IBM Data Science Professional Specialization - Coursera Clustering Neighbourhood in Pune

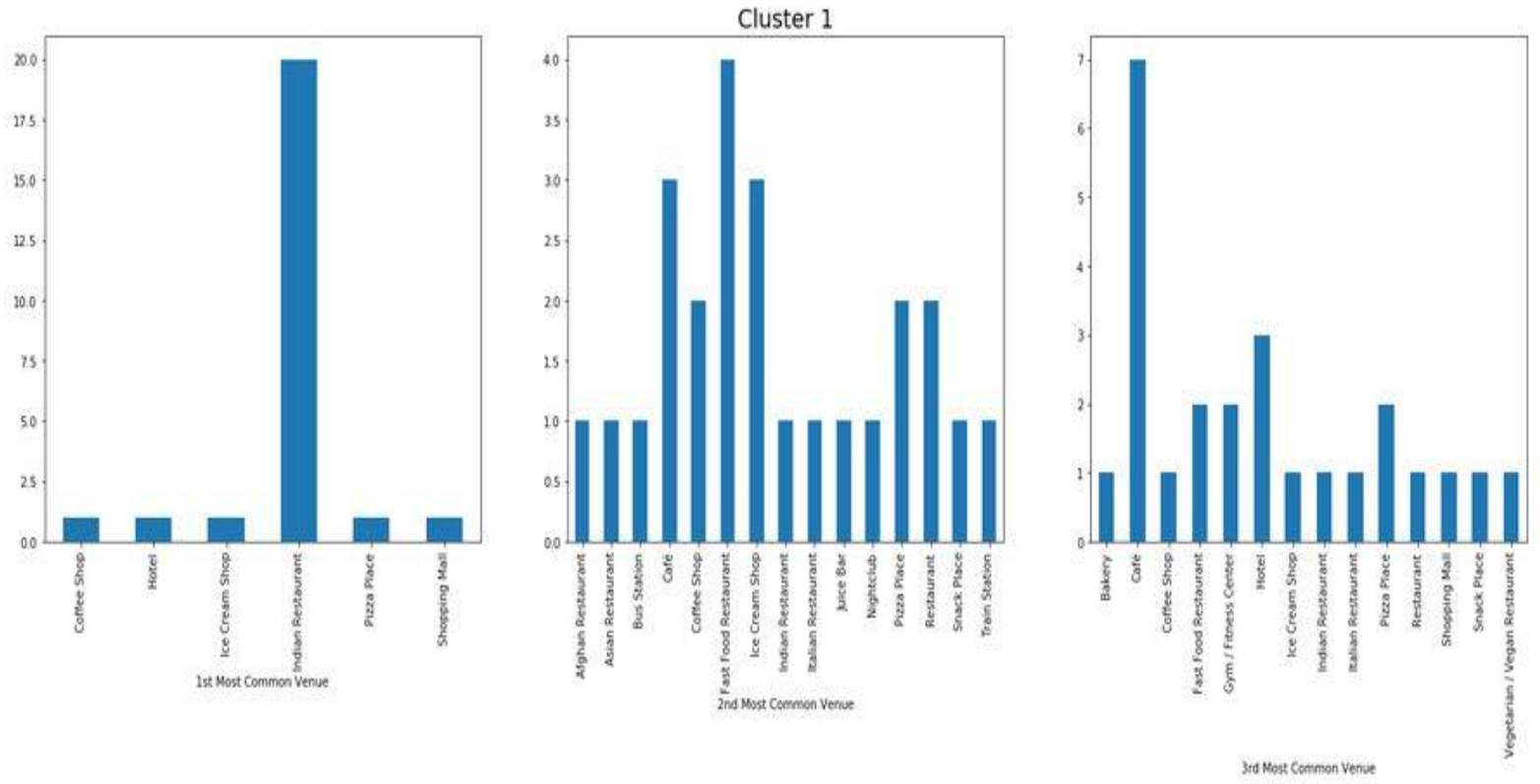
Clustering the Neighbourhoods

- The clustering of Neighbourhood helps us to find the relation / commonness between two or more localities of the city.
- These Relation can be used to determine the cost of living in the area, find a similar area where the user is living currently.
- Find a locality with required amenities close to the work place.
- Predicting the future of a new locality based on current similar developed one

