Machine Learning for Cities

Predicting Gentrification via Socioeconomic Changes:

Evidence from New York City 2012-2019

Presented by:

Alec Bardey (ab9732)

Max Magid (mmm9940)

FNU Sonam (ss15624)

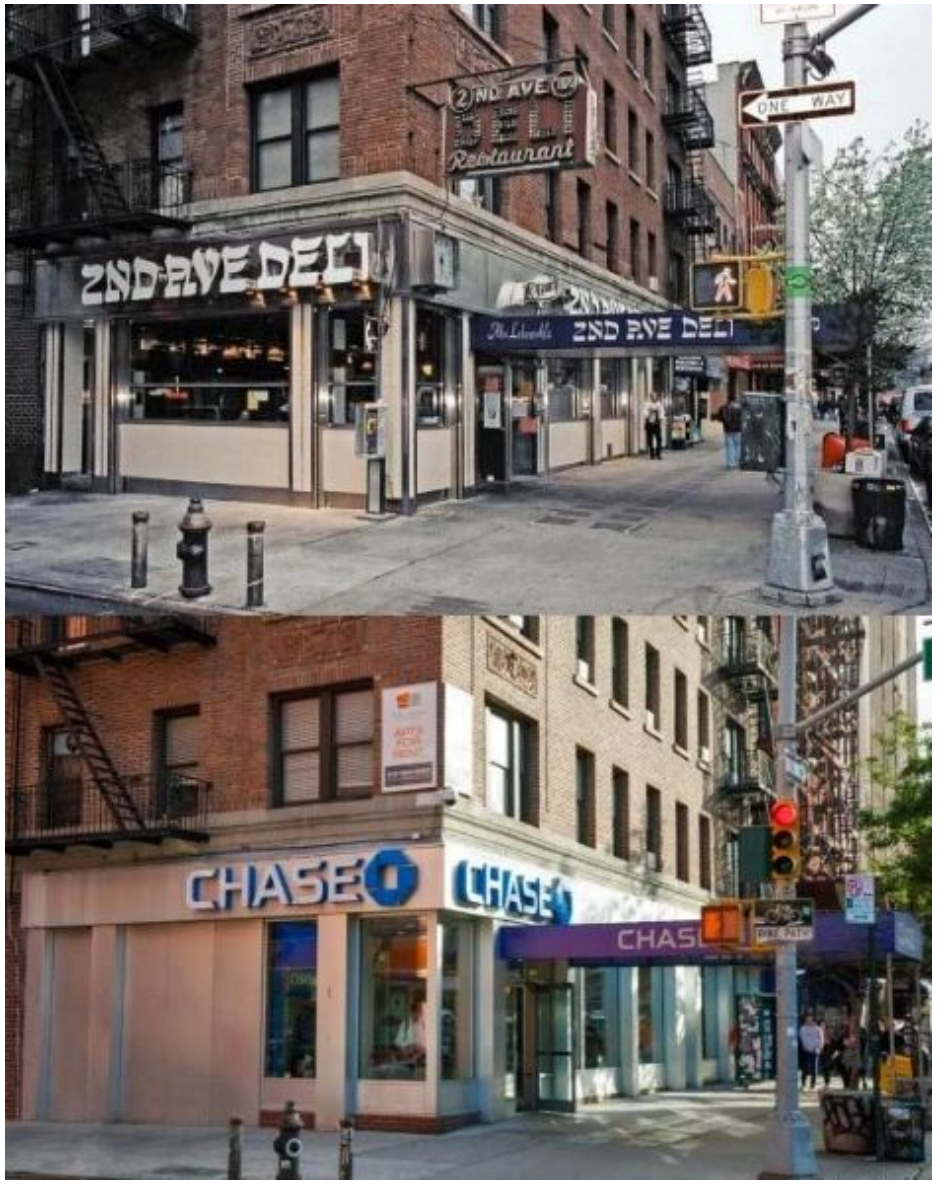
Suraj Sunil (ss14449)

Introduction - What is Gentrification?

