# **The Battle of Neighborhood: Bangalore VS Hyderabad**

Siddharth Datthathreya

Email: siddharthdathathreya@gmail.com

## 1. Introduction

### 1.1 Background

Hyderabad and Bangalore, they are the metropolitan cities of my country called India. They are known for their technological advances and development in the IT Sector. The people who wants to come to this place is mainly for the jobs and luxurious lives. For that we need to know what are the frequent venues and where it will be. This places are also known for the rich hotels and every corner is having restaurants/ street foods. As I have lived my life in both of these states , I would like to recommend this is the best places to live, here is why:

1. You'll have great society
2. Meet other like-minded travelers and make new friends around the world!
3. So many opportunities that you can grab and many restaurants and many, many more.

### 1.2 Business Understanding/Problem Description:

As the people doesn't know where to live or start to work in a country, I am representing in this report how to start and which colony is having all the resources such as utilities, groceries, gym, IT services etc. And one more thing

* Which is the good place to start?
* Where we can we get all the resources as I mentioned above?
* In the given neighborhood, is there any good rating services like online-ordering food is available?

As people who migrate from other states, other countries, these two are the most recommended places stated by the TOI in our country for the luxurious life and cleanliness. The Bangalore is also known as Silicon City of India. Hyderabad is getting developed in much more stages nowadays. I will try to answer the following questions

* Where is the good place to start in Hyderabad or Bangalore?
* Where are the *'amenities'* located?
* Recent activities and Most common venues
* Suggests the best place to start either in Hyderabad or in Bangalore.

### 1.3 Target Audience

This project will serve two groups of audience:

1. **Travelers:** Help them make an informed decision while choosing a place by providing an in-depth analysis of place and their neighborhood.
2. **Non Residents:** Provide useful information and models which can help them where to open their first/next place to live.

## 2. Data Requirements

I analyzed in this project: **Hyderabad and Bangalore**. Following are the datasets used in the project:

1. [Hyderabad Pincode](https://finkode.com/ap/hyderabad.html)
2. [Bangalore Pincode](https://finkode.com/ap/bangalore.html)

#### 2.1 Hyderabad Pin code:

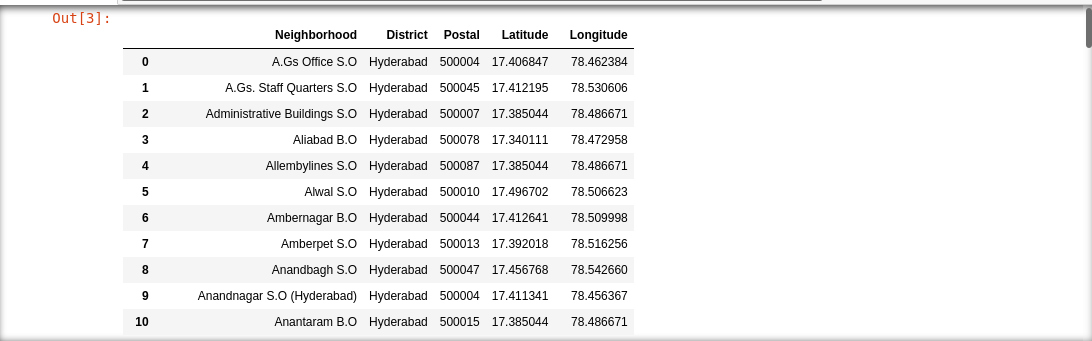
Using the web scrape method using beautifulsoup I have extracted the following columns:

* Neighborhood
* Pin code
* District

Below is a snapshot of the Hyderabad dataset:



After getting the latitudes and longitudes the values will be like:



#### 2.2 Bangalore Pin code:

Using the web scrape method using beautifulsoup I have extracted the following columns:

* Neighborhood
* Pin code
* District

Below is a snapshot of the Hyderabad dataset:



After getting the latitudes and longitudes using Google maps API the values will be like:



## 3. Methodology:

### **3.1 Data Collection:**

* The Hyderabad and Bangalore pin code were scraped from the respective sites. (<https://finkode.com/ap/hyderabad.html>) and (https://finkode.com/ka/bangalore.html)
* We used Foursquare API to get the venues around the hostel.
* We acquired the latitudes and longitudes data by using Geocoding API from google cloud.