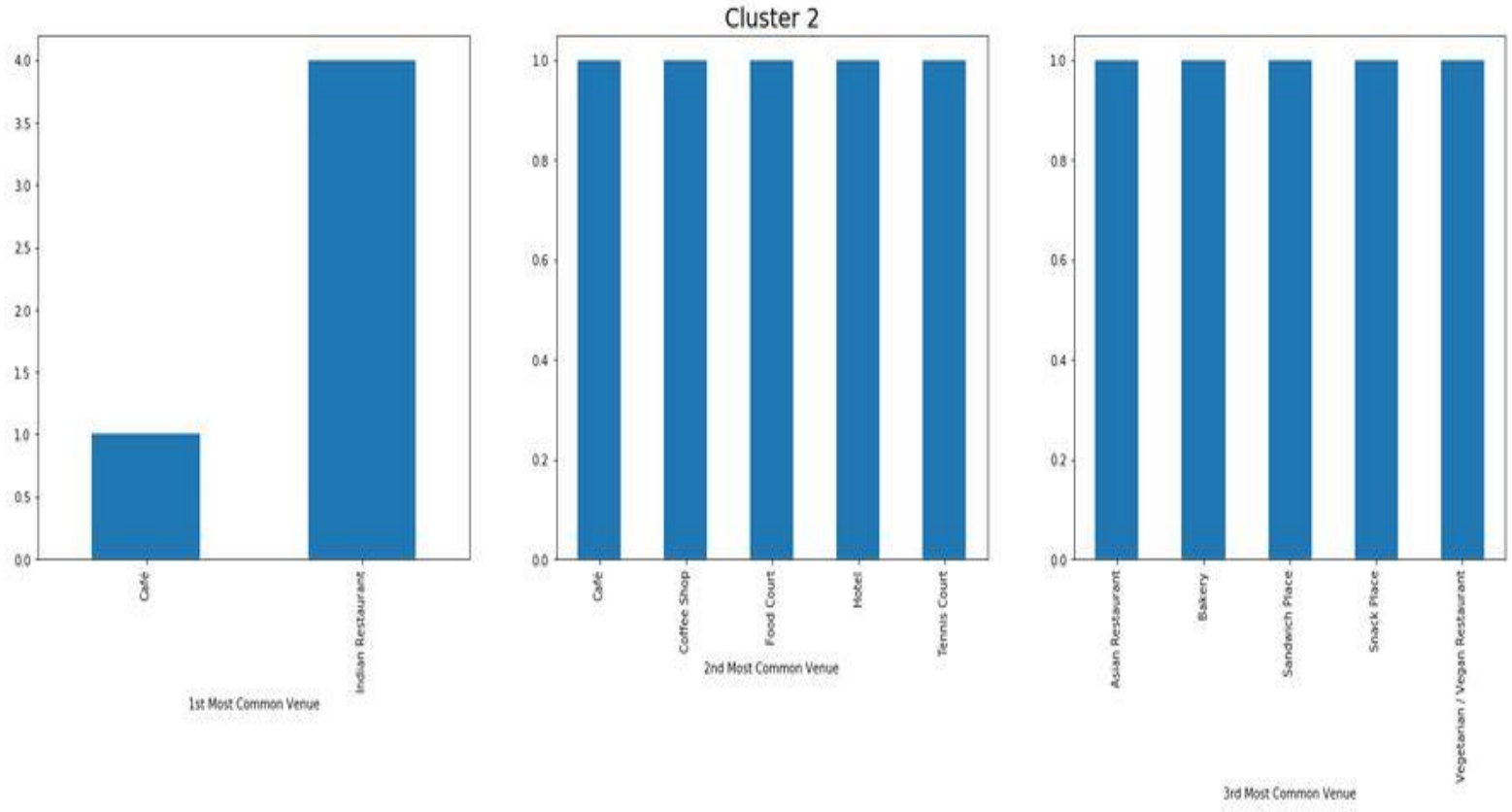
Data Acquisition and Cleaning

- The data about the neighbourhood was scraped from Wikipedia and was stored in a list .
- The data about the price of land per sqr feet was scraped from 99acres and were then added to the neighbourhood data.
- The location of each neighborhood was acquired using Google Maps.
- All the data was combined and a dataframe was created.
- Using FourSqure API , the top 10 most visited places in each neighbourhood was acquired and hence the data frame was ready
- The final count of neighbourhood was 57 .