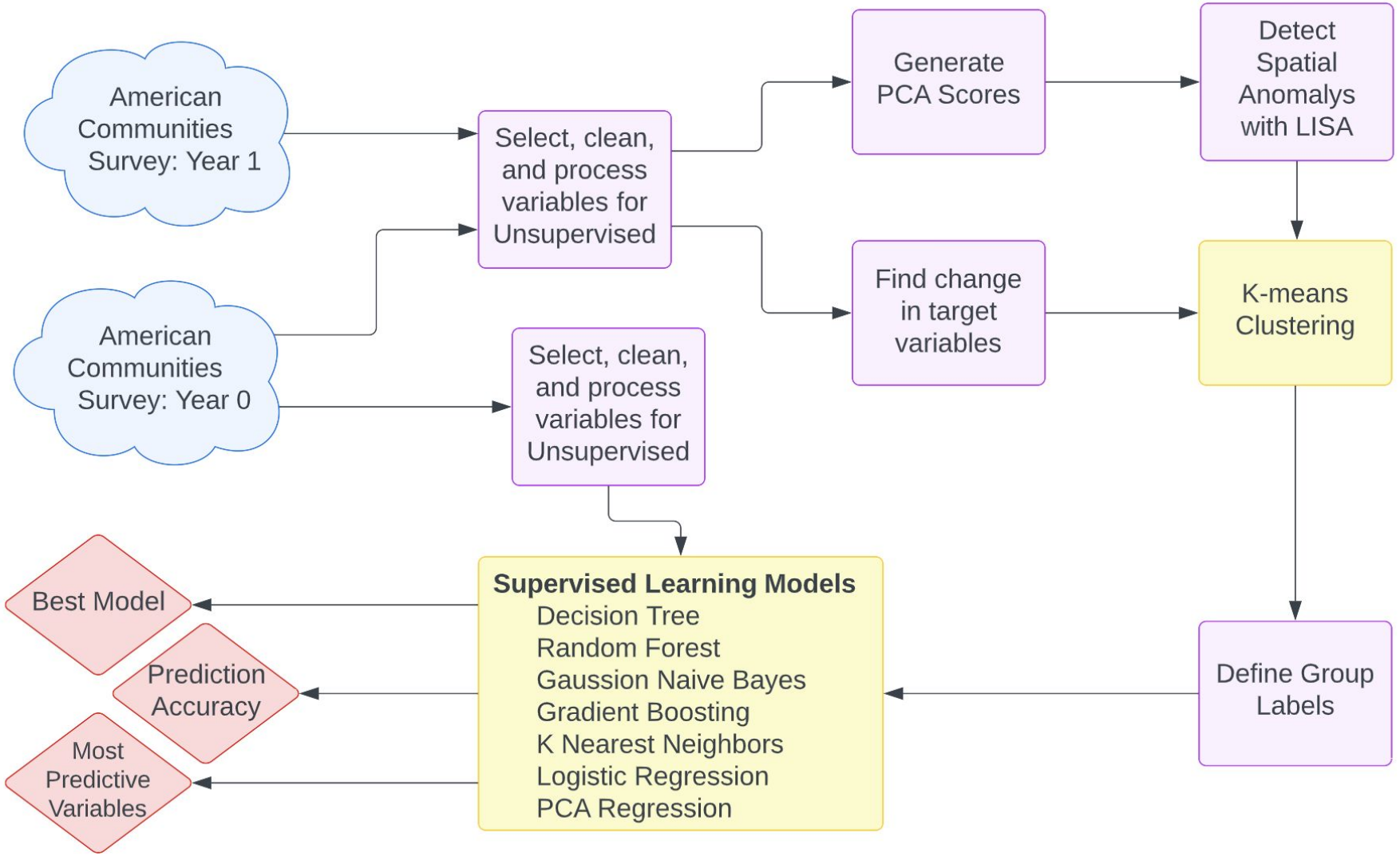
- <u>Inherent definitional ambiguity</u>
  - Urban renewal vs. displacement
- <u>Disproportionate negative impacts</u>
  - Thwarts economic mobility
  - Perpetuate racial inequality
  - Worse health outcomes
- Unique history in America + NYC

Background - What has been done?