### **3.2 Analytic Approach:**

I took two approaches in the project.

Firstly, I used exploratory data analysis(EDA) to find the given dataset is reliable or not by using the data preprocessing methods and normalize the given json. Directly, by using the BeautifulSoup package, we are able to retrieve the data from the given site as mentioned above. After retrieving, I used the Geocoding API from Google maps with the relevant credentials and found the latitudes and longitudes. Later I plotted those in the map using Folium package.

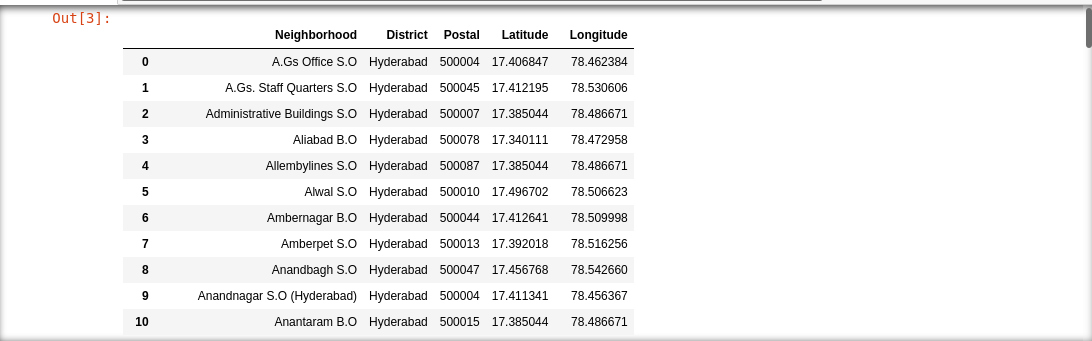
Secondly, I used prescriptive analytics to help finding a perfect place to live whether it might Hyderabad or Bangalore. I will use clustering(K-Means). I will find the most common venue of a particular place where we / the people can actually live in that place with luxurious life with all the amenities present in the neighborhood.

**4. Analysis**

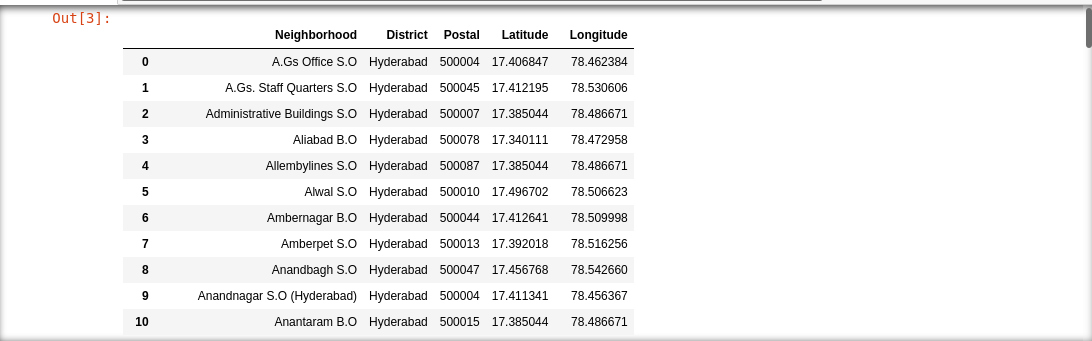
### **4.1 Exploratory Data Analysis**

First, we have to gather the places as the dataset regarding the Hyderabad and Bangalore is not easily found. So, I have acquired the data from the sites mentioned above and filled it with latitudes and longitudes using the Geocoding API from Google Maps.

The below dataset is about Hyderabad district and it is acquired from the site mentioned above:



The below dataset is about Bangalore district and it is acquired from the site mentioned above:

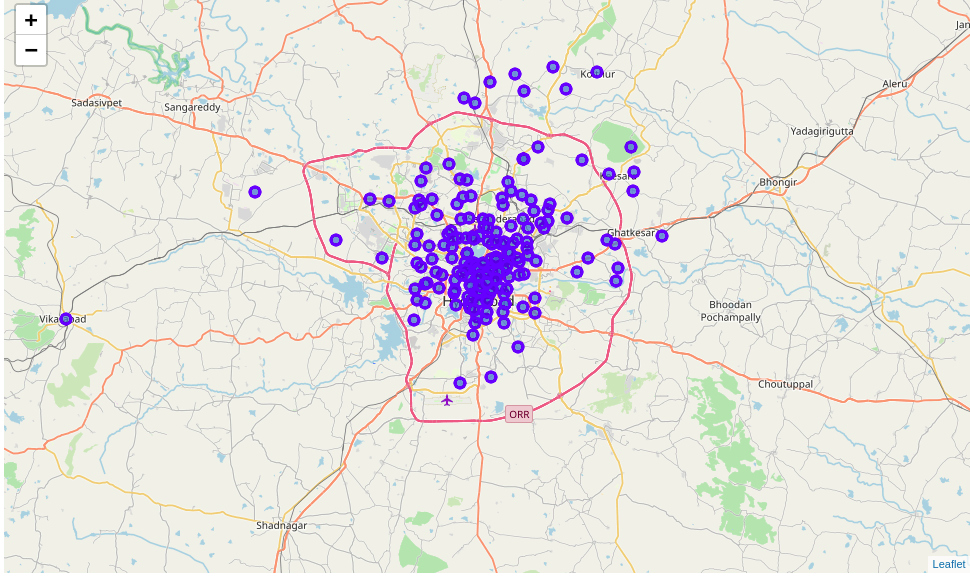


To acquire those datasets, first we need to scrape the data from the table using BeautifulSoup (bs4 package) and converting it into html parser. Later we list it in the form of the lists and store it in a variable. After acquiring the lists, convert those lists into data frame using the pandas' package. Remember we are using python language for the analysis of data. Later we import the json library and google maps package for accessing the latitudes and the longitudes of the given dataset as we have seen above. I used the credentials key, got from the google cloud console. It will take five minutes to gather all the latitudes and longitudes. When we are gathering the data, it will be stored in the form of JSON. We need to convert that into list using the json package. And store it in the data frame. The given data will be formed as we are seeing above.

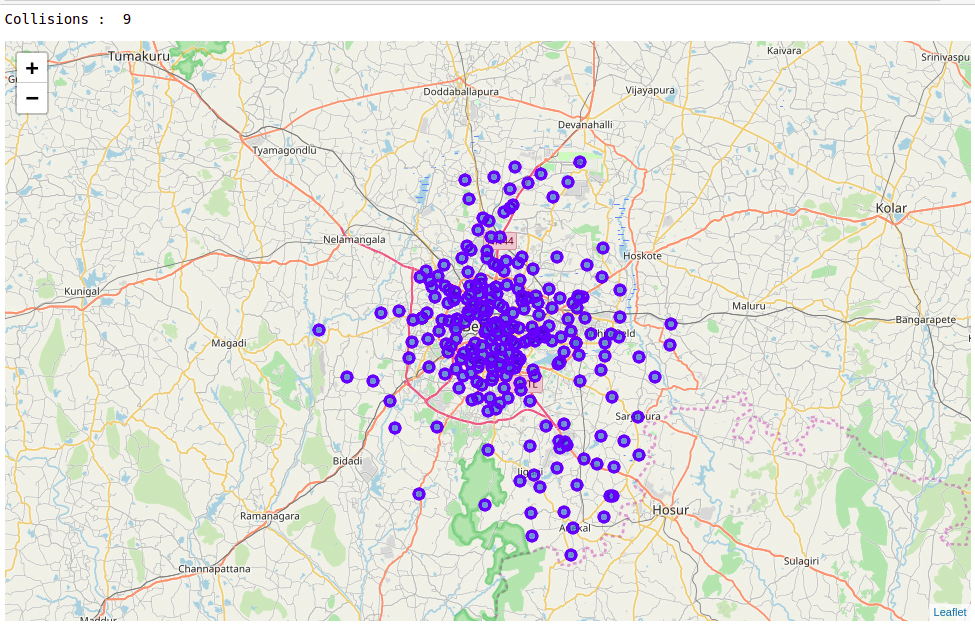
After the analysis, we are finding the number of collisions in the given data set. A **collision** or **clash** is a situation that occurs when two distinct pieces of data have the same hash value, checksum, fingerprint, or cryptographic digest. For the Hyderabad data set we got 14 and for the Bangalore data set we got 9. The total number of rows of Hyderabad and Bangalore data set are 202 and 269 respectively. As I am using the pin code of the separate districts, I will be getting these values.

Later we will plot those formed data into the map using the Folium API.

