Applied Data Science Capstone

Segmenting and Clustering areas in Pune for new Gyms

Yuvraj Thakare

IBM Data Science Professional Certificate

**Introduction:**

Pune is very populated city and it has a lot of educational institutions.Therefore it has a large student population which comes to pune for educational purpose. It also has large IT sectors, attracting large amount of employees all around the India.

Considering all these factors there is more of youth population in Pune. In today’s generation, youth is giving importance to “Healthy Living” and more into Gym’s and workouts stuff. One more reason behind this is Social Media, youth is constantly active on social media posting pictures, making videos, etc. Hence, to look good fitness is really important. There is a lot of demand for gyms in pune but accordingly proper professional gyms are really less as compared to people.

That’s why through this project I’m trying to segment the neighborhoods of pune according to the availability of gyms to help the new gyms to open in locations where gyms are very less and have great potential for business.

**Data :**

• List of Neighborhoods in Pune .This basically sets a scope of our

project .

• Area Markers i.e Longitudes and Latitude of the Scope and also

the neighborhoods targeted for analysis.

• Venue data , for calculation of customer occurrences and

availability of specific venue in an area.

**Sources for Data Availability:**

• First and foremost Source for list of neighborhood in Pune can be

made available through Wikipedia . A csv file is created containing

the list which is further merged with area markers.

• The Coordinates of Neighborhood can be extracted from Python

Geocoder Package.• Venue data can be extracted from Foursquare.com site , which is

prestigious site four venue details , ratings etc. A Developer

account needs to be created so as to avail API calls.

Thank

You.