

The Battle of the Neighborhoods: Pune City

1. Introduction: -

In this project we will try to find out an optimal livable location in a city. By livable location here it means the number of facilities, nearby venues provided in the vicinity of a location. This report specifically targets stakeholders who are interested in finding a livable location in Pune city, India.

Also, the stakeholder would choose a particular location which is closest to its workplace. Hence, we would categorize various locations in the city on the basis of their distance between **the most prominent workplaces/ industrial zones in the city** and **nearby venues to those locations**.

We would cluster all the important/livable and recommended regions where someone can live which have the most number of amenities/venues in the locality. User/Stakeholder can choose according the location nearest to his/her office/campus.

For example: - Hinjewadi and Magarpatta are two IT company zones which are situated in opposite direction to each other, If someone lives in a location near to Hinjewadi and his/her company moves to Magarpatta, which location should be suggested having similar facilities/ venues compared to his/her current living location.

We would identify areas with most promising characteristics and their advantages will then be clearly expressed, so that best possible final location may be chosen by our stakeholders.

2. Data Acquisition and Cleaning: -

2.1 Data Sources: -

We have taken latitude and longitude of most prominent pune locations from [PMC Open Data Store](#) official website and [Kaggle Dataset](#). The data derived from kaggle dataset do not have latitude longitude information, hence we have used geopy library to fetch latitude and longitude values for such locations.

The data of most prominent Industrial areas in the city is also taken from [PMC Open Data Store](#). Once, latitude and longitude information is fetched from geopy library we have 2 datasets.

1: - Locations in Pune

2: - Industrial areas in Pune

Also FourSquare API is used to locate all the venues in 1K.M area of those locations. We have calculated the distance of each location from all industrial areas also for our analysis.

2.2 Data Cleaning: -

Data downloaded was from two different sources, it was combined into one table. Their were some locations whose latitude, longitude location couldn't be identified through geopy library. Those locations were deleted from our cleaned dataset. Some multiple entries existed for locations common in both dataset, hence only one copy was used.

The distance of each location from industrial areas is much larger as compared to frequency of each venues in the location, as a result of which all these features are normalized for our analysis.

3. Data Analysis: -

Following two different approaches are followed to identify similar locations in Pune city: -

1. All venues in 1 K.M area is used to identify the similarity between locations in the city.
2. All venues in each location along with their distance from all industrial areas is taken to identify the similarity between locations in the city.

Out of 197 locations used in our analysis, 10 clusters will be formed in each analysis phase.

The map of Pune city superimposed by all the locations used in our analysis is as follows: -

