Explore and Compare two Cities

24ᵗʰ May 2021

INTRODUCTION

Suppose someone is living in Noida and is moving to Thane. The person wants to know which areas in Thane would be most suitable for him/her to live and which areas to avoid.

In this project we will solve this problem.

This project will explore residential areas in two cities based on the most popular locations in the area. This information will be used to create sets of areas/neighborhoods which are most similar or not similar.

This will enable anyone moving between the cities to understand the similar areas to live in.

DATA

Following data has been used for the project:

- List of residential sectors for Noida has been extracted from url - "https://www.findeasy.in/noida-sectors/"
- List of residential areas for Thane and Navi Mumbai has been extracted from following urls -

  "https://www.magicbricks.com/localities-in-thane"

  "https://www.magicbricks.com/localities-in-Navi-Mumbai"

- The neighborhood data for residential sectors has been extracted from Foursquare API - "https://api.foursquare.com/v2/venues/explore"
- The latitude and longitude for the residential sectors has been extracted using geolocator from Nominatim.
- The areas for which the location data was not available with Nominatim were extracted manually and updated to the data.

METHODOLOGY

The project involved multiple steps like Exploratory Data Analysis, Model building etc. The actions performed in these steps are described below.

Exploratory Data Analysis:

In this project we have explored two different cities. To enable this following steps were performed as part of data extraction and analysis -

- We used multiple libraries in python to extract and parse the list of residential neighborhoods in Noida and Thane.
- This list of residential neighborhoods was then used to get the location information or in other words to fix the pin as the starting point. This was achieved using the "geolocator" library in python. Geolocator uses the "Nominatum" database to get the location data.
- The pin(with latitude and longitude) data was used to get the data for nearby venues and venue categories from "Foursquare".

  Background to Foursquare:

  Foursquare City Guide, commonly known as Foursquare, is a local search-and-discovery mobile app developed by Foursquare Labs Inc. The app provides personalized recommendations of places to go near a user's current location based on users' previous browsing history and check-in history.

- While using an API provided by Foursquare we have to decide how far out from the pin we have to search for restaurants, etc. (Venues). In other words we are deciding the radius of search.
- The search radius was limited to "2000" meters to get the relevant results.
- This data was now modified to be used to generate the ML model. The details are explained in the next section.

ML Model Creation:

For this project we have used the "k-means" clustering model. This model will help us to separate the neighborhoods based on the frequency of the different types of venues around the neighborhood.

To achieve this the following steps were performed -

- Venue Categories are a set of categorical values. These values have to be converted to numerical values so that we can use them to train the model.
- We have used one-hot encoding method to convert the categorical value to numerical values.
- The dataframe containing venue categories contained values as below -

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Sector 11	28.598112	77.333814	Fresh Food Factory - Accha Khao Accha Khilao	28.580971	77.317292	Indian Restaurant
Sector 11	28.598112	77.333814	Lakshmi Coffee House	28.57014	77.332734	South Indian Restaurant
Sector 11	28.598112	77.333814	DLF Mall Of India	28.567503	77.321362	Shopping Mall
Sector 11	28.598112	77.333814	PVR Superplex	28.574257	77.353622	Multiplex
Sector 11	28.598112	77.333814	Berco's Garden Restaurant	28.594078	77.338786	Chinese Restaurant

- After one-hot encoding, we created new dataframe and the value look like below -

Neighborhood	ATM	American Restaurant	Arcade	Asian Restaurant	Athletics & Sports
Sector 11	0	0	0	0	0
Sector 11	0	0	0	0	0
Sector 11	0	0	0	0	0
Sector 11	0	0	0	0	0
Sector 11	0	0	0	0	0

- We then group the values by Neighborhood and calculate the mean for each column. This is done so that we can understand the weightage of each venue type for a neighborhood location -

Neighborhood	ATM	American	Arcade	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery