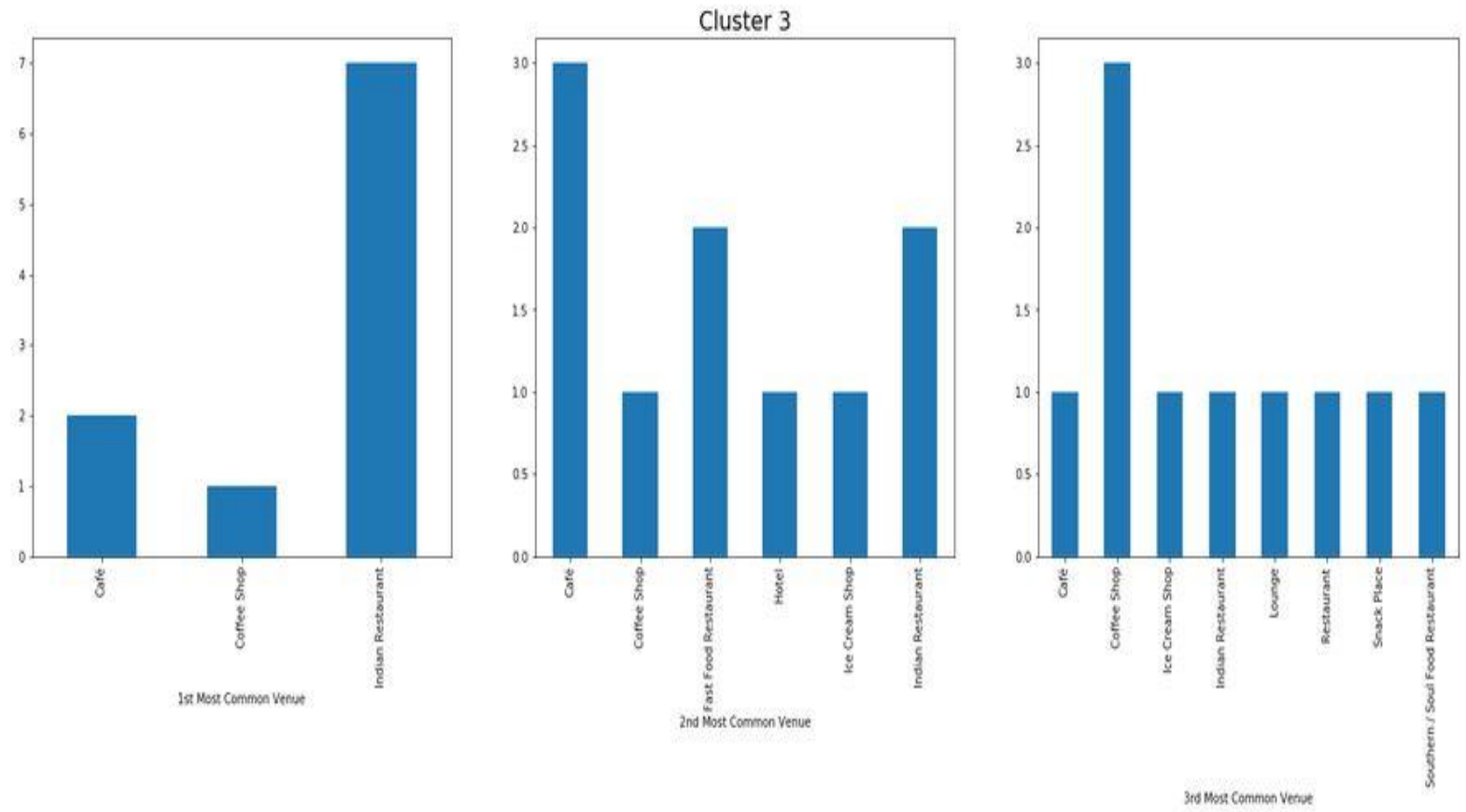
Cluster-wise Top 3 Places Visited
Cluster 1