- Predictions with unstructured data
  - AirBnB reviews - Jain et al. (2021)
  - Changes in satellite imagery - Ilic et al. (2019)
- Clustering
  - K-means - Knorr (2019)
- Supervised learning
  - Thackway et al. (2021)
- <u>Research Question:</u> How can we identify and predict gentrification in New York City using scalable data?
- <u>Data</u> - American Community Survey



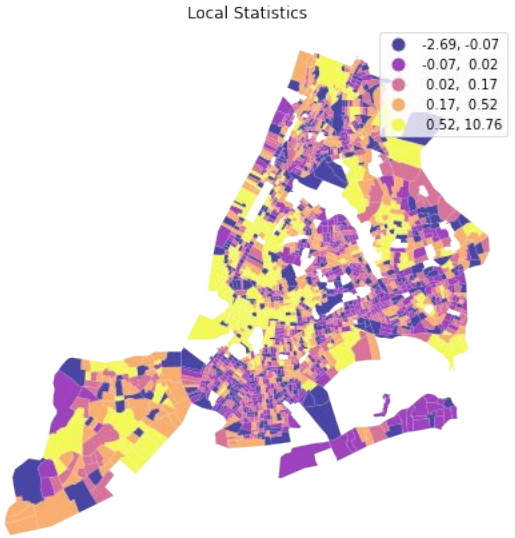
Methodology Overview

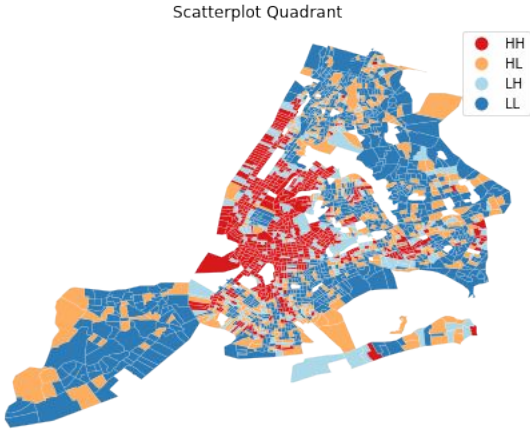


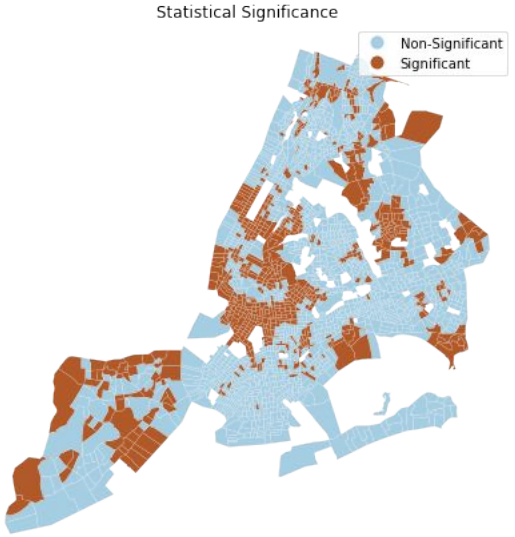
NYU CUSP

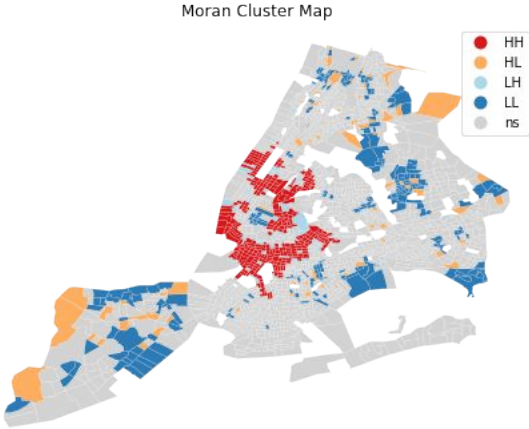
Methodology Unsupervised

- Used Principal Component Analysis to find a single gentrification score
- Subtracted the gentrification scores per tract between Year 0 and Year 1
- Generated clusters of significant positive or negative spatial autocorrelation
- Clustered census tracts by variables including spatial autocorrelation
- Interpreted clusters based on average characteristics to create labels



