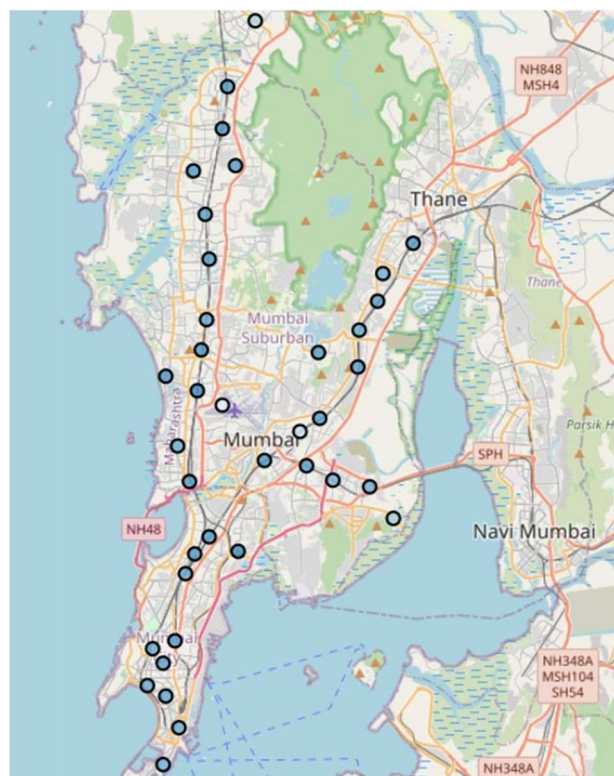
Methodology

Explanatory data Analysis

Started with the creation of the map on based of the importance of the location, This will help us to find whether the data received is correct and all the places is in desired region and also to understand the data. Like in Delhi data frame it can be clearly seen that the places in central Delhi and some places of south Delhi have more importance than others. Similar case for Mumbai data frame.

