Deliverable 2 Report - Team 1

Municipalities Analysis:

Evictions number normalization(divided by total number of people):

Top 10 municipalities after normalization:

Fall River, Worcester, New Bedford, Fitchburg, Gardner, Webster, Framingham, Rowe, North Adams, Marlborough.

Compared with top 10 municipalities with most evictions number:

Worcester, Boston, Fall River, New Bedford, Springfield, Framingham, Fitchburg, Brockton, Lynn, Lowell.

Gardner has 51 evictions with 20610 population;

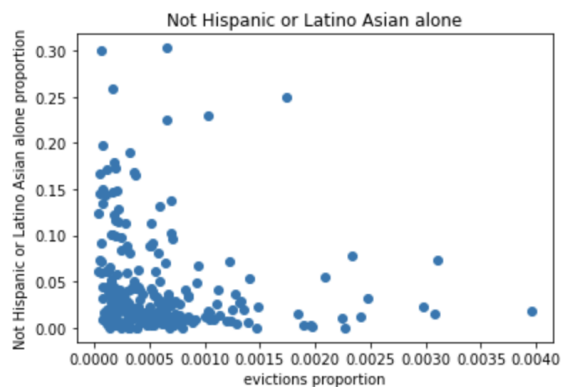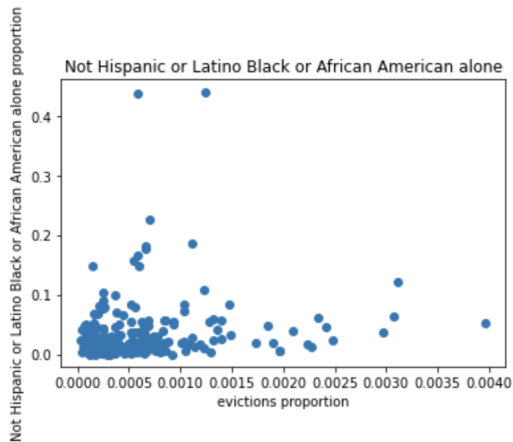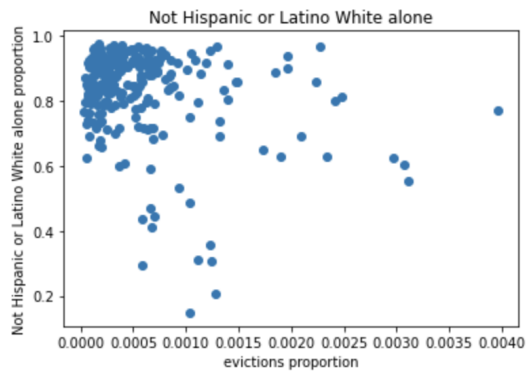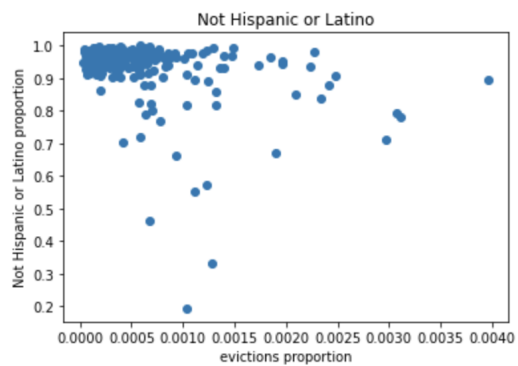Webster has 41 evictions with 16973 population;

Rowe has 1 eviction with 441 population;

North Adams has 29 evictions with 12959 population;

