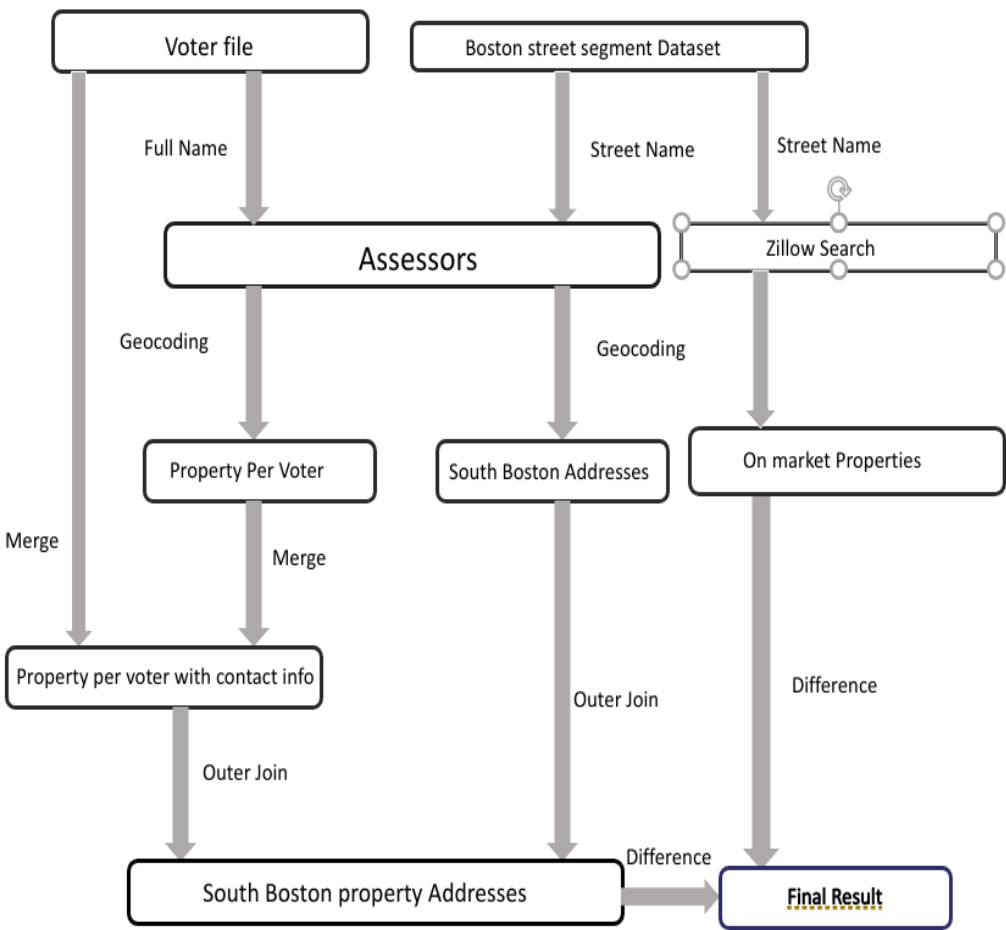


Arezoo Sadeghi, Anastasiia Khudokormova, Xiaoyan Ge

## Introduction

South Boston Neighborhood Development Corporations is a group of local residents committed to maintaining South Boston's character as an affordable, safe, working class community. Main goal for this project is to identify the buildings that are not on the market in order to contact their owners for further cooperation.

## Data Flow



## Analysis

In order to answer the question about correlation between the age and number of properties owned by a person, we have taken the “property per voter with contact info” data set and aggregates the results by owner’s full names to count the number of properties they have. We have done correlation and regression analysis for different age ranges.