The below plotted map is for Hyderabad district:



The below plotted map is for Bangalore district:

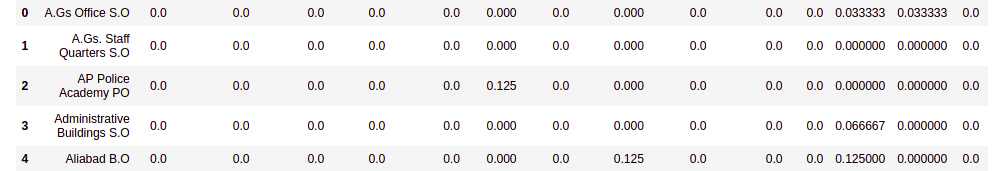


To access the venues from the exact pin code address we need to use the FourSquare API. FourSquare API allows us to access the venues and most common venues based on the people’s ratings and contact numbers and a lot more. We can use this for acquiring the data regarding the neighborhood places for the amenities and such. To activate the FourSquare APi, we need to use the credential ID and credential key which will be provided from FourSquare developer's console. We have to login to access those details or sign up. We need the category table from the pin code so I have acquired the data set.

After acquiring the data set, proceed to store it in another data frame so that the original data set won’t be lost. The explored dataset will be converted into one-hot encoding for the prediction analysis therefore we will convert it. The below are the datasets which are converted into one-hot encoding I.e. acquired from the explored data frame.

The below is for Bangalore and Hyderabad one-hot encoded data set:

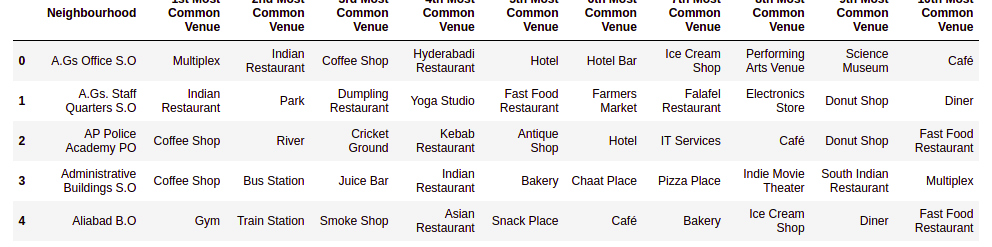




We have converted it into one-hot encoding because we have to find the prediction at which place the most common venue will be there for the clustering the labels. That is why, we have converted the categorical table into one-hot encoded form.

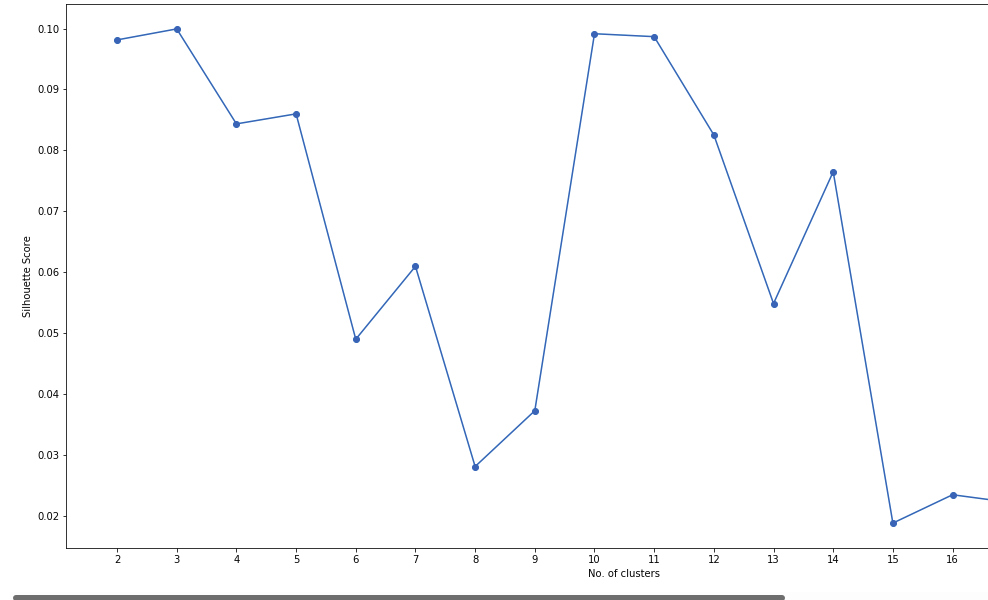
After that we will find the places which are having the most common venue by using the function created in the code and we will get the Bangalore and Hyderabad data set as seen below:

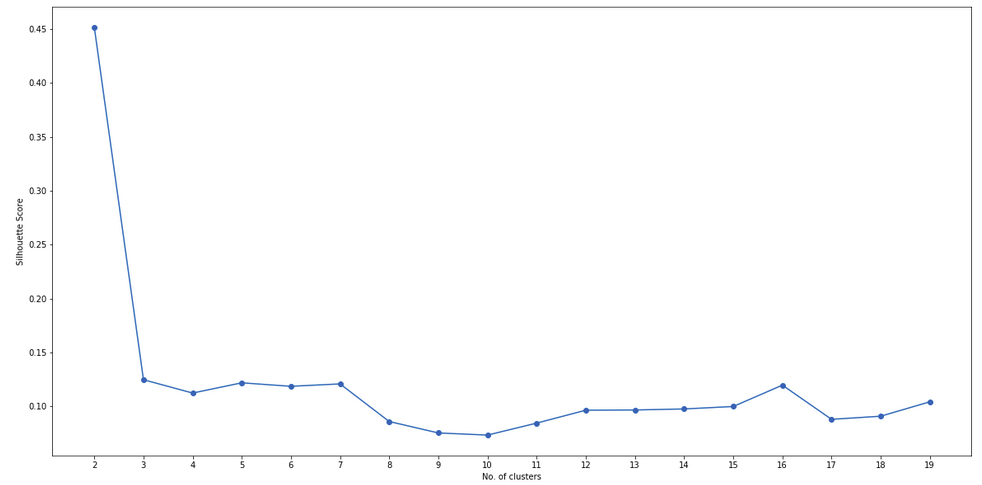




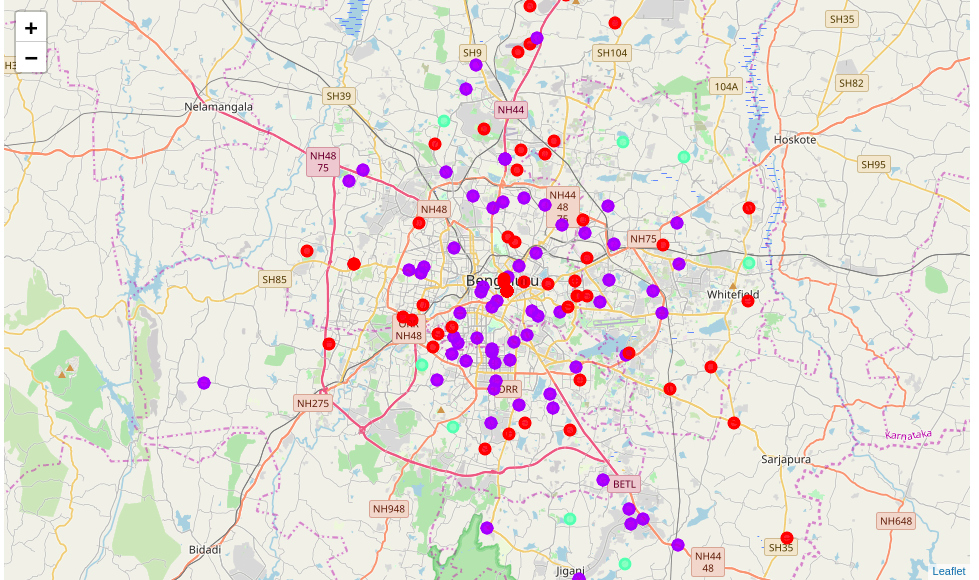
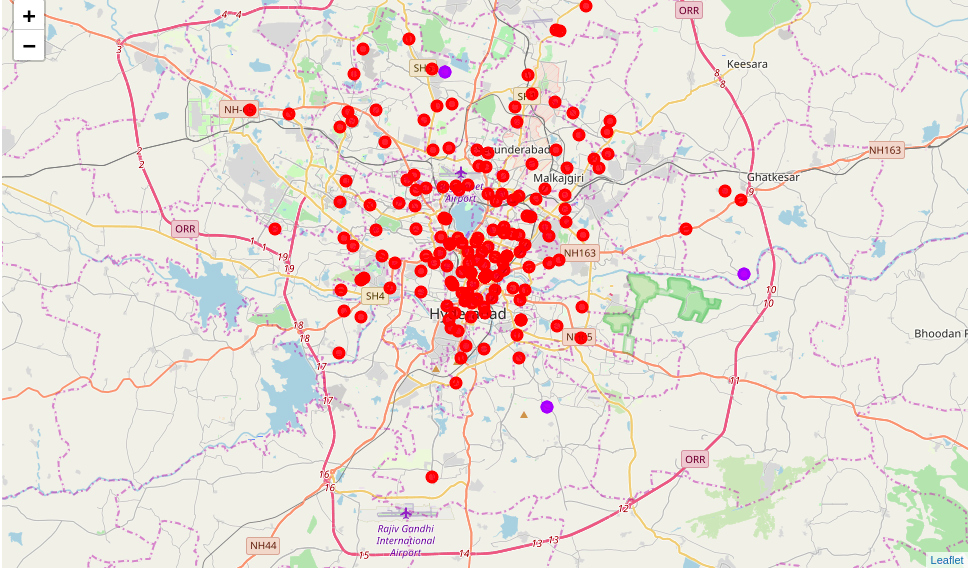
### **4.2 Clustering**

