Gentrification Analysis and Prediction

In Association with AI for Good

Rohan Suresh, Dilshan Pathirana, Jeffrey Wirjo, Karam Mukhtar, Shyamkarthik Rameshbabu, Yoon Sung Hong

Overview

Problem Statement

High Level Goals

Architecture

Learning Path

Solution

Next Steps

What is Gentrification?

gentrification noun

gen·tri·fi·ca·tion | \ˌjen-trə-fə-ˈkā-shən\

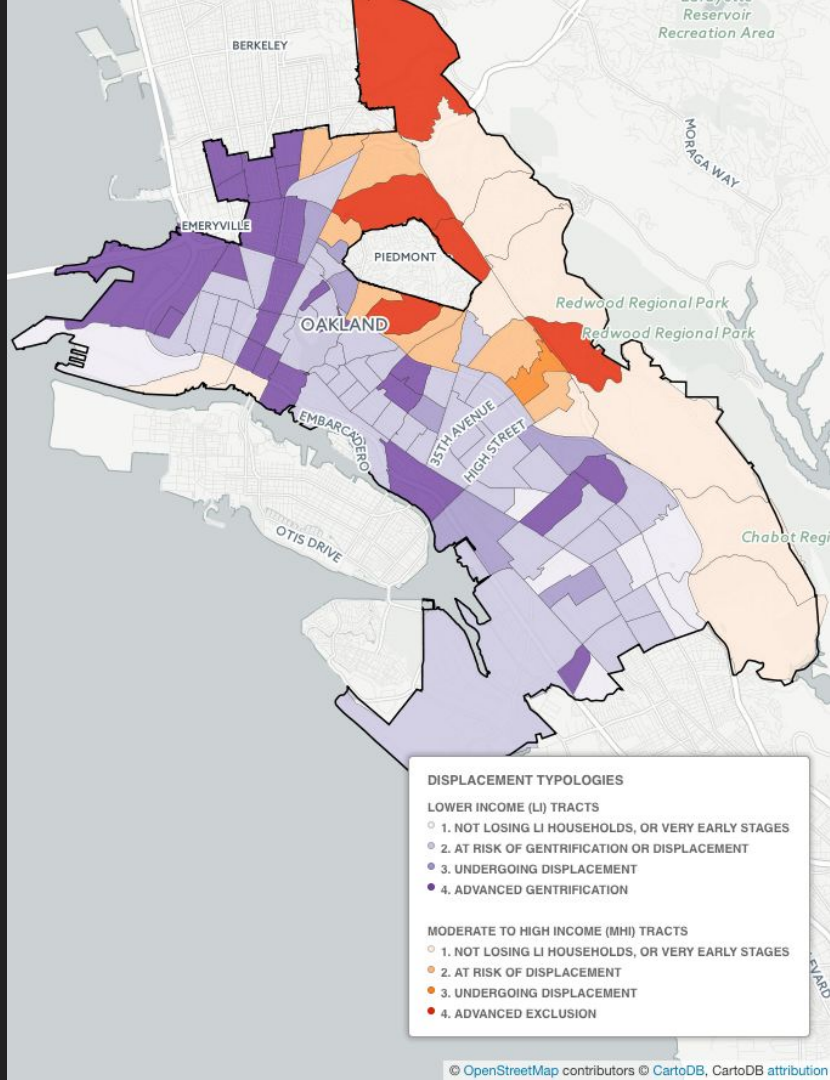
Definition of *gentrification*

: the process of repairing and rebuilding homes and businesses in a deteriorating area (such as an urban neighborhood) accompanied by an influx of middle-class or affluent people and that often results in the displacement of earlier, usually poorer residents

Problem Statement

- Policy makers frequently have inaccurate understandings of relevant factors that impact gentrification.

