**Coursera Capstone**

## Capstone Project - The Battle of Neighborhoods

**Project Report**

**By : Tushar Nirmal**

# Problem Statement

Pune is the [second largest](https://en.wikipedia.org/wiki/List_of_cities_in_Maharashtra) city in the [Indian state](https://en.wikipedia.org/wiki/States_and_union_territories_of_India) of [Maharashtra](https://en.wikipedia.org/wiki/Maharashtra), after [Mumbai](https://en.wikipedia.org/wiki/Mumbai). It is the ninth [most populous city](https://en.wikipedia.org/wiki/List_of_most_populous_cities_in_India) in the country with an estimated population of 3.13 million .

According to the [2011 census](https://en.wikipedia.org/wiki/2011_Census_of_India), the urban area has a combined population of 5.05 million while the population of the metropolitan region is estimated at 7.27 million.

The city is considered to be the cultural capital of Maharashtra.  The city has emerged as a major educational hub in recent decades, with nearly half of the total international students in the country studying in Pune. Research institutes of [information technology](https://en.wikipedia.org/wiki/Information_technology), education, management and training attract students and professionals from India and overseas.

Therefore, in addition to its rich heritage and cultural value combined with multi-ethnic population ,it is home to great food restaurants with touch of every flavour and a page taken from every culture’s recipe . It has already become a very interesting place for food lovers. Whether it is working professionals or students or people who are travelling, everyone loves good tasty food.

The target of this project is to use Foursquare location data to determine what might be the best neighbourhood in Pune to open a restaurant. As fast food demand is increasing and taste resonates with maximum people , we will be focusing on finding the best neighbourhood for opening a fast food restaurant .

People who are looking to open restaurant in Pune will be interested in this project . As this project gives the best location to open a fast food restaurant for the client , It can turn into most profitable and visited by as many people as possible . Also , based on client’s requirement , same project can be applied for any category of venue . Such wide use can be taken advantage of by the client and thus is a very useful project as per current business and marketing opportunities.

# Data Description

Following sources of data will be used for getting the required information for our capstone project : -

1. **The list of neighborhoods of Pune, India from Wikipedia url :**

**“https://en.wikipedia.org/wiki/List\_of\_neighbourhoods\_in\_Pune"**

1. **The coordinates ( latitude , longitude ) of the neighborhoods of Pune from Open street Map APIs from url –**

**“http://nominatim.openstreetmap.org”**

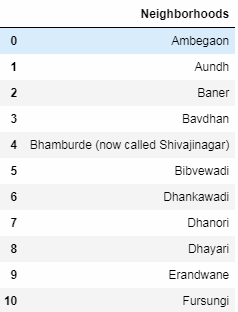
1. **From foursquare.com , we need to determine the proximity of various amenities as per requirement .Mainly we will need the following data :**

* **Details of all the venues of Pune neighborhoods within 1000 meters of the coordinates**
* **Details of office venues of Pune neighborhoods within 1000 meters of the coordinates .**
* **Details of high school venues of Pune neighborhoods within 1000 meters of the coordinates .**
* **Details of university venues of Pune neighborhoods within 1000 meters of the coordinates .**

**Methodology**

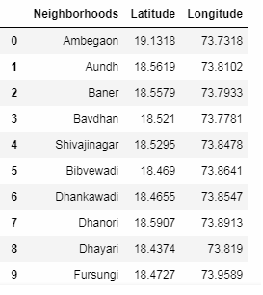
* The business solution provided is divided in the following steps -