		Restaurant					
Sector 100	0	0	0	0	0	0	0.25
Sector 104	0	0	0	0	0	0.029412	0
Sector 105	0	0	0	0	0	0	0.142857
Sector 107	0	0	0	0	0	0	0.125
Sector 108	0	0	0	0	0	0	0.076923
...	...	...	...	...	...	...	...
Sector 79	0	0	0	0	0	0.029412	0
Sector 82	0	0	0	0	0	0	0.1
Sector 92	0	0	0	0	0	0	0
Sector 93	0	0	0	0	0	0	0.076923
Sector 99	0	0	0.111111	0	0	0	0

- This data is now used to generate clusters/sets using k-means clustering.
- We tried multiple cluster sizes and decided to use a cluster size of "6".
- The mean data for each neighborhood is used to decide the 10 most common venue categories for each neighborhood.
- Exploring the two cities helps us understand the similar areas in each city and also generate data for the next step.
- The data generated using Foursquare API for Noida and Thane was merged and clustered or sets were generated.
- The clusters generated from the data provided us the list of similar areas in two cities.
- Each cluster generated represents the similar areas a person can look to live if he/she is planning to move between the cities.
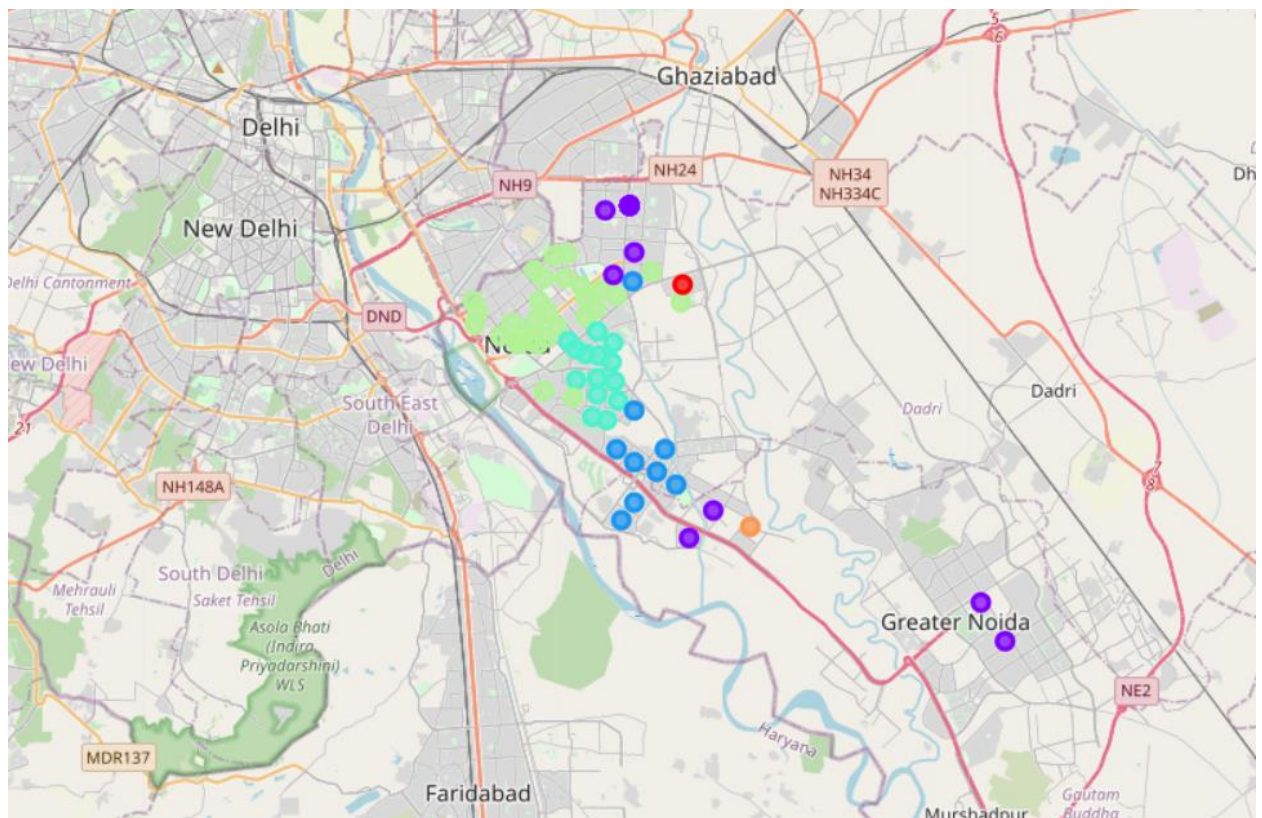
RESULTS

In this section we will show the data generated from exploring two cities and also after comparing the data for two cities(Noida and Thane).

