The University of Chicago

MACS 30122, Winter 22

Group Project

Group “None”: Chongyu Fang, Feihong Lei, Juno Wu, Coco Yu

Github: git@github.com:cs-ssa-w22/final-project-none.git

**Progress Report**

**Goal of Project**

Despite the fact that neighborhood security represents an important factor in individuals’ choice of housing, average rental websites do not reflect crime rates information, probably due to a conflict of interests with landlords. The main purpose of this project is to provide an aggregated rental information in Chicago City incorporating both rental information and past neighborhood crime rates in 2021 for future tenants. The final output will be an interactive map where users can click on geographical points to access both crime rates and housing features including bedroom numbers, bathroom numbers, price range, amenities and other features. With the aggregated information and interactive map, we wish to provide a user-friendly interface to facilitate future tenants’ decision-making on housing.

**Data**

https://www.apartments.com. The website does not provide a public API for housing information. We will use html scraper to gather data.

https://data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6. The website provides a csv file for all incidents in 2021.

**Tasks**

**Task 1.** Data Collection (Chongyu Fang & Feihong Lei)

Subtask 1: grab information from the rental website using html scraper.

* The specific location of each posting: longitude and latitude for later mapping with crime data.
* The street address and the name of the property
* The price range of the property
* The size range of the property
* The floor plans of the property (#of bedrooms and # of bathrooms)
* A specific description of the property (including some special characteristics of this property)

Create a csv output file compiling features discussed above.

Subtask 2: Export the crime rates information in the csv format.

From crime rates websites, we can get a csv file with about 200, 000 rows. Each of the incidence has a lot of information. But what we will mainly use is its exact location (longitude, latitude), the type of the crime. Also, the unique types of crime type is about 30, and we will group them into 17 groups so that we can use CHI score to weigh each type of crime differently based on its severity later in our analysis.

**Task 2**. Data Cleaning (Feihong Lei)

Subtask 1: For each of the house listing, calculate police report incidents in the neighborhood (defined as a 3-mile radius to the housing location) and add an extra column to the housing csv file.

* First loop through the list of properties
* Then loop through the crimes that happened within the range of its location
* Add the crimes together based on their different weights
* Get the final score that represents the crime severity around the property

(This neighborhood can be calculated by the longitude and latitude we have found for each property and crime.)

**Task 3**. Data visualization (Juno Wu & Coco Yu)

Subtask 1: create an interactive map using the folium package in python. Pinpoint all the datapoints on the map and attach aggregated information on the markup.