Marlborough has 83 evictions with 39736 population.
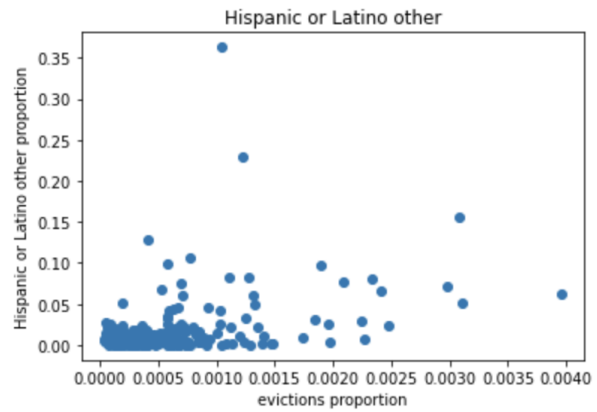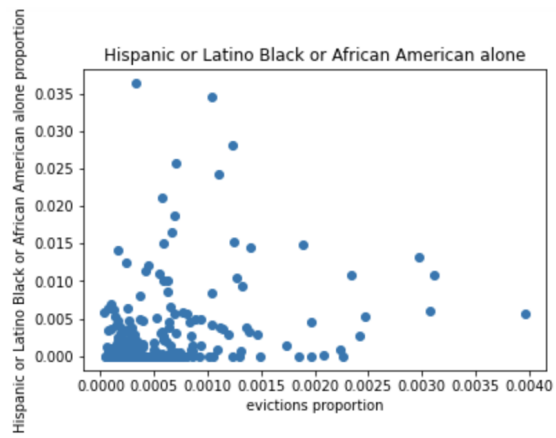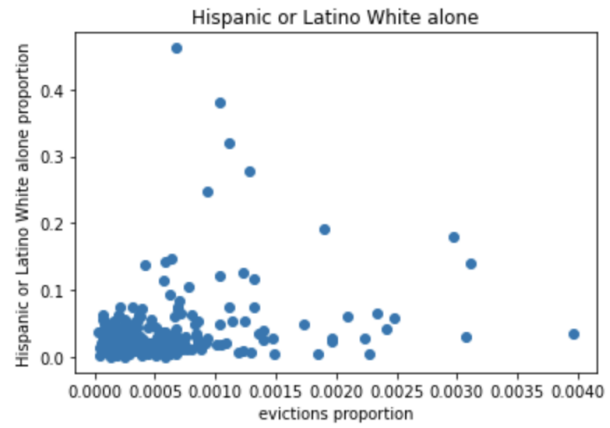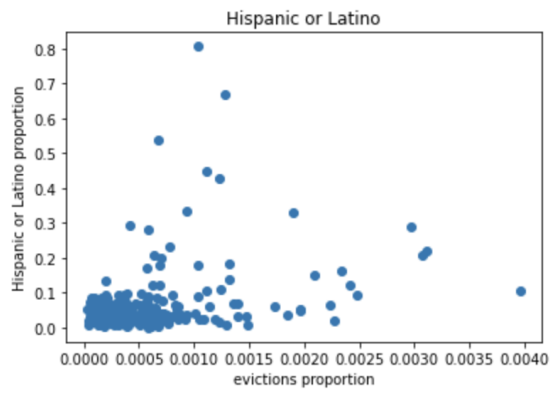
Races proportion in municipality:

Not Hispanic or Latino:

In Not Hispanic or Latino people, it seems not much correlation exists between races and evictions proportion in municipality.

Hispanic or Latino:

Hispanic or Latino

Hispanic or Latino White alone

Hispanic or Latino Black or African American alone

Hispanic or Latino other

Preliminary Analysis 2:
Eviction Dataset: (more visualizations:
https://colab.research.google.com/drive/1Nhj_8xSEFw-1e_6bpKz1lFI7EflPfjPB?usp=sharing)

Realsttype column analysis: The larger-sized warehouses do indeed have many rooms.

```
df_non_residential['num_rooms'][df['realesttyp'] == 12]          df_non_residential['num_units'][df['realesttyp'] == 12]

0.0      53                                                       0.0      56
16.0      5                                                       173.0     2
3.0       2                                                       1.0       2
531.0     2                                                       96.0      2
20.0      2                                                       177.0     2
39.0      1                                                       150.0     2
12.0      1                                                       82.0      1
30.0      1                                                       71.0      1
8.0       1                                                       40.0      1
6.0       1                                                       250.0     1
895.0     1                                                       60.0      1
240.0     1                                                       117.0     1
24.0      1
```

