RACE AND GENTRIFICATION WITHIN THE US

Damian Franco, Nicholas Livingstone, Datenzing Tamang
CS 467/567 Principles and Applications of Big Data

THE UNIVERSITY OF NEW MEXICO

Background: In recent years, affluent households moving into disinvested areas or neighborhoods has rapidly increased, altering the racial, socioeconomic, and institutional makeup of many urban communities in the US. This transition of low-value neighborhoods into high-value areas has caused long-term residents and businesses to be displaced due to increasing rents, mortgages, and property taxes. We explore various factors contributing to gentrification in the US and try to see if any correlation exists between them.

Hypothesis: There is a correlation between racial demographics of a county and whether that county is considered gentrified.

Methods

We utilized a criteria for gentrification defined by the National Community Reinvestment Coalition[1]. A county can be considered eligible for gentrification or gentrified based on the following requirements:

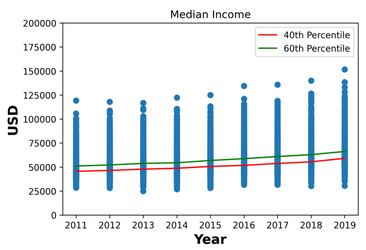
	Eligible	Gentrified
Population	> 500	
Median Home Value (Percentile)	< 40%	> 60%
Median Household Income (Percentile)	< 40%	Increased
Education (Percentile)		> 60%

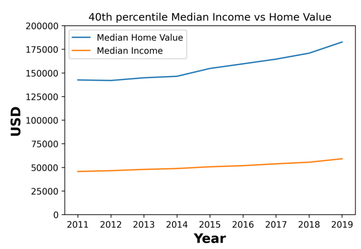
We explored US Census[2-5] datasets covering every US county from 2010-2019 for the following categories:

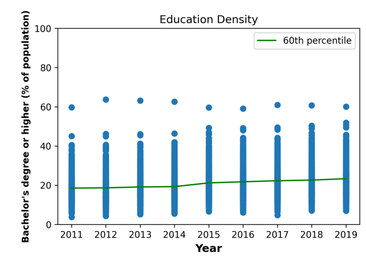
- Housing Price
- Income
- Education
- Population

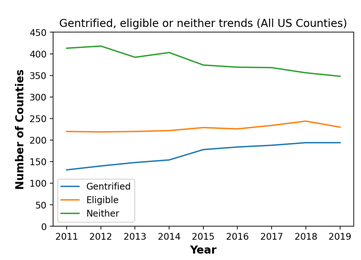
Data Trends

Initial findings showed a clear increasing trend of gentrified counties within the US. Additionally, we found that the median home value has been increasing at a greater rate than median income.



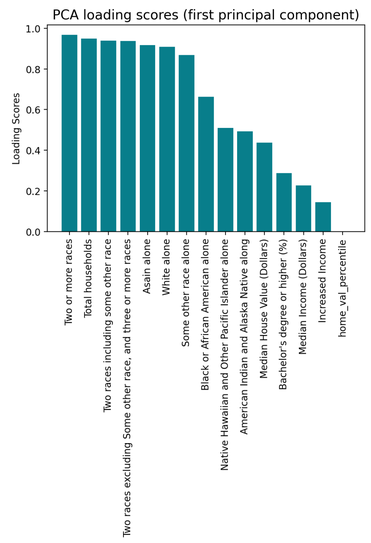


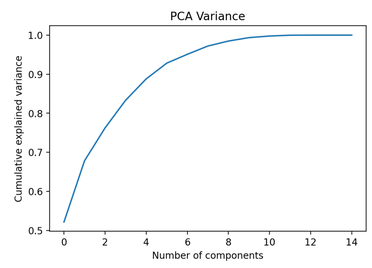




Feature Analysis

These figures show that the 'Two or more races', 'Total households', 'Two races including some other race' and 'Two races including some other race and three or more races' accounts for the largest amount of variance in our data from 2010-2019.





Models

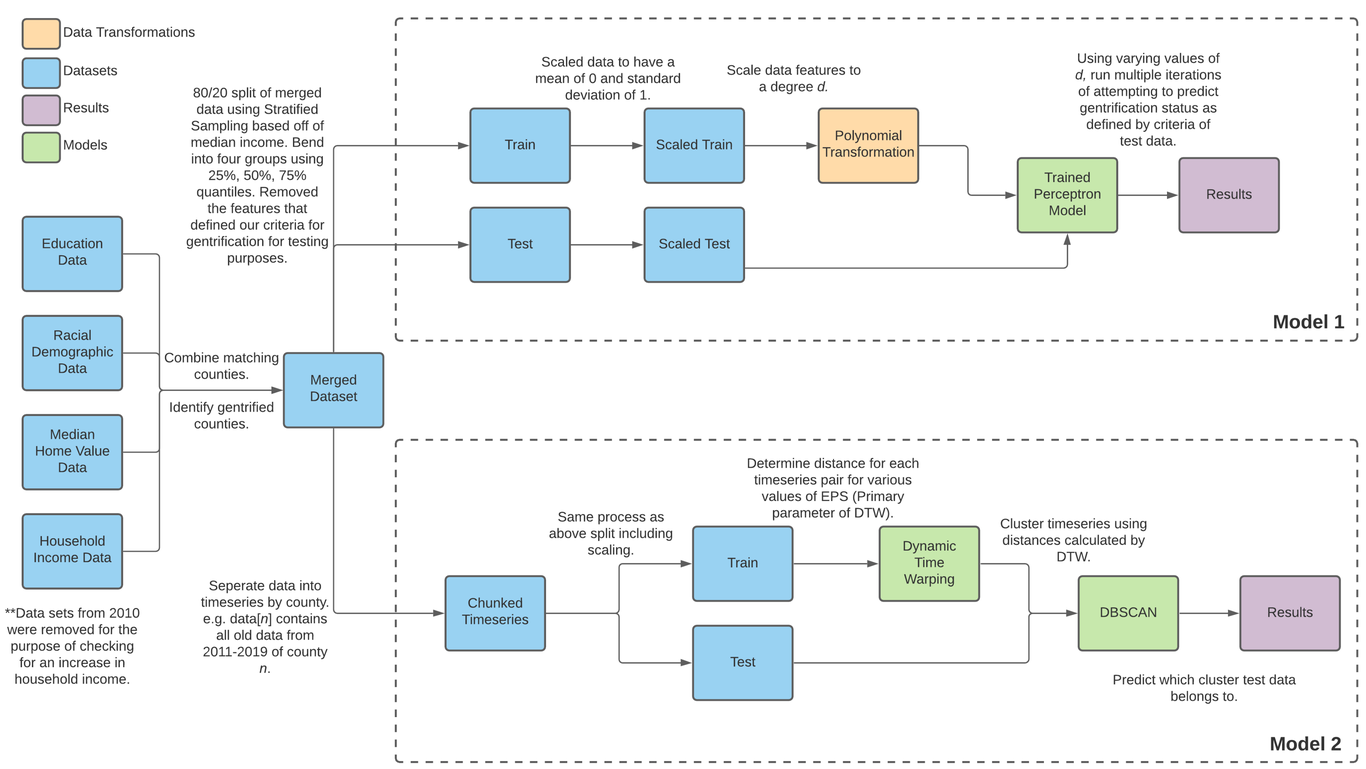
Our team developed two different machine learning models for data analysis. First, datasets are combined, and gentrified counties are identified. Then, the features used for identification were removed. This left only racial demographic data. Both models use the same test/train split and scaling methods (described in the model diagram).