Noida:

- The data for top ten venue categories for each neighborhood was generated. Including only limited here due to space constraints and for clarity -

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Sector 100	Golf Course	Gym / Fitness Center	Bakery	Market	Dry Cleaner
Sector 104	Fast Food Restaurant	Indian Restaurant	Café	Food	Coffee Shop
Sector 105	Pizza Place	Pharmacy	Bakery	Market	Diner
Sector 107	Pizza Place	Thai Restaurant	Gym / Fitness Center	Bakery	Market
Sector 108	Pizza Place	Café	Golf Course	Snack Place	Pharmacy

- A map was generated using "folium" library to show the resulting clusters as a visual aid to understand how the city is divided into different sets -

Thane:

- The data for top ten venue categories for each neighborhood was generated -

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Airoli, Navi Mumbai	Fast Food Restaurant	Pizza Place	Indian Restaurant	Gym	Department Store
Anand Nagar, Thane	Indian Restaurant	Fast Food Restaurant	Ice Cream Shop	Dessert Shop	Pizza Place
Balkum village, Thane	Pizza Place	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Shopping Mall
Belapur, Navi Mumbai	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Dessert Shop
Bhayanderpada, Thane	Café	Indian Restaurant	Pizza Place	Fast Food Restaurant	Multiplex

- A map was generated using "folium" library to show the resulting clusters as a visual aid to understand how the city is divided into different sets -
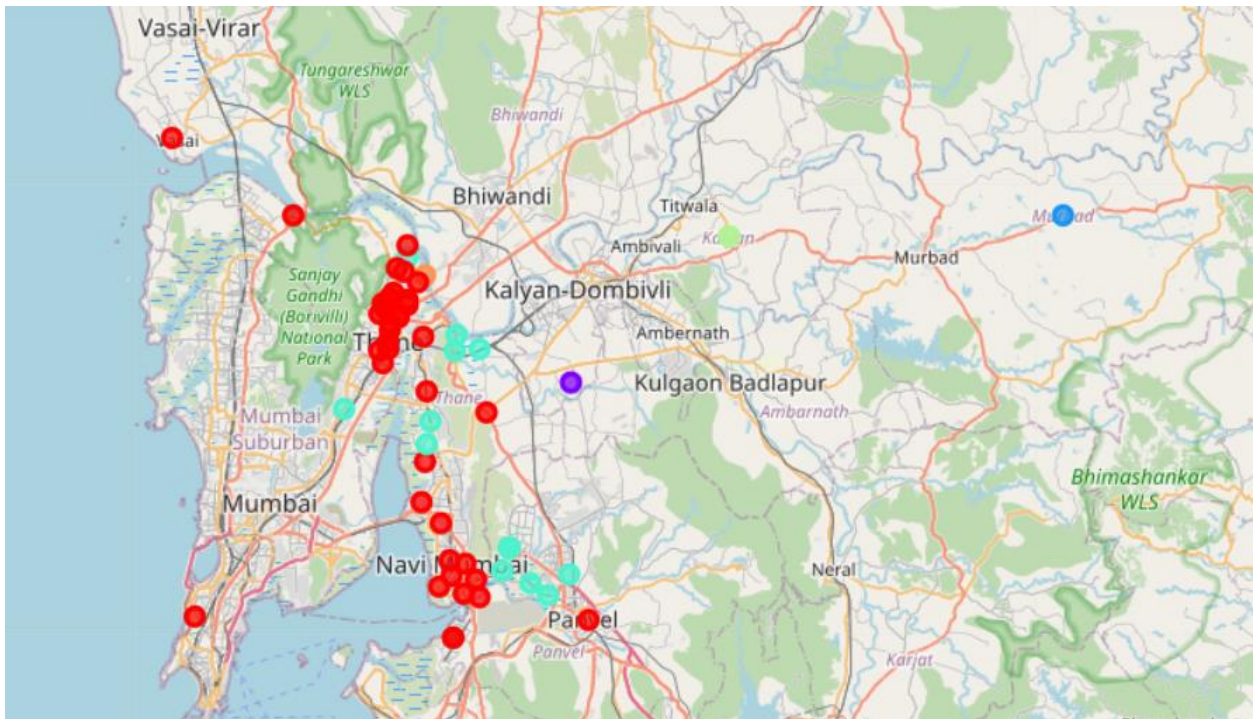
Comparing Neighbourhoods in Noida and Thane:

- Comparing the two cities we decided to generate 10 clusters/sets.
- We analysed a total 157 areas.
- These clusters contain similar neighborhoods from Noida and Thane along with ten venue categories in both cities.
- Here we will take three clusters/sets as an example which will help us to draw our conclusions.

Cluster 0:

Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Sector 11	Noida	28.59811245	77.33381362	0	Indian Restaurant	Market	Stadium	Sandwich Place	Pizza Place
Sector 12	Noida	28.59511835	77.33848623	0	Indian Restaurant	Stadium	Pizza Place	Restaurant	Electronics Store

Sector 14	Noida	28.58723 095	77.30822 174	0	Indian Restaurant	Hotel	Plaza	Bakery	Department Store
Sector 15	Noida	28.58279 785	77.31022 209	0	Indian Restaurant	Hotel	Chinese Restaurant	Sandwich Place	Café
Sector 15a	Noida	28.57799 69	77.30850 904	0	Indian Restaurant	Café	Chinese Restaurant	Restaurant	Fast Food Restaurant
Sector 17	Noida	28.57374 69	77.32003 759	0	Indian Restaurant	Café	Coffee Shop	Chinese Restaurant	Fast Food Restaurant
Sector 19	Noida	28.57839 66	77.32476 922	0	Indian Restaurant	Café	Coffee Shop	Chinese Restaurant	Sandwich Place
Sector 21	Noida	28.58682 275	77.33616 219	0	Indian Restaurant	Market	Sandwich Place	Stadium	Pizza Place
Sector 22	Noida	28.59520 28	77.34727 924	0	Indian Restaurant	Pizza Place	Gym	Coffee Shop	Shopping Mall
Sector 23	Noida	28.59235 44	77.35381 332	0	Indian Restaurant	Pizza Place	Gym	Vegetarian / Vegan Restaurant	Coffee Shop
Sector 25	Noida	28.58300 83	77.33981 906	0	Indian Restaurant	Market	Multiplex	Sandwich Place	Hotel
Sector 26	Noida	28.57946 555	77.33512 252	0	Indian Restaurant	Chinese Restaurant	Multiplex	Sandwich Place	Coffee Shop
Sector 27	Noida	28.57467 68	77.32872 811	0	Indian Restaurant	Café	Coffee Shop	Chinese Restaurant	Sandwich Place
Sector 28	Noida	28.57160 795	77.33182 83	0	Indian Restaurant	Chinese Restaurant	Café	Coffee Shop	Sandwich Place
Sector 29	Noida	28.57083 15	77.33582 067	0	Indian Restaurant	Café	Coffee Shop	Chinese Restaurant	Pizza Place

