Garima Singhal

06/01/2020

Suggesting neighborhoods in Chicago: Per Capita Income and Lower Crime Rate

Introduction:

Chicago, Illinois is one of the most populous big cities in the United States. It is famous for its bold architecture, consisting of skyscrapers such as Willis Tower, John Hancock Center, and the Tribune Tower. The city is renowned for its museums, and art galleries. It is also one of the greatest hubs for business, education, industry, culture, transportation, and a lot more. The city is highly diversified with people from all backgrounds and cultures making it the most balanced economy in the United States.

After moving to United States about 5 years ago, I always had the dream of working for a firm in a big city. So, when I got an offer from a technology company in Chicago, I was excited to move. However, when I was looking for places to live in Chicago, I found out that not every neighborhood in the city is considered safe and sound. This led me to do intensive research for finding a good and safe neighborhood in the city.

Business Problem:

Chicago being one of the cities embodying highly paid corporate jobs, we see that more people are moving into the city every day. Considering Chicago, which has an overall crime rate higher than the US average, it becomes challenging to find a good and safe neighborhood.

In 2016, the city saw a surge in gun violence with 762 murders, 3550 shooting incidents, and 4331 shooting victims which was more than the number of murders in New York City and Los Angeles, combined. The estimated number of homicides in Chicago increased by 52% in 2016. Most of these killings happened in five mostly black and Latino neighborhoods on the south and the west side of city.

To buy an apartment or a house, deciding which neighborhood you should choose is one of the most important decisions. Safety is the foremost priority when it comes to finding the right neighborhood and income being the second most important. As I have seen from my personal experience the process of finding a safe neighborhood based on your annual income can be tiring.

The aim of this project is to find a safe neighborhood based on the crime rate and per capita income in various neighborhoods across the city of Chicago. The goal of this project is to help new people move into the city and help them find a neighborhood which is safe, has a low crime rate and fits into their budget.

Data Section:

The data required for this project is a combination of three data sources. The first source of data is a Wikipedia Page that contains the list of the Chicago community areas. The dataset consists of following columns:

Column Name	Description	Type
Serial Number		Number
Community Area Name		Plain Text
Neighborhood	Name of the neighborhood in the Community area	Plain Text

The second data source for the project will use the Chicago Crime Data that shows the crime per community area in Chicago. The dataset consists of the following columns:

Column Name	Description	Type
ID	Unique identifier for the record.	Number
Case Number	The Chicago Police Department Record Number	Plain Text
Date	Date when the incident occurred.	Date & Time

Block	The partially redacted address	Plain Text
IUCR	The Illinois Uniform Crime Reporting code.	Plain Text
Primary Type	The primary description of the IUCR code.	Plain Text
Description	The secondary description of the IUCR code.	Plain Text
Location	Description of the location	Plain Text
Arrest	Indicates whether an arrest was made.	Checkbox
Domestic	Indicates whether the incident was domestic related as defined by the Illinois Domestic Violence Act.	Checkbox
Beat	Indicates the beat where the incident occurred.	Plain Text
District	Indicates the police district where the incident occurred.	Plain Text
Ward	The ward (City Council district) where the incident occurred.	Number
Community Area	Indicates the community area where the incident occurred.	Plain Text
FBI Code	Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).	Plain Text
X Coordinate	The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection.	Number
Y Coordinate	The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection.	Number
Year	Year the incident occurred.	Number
Updated On	Date and time the record was last updated.	Date & Time
Latitude	The latitude of the location where the incident occurred.	Number
Longitude	The longitude of the location where the incident occurred.	Number
Location	The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal.	Location

The third data source, the Chicago Census Data – Selected socioeconomic indicators in Chicago, 2008 – 2012 will be used. This dataset contains a selection of six socioeconomic indicators of public health significance and a "hardship index", for each community area. The dataset consists of the following columns:

Column Name	Description	Type
Community Area Number		Number
COMMUNITY AREA NAME		Plain Text
PERCENT OF HOUSING CROWDED	Percent occupied housing units with more than one person per room	Number
PERCENT HOUSEHOLDS BELOW POVERTY	Percent of households living below the federal poverty level	Number
PERCENT AGED 16+ UNEMPLOYED	Percent of persons over the age of 16 years that are unemployed	Number
PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA	Percent of persons over the age of 25 years without a high school education	Number
PERCENT AGED UNDER 18 OR OVER 64	Percent of the population under 18 or over 64 years of age (i.e., dependency)	Number
PER CAPITA INCOME	Community Area Per capita income is estimated as the sum of tract-level aggregate incomes divided by the total population	Number
HARDSHIP INDEX	Score that incorporates each of the six selected socioeconomic indicators (see dataset description)	Number

We will use web scraping technique to extract the data from the Wikipedia page, with the help of python requests, and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package while will provide us the latitude and longitude coordinates of all the neighborhoods.

The next step in the project will be using the Foursquare API to get the venue data for those neighborhoods. Foursquare consists of one of the largest databases of 105+ million places

and is used by almost 125,000 developers. This project will use multiple data science skills, such as Web Scraping, working with Foursquare API, data cleaning, data wrangling, machine learning algorithm: K-means clustering, and data visualization using the Folium package.