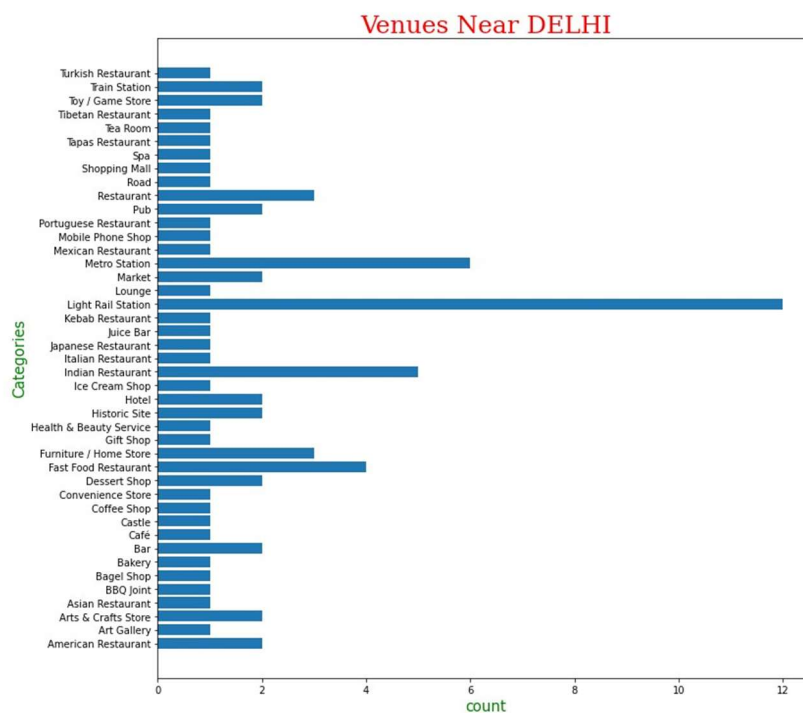


Map 1: .Delhi Neighbourhood

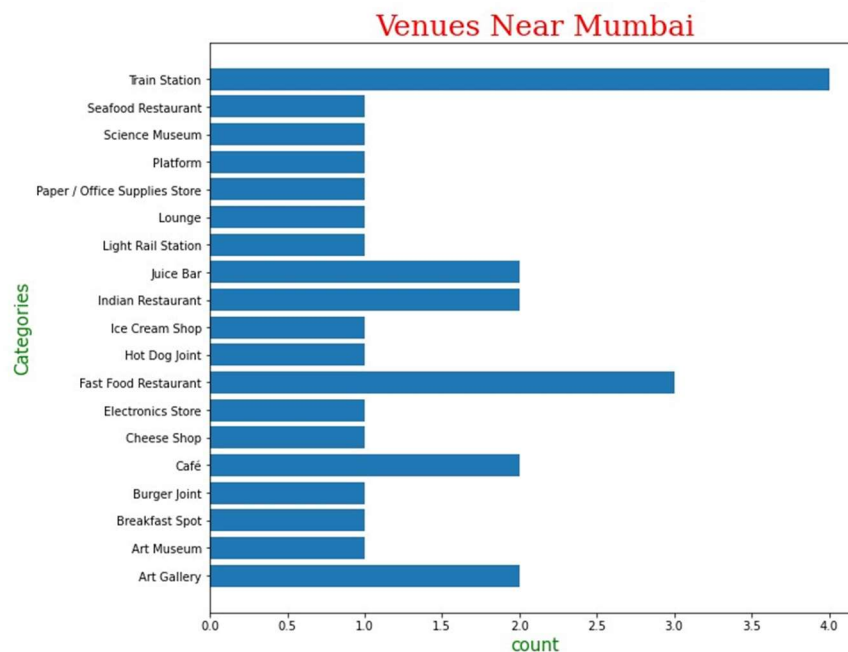


Map 2: Mumbai Neighbourhood

The received from Foursquare API need to be visualised now , in order to do this fir we grouped the data based on the category of the venue and visualized it in form of vertical bar chart, this will help us to see the dominance of the similar kind of venues in both the cities.



Graph 1: types of venues near neighbourhood of Delhi



Graph 1: types of venues near neighbourhood of Mumbai

To view the most visited place in cities we hot encoded the data based on the categories of the venues. This will help us to find the most visited venue if the neighbourhood have two of more venues nearby.

	Neighborhood	Most Visited place
0	Adarsh Nagar	Light Rail Station
1	Badarpur	Train Station
2	Chandni Chowk	Historic Site
3	Chattarpur	Furniture / Home Store
4	Civil Lines	Italian Restaurant

Data frame head 1: Top visited place in neighbourhoods of Delhi

	Neighborhood	Most Visited place
0	Andheri	Fast Food Restaurant
1	Bandra	Train Station
2	Byculla	Art Gallery
3	Colaba	Art Gallery
4	Dadar	Breakfast Spot

Data frame head 1: Top visited place in neighbourhoods of Mumbai

Data cleaning

When plotting points on the map it can be clearly viewed that there are some neighbourhood that have NULL columns, we can remove that rows or let it to be handled by exception handling which is being used to retrieve the data from Foursquare API

Similarly the case for data we get from foursquare API have no nearby venues for some of the places which is handled by exception handling also.

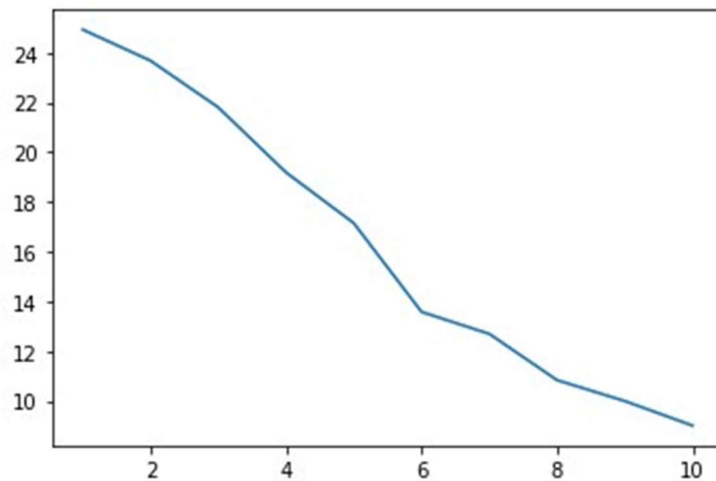
At last to train the data we have uneven number of columns for the cities. To make it equal we manually imputed some of the columns in both of the data frame for it can be used in training and prediction.

```
mumbai_needed_columns=list( set(delhi_grouped)-set(mumbai_grouped))
delhi_needed_columns= list(set(mumbai_grouped)-set(delhi_grouped))
mumbai_grouped[mumbai_needed_columns]=0.0
delhi_grouped[delhi_needed_columns]=0.0
```

Code 1: Data cleaning (imputation of columns)

Modelling

For the modelling of the data we use clustering algorithm called KMEANS. This algorithm forms cluster on the basis of K – neighbours. Choosing a K for the model is a crucial part and to do this we use the concept of ELBOW POINT.