However,

```
df_non_residential['initiating'].value_counts()

Efiled SP Summons and Complaint - Non-payment of Rent    230
Summary Process - Residential (c239)                      44
Efiled SP Summons and Complaint - Cause                   43
SP Summons and Complaint - Non-payment of Rent            11
Efiled SP Summons and Complaint - No Cause                 4
SP Transfer- No Cause                                      1
Efiled SP Summons and Complaint - Foreclosure             1
SP Summons and Complaint - Cause                           1
SP Summons and Complaint - No Cause                        1
SP Transfer - Non-payment of Rent                          1
Name: initiating, dtype: int64
```

There are 44 evictions that contradict my conclusion in the non-residential part of the dataset. They are clearly residential, which made me conclude that I misunderstood the column meaning.
To conclude, only main and most important columns will be used.

We did some preliminary analysis, as well as some research, into the owners who filed more than 20 evictions. For example, a higher proportion of evictions from these owners where filed by corporations rather than individuals (right side being for the whole dataset for comparison):
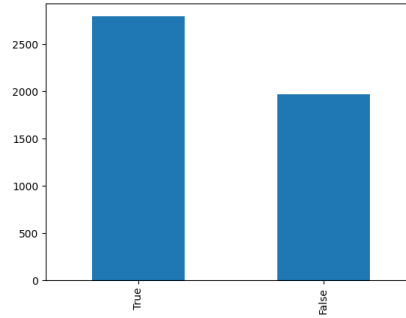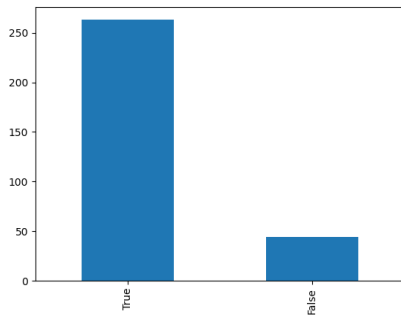
The municipalities represented by these evictions was also limited:
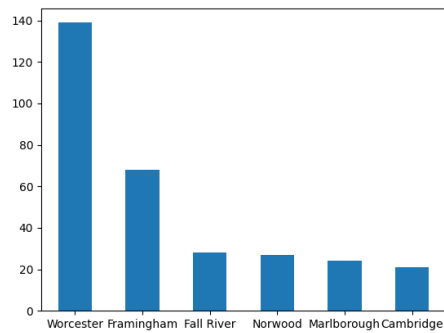
There are only a small handful of of these owners, so we sought to find some more information about some of them:

```
owner_name
WORCESTER HOUSING AUTHORITY        76
JEFFERSON AT EDGEWATER HILLS       44
LINCOLN STREET REALTY COMPANY      41
GPT-RG FALL RIVER LLC              28
WINDSOR GARDENS PROPCO LLC         27
FRAMINGHAM HOUSING AUTHORITY       24
GS STONEGATE PROJECT OWNER LLC     24
600 MAIN STREET,LLC               22
ARCHSTONE NORTH POINT I LLC,      21
```

Both the Worcester Housing Authority and the Framingham Housing Authority are organizations particularly focused on providing affordable housing that are both managed by boards of commissioners.

Windsor Gardens, 600 Main Street, and Archstone North Point are particular apartment buildings which lie in Norwood, Worcester, and Cambridge respectively.

We couldn't find much specific information on the other owners, but they all have relatively corporate-sounding names.

