### Capstone Project – Student Hostel Near Engineering Colleges in Hyderabad

### Introduction / Business Problem

### Hyderabad is the capital and largest city of the South Indian state of Telangana. It has a population of over 20 million people spread across 650 square kilometres. Being one of the top IT services companies’ destination in the country, Hyderabad is home to 131 engineering colleges and attracts thousands of students from across the country.

### One of my friends is interested in setting up a student hostel near engineering colleges and tasked me on zeroing an area with highest density of engineering college, so that he has a flourishing business.

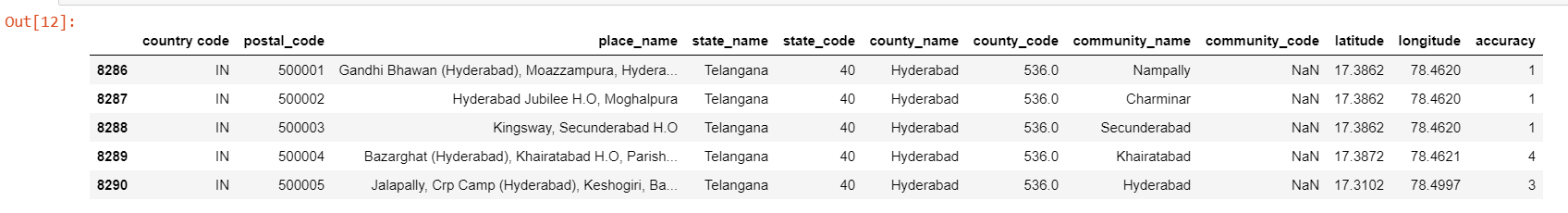
### Through this project I will explore Hyderabad for engineering colleges and find area/neighbourhood with highest density of engineering colleges.

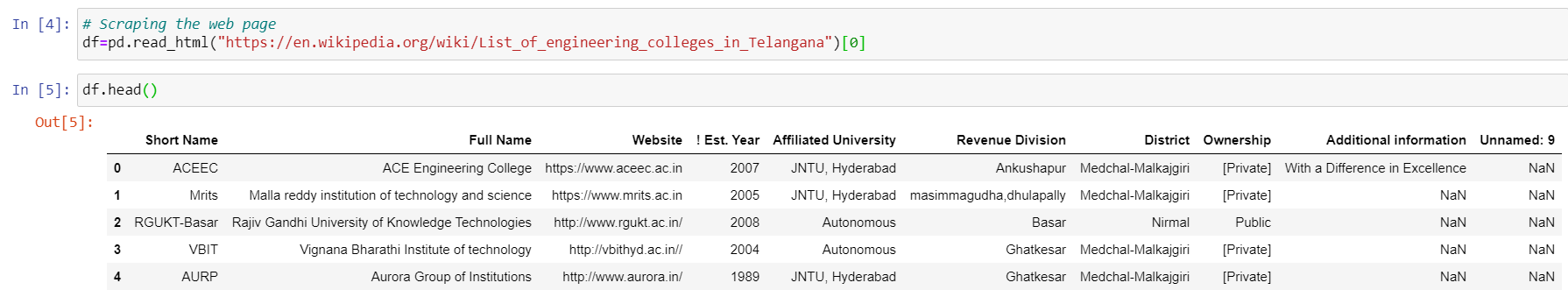
### The Data Section

* I made use of a python library called ***pgeocode* (**[**https://github.com/symerio/pgeocode**](https://github.com/symerio/pgeocode)**)** which is a high-performance off-line querying of GPS coordinates, region name and municipality name from postal codes. Distances between postal codes as well as general distance queries are also supported. The used GeoNames database includes postal codes for 83 countries.