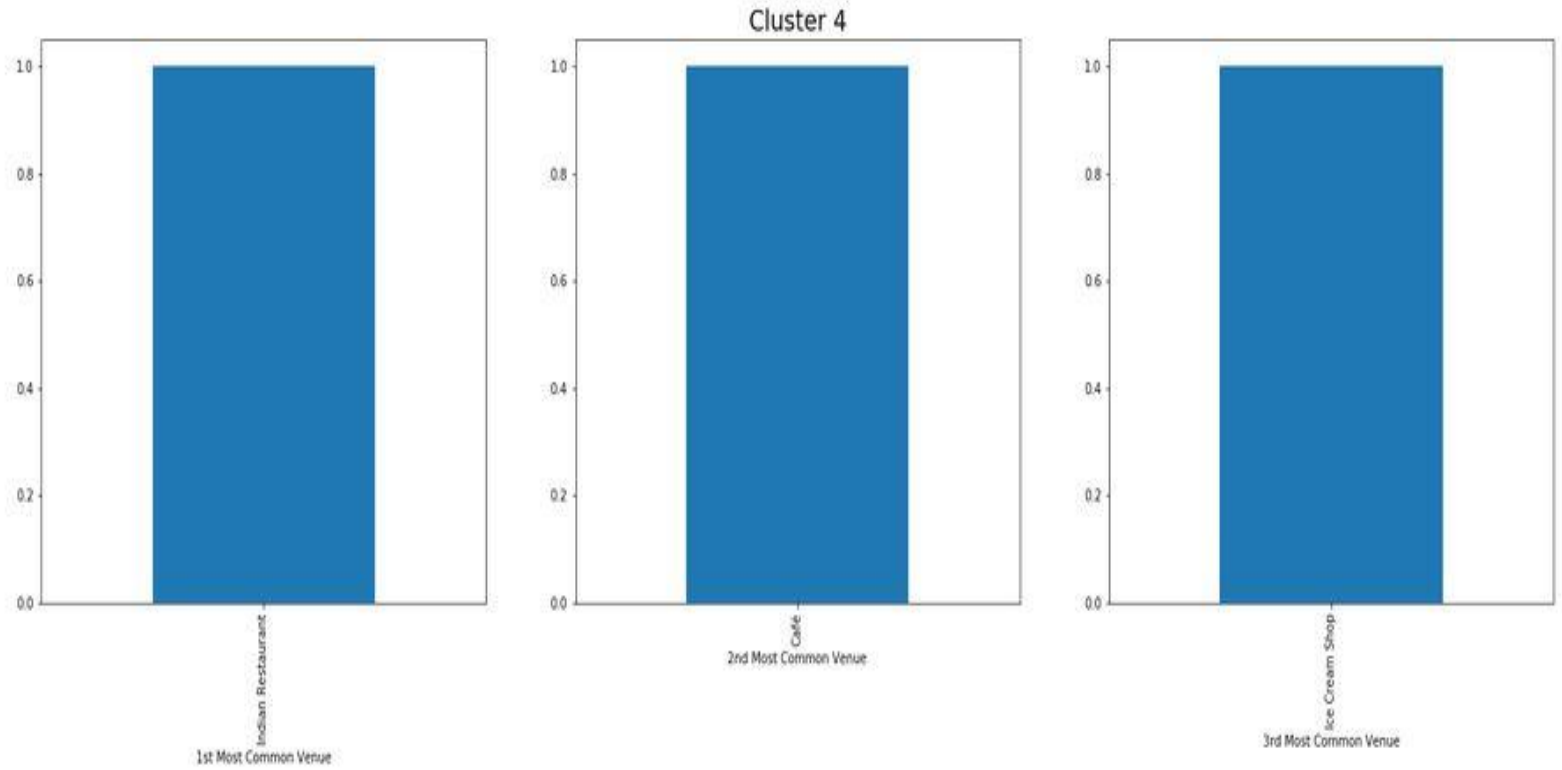
Cluster 2


Cluster 2
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Café
Indian Restaurant
1st Most Common Venue
1.0
0.8
0.6
0.4
0.2
0.0
Café
Coffee Shop
Food Court
Hotel
Tennis Court
2nd Most Common Venue
1.0
0.8
0.6
0.4
0.2
0.0
Asian Restaurant
Bakery
Sandwich Place
Snack Place
Vegetarian / Vegan Restaurant
3rd Most Common Venue

