**A. Introduction**
A.1. Description of problem Neighborhoods of Bangalore & Mumbai:

Mumbai and Bangalore are the major employment centers in India. Mumbai with a population of 280 Million (Urban conglomerate) is touted as commercial capital of India for ages. Bangalore on the other hand is coming of age on its own as the silicon valley of India and recently as start up capital of India.

Both cities have their advantages and disadvantages. With this project we will try and sort out the neighbourhoods in both cities with best of amenities for social life of a young, urban employee to enable a prospective employee zero in on a suitable based on select parameters.
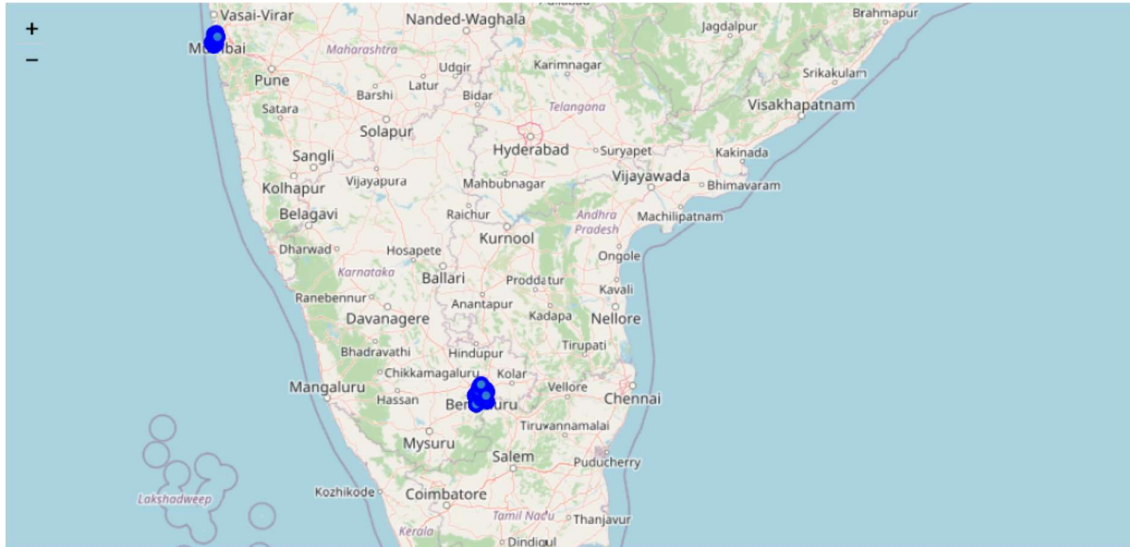
A.2. Data Description
Data is gathered from following sources;
- We have relied on Pincode and geometrical coordinates available on ttp://www.geonames.org/export/zip/. We download the file for India. The file is uploaded to lab folder for easy access..
- **Use Foursquare API** to get the most common venues of given neighbourhoods in these two cities.

B. 1 Methodology
A total of 258 pincodes for the city of Bangalore and 146 Pincodes for the city of Mumbai were found. We will wrangle this data and make it useful through steps outlined in the report.

B.1a Data for Bangalore

	Pincodes	Area	StateUT	Region	City	Lat	Long
0	560001	Bangalore G.P.O.	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
1	560001	Legislators Home	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
2	560001	Mahatma Gandhi Road	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
3	560001	Vidhana Soudha	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
4	560001	Rajbhavan (Bangalore)	Karnataka	Bangalore	Bangalore North	12.9914	77.5944
...	...	...	...	...	...	...	...
253	560110	Ullalu Upanagar	Karnataka	Bangalore	Bangalore	13.1791	77.0942
254	562130	Tavarekere (Bangalore)	Karnataka	Bangalore	Bangaloresouth	13.9288	77.0614

	Pincodes	Area	StateUT	Region	City	Lat	Long
255	562157	Bettahalsur	Karnataka	Bangalore	Bangalore North	13.1438	77.6169
256	562162	Madanayakanahalli	Karnataka	Bangalore	Bangalore North	17.8647	77.8229
257	562162	Dasanapura	Karnataka	Bangalore	Bangalore North	17.8647	77.8229

B.1b Data for Mumbai :