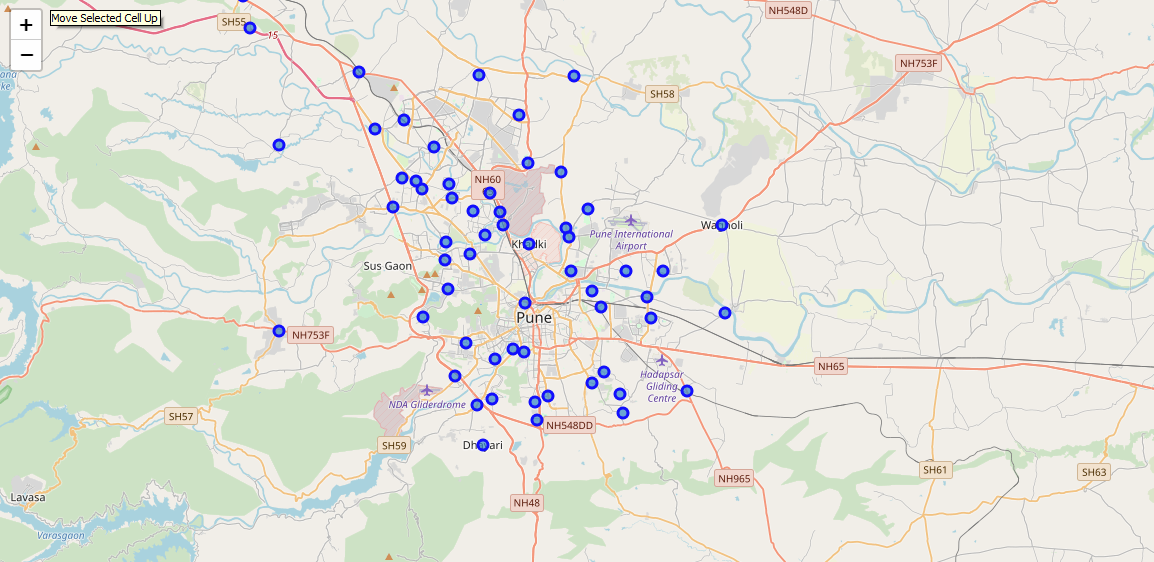
Firstly , we need to have the name of all the neighborhoods of Pune in order to find the best one for opening a restaurant . Same can be obtained from a Wikipedia url which contains all the names. These names are fetched using beautifulsoup package and is stored in a dataframe . After successfully fetching the neighborhoods name from Wikipedia , the dataframe will look as :



* Step – 2

The dataframe obtained above contains names of neighborhoods of Pune . To get the latitudes and longitudes of these neighborhoods , we will use the Open street Map APIs . Once we get all the coordinates , we will add these coordinates to our dataframe . There are few neighborhoods for which we are not able to fetch coordinates . We will manually add these coordinates by searching on google . Once that is complete , dataframe will look as :





* Step – 3

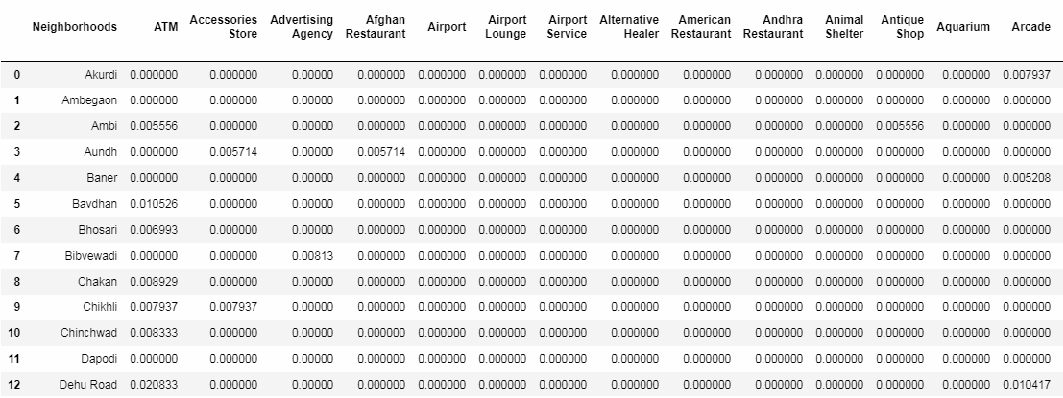
Using Foursquare.com , we will find proximity of various amenities from each neighbourhood of Pune . The resulting dataframe will contain following columns – neighborhood name , neighborhood latitude and longitude , venue name ,venue latitude and longitude and venue category .



