# Capstone Project -

# The Battle of Neighborhoods

## *Contents*:



---

# *Introduction / Business Problem*:

I am an IT professional and I have been working in India for the past 15 years. Recently, I have been transferred to Pune City and I am in search of a location, which would be closer to my office but also should be well developed and having a lot of amenities. My office location is going to be Hinjawadi. Since the cost of the rented house was to be borne by my employer, the primary focus during search was on amenities around and the distance for transportation to the office.

After discussions with a few friends and a couple of searches on the web, it came to my understanding that there are two localities in Pune, which have been recently developed with modern amenities and the rate of development in these areas was fastest within the city. These areas were Aundh and Kothrud.

Therefore, I would primarily compare these two areas on two parameters.
(1) Number of amenities around the locality and
(2) Distance from my workplace

***Target Audience***: At present, the target audience for such problems is mainly me and my family members and also a group of persons from my company with their families, who are also relocated to Pune from other areas.

---

*End of Introduction / Business Problem Section*

## *Data*:

Datasets that would need to be referred to solve the above problem include ...

1) For understanding the amenities / venues in these two localities, we would be using 'Foursquare location data' as a data source. Both these locations would be explored for various amenities or venues within a specified radial distance, say 4 kms through API calls placed on 'Foursquare'.

   FourSquare is a local search-and-discovery app which uses location-specific data to help users find the best restaurants, bars, shops, entertainment, parks and nightlife in their area. When a user visits the location they can 'check-in', which tells their friends where they are, but they can also leave tips and reviews and even upload photos they've taken.

   Young people put great trust in peer reviews, particularly those written by their friends, which is why FourSquare is such a popular resource that is currently used by an estimated 55 million consumers around the world.

   I would be using the following request to explore the venues around desired localities. GET 'https://api.foursquare.com/v2/venues/explore....'.

   This request returns a list of recommended venues near the current location. As response to this request following information is retrieved from Foursquare in the form of a 'json' file, which could be further fetched in a dataframe.

| *Field* | *Description* |
|---|---|
| id | A unique string identifier for this venue. |
| name | The best known name for this venue. |
| location | An object containing lat & lng - latitude and longitude |
| categories | An array of categories that have been applied to this venue. |

The number and types of categories could then be used to compare the two localities to make an informed decision while choosing any of them.

*For Example*: following data is fetched from 'Kothrud, Pune' location from Foursquare, fetched in a dataframe and first five records are displayed.

```
df.head()
```

| | id | name | categories | lat | lng |
|---|---|---|---|---|---|
| **0** | 0 | Hidden Place - The Hangout | Pub | 18.509107 | 73.812280 |
| **1** | 1 | Cafe Coffee Day | Café | 18.500140 | 73.814254 |
| **2** | 2 | Barometer | Café | 18.498824 | 73.819240 |
| **3** | 3 | Endurance Fitness Club | Gym / Fitness Center | 18.499235 | 73.821117 |
| **4** | 4 | CCD | Café | 18.507830 | 73.808498 |

2) Location coordinates of office locality and proposed localities would be derived from 'Geopy'.

3) For knowing the travel distance to my office, I would depend on 'Geopy'. Geopy can calculate 'Geodesic distance' between two points using the function 'geopy.distance.distance'. The geodesic distance is the shortest distance on the surface of an ellipsoidal model of the earth. Another method, 'Great-circle distance' uses a spherical model of the earth, using the mean earth radius. However this method results in an error of up to about 0.5%.