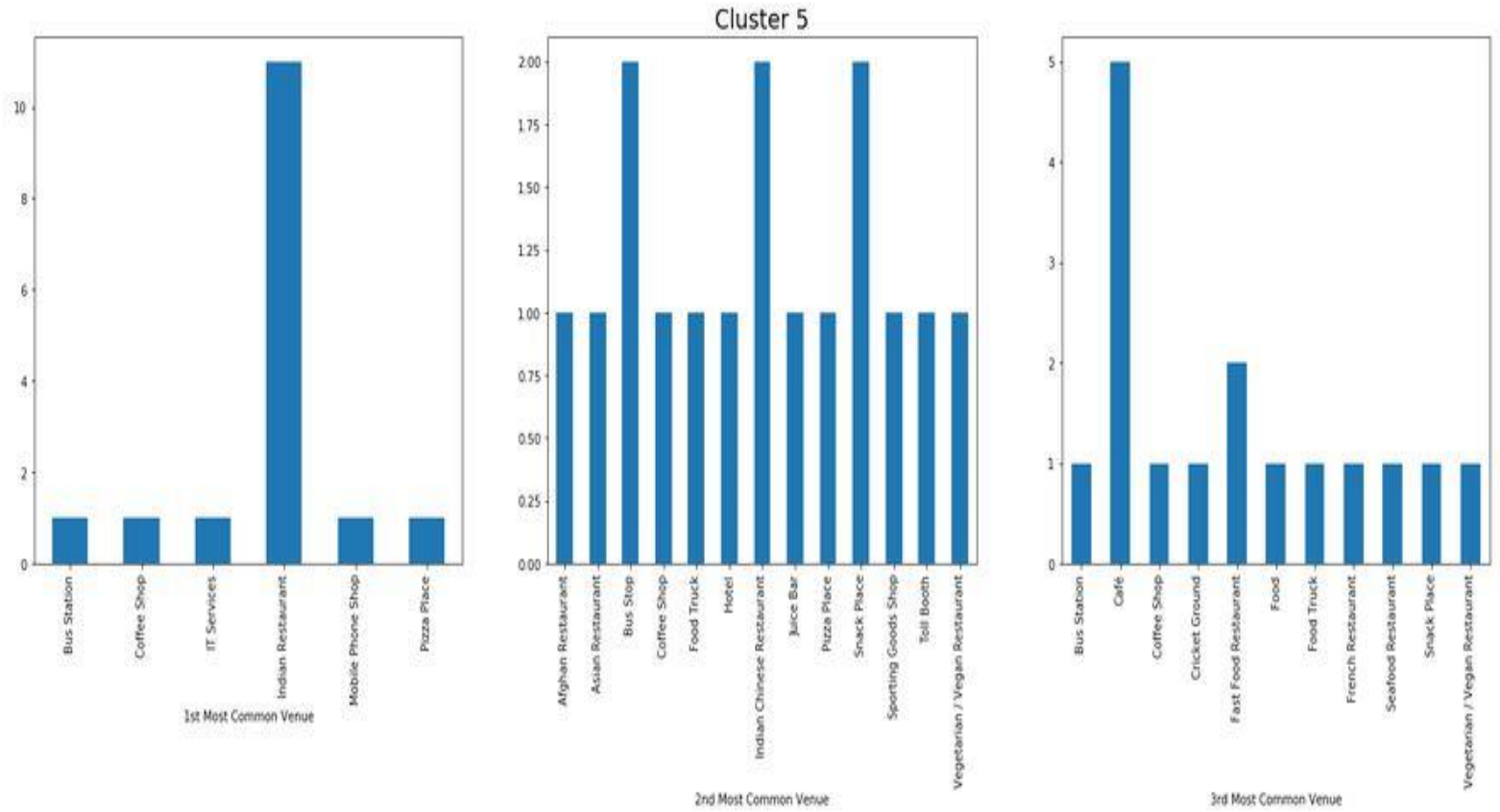
Cluster 3


Cluster 3
Café
Coffee Shop
Indian Restaurant
1st Most Common Venue
Café
Coffee Shop
Fast Food Restaurant
Hotel
Ice Cream Shop
Indian Restaurant
2nd Most Common Venue
Café
Coffee Shop
Ice Cream Shop
Indian Restaurant
Lounge
Restaurant
Snack Place
Southern / Soul Food Restaurant
3rd Most Common Venue