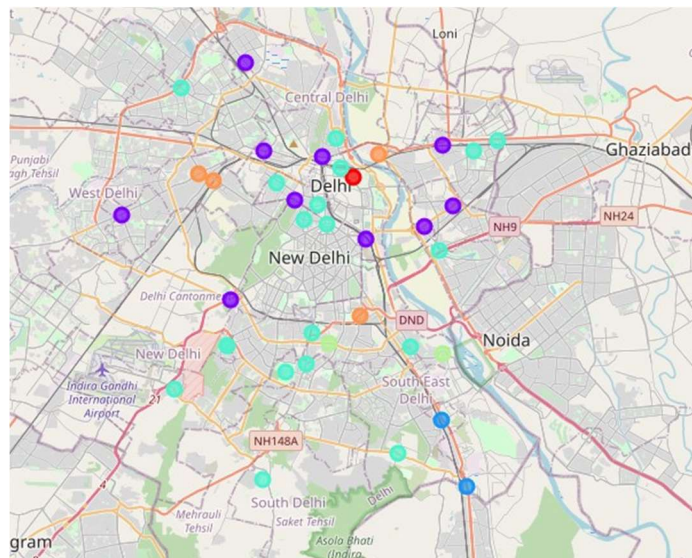
Graph 3: Selecting K (Elbow Point Method)

Increasing the K will always decrease the error but sudden decrease or increase which creates a knee of the curve. That point is taken as value of K. KMEANS is chosen because it guarantees convergence. And can easily adapt to new examples

Results

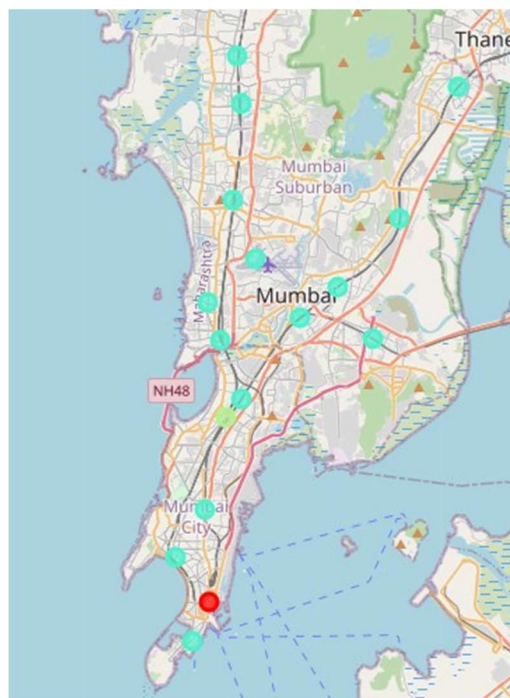
The Labels are provided to the data and it is separated into different clusters. We can map it with different colours according to the clusters they belong to.

The clusters are formed for Historic sites, Light Rail Station, Train Station, Restraints and different stored and services, Metro Stations, Art & Craft Stores.



Map 3: Clustered Data Delhi

When the model used for prediction on Mumbai data it classify it in clusters for seafood restaurants. Restaurants and different stored and services, Breakfast Spot.



Map 4: Clustered Data Mumbai

Discussion

While training and prediction it is observed that a cluster have multiple categories of venues in same cluster, it can be due to the technique we used while data cleaning. Performing Dimensional Reductionality may solve this issue

Conclusion

In search for better life one may be ready to move at different place. And doing this is not an easy task, and getting same facilities as one used to get at earlier place is not easy, however from this project it can clearly be seen that one can easily get a place near any Restaurants and can while they also have to sacrifice other facilities that they used to get. Choosing a perfect place is a trade-off. One may or may not find a place like before.