Sector 30	Noida	28.57330205	77.33917321	0	Indian Restaurant	Chinese Restaurant	Coffee Shop	Café	Multiplex
Sector 31	Noida	28.57720495	77.34370348	0	Indian Restaurant	Multiplex	Pizza Place	Sandwich Place	Bakery
Sector 33	Noida	28.58839945	77.35633363	0	Indian Restaurant	Gym	Convenience Store	Multiplex	Pizza Place
Sector 34	Noida	28.58372415	77.36046749	0	Pizza Place	Gym	Indian Restaurant	Multiplex	Vegetarian / Vegan Restaurant
Sector 37	Noida	28.5707841	77.3271074	0	Indian Restaurant	Café	Chinese Restaurant	Sandwich Place	Coffee Shop
Sector 44	Noida	28.5531837	77.3381716	0	Chinese Restaurant	Indian Restaurant	Pizza Place	Sandwich Place	Coffee Shop
Sector 45	Noida	28.55060205	77.35051279	0	Sandwich Place	Golf Course	Chinese Restaurant	Indian Restaurant	Thai Restaurant
Sector 52	Noida	28.5885198	77.36890648	0	Pizza Place	Gym	Vegetarian / Vegan Restaurant	Indian Restaurant	Convenience Store
Sector 55	Noida	28.6043261	77.34802934	0	Indian Restaurant	Pizza Place	Gym	Food	Vegetarian / Vegan Restaurant
Sector 56	Noida	28.60387165	77.34257203	0	Indian Restaurant	Coffee Shop	Pizza Place	Restaurant	American Restaurant
Sector 70	Noida	28.59763785	77.38432381	0	Indian Restaurant	Pizza Place	Gym	Convenience Store	Vegetarian / Vegan Restaurant
Sector 74	Noida	28.5707841	77.3271074	0	Indian Restaurant	Café	Chinese Restaurant	Sandwich Place	Coffee Shop
Sector 120	Noida	28.58623125	77.39659348	0	Indian Restaurant	Recording Studio	Dry Cleaner	Wine Shop	Donut Shop

Sector 128	Noida	28.57078 41	77.32710 74	0	Indian Restaurant	Café	Chinese Restaurant	Sandwich Place	Coffee Shop
Sector 168	Noida	28.57078 41	77.32710 74	0	Indian Restaurant	Café	Chinese Restaurant	Sandwich Place	Coffee Shop
Thane West	Thane	19.19432 94	72.97017 79	0	Indian Restaurant	Fast Food Restaura nt	Ice Cream Shop	Coffee Shop	Seafood Restaurant
Ghodbunder Road	Thane	19.28821 32	72.89557 45	0	Indian Restaurant	Restaura nt	Food	Gym / Fitness Center	Food Court
Majiwada	Thane	19.21302 51	72.97848 52	0	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Pizza Place	Seafood Restaurant
Hiranandani Estate	Thane	19.25685 595	72.98427 052	0	Grocery Store	Bus Station	Café	Bakery	Food
Manpada	Thane	19.23141 56	72.97381 56	0	Indian Restaurant	Pizza Place	Coffee Shop	Dessert Shop	Bakery
Balkum village	Thane	19.22169 075	72.98449 242	0	Pizza Place	Coffee Shop	Indian Restaurant	Fast Food Restaurant	Bakery
Naupada	Thane	19.18904 69	72.96923 39	0	Indian Restaurant	Fast Food Restaura nt	Ice Cream Shop	Dessert Shop	Chinese Restaurant
Anand Nagar	Thane	19.17855 6	72.96600 4	0	Indian Restaurant	Fast Food Restaura nt	Ice Cream Shop	Dessert Shop	Pizza Place
Waghbil	Thane	19.26596 68	72.98477 05	0	Pizza Place	Café	Grocery Store	Food Court	Department Store
Teen Hath Naka	Thane	19.18837 79	72.96358 94	0	Indian Restaurant	Fast Food Restaura nt	Ice Cream Shop	Dessert Shop	Coffee Shop
Dhokali	Thane	19.22557 86	72.98454 36	0	Indian Restaurant	Coffee Shop	Pizza Place	Fast Food Restaurant	Shopping Mall

