BU Spark!
Fuqing Wang, Tian Chen, Zixiang Wei and Xiao Lu

Cambridge Eviction Study Project Proposal

Background:
Over the past two years, the City of Cambridge has been collecting and analyzing data from the state of Massachusetts online court dock system in order to achieve a better understanding of the factors that cause eviction both demographically and economically.

Goals:
Our project aims to explore research questions on what causes evictions under market-rate housing stock and where they might occur in the future. Strategically, our goals for this project are twofold. First, we are going to update and optimize the existing data collection and analysis procedures based on current data scraping and collecting tools. Second, we will develop an supervised-learning analytic model that demonstrates the relation between eviction and different social aspects(demographics, housing conditions, spatial characteristics, and personal reasons) both at present and in the future.

Plans, approaches:

1. Data collecting: review existing data and add new data to current data set.
2. Data cleaning: use python scripts to filter noisy data entries(such as data with missing values, duplicate data entries and data with discrepancies)
3. Data analysis:
    a. First, normalize all data entries
    b. Second, use algorithms such as linear regression or logistic regression to find the most weighted features(house prices, market affordable rates and criminal rates, etc). This part of analysis gives an explanation of under what conditions eviction occurs.
    c. Third, use geographical features(longitudes, latitudes and zip codes) and obtain a distribution of where evictions happened. Then we will use the regression algorithm again, and predict regional distributions of eviction in Cambridge in the following 5 years. This part of analysis gives an explanation of where eviction might happen in the future.
4. Data visualization:
    a. Use graphing tools such as python numpy to give a graph view of what factors caused eviction and how much each factor contribute to eviction
    b. Use Google map api to create a map view of eviction distributions in Cambridge from the past 5 years to the future 5 years

Non-goals, extensions:

As an extension of our proposed goals, one of our non-goals is to implement visualization tools that show the density and distribution of evictions and how these might change over time. On top of that, we would like to apply unsupervised-learning techniques to our existing model in order to achieve a deeper understanding of what causes evictions and possibly ways to prevent them, if time and datasets permitted.

Product:
By the end of this semester, we will propose an automated data analytical model as well as integrated visualization tools which could present analytical results right on the screen. Ideally, our analytical model should be scalable subject to adding additional data source and tools for data manipulation in the future.

Open questions:

1. Intuitively, arrears of rent is the primary factor that causes eviction, what are the other factors that cause eviction?
2. Is eviction really geographical related? If it is, how does geographical factors affect eviction rates?
3. Given various data sources to use, how do we correlate different data sources?
4. Is it necessary for us to rework on the scraping tool to make it platform independent? Regarding current scraping tools, we see it is strictly coupled with microsoft system - it only runs on Windows 7 and it uses microsoft access which microsoft database driver to run, what is the primary reason for this?
5. We are missing the data folder which has data from previous years, are those data already saved in the database?
6. The current scraping tool only scraps the downloaded html page, are we looking for further optimization for this tool? Current scraping process is redundant, is there a way of direct data access?

Dataset:
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
census.data.gov (replacing Factfinder, hosts most recent 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