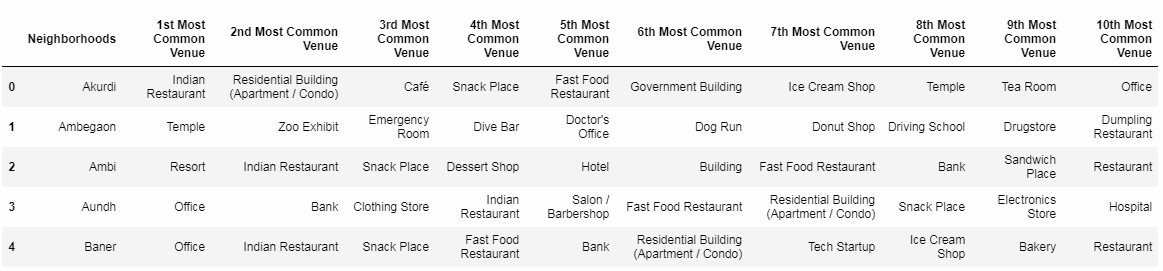
* Step – 4

We want to know the which category is the most common in each neighborhood . To know the top most category for a neighbourhood , we will do it in two sub steps :

* + - Firstly , we will group by the dataframe by neighborhood and convert each category in the column “Venue Category” from categorical to numerical (0 if category is not present for that neighborhood and 1 if present ) and then taking the mean for all the columns for each neighborhood . As a result , the resulting dataframe indicates the frequency of that venue’s category for that neighborhood . The greater the value , the greater the number of entries for that category . The dataframe will look like –

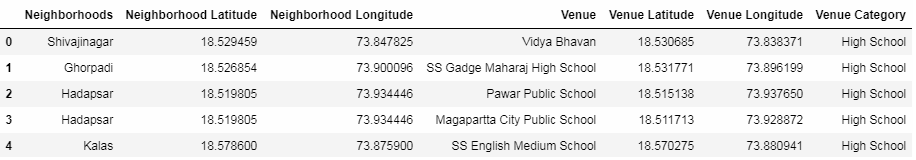


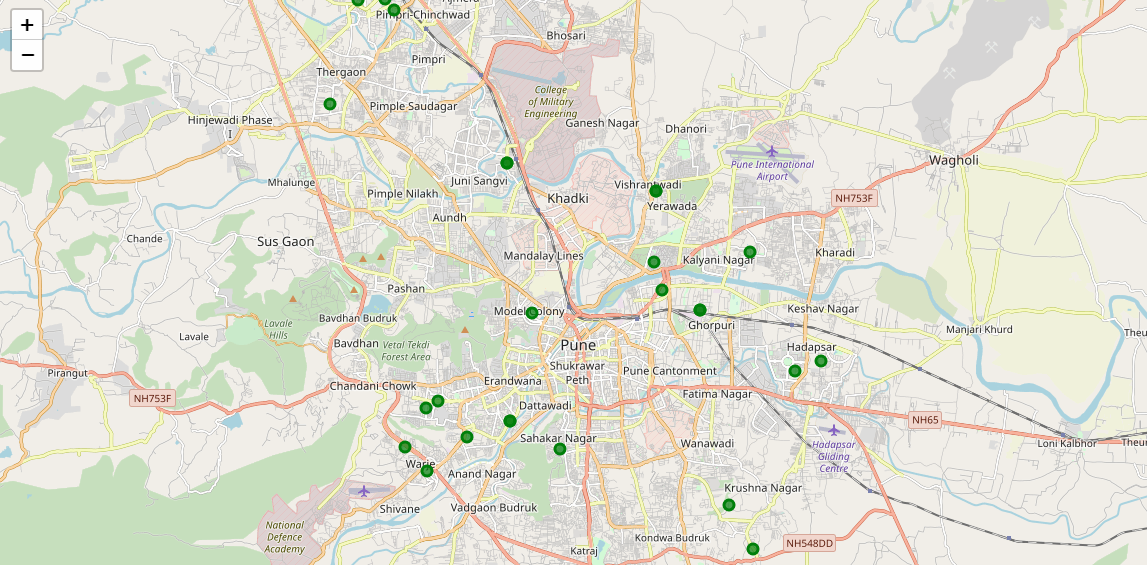
* + - Secondly , for each neighborhood , we will sort the row in the descending order in order to get the most common venues (we have written a function which returns the values of the most common venues for each neighborhood). Therefore the resulting dataframe will look like –



