

Segmenting and Clustering Neighborhoods in Chicago, Illinois

Wendy Hu

Introduction/Business Problem

Chicago, on Lake Michigan in Illinois, is the third largest city in the U.S and one of famous international financial centers in the world. The city is also renowned for its museums, including the Art Institute of Chicago with its noted Impressionist and Post-Impressionist works. According to 2018 Time Out City Life Index, Chicago was ranked first in the global quality of life survey of 15,000 people in 32 cities.

So, despite Chicago is famous for crime capital, there are many people considering to settle in this city. When they choose the community and apartment they will live in, safety and neighborhood facility can be top determinant factors. For people who consider to moving in this city, they are willing to read reports in answering the following questions:

1. What community is the safest in Chicago?
2. Is crime rate connected with average income and education?
3. Using Foursquare data, how different are venues distributed in neighborhood of different crime rate?

Data

Chicago Socioeconomic Indicators:

<https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv>

This dataset contains a selection of six socioeconomic indicators of public health significance and a “hardship index,” by Chicago community area, for the years 2008 – 2012. The indicators are the percent of occupied housing units with more than one person per room (i.e., crowded housing); the percent of households living below the federal poverty level; the percent of persons in the labor force over the age of 16 years that are unemployed; the percent of persons over the age of 25 years without a high school diploma; the percent of the population under 18 or over 64 years of age (i.e., dependency); and per capita income.

Chicago Public Schools:

<https://ibm.box.com/shared/static/f9gijv1gimxxzycdhplzt01qtz0s7ew7.csv>

This dataset shows all school level performance data used to create CPS School Report Cards for the 2011-2012 school year.

Chicago Crime Data:

<https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv>

This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system

Venue Data:

<https://foursquare.com/>

Venues data can be extracted from this website. We can utilize this website to analysis what venues distributed around different neighborhoods.