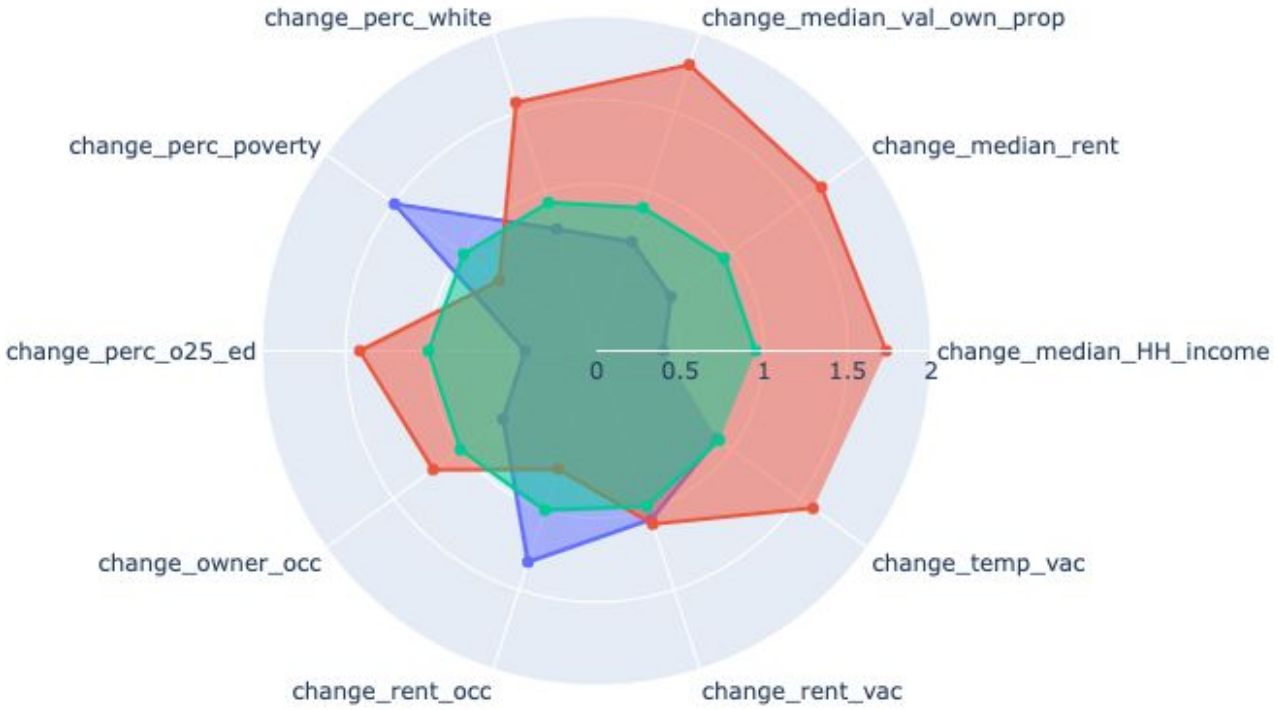


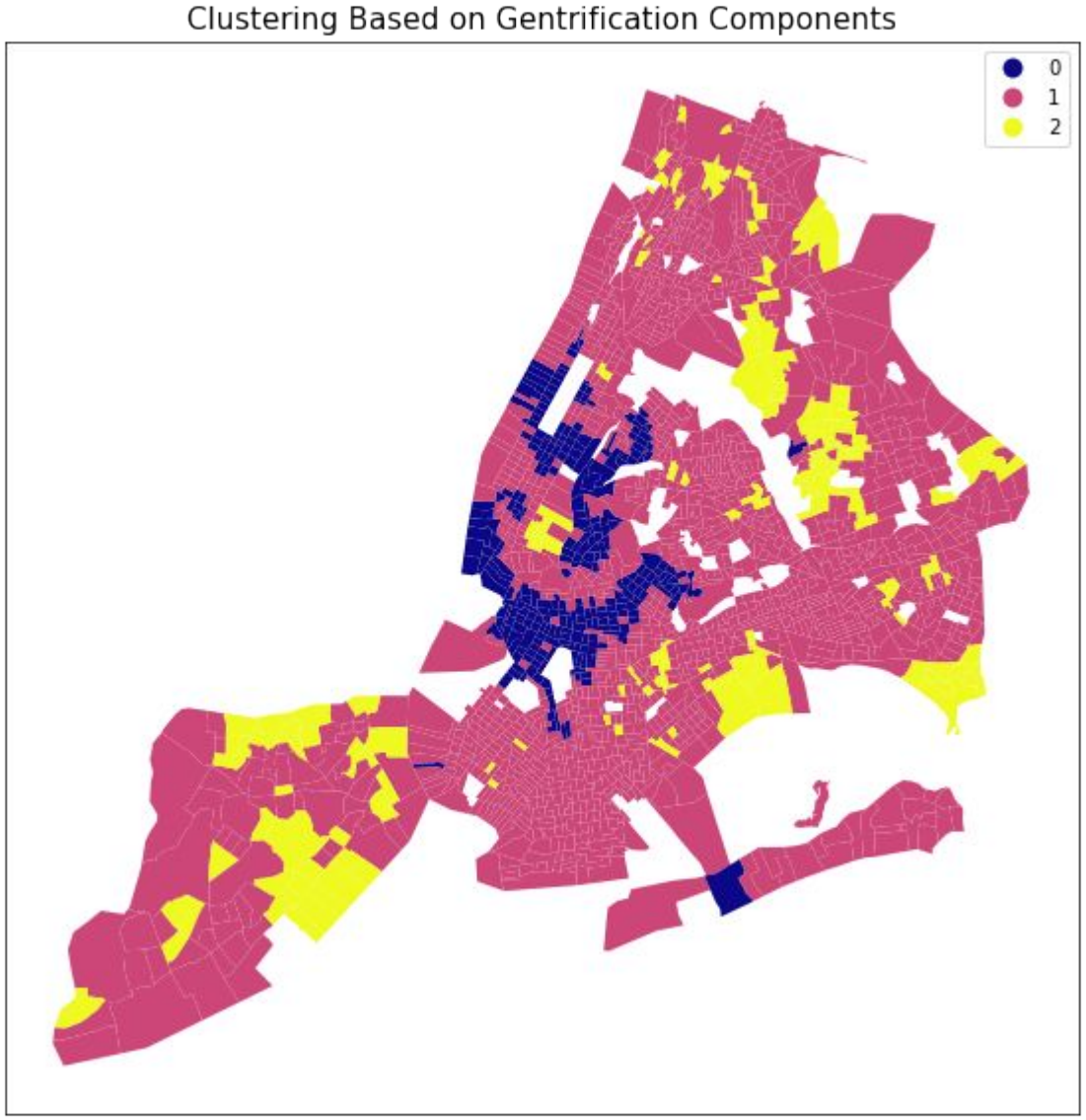




NYU CUSP

Methodology Unsupervised





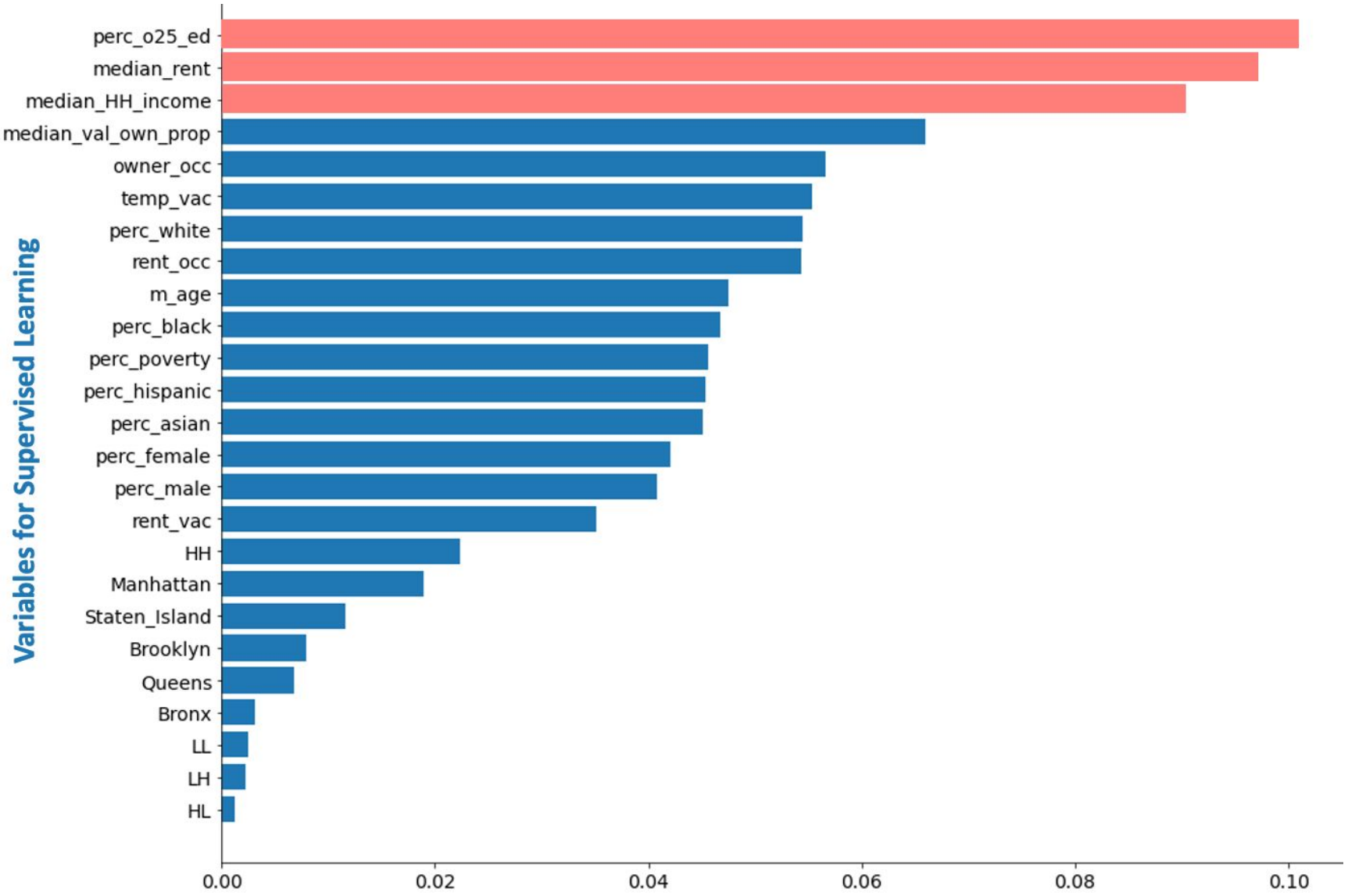
NYU CUSP