Cluster 4

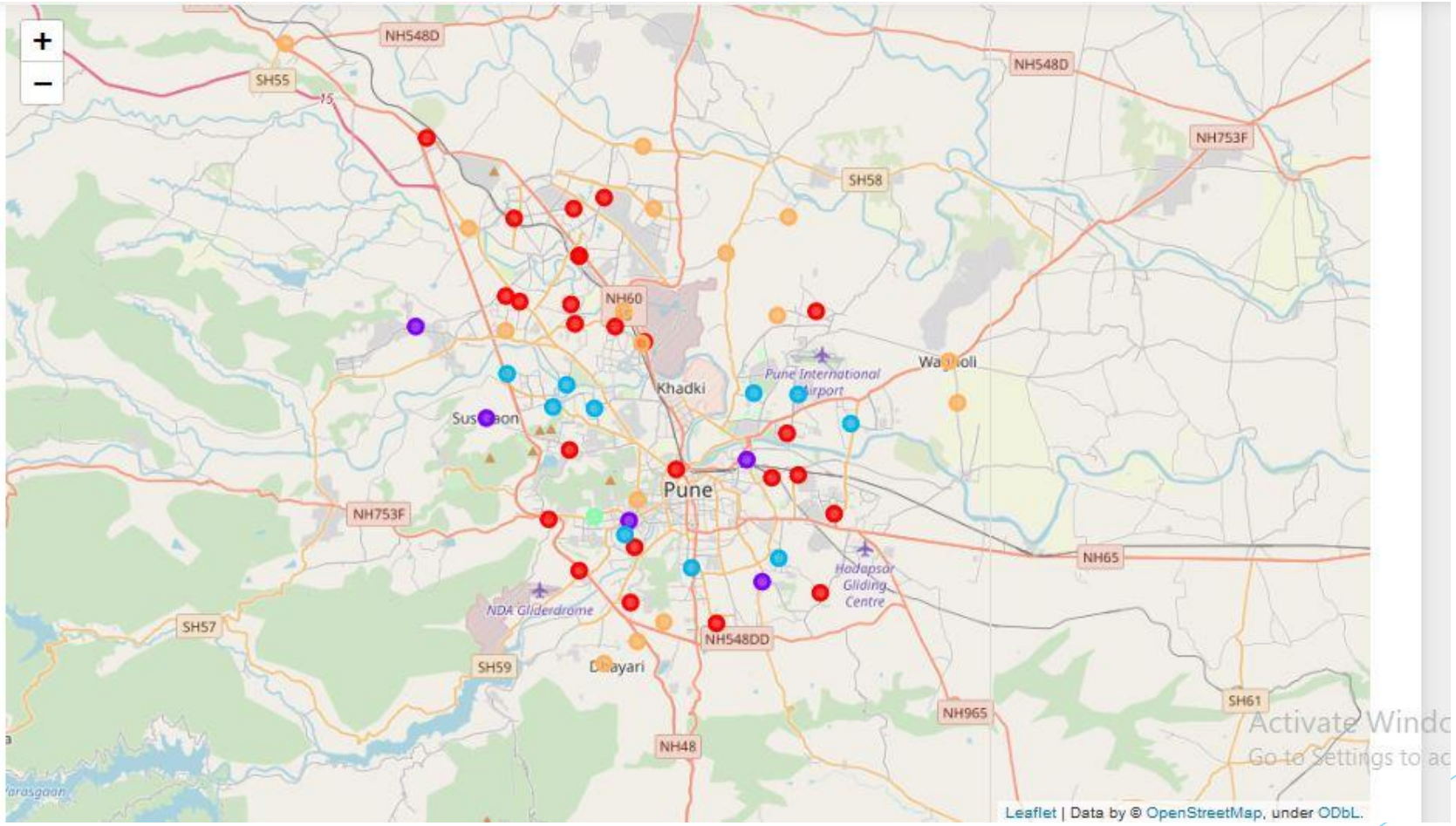

Cluster 4
1.0
0.8
0.6
0.4
0.2
0.0
Indian Restaurant
1st Most Common Venue
Cafe
2nd Most Common Venue
Ice Cream Shop
3rd Most Common Venue