I used '*index* postal\_codes' function which creates a data frame of unique postal codes of a given country. The data frame consists of following columns:

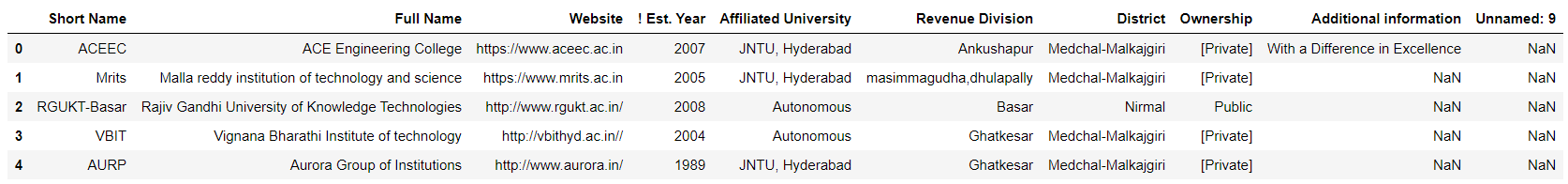
* + country code: iso country code, 2 characters
  + postal code: postal code
  + place name: place name (e.g. town, city etc)
  + state\_name: 1. order subdivision (state)
  + state\_code: 1. order subdivision (state)
  + county\_name: 2. order subdivision (county/province)
  + county\_code: 2. order subdivision (county/province)
  + community\_name: 3. order subdivision (community)
  + community\_code: 3. order subdivision (community)
  + latitude: estimated latitude (wgs84)
  + longitude: estimated longitude (wgs84)
  + accuracy: accuracy of lat/lng from 1=estimated to 6=centroid
* Using this function, I first I obtained the postal code details for the entire country India and then queried the data to retain the postal codes of Hyderabad. The sample data is shown below:



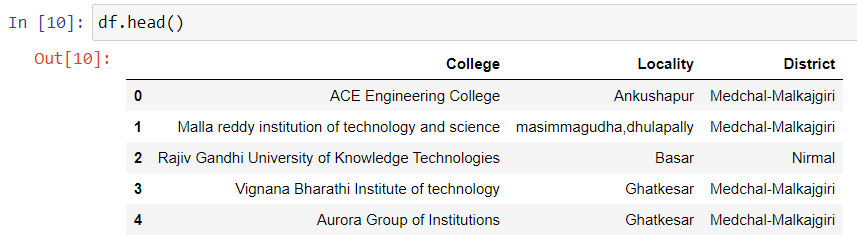


### Methodology

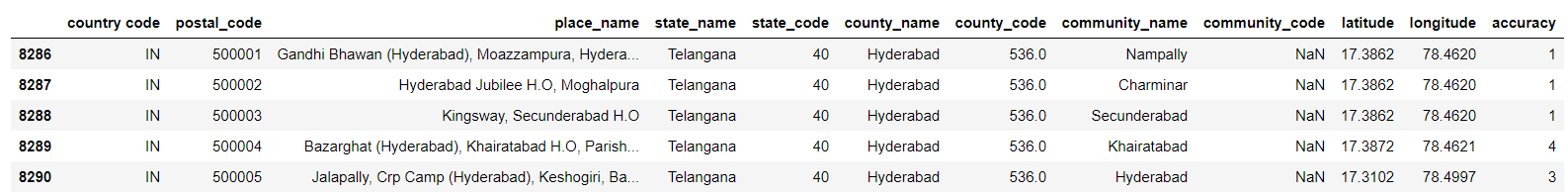
* I scrapped web to get the list of engineering colleges in Hyderabad. The data was pulled from <https://en.wikipedia.org/wiki/List_of_engineering_colleges_in_Telangana>



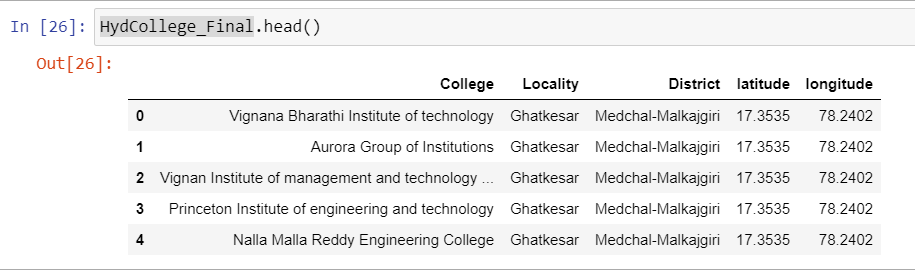
* I then dropped unnecessary columns and rows, renamed columns to get list of colleges with localities as shown below:



* Since the college list did not had Geo Location codes, I used pgeocode library to get geo location coded for localities of Hyderabad.