Vartak Nagar	Thane	19.2155371	72.9627101	0	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Seafood Restaurant	Pizza Place
Vasant Vihar	Thane	19.2224897	72.9663424	0	Indian Restaurant	Pizza Place	Fast Food Restaurant	Coffee Shop	Shopping Mall
Pokhran Road	Thane	19.2191453	72.9674865	0	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Pizza Place	Shopping Mall
Shilphata	Thane	19.1430394	73.0467879	0	Indian Restaurant	Tea Room	Fast Food Restaurant	Art Gallery	Train Station
Kapurbawdi	Thane	19.2142578	72.9778462	0	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Pizza Place	Shopping Mall
Patlipada	Thane	19.2493501	72.9760694	0	Indian Restaurant	Bakery	Ice Cream Shop	Café	Bus Station
Hiranandani Meadows	Thane	19.22734225	72.9718477	0	Indian Restaurant	Coffee Shop	Pizza Place	Fast Food Restaurant	Shopping Mall
Brahmand	Thane	19.2469649	72.9818412	0	Indian Restaurant	Coffee Shop	Bus Station	Café	Pizza Place
Kolshet	Thane	19.2389025	72.9932692	0	Pizza Place	Wine Shop	Thai Restaurant	Pub	Coffee Shop
Khopat	Thane	19.2033291	72.9717815	0	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Ice Cream Shop	Chinese Restaurant
Palghar	Thane	19.3453756	72.8007115	0	Seafood Restaurant	Indian Restaurant	Flea Market	Café	Bus Station
Jeejamata Nagar	Thane	18.9922889	72.81878651	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Café	Pub
Louis Wadi	Thane	19.20886565	72.97167426	0	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Ice Cream Shop	Pizza Place
Hiranandani Road	Thane	19.22734225	72.9718477	0	Indian Restaurant	Coffee Shop	Pizza Place	Fast Food Restaurant	Shopping Mall

Panvel	Navi Mumbai	19.0067566	73.04075004	0	Indian Restaurant	Fast Food Restaurant	Café	Italian Restaurant	Pizza Place
Nerul	Navi Mumbai	19.0336118	73.01813957	0	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Chinese Restaurant	Snack Place
Airoli	Navi Mumbai	19.1585147	72.99940185	0	Fast Food Restaurant	Pizza Place	Indian Restaurant	Gym	Bus Station
Seawoods	Navi Mumbai	19.0221923	73.01873757	0	Indian Restaurant	Café	Fast Food Restaurant	Coffee Shop	Sandwich Place
Kopar Khairane	Navi Mumbai	19.1056227	72.9988648	0	Fast Food Restaurant	Department Store	Indian Restaurant	Ice Cream Shop	Hotel
Vashi	Navi Mumbai	19.075784	72.9952364	0	Indian Restaurant	Ice Cream Shop	Coffee Shop	Hotel	Fast Food Restaurant
Sanpada	Navi Mumbai	19.0607337	73.0116775	0	Indian Restaurant	Fast Food Restaurant	Hotel	Chinese Restaurant	Café
CBD Belapur	Navi Mumbai	19.01898695	73.03909472	0	Indian Restaurant	Fast Food Restaurant	Café	Pizza Place	Restaurant
Belapur	Navi Mumbai	19.0088205	73.028701	0	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Dessert Shop
Uran	Navi Mumbai	19.0307053	73.0300046	0	Fast Food Restaurant	Indian Restaurant	Coffee Shop	Diner	Garden
Sector 36	Navi Mumbai	19.01371895	73.00999035	0	Café	Shopping Mall	Fast Food Restaurant	Sandwich Place	Indian Restaurant

- This is a big set with 66 neighborhoods.
- This set contains mixed neighborhoods from Noida, Thane and Navi Mumbai(part of Thane district).