* Step – 5

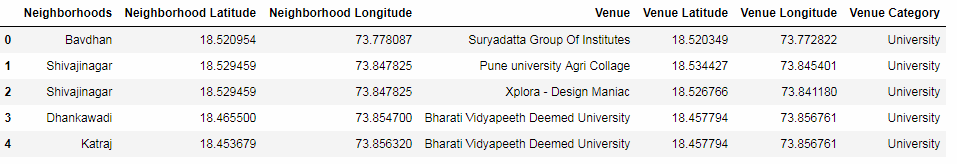
Using Foursquare.com , we will find proximity of all the high schools from all the neighborhood of Pune . The resulting dataframe will have the coordinates of high schools along with coordinates of neighborhood –

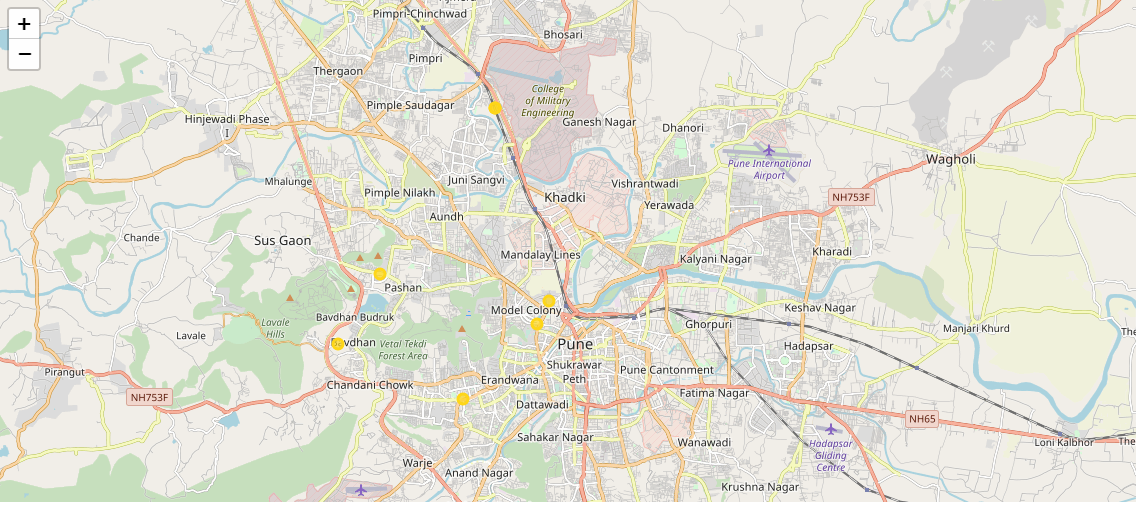




* Step-6

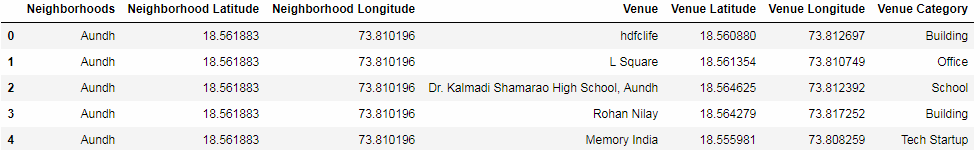
Using Foursquare.com , we will find proximity of all the university from all the neighborhood of Pune . The resulting dataframe will have the coordinates of universities along with coordinates of neighborhood –

