

2. A description of the data and how it will be used to solve the problem

2.1 Data Acquisition and Description

The data required for the Location analyst to find the similar type of neighbourhood of Pune city are:

- Locations by ZIP code
- Positional coordinates
- Neighbourhoods venues
- Population database.
- Hospitals and clinics location information

Locations by ZIP code:

- This data is required for knowing the different areas in Pune city.
- The location zip code can easily find by googling. The <https://www.mapsofindia.com/pincode/india/maharashtra/pune/> web page having table which contain location name, postal code, state and District.

Positional coordinates

- This data is useful to determine the geographical location when we visualize the places and also require for getting the venues for that places.
- Positional coordinate data of each postal code can be achieved using Geocoder or <http://download.geonames.org/export/zip/> web site.

Neighbourhoods venues

- This data is very important to find out what facilities are available in each place.
- To gather this data, the Foursquare API was used. This API provides real-time geographical data of almost anywhere on Earth.

Population database.

- Using this data, we can calculate Population density. So we know the area is upscale resident or not.
- This data we can get from <https://www.censusindia2011.com/maharashtra/pune-population.html> web site.

Hospitals and clinics location information

- This data we use to find out how many medical facilities are available in each place.

This data set we get from <https://pin-code.org.in/hospitals/listing> and

www.unipune.ac.in/admin/circular/List%20of%20Hospitals.xls

2.2 Uses of Data

Gathering all data then cleaning and prepared ready to use data we can use for clustering. The idea was to cluster and determine similar neighbourhoods across Pune city based on the types of venues that they offered, residence density and hospitals, clinics facility.