After the above step, we have to find the K value for the the clustering analysis of the K-mean Clustering algorithm. Therefore, we will be using the silhouette method to find the optimized k-value. For the Bangalore data set we will get 3 as the optimized value, whereas for the Hyderabad data set we will get 2 as the optimized value for the K. We will find it by plotting it in the graph as shown below for the Bangalore and Hyderabad data set respectively.

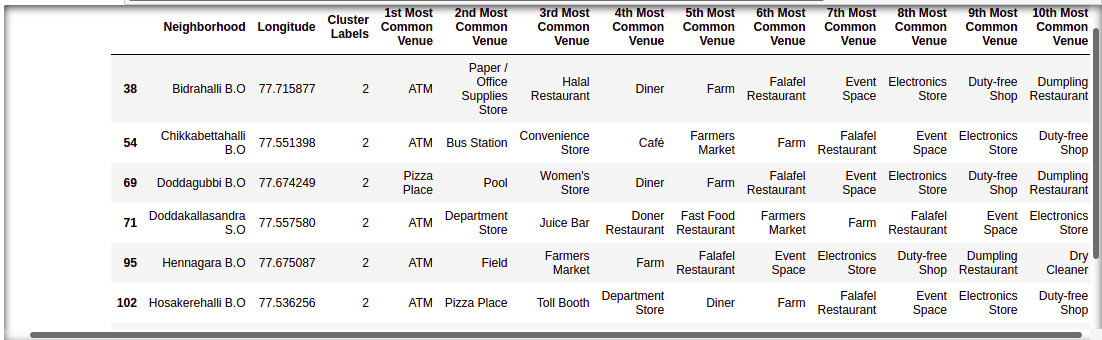




Now we will geo-visualize the data using the K-means Clustering algorithm for both the data sets Hyderabad and Bangalore respectively:

We will see that the cluster labels are maximum up to 2 for Bangalore whereas for Hyderabad it is just 1 because of the places are very highly near to each other and the recommendation venue is very high in those places.

Below is the final cluster data set for the Bangalore and Hyderabad:





**5. Result**

We got a glimpse on where to start to live in these cities which I have taken as my project I.e Hyderabad and Bangalore. In Hyderabad, the best place to start is, having three places whereas, in Bangalore there are eight places to start with. All of the mentioned places are having all the amenities and IT sectors are booming in both of these districts. I conclude that no matter, where we start, start from a place where it is comfortable for you in terms of language, society etc. which is of course, out of presentation topic. The above-mentioned places are best places to start whether it is Bangalore or Hyderabad.

**6. Discussion**

According to the above analysis, the best place to start to live in Bangalore are Bidrahalli, Chikkabettahalli, Doddagubi, Doddakallasandra, Hennagara, Hosakerahalli, Hulimangala and Kadugodi. And the best place to start to live in Hyderabad are Badangpet, Jeedimetla and Pratapsingaram.In those above places all of the required essentials are available in the given neighborhood.