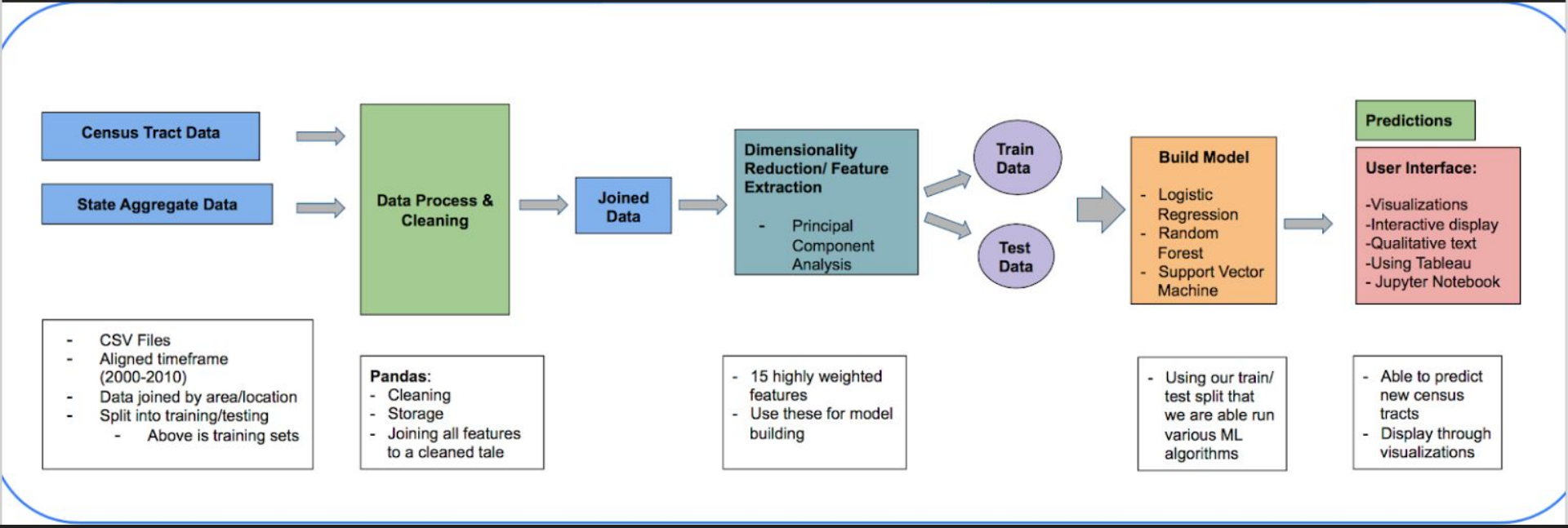
- Result:
  - Forced displacement
  - Discriminatory behavior
  - Exclusion of:
    - low-income communities
    - people of color



High Level - Goals

- **Research** - Identifying the appropriate data source and methods
- **Data Collection** - Areas both gentrified and non-gentrified
- **Feature Selection** - Extracting the significant variables for models
- **Modelling** - Build model to predict the percent change of Y-Labels

Solution Architecture



Learning Path - *An Iterative Process*

Obstacles:

- Predict gentrification using binary classification
- “Correlation is not causation”
- Focus on Midwest Region
- Granularity at county level

Pivot necessary! Needed to nail down our features

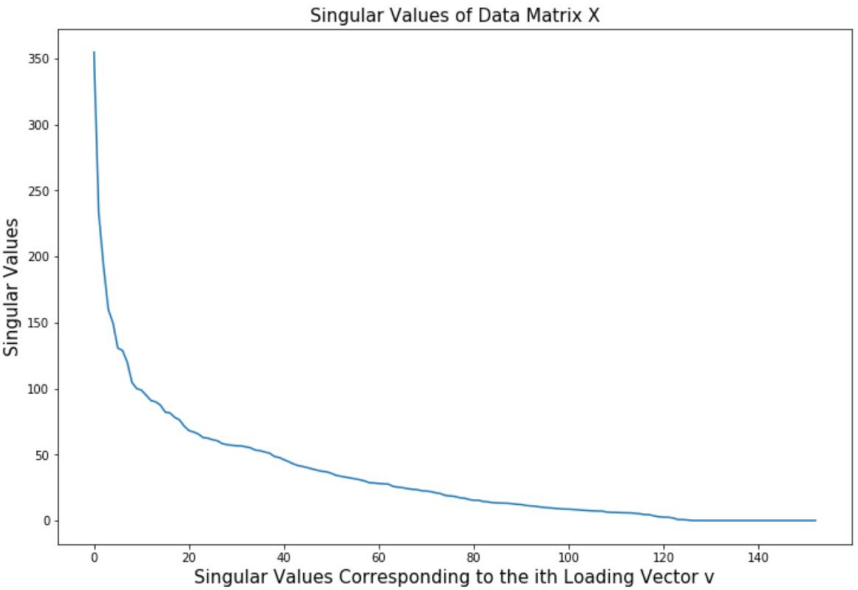
Response:

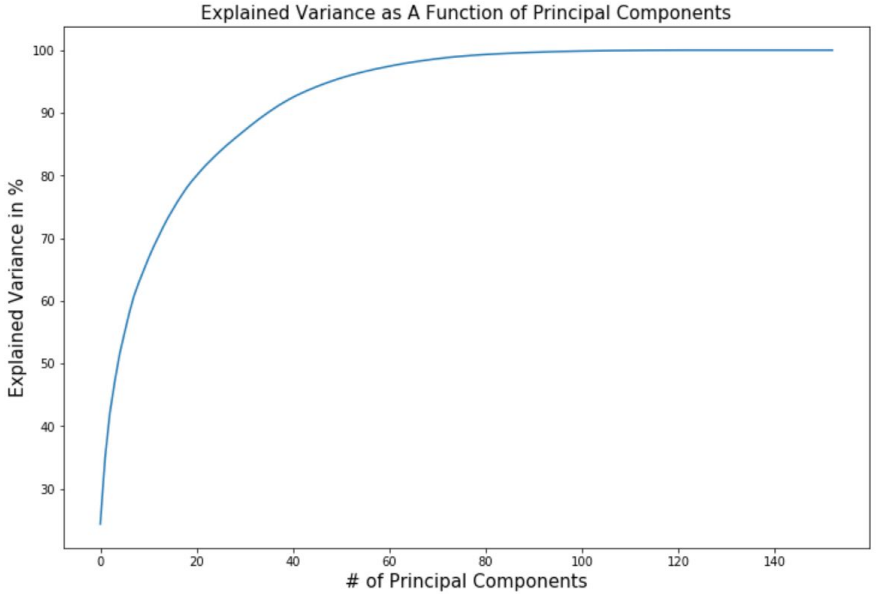
- Comprehensively define how we measure gentrification
- Select areas of isolated change
- Assumption of Midwest unsupported claim
- Found data at a city/ zip-code level
- Found data at census tract level

Final Solution

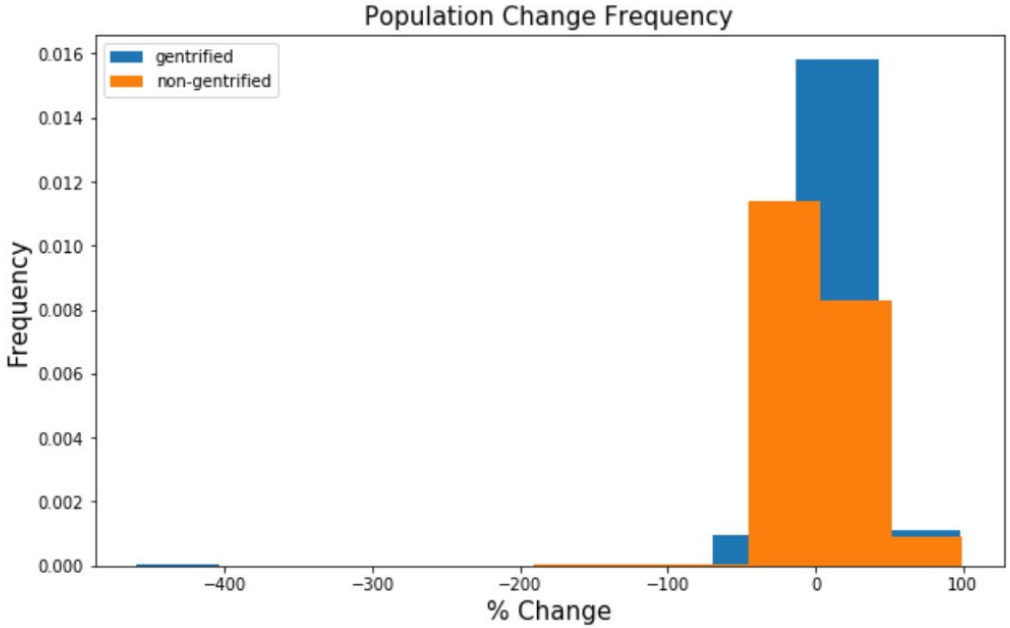
- **Collected data (census tract) for a feature set of 2000 and 2010**
  - Metros chosen from academic research papers
- **Labels are percent change of**
  - Renter occupied units
  - Population density
  - Minorities population (Hispanic/Black)
- **Ran Principal Component Analysis (PCA):**
  - Determine most relevant a subset of features
- **Model Tuning - Train/Test Accuracy:**
  - Built a model that predicts changes in our labels