Methodology Supervised

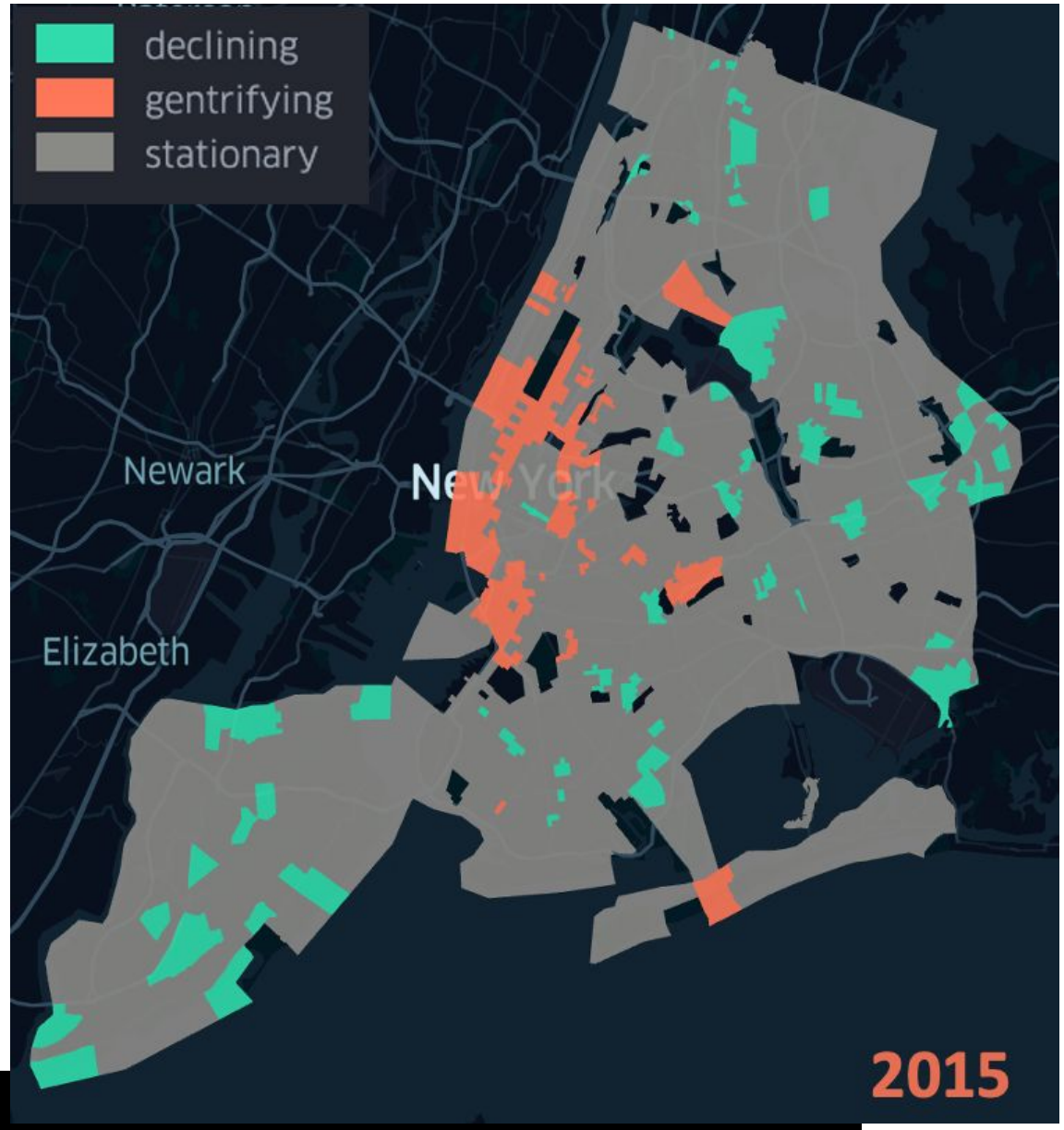
- Conducted unsupervised method for 2012-2015, 2014-2017, and 2016-2019
- Using 2012 and 2014 datasets with labels from unsupervised method as training dataset, 2016 with labels as test data set.
- Added additional demographic information to our training and testing dataset
- Conduct several supervised methods and evaluate their performance by looking at precision and recall of gentrifying areas.
- Fine tune the features on our best method
- Find most important features