Here is an *example* of distance.distance usage:

```
>>> from geopy import distance
>>> newport_ri = (41.49008, -71.312796)
>>> cleveland_oh = (41.499498, -81.695391)
>>> print(distance.distance(newport_ri, cleveland_oh).miles)
```

538.39044536

---

*End of Data Section*

# *Methodology*:

**Following libraries are imported and used.**

1) **numpy:** library to handle data in a vectorized manner
2) **pandas:** library for data analysis
3) **Json:** library to handle JSON files
4) From **geopy.geocoders** import **Nominatim:** to convert an address into latitude and longitude values
5) from **geopy** import **distance:** to calculate geodesic distance between two given coordinate sets
6) import **requests:** library to handle requests
7) from **pandas** import **json_normalize:** to transform JSON file into a pandas dataframe
8) **Matplotlib** and associated plotting modules to plot charts
9) import **folium**: library for map rendering

## Geopy.geocoders:

The office is located in the **'Hinjawadi'** area in the West of Pune City. From preliminary screening through friend circle and web-search, two localities were already chosen, **'Kothrud'** and **'Aundh'**. The latitudes and longitudes were derived for all these three locations from geopy.geocoders.

## Foursquare:

- A 'Developer' account was created on 'Foursquare.com' and sought 'Client_ID' and 'Client_Secret' codes, which are required as an input while making API calls to 'Foursquare'. The current date is used for providing the 'Version'.

- Since my main aim was to compare the two localities based on the development in the areas and amenities available in the vicinity, API calls were placed on 'Foursquare' for exploring the types and number of venues available within a radius of 4 kms in respect to the lat-long coordinates fetched from 'geopy'.

- The calls on each locality were placed in loop, fetching 100 venues on each call in order to fetch all venues within the desired region.

- The ‘json’ files retrieved through calls were normalised and then converted into dataframes for both the locations. Only the desired features like name of the venue, categories and their latlongs were fetched from the entire data.

### Folium:

The fetched venues were then plotted on a map in different colours using ‘folium’ to visualise the distribution and population of venues in the two localities.

### Matplotlib:

The two dataframes were merged on ‘categories.value_counts’ to understand the number of each type of venue category for comparison. The data was then sorted on the number of categories and plotted on a ‘bar-chart’ to visually compare the top 15 categories using ‘matplotlib’.

### geopy.geodesic:

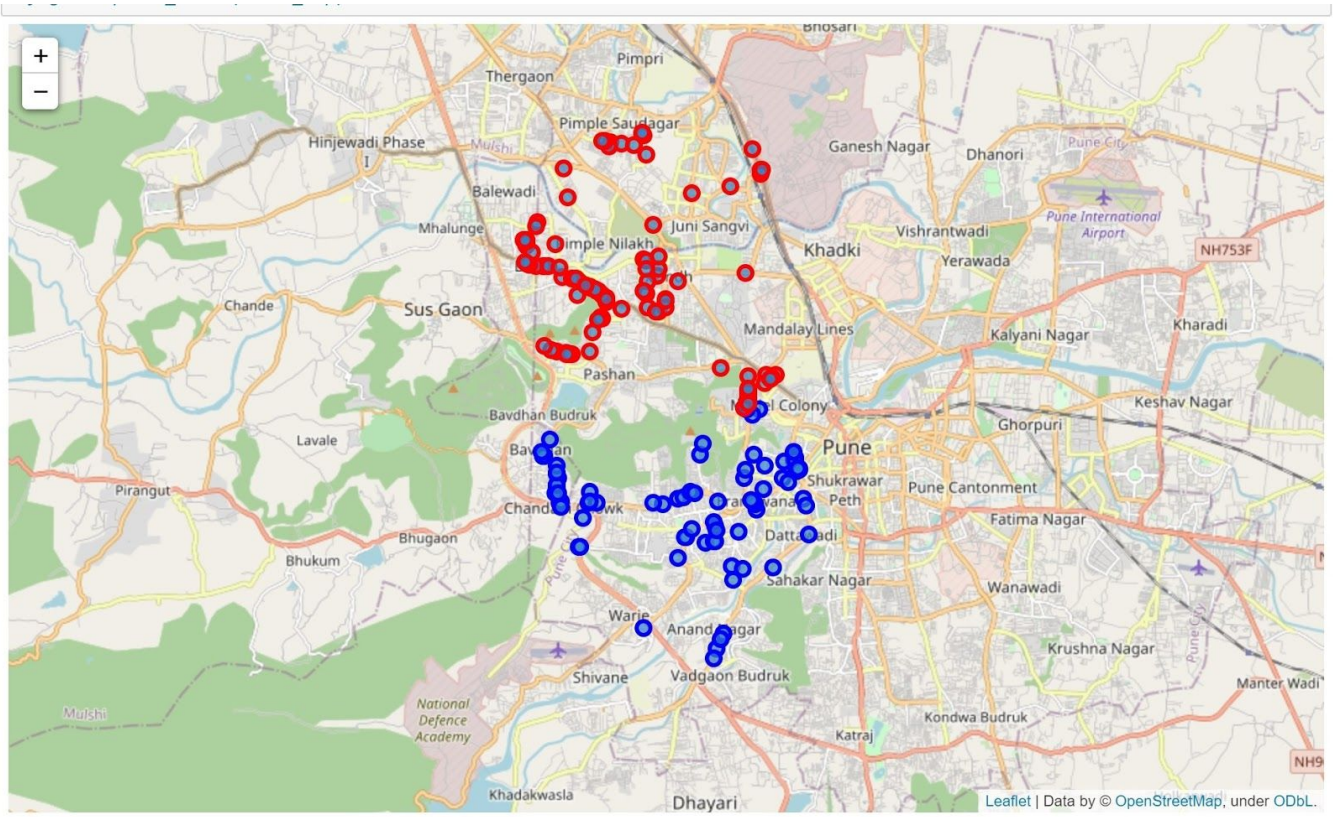
To calculate the distance of the two localities from the office, Geopy’s geodesic ‘distance.distance’ function was used.

