Early Insights Report

Introduction

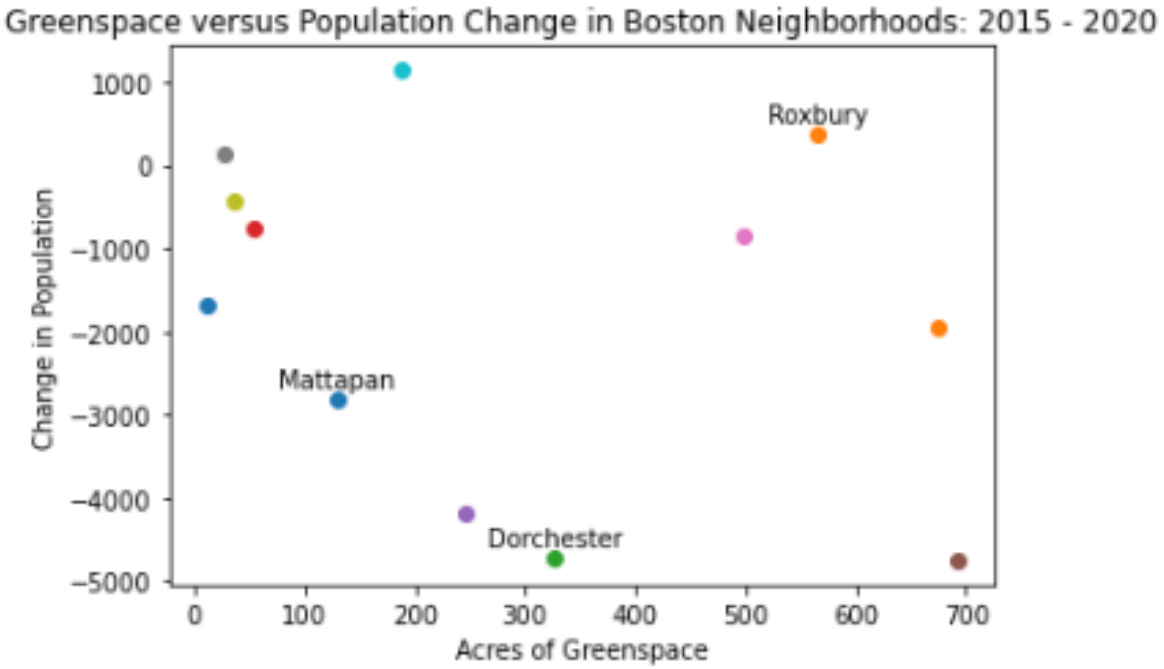
For our early insights, we gathered data on open spaces and population data for Boston neighborhoods in order to discover whether there is a correlation between the creation of greenways and new parks, and displacement in Boston.

Data Collection and Preprocessing

First, we gathered data from the 5-year American Community Survey (ACS) and the decennial census for Boston neighborhoods to gain insights into population change. Both of these datasets contained information about the racial demographics and population of various neighborhoods in Boston. We then had to clean both datasets and remove rows with insignificant details in order to properly process the data in our code. We compared the 2015-2019 census data to the 2020 census data to calculate the estimated change in population in Boston neighborhoods. The U.S. Census Bureau suggests that when doing comparisons with 5-year estimates, you should not compare overlapping time periods, so comparing the ACS from 2015-2019 to 2020 is a valid way to see the change in population. In addition to these datasets, we found another dataset on open spaces in Boston. This dataset contains the number of acres of each open space, the district it belongs to (Mattapan, Roxbury, etc.) and categorizes them into the types of space they are (Parks, Playgrounds & Athletic Fields, Community Gardens, etc.) Since our focus is on greenways, we filtered out “Cemeteries & Burying Ground” and “Malls, Squares & Plazas,” since they are not considered greenways. With these 3 datasets, we created some plots in an attempt to answer our questions.

Data Analysis

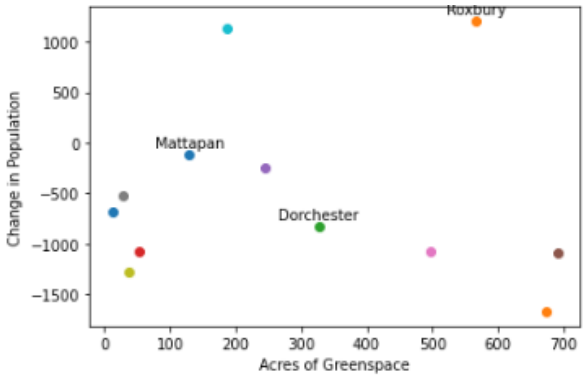
In order to analyze our data, we created a Jupyter notebook to process the data and create plots. We wanted to gain insights into each neighborhood so we calculated the population change for each. To calculate population change, we compared the population numbers from the ACS dataset with corresponding data from the 2020 census and subtracted the ACS population from the 2020 census population. As for the open space dataset, we decided to find the total amount of land that the greenways in Boston covered, so we summed up the acres of each individual greenway with respect to their district. With these 2 steps done, we had the population change and space taken up by greenways in a particular neighborhood, so we created a plot to visualize our data. On the next page you can view the plots that we created.




```
Correlation Coefficient: -0.30004374096415903
```

