

Our project tackled the problem of gentrification in the city of Chelsea. We analyzed several data-sets that showed that there is an increased demand for housing in this area and is seeing a direct correlation between demand and growing rent prices. Because this area has long been populated with low income families/individuals, it is no surprise that the increasing rent prices is putting a great deal of stress on the native residents of this area and although the City doesn't want to deter the influx of new people populating this area, it also doesn't want these increasing rent prices to drive out the current existing residents and wants to ensure that they have affordable housing to those that can't keep up with the inflated prices.

We wanted to use specific data-sets that accurately showcased this process of gentrification and adequately brought to light in what ways were the native residents of this area being directly affected by this. For project 1, we used the following data sets:

- ChelseaAssessorsDatabase2018.json
- 24a90fa2-d3b1-4857-acc1-fbcae3e2cc91.json
- income-in-the-past-12-months.json
- City_ZriPerSqft_AllHomes.csv
- Metro_Zhvi_Summary_AllHomes.csv
- Housing-data.xls

This data was collected strictly on the basis of the fact that we know they would serve to be very helpful in analyzing and answering the question of "How is Chelsea being gentrified and what effects is it having on the native residents of this area?" for our project 2.

For Project 2 we created 3 different scripts to ingest the following data: 'income-in-the-past-12-months', 'City_ZriPerSqft_AllHomes.csv', and 'labor-force-and-unemployment-data-chelsea-2001-2017.csv.xlsx'. We developed the income_data.py script that ingested the first given data-set was also utilized in our statistical analysis. We also implemented a script called price_per_sqrt_data.py that was used to ingest the second given data-set into our database. We were able to

transform this data by removing several of the rows and selecting those that are relevant to the question we are trying to answer. Finally, our last script, `unemployment_data.py`, was used to find the unemployment rates of the residents that populated Chelsea, MA and ingest the last data-set and place it into our database. A transformation was also conducted on this data by removing unnecessary columns that didn't adequately showcase what this process of gentrification looks like in Chelsea. We also created a `statistical_analysis.py` script that finds the correlation between unemployment rates and average home prices per square foot and computes the p-value. We wanted to determine if there was any correlation between the unemployment rates and home values and ultimately found that there was actually a strong negative correlation (~-82%). This conclusion is probable because when people have more disposable income, the demand for homes goes up along with the prices. The p-value for this analysis was relatively low at .001 which indicates it was a significant and unique result. We also created a script that performs the optimization and constraint satisfaction algorithms on our datasets. We used the income in the past 12 months dataset and found how much a new home should cost for at least 50% of people to be able to afford it. We used the Z3 library to do this but also devised a brute force solution since the Z3 route was returning irregular results. In our dataset we had the percentage of people in a certain income bracket, so we used this to create several uniform distributions of individuals so we could simulate it.

Additionally, through our findings we saw that because the rent prices in Boston are also increasing there is a large influx of people moving to Chelsea, which is neighboring city to Boston and this is ultimately also a contributing factor to Chelsea's growing rent prices. Also, we saw that moving forward this trend of gentrification is only going to continue so we believe that a probable solution to ensure that low income families can keep living in this area is to increase the amount of affordable housing units per building.

Finally for project 3 we made a visualization of areas that are heat mapped by assessor price vs. the actual price. Our goal here was the ultimately showcase the inflated prices of certain areas in Chelsea that are clear indications of the gentrification process that the city is undergoing; as well as the areas that are considered less desirable in terms of living and do not obtain the inflated rent prices. For our first visualization we plotted the percentage in selling price vs. the actual appraisal price to identify the higher in demand areas; this was the indicate the process of gentrification and displacement of the native population that resides in Chelsea. For our second visualization, we plotted the price per sqft of various living areas to find hotspots where the cost of living is increasing.