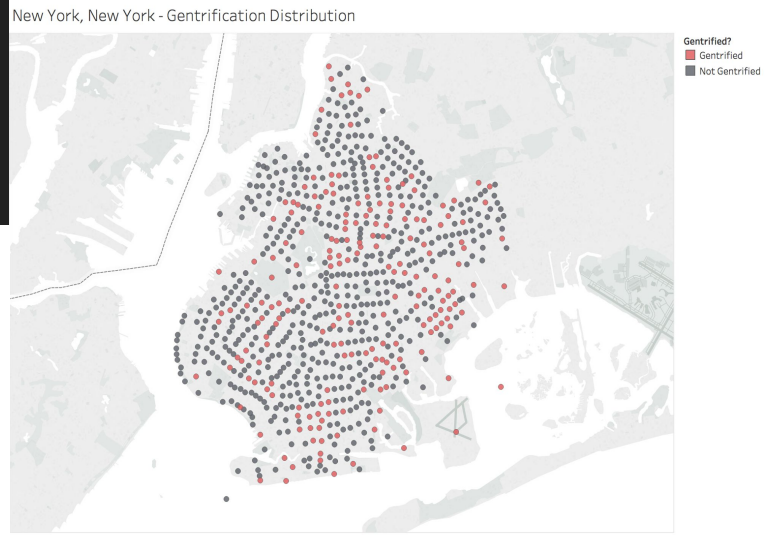
Feature Selection Techniques

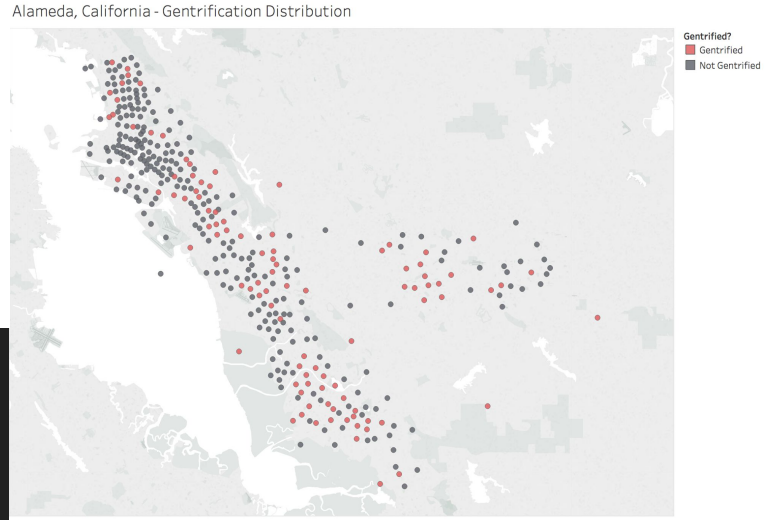




Findings and Visualizations

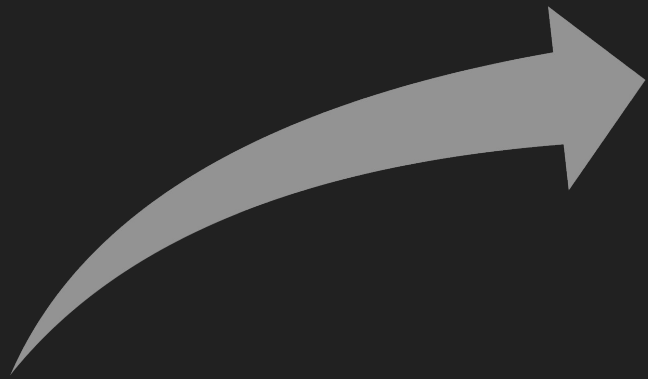




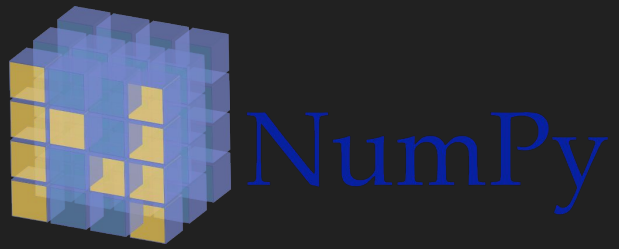


Use Cases/ Next Steps

- Create UI for an interactive model
  - Public Education
  - Scale to community
- Crucial in policy decisions making
- Better prepare for imminent gentrification
- Building block using Machine Learning
  - Collect/train on more data
  - 2000-2010 => current day



Technologies Used



Slide References

https://www.eastbayexpress.com/SevenDays/archives/2015/08/28/uc-berkeley-researchers-map-oakland-gentrification-and-displacement

https://gjplp.org/2017/09/05/examining-the-negative-impacts-of-gentrification/

https://www.python.org/community/logos/

https://softwareengineeringdaily.com/2018/07/18/an-overview-of-jupyter-notebooks/

https://www.bluedata.com/blog/2017/06/large-scale-data-science-operations/

https://geo-python.github.io/2018/lessons/L5/pandas-overview.html

https://en.wikipedia.org/wiki/NumPy#/media/File:NumPy_logo.svg

http://cs229.stanford.edu/proj2016/report/AhmariKarimiXuMistele-GentrificationCausationAndIdentification-Report.pdf

https://trace.tennessee.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1428&context=utk_gradthes

https://eig.org/news/opportunity-zones-map-comes-focus

https://communityinnovation.berkeley.edu/sites/default/files/developing_a_new_methodology_for_analyzing_potential_displacement.pdf?width=1200&height=800&iframe=true