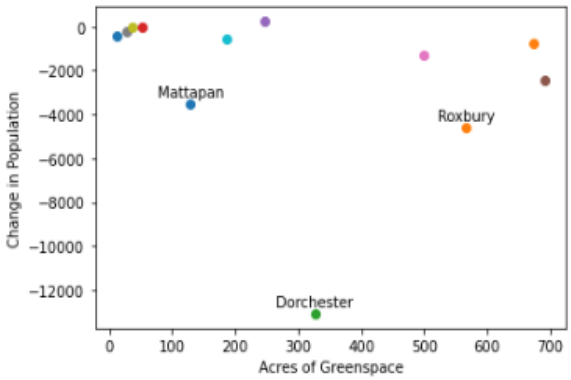
This is the plot that we created for all populations combined and it had a correlation coefficient of -0.3, which is a weak negative correlation. The general trend seems to imply that more greenspace results in a decrease in population, but this cannot be concretely stated. However, an interesting observation is that Roxbury seems to be an outlier. It actually sees an increase in population even with such a large amount of greenspace. Now since we also had data on the distribution of races, we also created plots for every race.

White



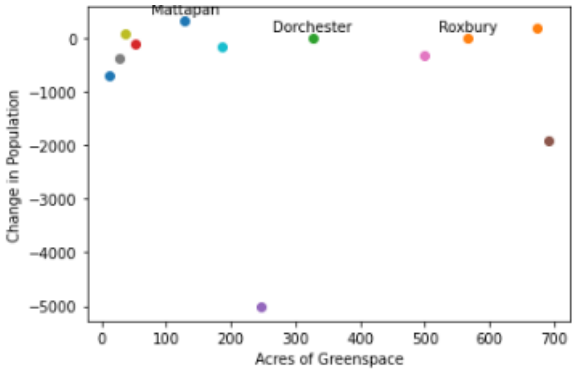
Correlation Coefficient: -0.07755391322386153

African American



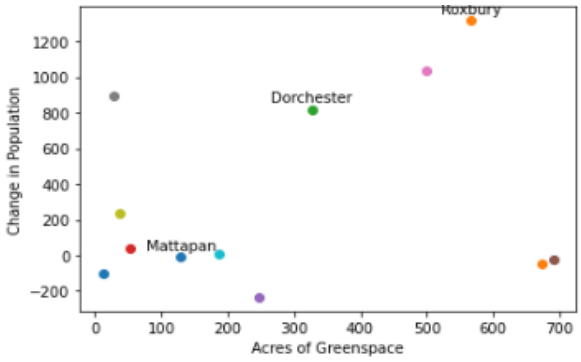
Correlation Coefficient: -0.24139106260161658

Hispanic



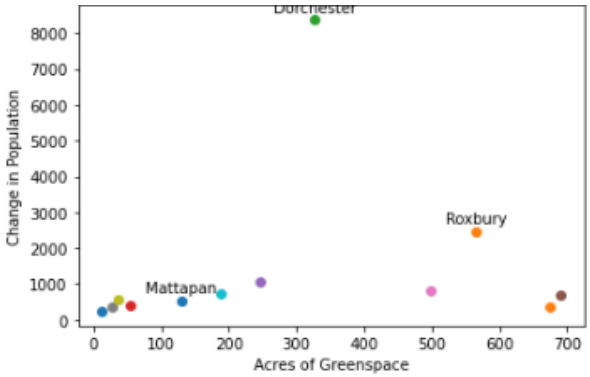
Correlation Coefficient: -0.06984862793003781

Asian



Correlation Coefficient: 0.21481525566943493

Other races



Correlation Coefficient: 0.15605657810809356

For white and Hispanic population, we can see that there is practically no correlation. For the African American population, there seems to be a slight negative correlation while for the Asian population, there seems to be a slight positive correlation. Then for the remaining races that are unidentified, there seems to be a very slight positive correlation. Upon analyzing these plots, we noticed there were a couple of outliers. In the white population, Roxbury is actually seeing an increase in population similar to the plot for all populations. Then in the African American plot, we see that Dorchester had a huge decrease in population, while the other neighborhoods had a slight decrease. Then for other races, we once again see that Dorchester is an outlier, but this time having a dramatic increase in population. While we are not sure of the reasons for this currently, we will dive deeper into this to see what is happening in these specific neighborhoods.

Conclusion

Based on our analysis of the data, it seems that there is no clear correlation between the creation of greenways and new parks, and displacement in Boston. The correlation coefficient for the general population is -0.3 which indicates a weak negative correlation. However, when we look at the plots for specific races, they do not seem to be consistent. As mentioned earlier, white and Hispanic populations have close to no correlation, the African American population has a slight negative correlation, and the Asian population has a positive correlation. Once again, these are all pretty weak correlation coefficients, so we cannot draw any definitive conclusions. In addition to this, we have two major outliers: Roxbury and Dorchester. Roxbury seemed to have an increase in population regardless of the increase in greenspace while Dorchester had major fluctuations compared to the other neighborhoods. This then leads us to the question of what could possibly be causing these outliers and disparities between races. Both of these could be an indication that gentrification is happening. It is probable that the creation of greenways and parks are causing the costs to live in a neighborhood to rise, forcing lower income individuals to move out. Our next step is to determine whether this is the case, so we will need to analyze datasets on housing prices, rent, income, and other relevant factors with a larger focus on the aforementioned outliers, Dorchester and Roxbury.