Finding the best community area in Chicago to live in

Aparna Subhakari R S

Contents

- ▶ Chicago is the most beautiful place to live in. One of the main considerations of a move to Chicago is affordability. Other reasons being :

    - ▶ Most populous and popular city

    - ▶ Higher education opportunities in the top universities in USA

    - ▶ Job opportunities with high salaries in the steady job market of Chicago

    - ▶ International hub for finance, culture, commerce, industry, technology, telecommunications, and transportation

    - ▶ One of the highest GDP s in the world

    - ▶ Surrounded by the beautiful Lake Michigan, and two rivers : the Chicago River in downtown and the Calumet River in South Side.

- ▶ The goal of this project is to find the best community area and neighbourhoods to live in based on the factors : 'Housing Availability', 'Employment rate', 'House Holds below poverty line'. The questions can be framed as :

  - ▶ What is the best community area to live in?

  - ▶ How community areas are categorized according to crowded housing or unemployment rate etc. features?

  - ▶ Does correlation exist between these features selected?

► list of neighbour community areas and neighbourhoods :
Scraped from the page `https:
//en.wikipedia.org/wiki/Community_areas_in_Chicago`

► Census data regarding housing, unemploymnet etc : Extracted
from the page `https://ibm.box.com/shared/static/
05c3415cbfbtfnr2fx4atenb2sd361ze.csv`

► Geo location data - Latitude and Longitude : Extracted from the
page `https://ibm.box.com/shared/static/
f9gjvj1gjmxxzycdhplzt01qtz0s7ew7.csv`

- Wikipedia data is a structured tabular data with no null values and empty rows.
- Data extracted from other Data Sets is also a well structured tabular data with many features but has empty rows i.e null values for features.
- Empty rows, Null values are eliminated.

► All the data frames are merged on a single column 'Community area' as the aim of this project is to find the best community area.

► Only required features are retained and redundant features are dropped in the main data frame.

- The final data frame is ready for processing.
- Final Data Frame:

|	Community_Area_Name	Neighbourhoods	Housing_Crowded_Percent	Below_Poverty_Percent	Unemployment_Percent	School_Type	Latitude	Longitude
0	ROGERS PARK	East Rogers Park	7.7	23.6	8.7	ES, ES, ES, ES, HS, ES	42.009665	-87.669949
1	WEST RIDGE	Arcadia Terrace,Peterson Park,West Rogers Park	7.8	17.2	8.8	ES, ES, ES, ES, ES, ES, HS, ES	42.003435	-87.697500
2	UPTOWN	Buena Park,Argyle Street,Margate Park,Sheridan…	3.8	24.0	8.9	ES, ES, ES, ES, HS, ES, ES	41.964346	-87.656053
3	LINCOLN SQUARE	Ravenswood,Ravenswood Gardens,Rockwell Crossing	3.4	10.9	8.2	ES, ES, ES, HS, ES	41.975867	-87.683254
4	NORTH CENTER	Horner Park,Roscoe Village	0.3	7.5	5.2	HS, HS, ES, HS, ES, ES, ES	41.946617	-87.691056
5	LAKE VIEW	Boystown,Lake View East,Graceland West,South…	1.1	11.4	4.7	ES, ES, ES, ES, ES, ES, ES	41.945247	-87.670957

Methodology
K-mean clustering

► K-means clustering is applied on the final data frame.
  ► No.of clusters :4
► Cluster Labels :

Out[25]:

|	Cluster Labels	Housing_Crowded_Percent	Below_Poverty_Percent	Unemployment_Percent	Community_Area_Name
0	0	3.514286	24.628571	19.035714	SOUTH SHORE, CHATHAM, AVALON PARK, SOUTH CHICA...
1	1	5.864286	18.764286	13.464286	EDISON PARK, NORWOOD PARK, JEFFERSON PARK, FOR...
2	2	5.641667	17.900000	15.883333	GARFIELD RIDGE, ARCHER HEIGHTS, WEST ELSDON, G...
3	3	4.184615	15.415385	7.807692	ROGERS PARK, WEST RIDGE, UPTOWN, LINCOLN SQUAR...
4	4	5.376190	28.842857	18.804762	NEAR WEST SIDE, NORTH LAWNDALE, SOUTH LAWNDALE...