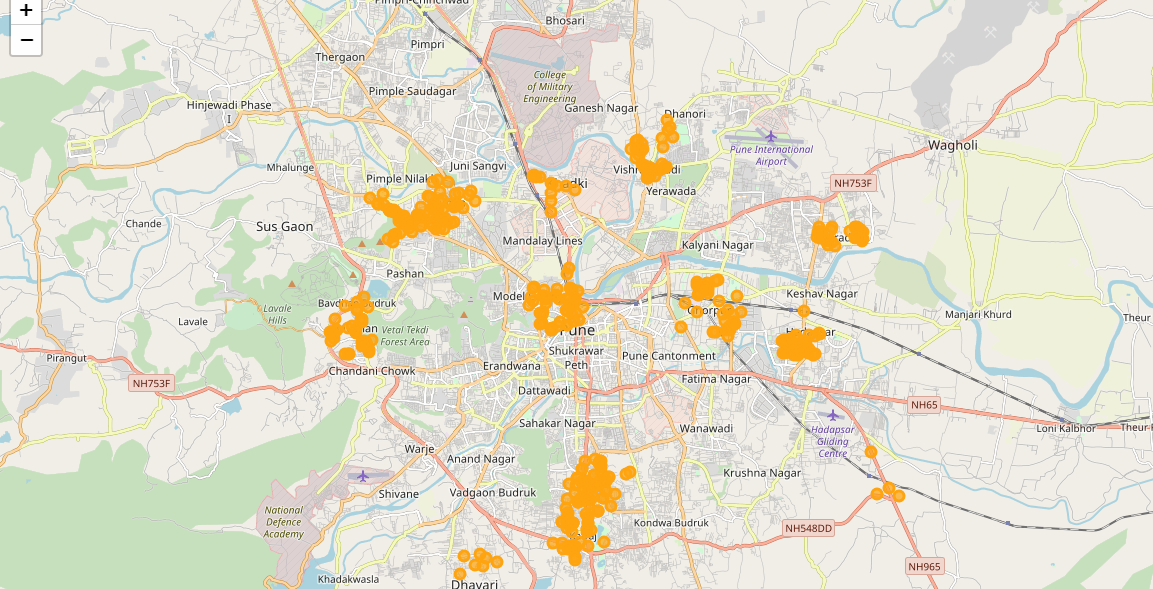




* Step – 7

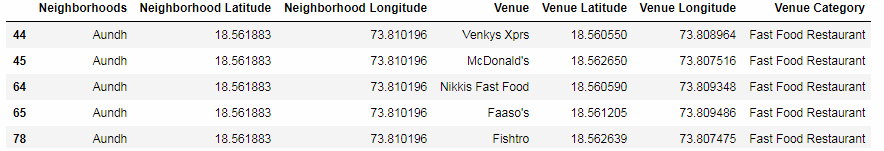
Using Foursquare.com , we will find proximity of all the offices from all the neighborhood of Pune . The resulting dataframe will have the coordinates of offices along with coordinates of neighborhood –