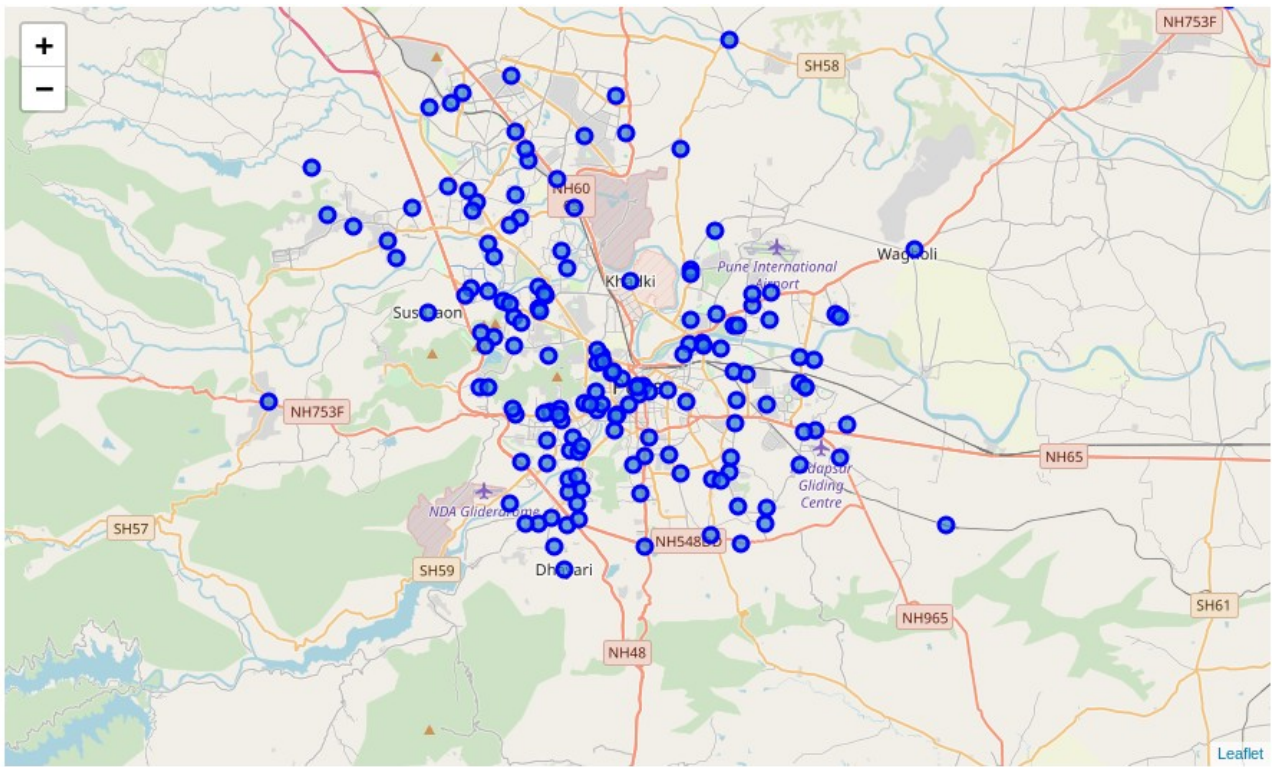


Illustration 1: Locations in Pune city

Now, let draw a map of Pune city superimposed by all the industrial areas in the city: -

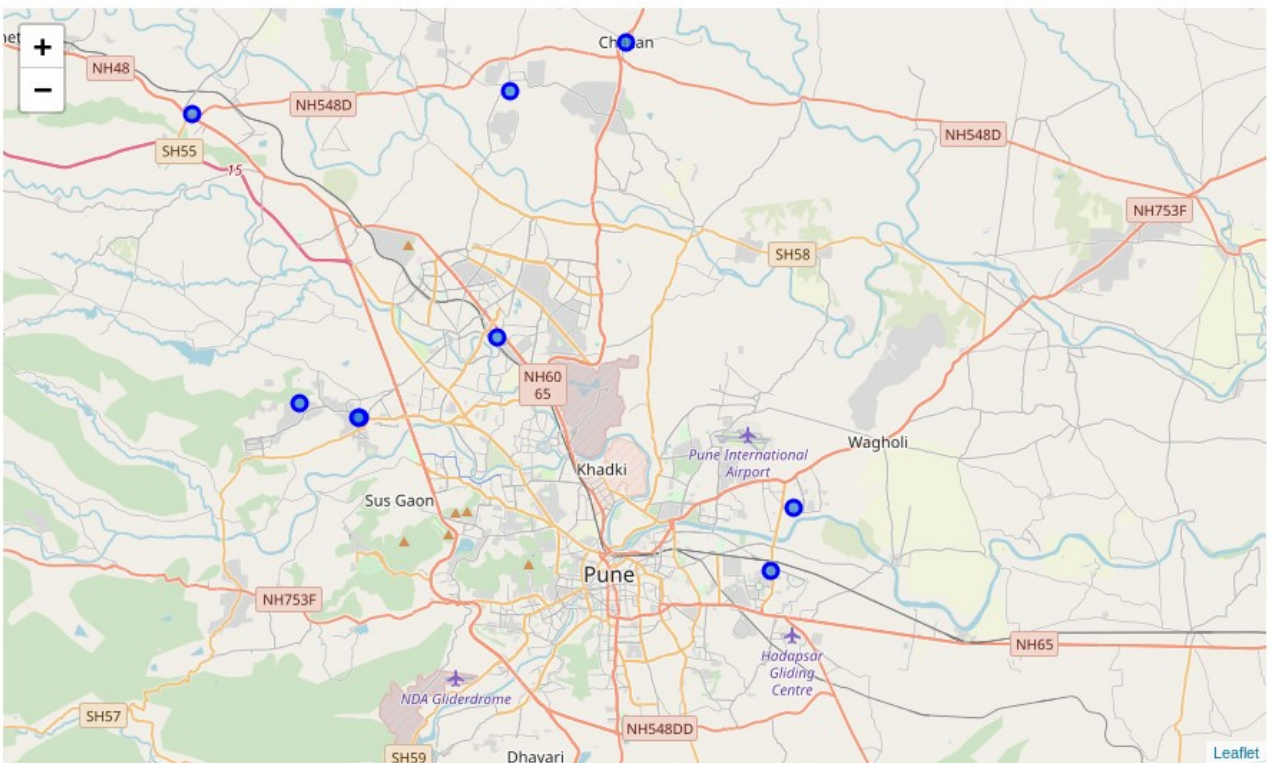


Illustration 2: Industrial areas in Pune city

3.1 Analysis Phase 1 (Similarity on the basis of venues in each location)

In this analysis phase foursquare API is used to determine all the venues in 1K.M geography of each location. K-Means clustering algorithm is used to classify locations in 10 clusters. All the locations where then given cluster labels from 0-9.