Techniques:

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^N \mathcal{N}(t_n | \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1})$$

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{t_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2.$$

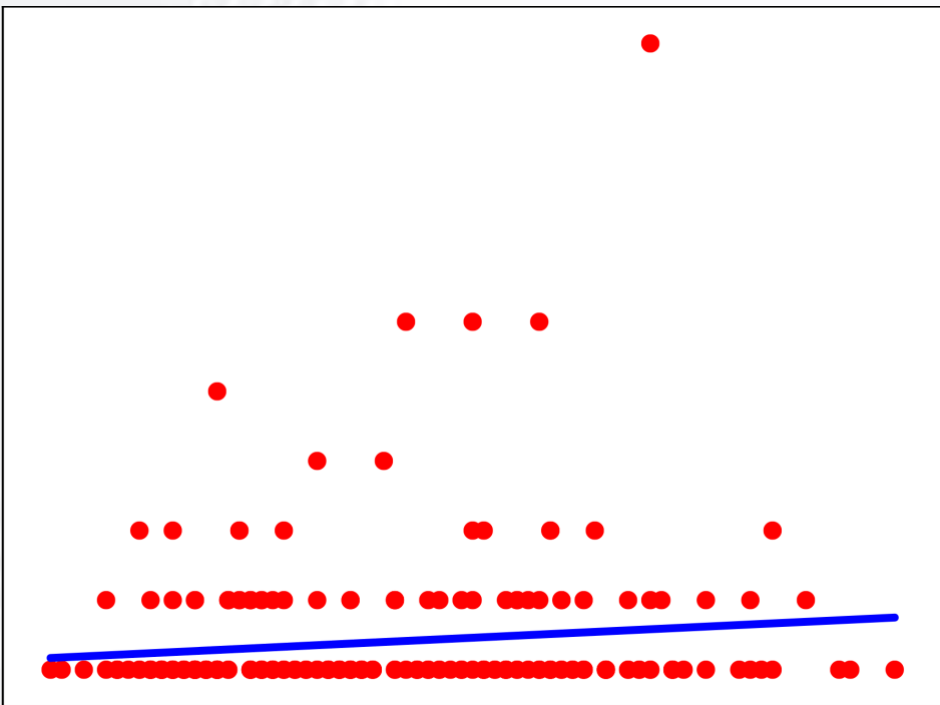
## Results

Correlation between age of owners and number of their properties

| Age Range | Correlation coefficient | P-Value |
|-----------|-------------------------|---------|
| 19-35     | 0.08669242032631995     | 0.464   |
| 25-45     | 0.0633716882769931      | 0.4545  |
| 20-45     | 0.0628646535726424      | 0.453   |
| 35-65     | 0.08957347833195095     | 0.2965  |
| 40-70     | 0.07076712290817734     | 0.453   |
| 30-60     | 0.10996008434546152     | 0.1635  |
| 40-95     | 0.06203316105818047     | 0.4585  |
| 50-95     | -0.03671870899042569    | 0.7155  |
| 30-95     | 0.10281159124264036     | 0.1225  |
| 19-95     | 0.1296640596971729      | 0.0415  |

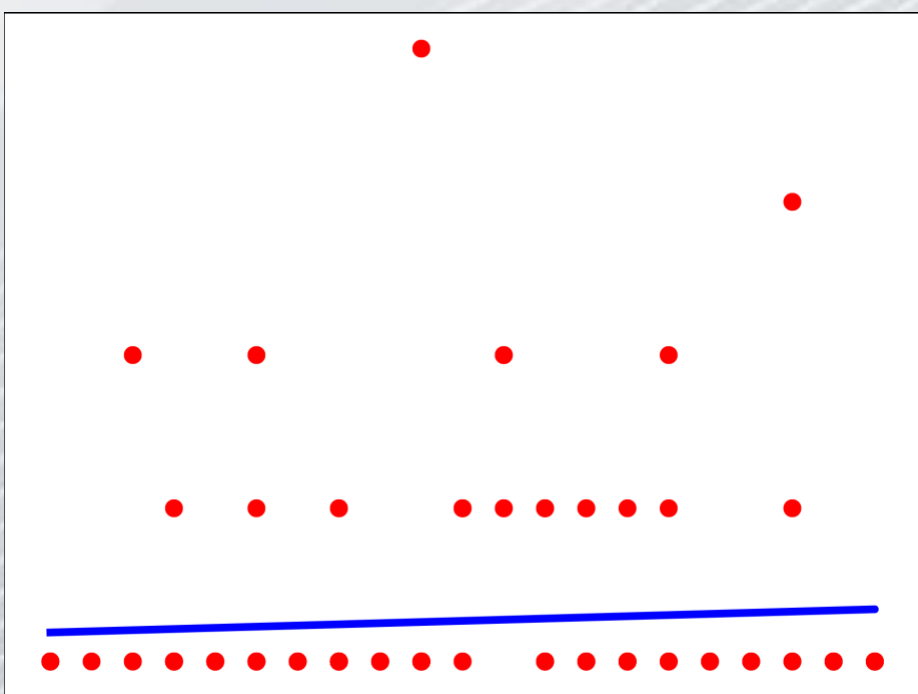
Regression results between age of owners and number of properties