	Pincodes	Area	StateUT	Region	City	Lat	Long
0	400001	Mumbai G.P.O.	Maharashtra	Mumbai	Mumbai	18.6291	72.8919
1	400001	Bazargate	Maharashtra	Mumbai	Mumbai	18.6291	72.8919
2	400001	Town Hall (Mumbai)	Maharashtra	Mumbai	Mumbai	18.6291	72.8919
3	400001	Tajmahal	Maharashtra	Mumbai	Mumbai	18.6291	72.8919
4	400001	Stock Exchange	Maharashtra	Mumbai	Mumbai	18.6291	72.8919
...	...	...	...	...	...	...	...
141	400098	Vidyanagari	Maharashtra	Mumbai	Mumbai	19.2355	72.8468
142	400099	Sahar P & T Colony	Maharashtra	Mumbai	Mumbai	19.2355	72.8468
143	400099	International Airport	Maharashtra	Mumbai	Mumbai	19.2355	72.8468
144	400099	Sahargaon	Maharashtra	Mumbai	Mumbai	19.2355	72.8468
145	400099	Airport (Mumbai)	Maharashtra	Mumbai	Mumbai	19.2355	72.8468

These pincodes contained erroneous entries, wrong geocoding etc. The datasets were therefore cleaned for errors, outliers etc and added to each other for visual effects. We pulled out data for both cities separately for this process. This will also help us save the data for cities separately for future use.

B.2 Data Cleaning:
Scatted plots, box plots provide us a bird view of distribution of datapoints for the cities indicating problematic datapoints.

We dropped outliers using interquartile ranges. After cleaning the data, we ended up with 270 localities (88 of Mumbai and 182 of Bangalore).

B.3 Generating Venues data and processing:

I utilized the Foursquare API to explore the neighbourhoods and segment them. The limit was set at **1000 &** the radius was set at 500 **meter** for each neighbourhood from their given latitude and longitude. Here is a head list of Venues with name, category, latitude and longitude from Forsquare API.

B.3a Bangalore:

	Neighborhood	Pincode	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bangalore G.P.O.	560001	12.9914	77.5944	Millers 46	12.991666	77.594207	Steakhouse
1	Bangalore G.P.O.	560001	12.9914	77.5944	Ujwal Bar & Restaurant	12.992280	77.594473	Indian Restaurant
2	Bangalore G.P.O.	560001	12.9914	77.5944	Alliance Française	12.991232	77.596723	Concert Hall
3	Bangalore G.P.O.	560001	12.9914	77.5944	Imperial Restaurant	12.991150	77.593837	Indian Restaurant
4	Bangalore G.P.O.	560001	12.9914	77.5944	Infinitea	12.987157	77.594835	Tea Room

B.3b Mumbai:

	Neighborhood	Pincode	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Girgaon	400004	18.95	72.8167	Palms - Catholic Gymkhana	18.949760	72.820019	Restaurant
1	Girgaon	400004	18.95	72.8167	Bachelors Juice House	18.950088	72.819185	Juice Bar
2	Girgaon	400004	18.95	72.8167	Bachelorr's Ice Creams	18.954113	72.815396	Ice Cream Shop
3	Girgaon	400004	18.95	72.8167	Mafatlal Hindu Bath & Boating Club	18.953616	72.815366	Harbor / Marina
4	Girgaon	400004	18.95	72.8167	Gola's	18.949307	72.819453	Food Truck

We also did a commonality test to figure out if there are common categories of venues that can be found in both cities. We plotted a Venn chart to reflect the output.

