City of Cambridge Evictions Study

Team Members:

Tiancheng Zhu
Kevin Peters
Drew Abram
Syahrial Dahler

Motivation/background of the project

This project is intended to analyze evictions in the City of Cambridge by utilizing the census data and eviction filings from the Massachusetts state court dockets. After collecting and analyzing the data, the client will have a better understanding of the potential contributing factors for evictions, the type of housing evictions most likely to happen, and the situations of the evicted tenants, etc. By conducting this analysis, the City of Cambridge will provide assistance to residents and evicted tenants more efficiently. The currently available datasets are court dockets provided by the City of Cambridge, MAPC, and census data of Massachusetts.

Goals, hypothesis, outcomes

Goals

1. Building a comprehensive database of eviction cases by implementing a data scraping script.
2. Carrying out in-depth data analysis to find patterns and provide valuable data to the client.
3. Finding the correlation among factors, such as demographics, property type and property management, with eviction.

Hypothesis

Ideally, the analysis would find patterns in eviction data to predict eviction outcome. Causes of evictions may be explained by classifying housing type where evictions happen, complex with the highest eviction rate, the financial condition of people getting evicted, and legal representation rate of eviction cases.

In terms of the primary cause of evictions, currently, the client believes that money could be the major reason.

Outcomes

The analysis would not only need to verify or deny that hypothesis, but it should also elucidate potential contributing factors for eviction.

End Result and Desired Product

By the end of the project, we would like to produce three products:

1. An analysis report about the eviction cases in the city of Cambridge. The report should be able to be produced anytime or to be produced periodically after the database has been updated. The analysis may include:
    - The pattern of eviction cases (based off property location, property management, renter profile)
    - Likelihood of a person to be evicted from the property based on his profile. Factors include race, income, and gender. The analysis also takes into account the property of residence.

2. A database containing the eviction cases in the city of Cambridge.
   This database would contain information about the evicted person profile and the property profile. The evicted person profile will include, among other factors:
    - Name
    - Gender
    - Age
    - Income
    - Race/Nationality
    - The type of lawyer defending him

   The property file will include, among other factors:
    - Address
    - Property Manager
    - Distance to public transport
    - Rent Price

3. A script that will be used to update the database. This script can be run every month to update the existing records and to insert new records of eviction cases.

Out of Scope (Non-Goals)
Additionally, eviction cases that are not documented in the court system are out of the scope of this project.

Methods

The approach for transforming the data includes first implementing a data scraper. Python scripts will parse data and match strings of keywords. The data will also need to be cleaned, particularly to correct mislabelled addresses.

Strategic questions and tentative methods of answering these questions are listed below:

1. Are there local demographic, physical housing conditions or spatial characteristics that make units more likely to be subject to eviction?

   Supervised learning will likely be implemented, using different demographics and the physical features of the building as classification groups. Various classification techniques can be explored to draw associations between data.

2. Are there distinct differences between housing that occur in the low-rise and small scale housing stock (developments under 25 units) compared to the newer, larger buildings?

   One approach could be to use supervised learning with classes of building types. Buildings would need to first be classified, perhaps by number of units and date of construction. Another approach could be to use unsupervised learning techniques to categorize residences by the patterns discovered by analyzing the data.

3. How does tenant representation by attorney affect outcomes?

   Supervised learning techniques will likely be implemented, using classes of represented and not represented, which can likely be extracted from court records.

Tools and Techniques

Difference algorithms, such as Levenshtein distance, can be used to match addresses in databases to the addresses in the city's Master address list. Affordable housing addresses in the city-provided master address list can be checked using the Google Places API. Clustering and logistic regression can be performed to examine the relationship between rent money and evictions. Affordable versus private and expensive versus cheap housing can be mapped using GIS software.

The Python FuzzyWuzzy library can be used as a starting point for Levenshtein distances and string matching. The Python Scikit library can be used for logistic regression and other data analysis techniques.

Open Questions and Uncertainties

The following questions are unresolved and may be answered through the course of the project:

1. Privacy: Is the available data considered open to the public?
2. Scope: Are all evictions conducted through the court system? If not, what is the percentage of eviction through court system vs. over all eviction cases?

Datasets

- MAPC Open Data Library (https://www.mapc.org/learn/data/)
- City of Cambridge Open Data Library (https://data.cambridgema.gov/browse)
- American Community Census data selected in consultation with city, FactFinder (https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t)
- U.S. Census data (https://data.census.gov/cedsci/)
- All Census Bureau APIs (https://www.census.gov/developers/)
- Census Bureau ACS API (https://www.census.gov/data/developers/data-sets/acs-5year.html)
- Court case dockets search for Massachusetts (https://www.mass.gov/search-court-dockets-calendars-and-case-information)