The map of the city drawn with each cluster location in a separate color according to the assigned cluster is as follows: -

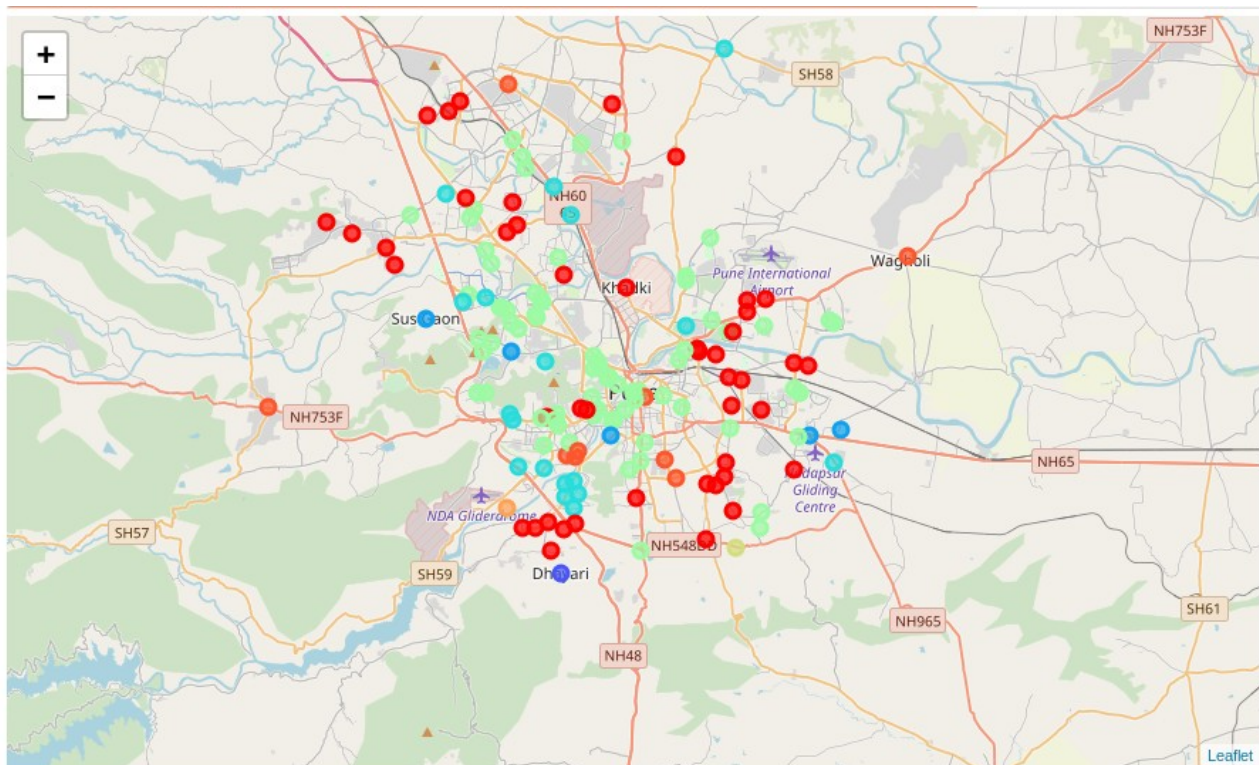


Illustration 3: Classification of locations through Analysis 1

We can clearly see that each area have similar type of locations. If our stakeholder wish to shift from one geographical area of the city to another he can make a decision using this analysis. Let's check some cluster locations of this analysis: -

Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Sinhagad Road	Indian Restaurant	Diner	Ice Cream Shop	Café	Bakery	Fast Food Restaurant	Gym / Fitness Center	Pizza Place	Cupcake Shop	Dosa Place
Warje	Coffee Shop	Indian Restaurant	Pizza Place	Fast Food Restaurant	Grocery Store	Diner	Eastern European Restaurant	Donut Shop	Dosa Place	Dumpling Restaurant
Paud Road	Indian Restaurant	Café	Sporting Goods Shop	Motorcycle Shop	Breakfast Spot	Bus Station	Diner	Sandwich Place	Cafeteria	Liquor Store
Fursungi	Fast Food Restaurant	Hotel	Indian Restaurant	Rock Club	Distillery	Farmers Market	Farm	Falafel Restaurant	Factory	Event Space
Prabhadevi Tech Park	Indian Restaurant	Fast Food Restaurant	Snack Place	Asian Restaurant	Shopping Mall	Lounge	South Indian Restaurant	Bistro	Café	Bar
Alandi	Fast Food Restaurant	Bus Station	Indian Restaurant	River	Distillery	Farmers Market	Farm	Falafel Restaurant	Factory	Event Space
Anand Nagar	Fast Food Restaurant	Bakery	Indian Restaurant	Coffee Shop	Snack Place	Ice Cream Shop	Pizza Place	Gym / Fitness Center	Diner	Falafel Restaurant
Balewadi Phata	Indian Restaurant	Fast Food Restaurant	Breakfast Spot	Café	Lounge	Vegetarian / Vegan Restaurant	Ice Cream Shop	Market	Malay Restaurant	Shopping Mall

Illustration 4: Cluster 1 of Analysis 1

Clearly, areas like Anand Nagar, Fursungi, Warje are distant areas in city, but similarities in their venues can be seen in our analysis.

Now, let's check another cluster and check the similarity of locations in it.

Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Karanjepul	Mobile Phone Shop	Zoo	Donut Shop	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Factory	Event Space	English Restaurant
Pimpalgaon - Daund	Mobile Phone Shop	Zoo	Donut Shop	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Factory	Event Space	English Restaurant
Pargaon	Mobile Phone Shop	Zoo	Donut Shop	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Factory	Event Space	English Restaurant

Illustration 5: Cluster 2 of Part 2 Analysis

Here also, Karanjepul, Pimpalgaon – Daund and Pargaon are distant areas but similar in the venues around them.

3.2 Analysis Phase 2 (Similarity on the basis of venues in each location and distance from Industrial areas)

In this analysis we have used our dataset of previous analysis and added 10 columns of the distance of each location from 10 industrial areas used in our analysis. Feature scaling is then performed and finally by using K-Means algorithm locations are classified into clusters.

Let's, check all cluster locations identified by different colors in the city as follows: -