Age range: 19-95



Regression results between age of owners and number of properties

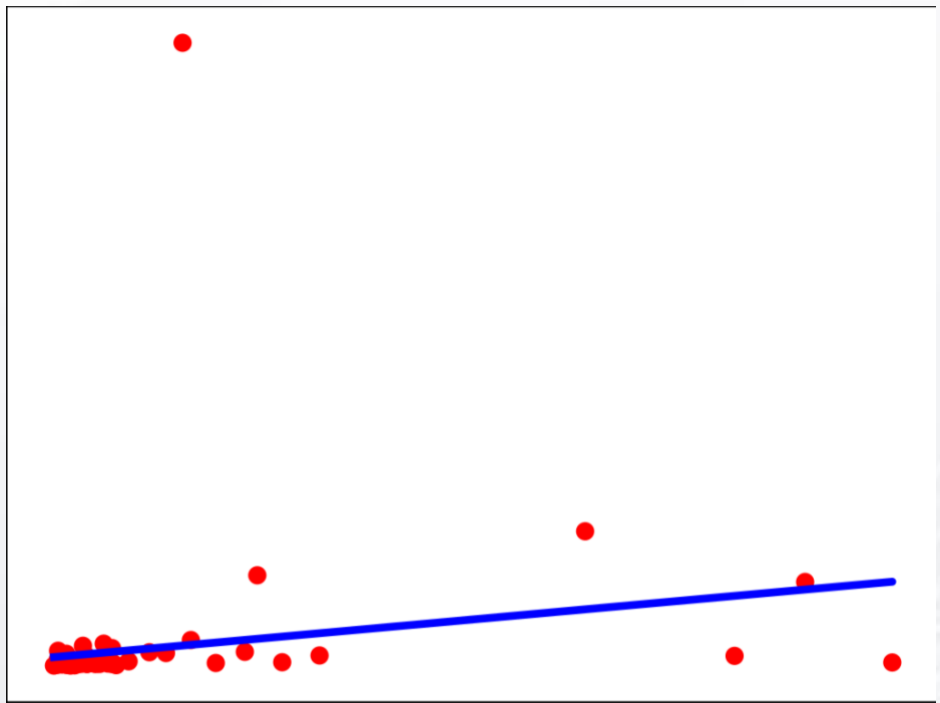
Age range: 25-45



## Crime Incident Analysis

In order to find correlation between the street’s safety and the value of the properties in that street we ran the algorithm on randomly selected subset of street data sets that divided Boston streets into two categories: safety and dangerous. As the number of streets with no crime rate records was big, we constrained the number of safe streets.

Regression results between crime rates and property values



## Conclusion

Finding all the **addresses** in South Boston without having false positives is a challenging task since scraping assessors using street names can give you a few addresses from around Boston with the same street names as in South Boston neighborhood. Filtering the final result by latitudes and longitudes require user-specified constraints.

## Data Sets

- Permit Database
- Crime Incident Database
- Street Names
- Voter File
- Zillow Search
- Assessors
- Googlemaps Geocoding

## Interactive Map