---

*End of Methodology Section*

# *Results*:

(1) Plotting venues around Aundh and Kothrud on the map



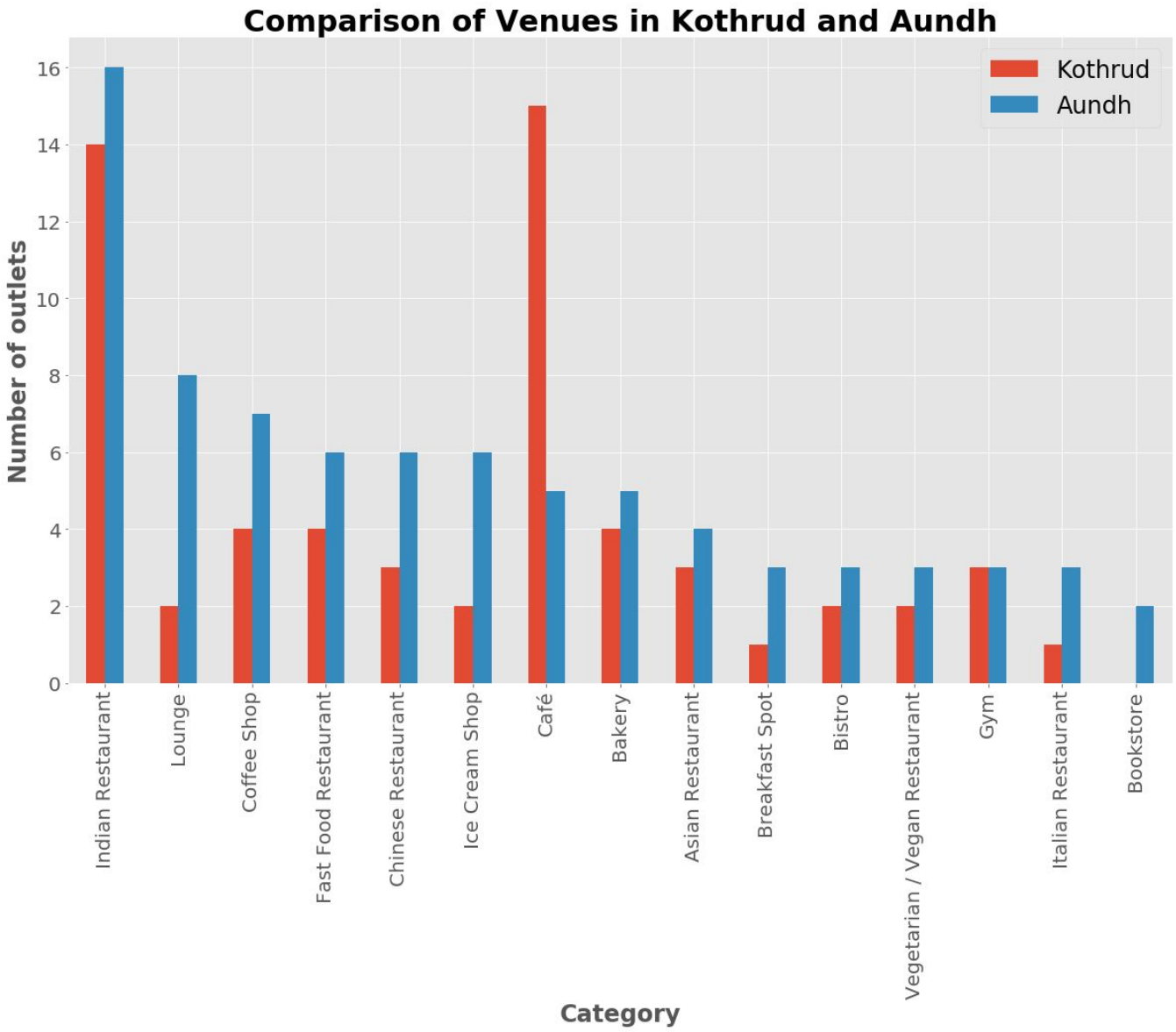

(2) Venues fetched around “Kothrud” from Foursquare

```
Total 96 venues fetched within a total radius of 4.0 Km for Kothrud Location

Total 96 venues fetched in Kothrud
```

(3) Venues fetched around “Aundh” from Foursquare

```
Total 100 venues fetched within a total radius of 4.0 Km from Aundh Location
Total 24 venues fetched within a total radius of 4.0 Km from Aundh Location

Total 124 venues fetched in Aundh
```

(4) Bar-chart comparing number of top-15 types of venues in Aundh and Kothrud



(5) Distance from office in Hinjewadi to both the proposed locations.

```
Distance from Hinjawadi to Kothrud:  12.149941753959334 km
AND
Distance from Hinjawadi to Aundh:  8.02677846599389 km
```

---

*End of Results Section*

# *Discussion*:

## ***Observations:***

From the above results, following are the observations …

(1) From the plot on the map, it is evident that the venues are more densely packed around Aundh than those in around Kothrud.
(2) Also it is observed that the Aundh Location is nearer to Hinjewadi compared to Kothrud Location.
(3) Total number of venues fetched in Aundh within a radius of 4 kms are more in numbers (124 venues) in comparison with those fetched in Kothrud (96 venues).
(4) The bar graph for top-15 venues also show that most of the types of venues are more in numbers in Aundh than those in Kothrud.
(5) The geodesic distance from the office to Aundh is lesser (8.0 kms) in comparison to Kothrud (12.1 kms).

## ***Recommendation:***

Now, we have the data to compliment our business problem.

(1) Both ‘Aundh’ and ‘Kothrud’ have modern amenities. However, ‘Aundh’ area has got more number of outlets in each category of venues.
(2) It is evident from map and geodesic distance that the ‘Aundh’ Location is nearer to ‘Hinjewadi’ compared to ‘Kothrud’ Location.

Looking at these results, it is recommended that the ‘Aundh’ area should be a preferred choice over ‘Kothrud’.

---

*End of Discussion Section*

# *Conclusion*:

As a solution to the business problem, defined in the Introduction section, it is evident from the analysis and results that ..

## ‘Aundh’ area is the preferred choice over ‘Kothrud’ area because -

(1) It has more number of various categories of venues and also the total number of venues as well to live a desired lifestyle.

(2) The travel distance to the office in ‘Hinjewadi’ would be shorter from ‘Aundh’ as compared to ‘Kothrud’, leading to shorter travel time and lesser fuel consumption.

---

*End of Report*