MA Geographical data:
Top 3 municipalities after normalization(as stated in municipalities analysis) (race):

Fall River

Worcester

New Bedford
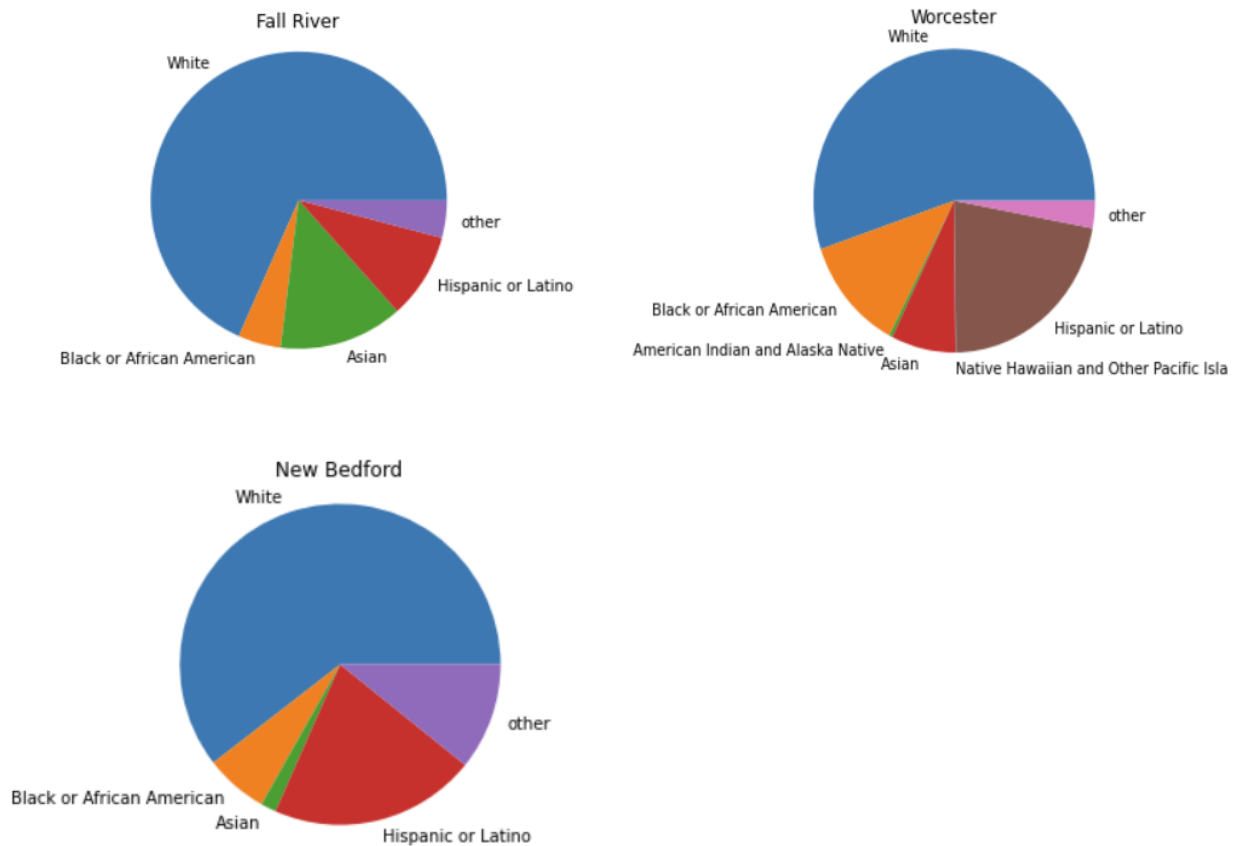
Socio-Economic factors :
The factors that were shortlisted for further analysis are :
 ● Income per household
 ● Occupation
 ● Education attainment

Data was collected from the census.gov website. These datasets contain features such as mean annual household income per municipality, number of people who graduated with a high school degree in these municipalities etc. We plan on performing an analysis of these features on the following municipalities : **Fall River, Worcester, New Bedford, Fitchburg, Gardner, Webster, Framingham, Rowe, North Adams, Marlborough.**

This is because the above-mentioned municipalities have a higher eviction count to population ratio. It would help understanding the overall scenario behind these places in order to understand the eviction numbers there.