Model 1

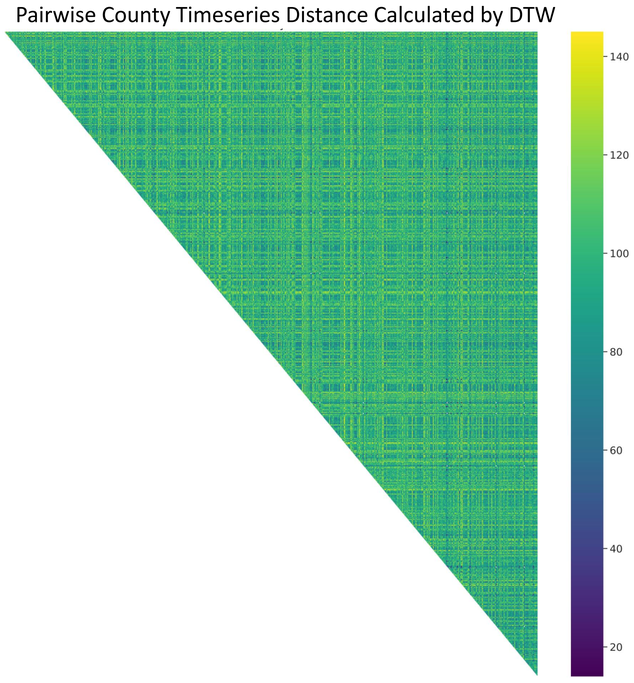
The data is first transformed to *d* dimensions. Then trained on a single perceptron model to predict whether the county id gentrified. We then attempt to predict gentrification status of the test set against the different values of *d*.

Model 2

The data is converted into timeseries for each county. We utilized Dynamic Time Warping to calculate distance between each county's timeseries. Then utilizing these distances, the counties are clustered and labeled. During this process, we compare against the test set and attempt to match them with an existing cluster over different EPS values.



Pairwise County Timeseries Distance Calculated by DTW



Model 1 Results

Degree (*d*)	Testing Score
1	.6724
2	.699
3	.6573
4	.7175
5	.6752
6	.6903
7	.6609
8	.6581
9	.6767
10	.6344

Model 2 Results

EPS	Counties Labeled	Unique Labels
2.0	1.0	1.0
13.0	3.0	2.0
15.0	3.0	1.0
28.0	10.0	2.0
33.0	14.0	1.0
38.0	14.0	1.0
39.0	15.0	1.0
58.0	98.0	1.0
62.0	142.0	1.0

Conclusion

- There is a moderate correlation between race and the gentrified areas.
- It's hard to precisely predict gentrification because there are a lot of different factors to be accounted for such as businesses, policy changes, immigration, etc.
- It is hypothesized that Model 2 failed because the timeseries data was short and varied.
- We hope to work with more complex models to produce better results.
- Our approach in Model 1 could potentially be used to identify other unknown correlated features in a different context.

Reference

1. Gentrification and Disinvestment 2020 - Jason Richardson, Director, Research & Evaluation, NCRC Bruce Mitchell PhD., Senior Research Analyst, NCRC Jad Edlebi, GIS Specialist, NCRC
2. U.S. Census Bureau (2020). Race, 2010-2019 American Community Survey 1-year estimates*. Retrieved from https://data.census.gov/cedsci/table?q=race
3. U.S. Census Bureau (2020). Median Value (Dollars), *2010-2019 American Community Survey 1-year estimates*. Retrieved from https://data.census.gov/cedsci/table?q=B25077%3A MEDIAN VALUE (DOLLARS)
4. U.S. Census Bureau (2020). INCOME IN THE PAST 12 MONTHS*, 2010-2019 American Community Survey 1-year estimates*. Retrieved from https://data.census.gov/cedsci/table?q=Income
5. U.S. Census Bureau (2020). Educational Attainment*, 2010-2019 American Community Survey 1-year estimates*. Retrieved from https://data.census.gov/cedsci/table?q=education