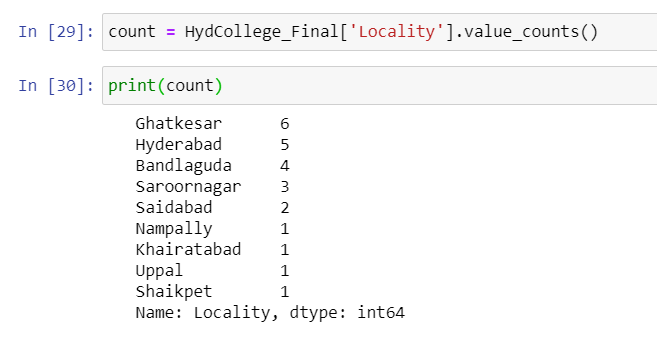
* I then merged college list with localities list to get geo location codes for colleges



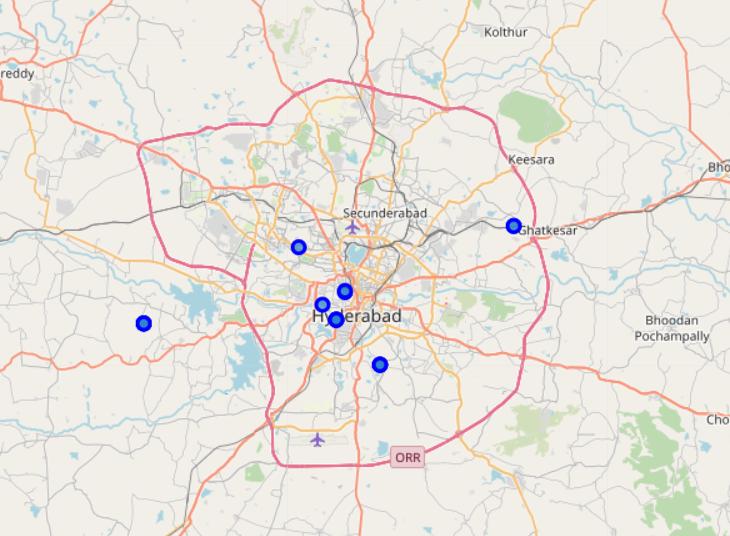
### Results

I scrapped the web for list of engineering colleges in Hyderabad and obtained an initial list of colleges. Since the geo location codes were not available, I used pgeocode library to get geo location codes for localities and merged it with the engineering colleges list. It was found that there were 104 localities in Hyderabad and 25 localities had engineering colleges.

Based on the analysis above, it is found that the Locality of Ghatkesar has highest density of engineering colleges at 6. Thus, it was suggested to set up student hostel in this locality.



I ended the study by visualizing the data on the Hyderabad map.



### Conclusion

The purpose of the project was to find areas in Hyderabad for highest density of engineering colleges for setting up a student hostel. Initially, the geographical coordinates of the localities of Hyderabad were extracted as geo codes were not directly available for engineering colleges. The dataframe of Hyderabad localities was merged with colleges list on location to get colleges geographical coordinates. The results were mapped using the folium library for visual representation.

From the results, Ghatkesar locality was found to have maximum college density. The project can be used for business ventures like cafes, pizza places, stationery stores, book stores which are allied business of colleges. The project can be further enhanced with exact location geo codes to find the radius distance with in which the business establishment can be set up.