▶ Visualization - Community Areas

Out[16]:

Results
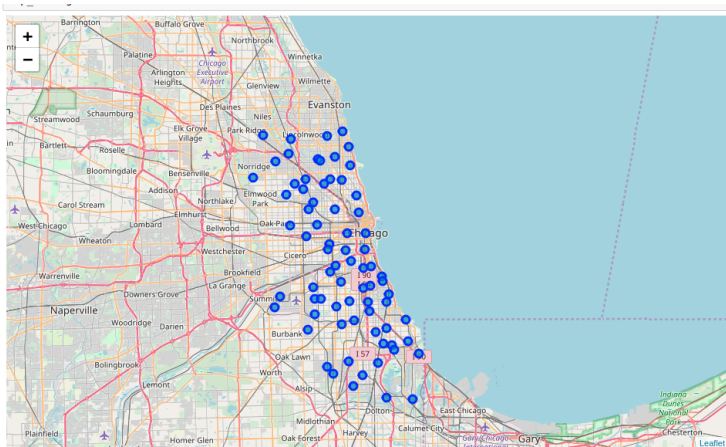
► Visualization - Clustered Community Areas
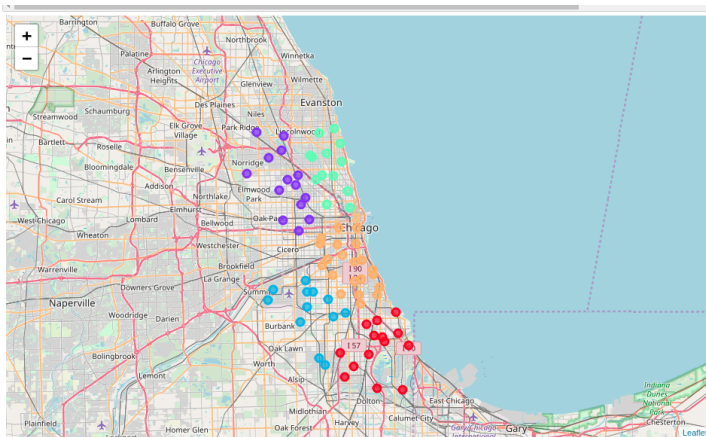
- ▶ Cluster 0
  - ▶ Features : Low percentage of crowded housing, High percentage of population below poverty, High percentage of unemployment rate
- ▶ Cluster 1
  - ▶ Features : High percentage of crowded housing , Medium percentage of population below poverty, Low percentage of unemployment rate
- ▶ Cluster 2
  - ▶ Features : High percentage of crowded housing , Low percentage of population below poverty, Medium percentage of unemployment rate
- ▶ Cluster 3
  - ▶ Features : Low percentage of crowded housing, Low percentage of population below poverty, Low percentage of unemployment rate
- ▶ Cluster 4
  - ▶ Features : Low percentage of crowded housing, High percentage of population below poverty, High percentage of unemployment rate

Observations

▶ Correlation Table :

Out[26]:

	Cluster Labels	Housing_Crowded_Percent	Below_Poverty_Percent	Unemployment_Percent
Cluster Labels	1.000000	0.317797	0.146386	-0.209088
Housing_Crowded_Percent	0.317797	1.000000	-0.078955	0.028639
Below_Poverty_Percent	0.146386	-0.078955	1.000000	0.840487
Unemployment_Percent	-0.209088	0.028639	0.840487	1.000000

▶ Negative correlation between percentage of crowded housing and percentage of population below poverty

▶ Highest correlation between percentage of unemployment and percentage of population below poverty

- **Final Result** : Best Cluster - Cluster 3
  - Low percentage of crowded housing, Low percentage of population below poverty, Low percentage of unemployment rate
  - **Community Areas** - Rogers Park, West Ridge, Up Town, Lincoln Square, North Center, Lake View, Lincoln Park, Near North Side, North Park, Albany Park, Avondale, West Town, Edgewater

▶ **Neighbourhoods** - East Rogers Park, Arcadia Terrace,Peterson Park,West Rogers Park, Buena Park, Argyle Street, Margate Park, Sheridan Park, Ravenswood,Ravenswood Gardens, Rockwell Crossing, Horner Park,Roscoe Village, Boystown, Lake View East, Graceland West, South East Ravenswood, Wrigleyville, Old Town Triangle, Park West, Ranch Triangle, Wrightwood Neighbors, Sheffield Neighbors, Cabrini–Green, The Gold Coast, Goose Island, Magnificent Mile, Old Town, River North, River West, Streeterville, Brynford Park, Hollywood Park, River's Edge, Sauganash Woods, Mayfair, North Mayfair, Ravenswood Manor, Belmont Gardens, Chicago's Polish Village, Kosciuszko Park, East Village, Noble Square, Polish Downtown, Pulaski Park, Smith Park, Ukrainian Village, Wicker Park, Andersonville, Edgewater Glen, Edgewater Beach, Magnolia Glen, Lakewood/Balmoral

Thank you!