We have some common venue categories in boroughs. We use unsupervised learning **K-means algorithm** to cluster the neighborhoods based on the venues present in them.
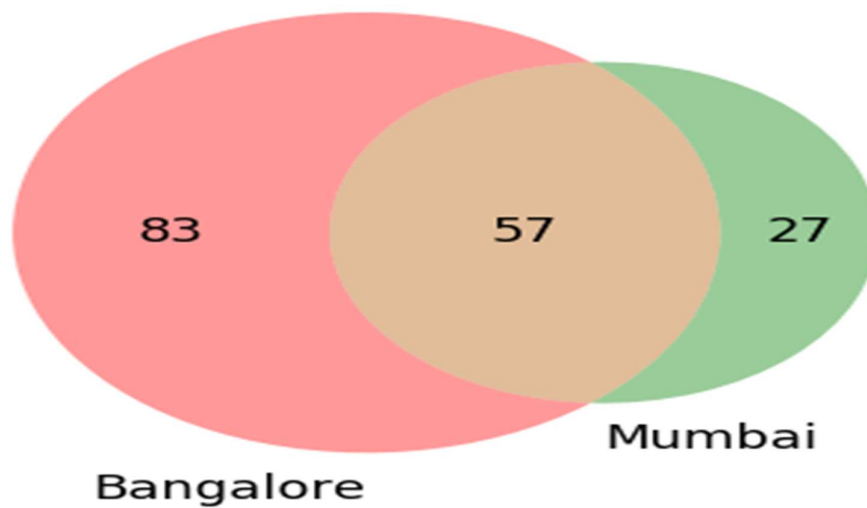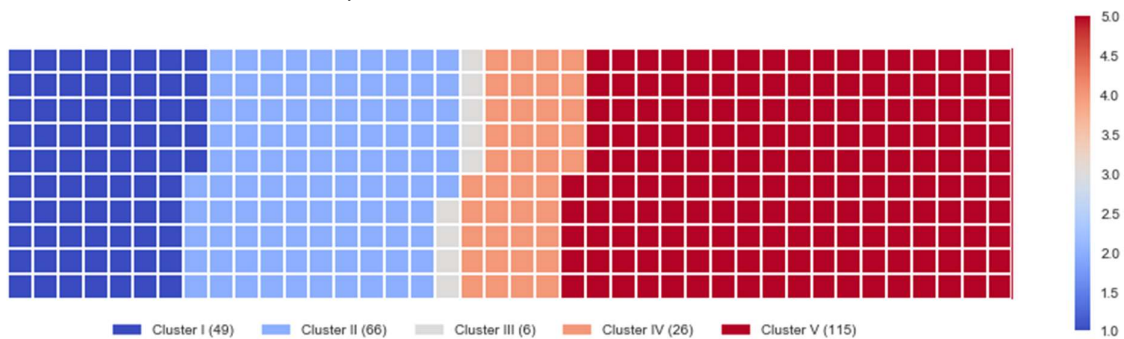
I used K-Means to cluster the neighborhoods into 5 clusters based on the chart generated below on consolidated dataframe (elbow method).
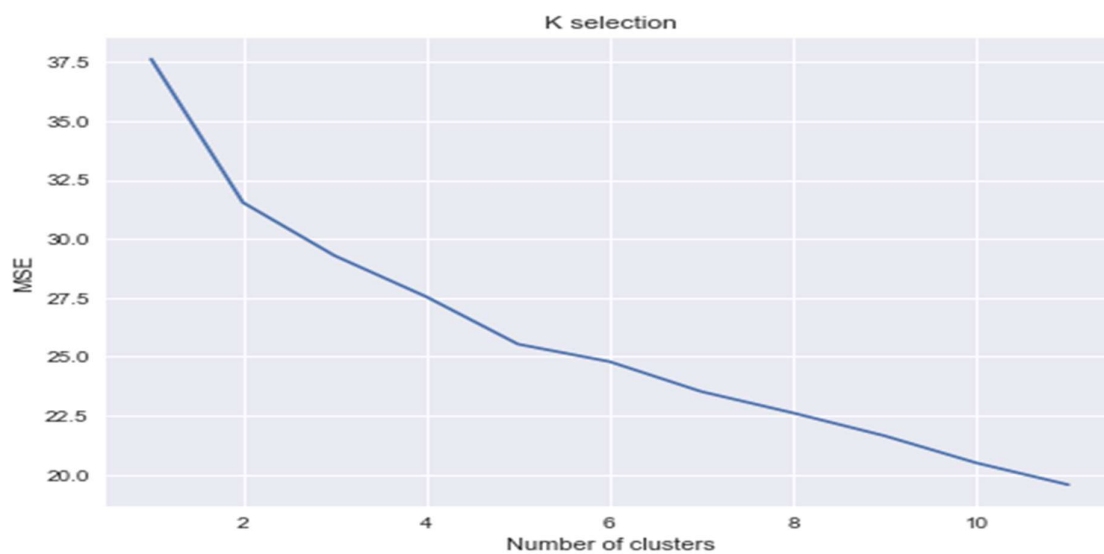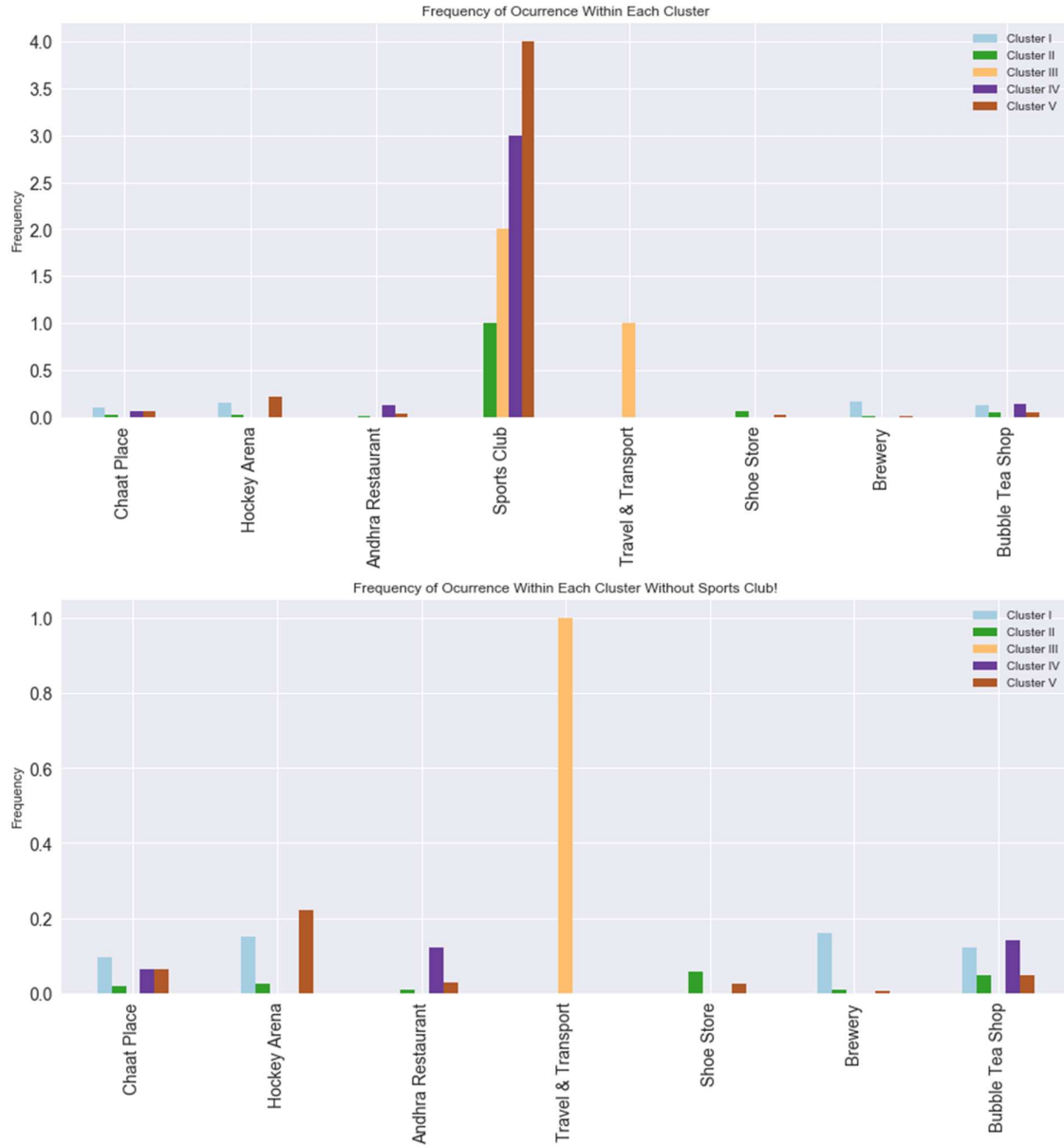

C. Results
The results are better illustrated by the waffle chart as below;

Neighborhoods falling in cluster V have most number of venues. As may be recalled from the venn chart above, there are about 57 categories of venues found in both cities.

D. Discussion

As mentioned above, both cities offer best of amenities and recreational facilities for modern day lifestyle. Key venue categories frequented by residents across clusters, are shown in the chart below.

Frequency of Ocurrence Within Each Cluster

Frequency of Ocurrence Within Each Cluster Without Sports Club!

F. Conclusion

Both cities are dotted with amenities necessary for modern lifestyle.

Neighborhoods falling in cluster 5 in both cities offer largest number of amenities (not account for travel and transport agencies!!). Main neighborhoods falling in cluster 5 are Adugodi, Agara, Banashankari II Stage, Bagalgunte, Banashankari III Stage in Bangalore and Naigaon, New Prabhadevi, Parel, Parel Naka and Prabhadevi in Mumbai.

Post filtering the neighborhoods based on the discussion in this report, one can zero in on a more suitable locality for relocation based on other parameters like proximity to their work location, rent etc.

G. References:
- [1] Pincode data
- [2] Forsquare API