Prediction of Optimal Location for setting up a Coaching Institute

Varun Deshpande.

Dec. 19 2019.

Background:

Pune, the second most biggest city in the State of Maharashtra, India is well known for its reputed colleges and universities which include colleges like Fergusson College, Savtribai Phule Pune University, Symbiosis College, etc. And with so many institutes scattered almost in all parts of city it is evident that whole city has huge scope for private coaching institutes to set up.

Problem:

There are various coaching that establish every year not that all run successfully. If a student needs to attend a coaching institute he or she will look into two things,

1. The Quality of Coaching in a institute.

2. The location of that institute.

The first problem can be easily solved by recruiting the best tutors but the second problem is something that always proves to be a hurdle. All good coaching institutes choose the best location to start their classes and it is usually near colleges and universities. But in case of Pune, where there are many colleges(More than 100), one needs to be sure of the location and it must be convenient for students of various colleges and universities to attend the classes. In this project we try to find a solution to this problem for the potential Coaching Institutes and same algorithm can be used for many other cities.

Data information:

To find a solution to this problem we will using the foursquare location data using the foursquare API. We will be using the location coordinates i.e. the Longitude and Latitude data available in the data to solve the problem. We will be extracting information about colleges in and around Pune. Pune has 2 major parts.

1. The Pune City

2. The Pimpri Chinchwad Area.

Mostly all the colleges are scattered in these two areas. We will extract the details about the colleges in these areas and get the location coordinates to further solve the problem.

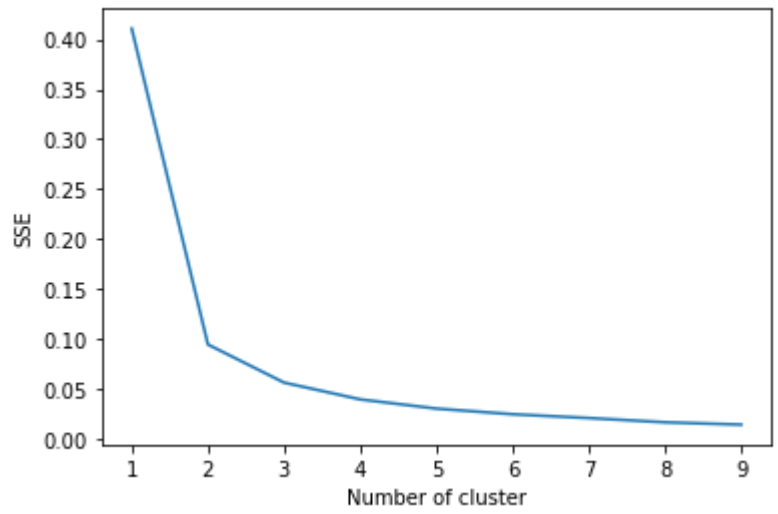
Methodology:

We use the technique of clustering specifically K means clustering method which will give us points where we can open the coaching centers as it will be the centroid of clusters formed by location coordinates of various colleges. But we don't yet know how many Coaching Centers we should open. For this we need to know appropriate No. of clusters. Which means a appropriate set of colleges such that any more set of colleges will not affect much to the distance traveled by a student to the class from College or University and any less number of set of colleges will drastically increase the distance. We will use the knee bend method to choose the appropriate No. of Clusters.

SSE is the average square of distance between a cluster centre and the elements in that cluster. More the distance, bad is the clustering. We plot No. of clusters Vs SSE for each No. of cluster. We then use folium graph to indicate the locations where the centroid lies and take decision accordingly.

Results:

We plot SSE Vs No. of clusters. The results is as follows



From the above diagram we see that the number of SSE reduces drastically at 2 but remains almost same after words. The Point where such a sharp difference is observed is called knee bend and we choose appropriate No. of clusters to be formed = the knee bend.

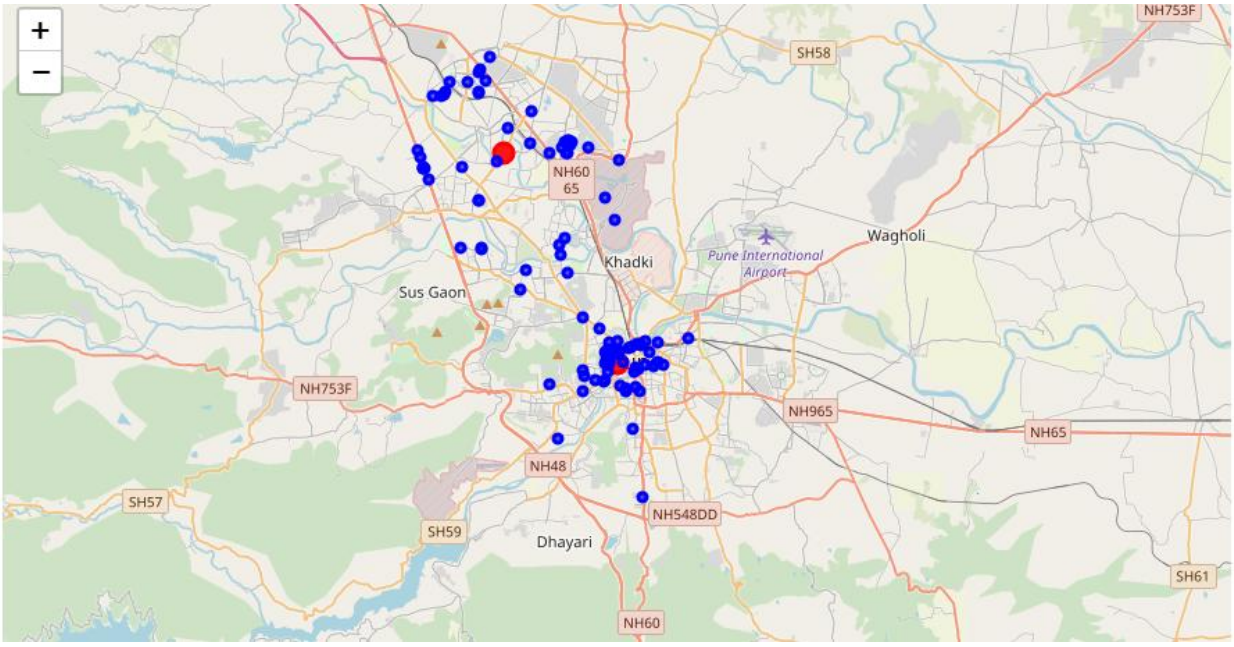
Hence, we will make 2 clusters.

We get the centroid coordinates of these two clusters as

Location	Latitude	Longitude
Centre 1 (Cluster 1)	18.52065405	73.84497271
Center 2 (Cluster 2)	18.61859758	73.78901255

Decision:

We use the coordinates we get from clusters and the coordinates of the colleges and mark it on the map. Thus from the map we can decide where to set up the coaching centers. We see that in map both the centers are such that they seem to be almost equidistant from all the colleges.



Conclusion:

I used the K means clustering to create appropriate number of clusters and also plotted them on the map which will make it easy for everyone who has to code to clearly spot where the two centers should be. We could use similar kind of algorithm in various other cities thus help the interested clients in choosing appropriate location.