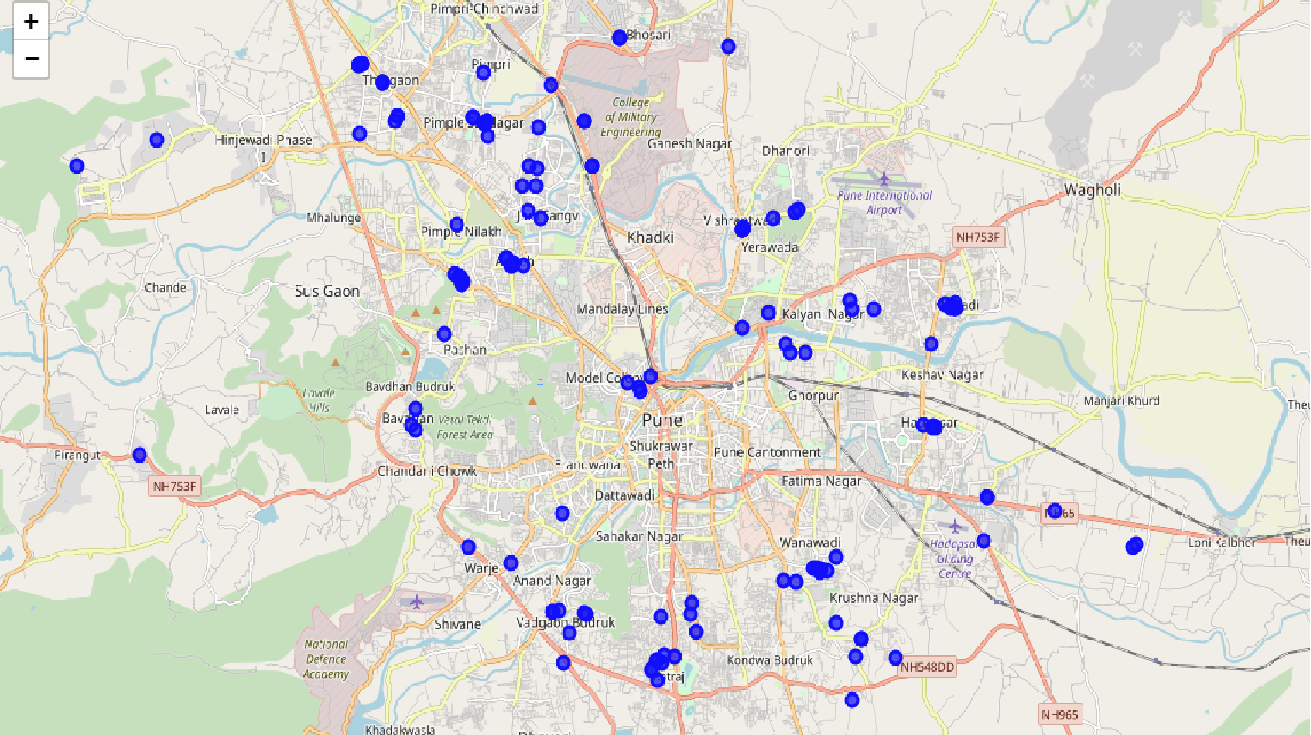




* Step -8

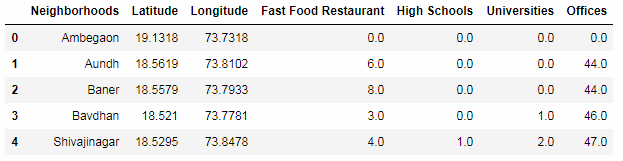
From the dataframe obtained in step-3 ,we fetch all the rows which has venue category as fast food restaurant . Thus the resulting dataframe will look as below –





* Step -9

We write a function to add columns (the value of each element is the count of the category for that neighborhood ) of above dataframes into the original dataframe (dataframe which contains coordinates and names of neighborhoods . )



* Step-10

When we multiply these rows with their weights and sum each row , we get a score for each neighborhood. The weights are defined as below :

* + - Schools have been weighted as 1 , since students are good customers.
    - Universities has been weighted as 1.5 , since college students are also good customers and have better spending ability as compared to school students.
    - Offices has been weighted as 2 , since employees are even better customers .

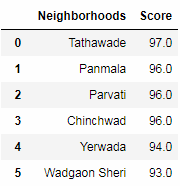
The neighborhood with the maximum score is our desired neighborhood for opening the Fast Food Restaurant

In order to conduct a similar analysis for any other category , we just need to change the category from “Fast Food Restaurant” to any other as mentioned by client .

Results :

The neighborhood with the maximum score is our desired neighborhood for opening the Fast Food Restaurant i.e. “Tathawade” with score of “97.0”

The second best option is “Panmala” with score of 96.0 .



**Discussion :**

Based upon the findings in the result section, the user can get the best neighborhood for opening a fast food restaurant .

For more optimization of the business solution , there are several recommendations which can be included in the future versions of this business problem –

* Several more categories can be included for calculating the score of a neighborhood along with schools , universities and offices . For example , there are several more categories like temples,banks ,clothing stores , airports/stations where people generally are present in large numbers. If our restaurant is present is vicinity to all these categories in addition to the ones included before , the accuracy of business solution can increase.
* The weights to these categories should be given as per their important and footfall of people around them who will be interested in going to a restaurant .
* Client can be asked to merge more than one categories to open a restaurant . For example , if in a neighborhood , two top most categories are Indian Restaurants and Chinese Restaurant . Therefore if a restaurant is opened with both the specialties , there will be extra changes of success because more people can visit the restaurant .

**Conclusion :**

In this project , I analyzed location based data and found the best neighborhood in pune for opening a fast food restaurant .Given the names of the neighborhoods , I learned to find its coordinates with the help of open street api . Given the coordinates of a point, I also learned to find various amenities near it with choice of distance from that point. By using various data cleaning and analysis steps , I was able to get the top most common categories of each neighborhood .

Most popular and common place where our customers could be present was taken into account separately and for each such category of venue, data was collected. The number of venues of every category in each neighborhood was calculated and combined together in the final dataframe.

Then building the scoring criteria by assigning weight to each category for which data was calculated separately was very crucial in finding the best neighborhood.

Thus the final business solution is very helpful for a client who is looking to open a new fast food restaurant in the Pune city .