Cluster 5

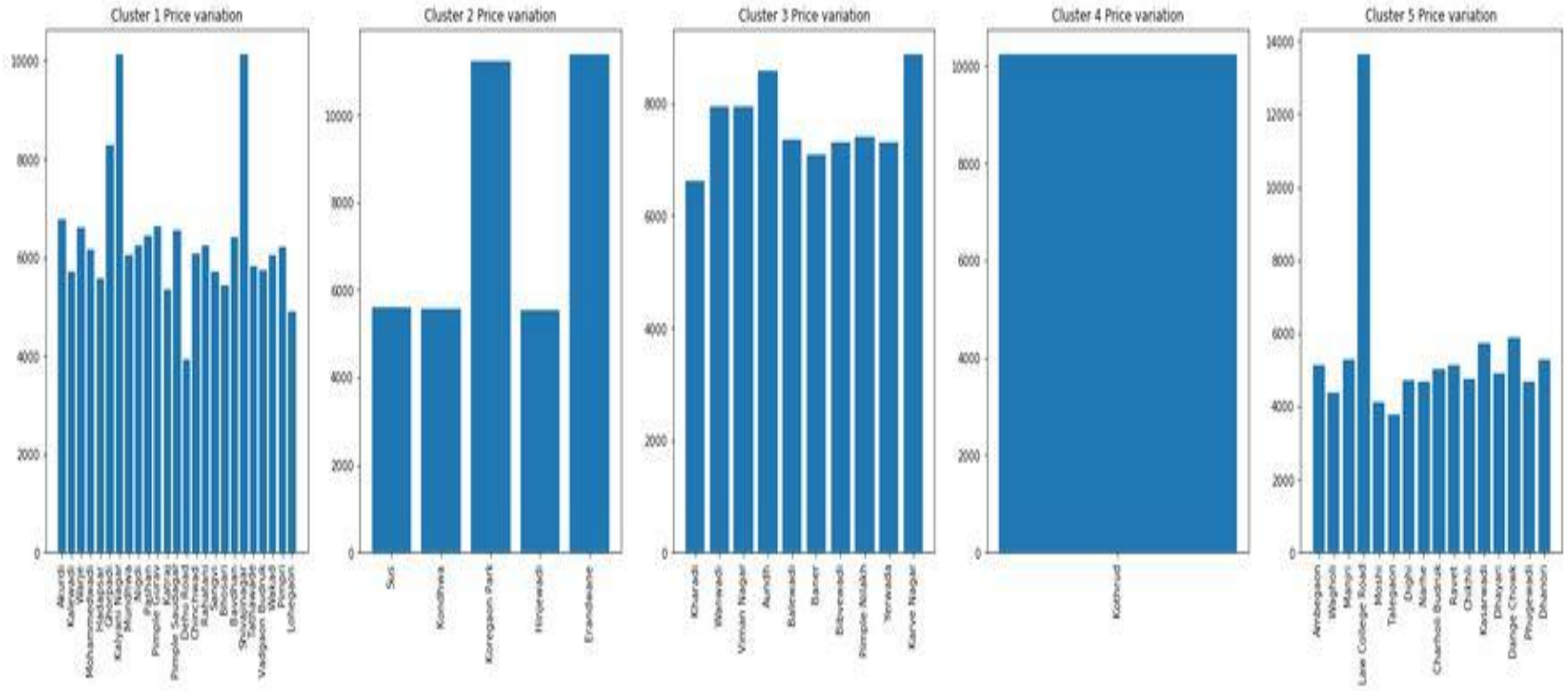

Cluster 5
Bus Station
Coffee Shop
IT Services
Indian Restaurant
Mobile Phone Shop
Pizza Place
1st Most Common Venue
Afghan Restaurant
Asian Restaurant
Bus Stop
Coffee Shop
Food Truck
Hotel
Indian Chinese Restaurant
Juice Bar
Pizza Place
Snack Place
Sporting Goods Shop
Toll Booth
Vegetarian / Vegan Restaurant
2nd Most Common Venue
Bus Station
Café
Coffee Shop
Cricket Ground
Fast Food Restaurant
Food
Food Truck
French Restaurant
Seafood Restaurant
Snack Place
Vegetarian / Vegan Restaurant
3rd Most Common Venue

Clustered Map

Out[345]:



Price Variation in each Cluster



Conclusion and future directions

- From the clustered maps, data set and visualisation,the following conclusions and observation can be drawn
- Cluster 1 and 5 are mostly located far off from the city center with the exception of 2-3 localities .These few localities however have their prices of lands. That is Cluster 1 and 5 if near to city centre have the high prices of land.
- Cluster 3 has a mid-priced land rate and are generally some where in middle of city center and outskirts of the City.They have a lot restraunts,cafe around them
- Cluster 1 has Shopping malls or other public entertainment places.Also Closer they are to city Centre, the higher are the prices of land
- Pune in general has a lot of restraunts and people love to visit them
- this model can made better by considering the following changes which I couldnt do in the current project due to lack of time.
- All the Restraunts and cafe be included in the same category
- Using more data about the city
- The distance from city center column being added to dataframe