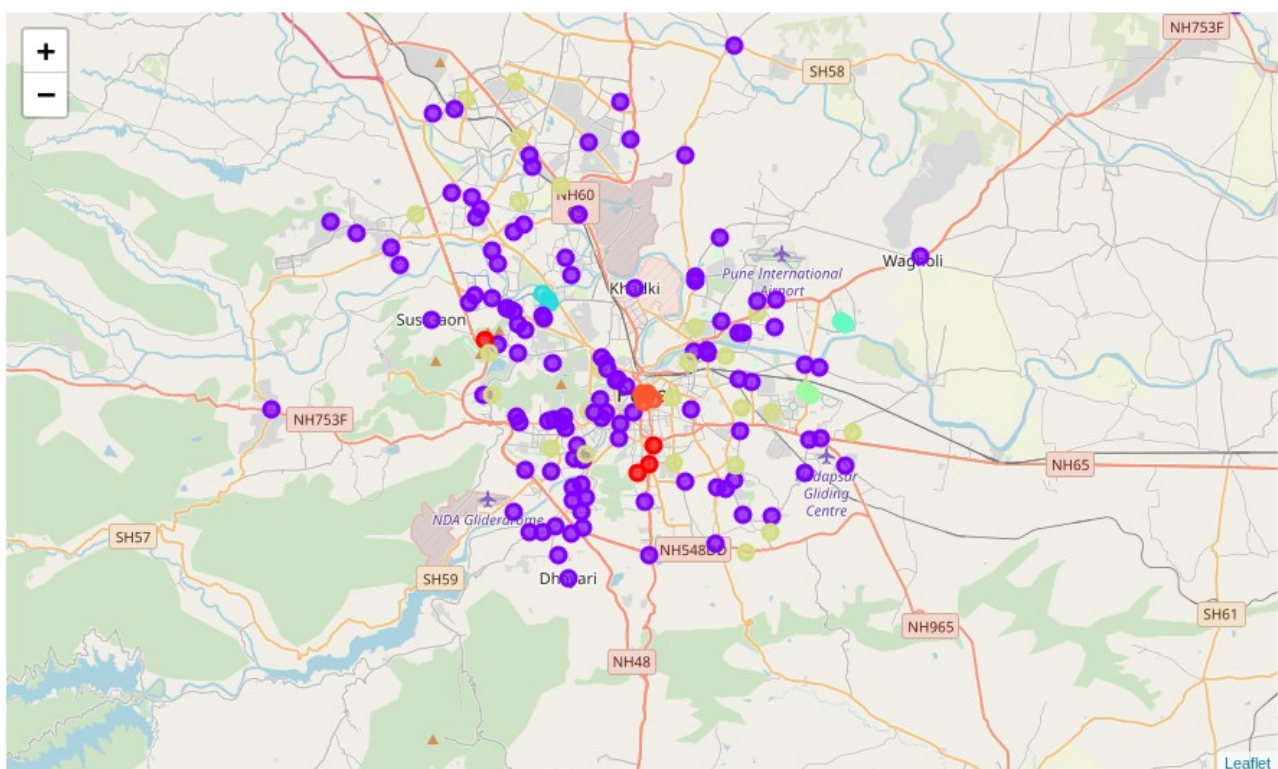


Illustration 6: Classification of locations through Analysis 2

The classification of locations is different in analysis 2 as compared to analysis 1. Thus, when distance from industrial areas is also considered for each location then the classification differs. These features have a great impact on our analysis.

Now, let's check some cluster locations from our analysis 2: -

Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	1st Most Common Industry	2nd Most Common Industry	3rd Most Common Industry
Satara Road	Indian Restaurant	Bakery	Ice Cream Shop	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Coffee Shop	Breakfast Spot	Shopping Mall	Fast Food Restaurant	Bistro	Kharadi Knowledge Park	Magarpatta City	Rajiv Gandhi InfoTech Park Hinjewadi Phase III...
Sahakar Nagar	Indian Restaurant	Southern / Soul Food Restaurant	Coffee Shop	Gym / Fitness Center	Ice Cream Shop	Shopping Mall	Breakfast Spot	Electronics Store	Bus Station	Bistro	Rajiv Gandhi InfoTech Park Hinjewadi Phase II	Rajiv Gandhi InfoTech Park Hinjewadi Phase I	Magarpatta City
Bhosle Nagar	Indian Restaurant	Coffee Shop	Southern / Soul Food Restaurant	Fast Food Restaurant	Asian Restaurant	Garden	Shoe Store	Tennis Court	Multiplex	Ice Cream Shop	Kharadi Knowledge Park	Magarpatta City	Rajiv Gandhi InfoTech Park Hinjewadi Phase III...
Pashan-Sus Road	Breakfast Spot	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Mountain	Italian Restaurant	Diner	Beer Garden	Ice Cream Shop	Food Court	Kharadi Knowledge Park	Rajiv Gandhi InfoTech Park Hinjewadi Phase III...	Magarpatta City

Illustration 7: Cluster 1 of Part 2 analysis

In this cluster also, Satara Road, Sahakar Nagar, Bhosle Nagar and Pashan-Sus Road are distant locations but have similarities in venues and distance from various industrial areas in the city.

Let's check another cluster from our analysis 2: -

Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	1st Most Common Industry	2nd Most Common Industry	3rd Most Common Industry
WTC-Kharadi	Coffee Shop	Pizza Place	Indian Restaurant	Irani Cafe	Café	Cafeteria	Pub	Go Kart Track	North Indian Restaurant	Diner	Magarpatta City	Rajiv Gandhi InfoTech Park Hinjewadi Phase I	Rajiv Gandhi InfoTech Park Hinjewadi Phase II
Eon Free Zone	Indian Restaurant	North Indian Restaurant	Coffee Shop	Fast Food Restaurant	Cafeteria	Café	Pub	Go Kart Track	Irani Cafe	Pizza Place	Kharadi Knowledge Park	Magarpatta City	Rajiv Gandhi InfoTech Park Hinjewadi Phase III...

Illustration 8: Cluster 2 of Part 2 Analysis

4. Results and Discussion: -

Our analysis shows that although there are various locations in Pune city. They can be classified into common clusters according to the venues around 1 K.M area of those clusters.

First we used all the locations in the Pune city and clustered them according to the frequency of venues around them and classified these 197 locations into 10 clusters.

In our second analysis, we choose 10 industrial locations and included them in our analysis also, to check for any difference in the clusters formed.

We found that the locations are classified differently in both clusters, the distance of each industrial location is also taken into consideration and then classified into clusters.

Although, the clusters are unevenly distributed in each cases, hence grouping most of the locations in 3-4 clusters only, which proves that Pune city has almost all similar set of venues available in each and every geography of the city. If a person/stakeholder tries to locate a place similar to another place at a far distant location, the probability of finding such an location is higher in our case.

5. Conclusion: -

The objective of this analysis was to classify similar locations of Pune city India. We have used two approaches to classify location across this city.

Although, the final decision of selecting a particular location will be made by our stakeholder 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), distance from work location, traffic on roads, levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.