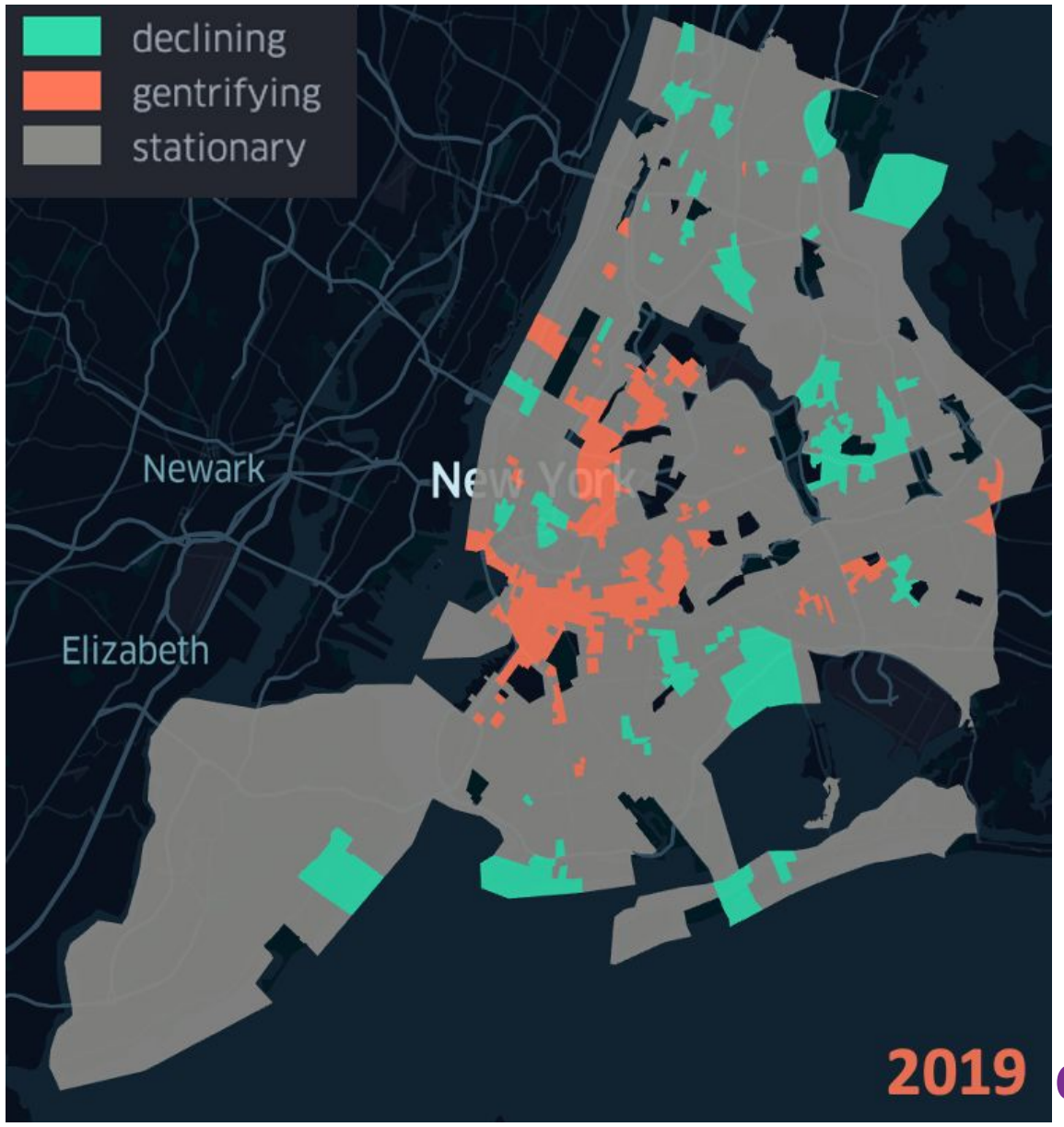
Results and Visualizations

Most important factors for predicting gentrifying areas



Results and Visualizations





Conclusion & Future Work

- Random Forest Model improves precision
- PCA Regression Model improves recall

- Our model worked well in predicting gentrification using both the methods

- More attributes can be studied further which increase the accuracy of the model for better results


Thank You