Census Tract Eviction Cases and Poverty Rate Analysis:
- Preprocessed the ACS economic characteristic data(2019, 5 years), and selected poverty rate feature to perform further analysis.
- Include the dataset by_census, which is a subset of our eviction filing dataset and contains only the name of each census tract and its corresponding number of eviction cases.
- Merged by_census with ACS dataset into 'poverty_rate_of_census.csv', so that the new dataset has the number of evictions and the poverty rate of each census tract.
- The scatter plot of num_eviction vs. poverty_rate is shown below.
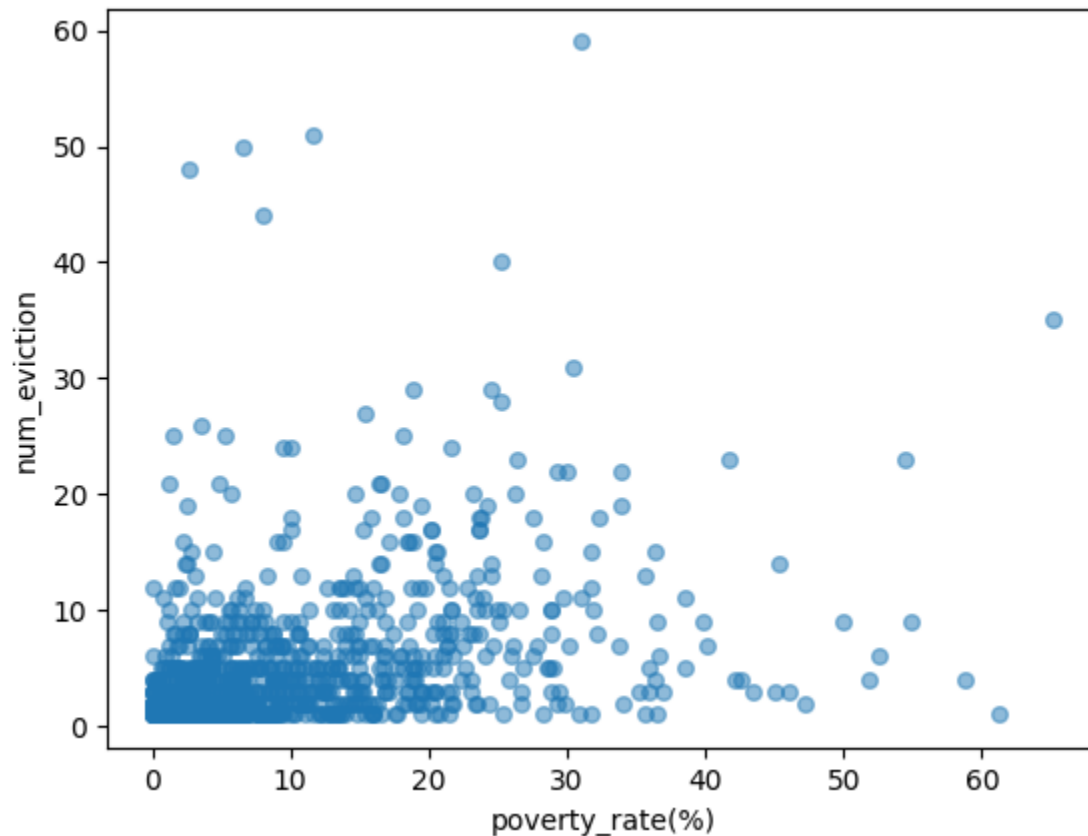
- Conclusions:
    a. Census tracts with low number of evictions generally have low poverty rate;
    b. Census tracts with high number of evictions have poverty rate approximately evenly distributed in 0%-40% without a specific pattern;
    c. The poverty rates of census tracts with an extremely high number of evictions can be as low as less than 5% and as high as 31%.
    d. Therefore, the poverty rate is a good indicator for a low number of evictions but might not be a key factor for a high number of evictions.
    e. The true reason behind a high number of evictions might be hidden behind property owners.

Further plan:

1. Normalize by number of renter households.
2. Visualize by GeoPandas and prepare for presentation.
3. Analyze the new data set.
4. Take 3 rates of cities with high evictions and compare them to lower rate cities.
5. Break up municipalities into different population brackets and compare the whole state to gateway cities.