Cluster 7:

Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Sector 35	Noida	28.5806152	77.35630307	2	Multiplex	Pizza Place	Stadium	Market	Sandwich Place
Sector 36	Noida	28.5719081 5	77.34853224	2	Bakery	Indian Restaurant	Multiplex	Market	Metro Station
Sector 39	Noida	28.5686223	77.35210967	2	Bakery	Gym / Fitness Center	Pizza Place	Multiplex	Café
Sector 40	Noida	28.5670239	77.3561036	2	Multiplex	Gym	Sandwich Place	Bakery	Pizza Place
Sector 41	Noida	28.5662237	77.36174561	2	Gym / Fitness Center	Multiplex	Bakery	Italian Restaurant	Food Truck
Sector 42	Noida	28.5580927	77.36117784	2	Café	Gym / Fitness Center	Multiplex	Bakery	Pizza Place
Sector 43	Noida	28.5575945 5	77.35167905	2	Pizza Place	Indian Restaurant	Multiplex	Café	Golf Course
Sector 46	Noida	28.5521626 5	77.36060566	2	Café	Golf Course	Lounge	Arcade	Pizza Place
Sector 48	Noida	28.5567102 5	77.36874453	2	Bakery	Café	Gym / Fitness Center	Pizza Place	Sandwich Place
Sector 49	Noida	28.5640158	77.36719145	2	Gym / Fitness Center	Multiplex	Bakery	Pizza Place	Food Truck
Sector 50	Noida	28.5711552	77.36748745	2	Bakery	Pizza Place	Multiplex	Gym / Fitness Center	Dessert Shop
Sector 51	Noida	28.5754068 5	77.36124549	2	Pizza Place	Bakery	Multiplex	Arcade	Vegetarian / Vegan Restaurant
Sector 99	Noida	28.5434589 5	77.35892385	2	Sandwic h Place	Market	Arcade	Coffee Shop	Café
Sector 100	Noida	28.5422873	77.36509586	2	Golf Course	Gym / Fitness Center	Bakery	Market	Wine Shop
Kalwa	Thane	19.197722	72.9978735	2	Seafood Restaura	Multiplex	Pizza Place	Dessert Shop	Wine Shop

					nt				
Mumbra	Thane	19.188536	73.0215253	2	Flea Market	Multiplex	Restaurant	Bakery	Donut Shop

- This set contains 16 values.
- This set contains neighborhoods mostly from Noida. Just 2 neighborhoods are from Thane.
- We have one more big cluster with 32 values with the same characteristics. But, this time the neighborhoods are mostly from Thane.
- We also generated a set of clusters which do not have any neighborhoods from both the cities -

Neighborhood	City	Latitude	Longitude	Cluster Labels
Kalyan	Thane	19.2716301	73.23586553	1
Kalyan West	Thane	19.2716301	73.23586553	1
Palava	Thane	19.16460915	73.11278411	4
Ulwe	Navi Mumbai	18.97686135	73.02013111	7
Ulwe Village	Navi Mumbai	18.97686135	73.02013111	7
Sector 53	Noida	28.5911007	77.3628053	8
Sector 61	Noida	28.59645805	77.36756438	8
Sector 66	Noida	28.6052229	77.3764073	8
Sector 71	Noida	28.59425315	77.37619464	8
Sector 82	Noida	28.5311541	77.38993356	8
Sector 92	Noida	28.5181771	77.39489696	8
Sector 93	Noida	28.5229507	77.38611653	8
Sector 105	Noida	28.5315704	77.36917722	8
Sector 107	Noida	28.545932	77.37656097	8
Sector 108	Noida	28.52660665	77.37679363	8
Sector 47	Noida	28.5498284	77.37006993	9

DISCUSSION

In this section we will discuss the results and observations we can see from the results.

- After exploring the cities Noida and Thane we can see that there is only one large cluster/set(66 neighborhoods) with mixed values from both Noida and Thane.
- The rest of the clusters either have one few values from another city or no values from another city.
- The neighborhoods in Cluster 0 are mostly the area situated near the city center in their respective cities.
- This shows most of the areas in Noida are not similar to areas in Thane.
- Only the areas situated near the city center are similar in Noida and Thane.

CONCLUSION

After analysing the data from both cities we can conclude the following -

- Someone living near the city center in one city should look to find a place near the city center in another city. Only then they will be able to get similar experience.
- If someone is not living in the city center in one city, they will have a different experience when he/she moves to the other city. This is because either they live near the city center or far from the city center they are moving from one set of neighborhoods to another.