

Capstone project 1st Week, 2nd part

Data

Data and data source

The following data is required for the successful implementation of the project:

- List of community areas of Chicago. Data source: https://en.wikipedia.org/wiki/Community_areas_in_Chicago
- Location data (geo-coordinates) of community areas of Chicago. Data source: geo-coordinates of community areas will be obtained by using Nominatim Geocoding service from Geopy library in the notebook.
- Top venues of community areas. Data source: this data will be obtained from Foursquare through an API.

Methodology, Explanation of data usage

1. The first steps in achieving the project objectives will be to cleaning, sorting and exploring data.
2. After that exploring the top venues in community areas by Foursquare will be the next step.
3. The K-means unsupervised machine learning technique will be used for creating clusters of community areas. In order to choose a relevant number of clusters silhouette scoring will be used.