Modeling Gentrification in NYC: Time Series and Neural Approaches

Tyler McKenzie & Ben Tunney

Northeastern University

Introduction

This project explores gentrification across a subsection of NYC sub-borough areas from 2005 to 2019. We apply time series techniques to analyze neighborhood changes, using median household income as a primary indicator. We apply a convolutional neural network in attempts to classify gentrification based on a variety of neighborhood-specific demographic features. The time series analysis presented on this poster focuses specifically on the Bushwick, and the neural network is applied to all neighborhoods.

Data

- 15+ datasets from NYC OpenData: income, education, race, housing, transit, age
- Community Districts were mapped to Sub-Borough Areas
- Time Series Focus variable: median household income (2005–2019)
- CNN Focus variable: Gentrification (binary)
  - Low-income Sub-Borough Area experienced income level increases higher than the median Sub-Borough Area [1]

References

1. Austensen et al. (2016). *State of New York City's Housing and Neighborhoods in 2015.* https://furmancenter.org/files/sotc/NYUFurmanCenter_SOCin2015_9JUNE2016.pdf
2. Reades et al. (2018). *Understanding urban gentrification through machine learning.* https://doi.org/10.1177/0042098018789054
3. U.S. Census Bureau (2023). *Identifying Gentrification using Machine Learning.* https://www.census.gov/content/dam/Census/library/working-papers/2023/demo/sehsd-wp2023-15.pdf

Time Series: Methodology

- Decomposed time series into trend, seasonal, residual
- Analyzed ACF and PACF to guide lag selection for ARMA model
- Fit ARMA(1,1) using least squares based on diagnostics

$$y_t = \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

Example ARMA(1,1) structure

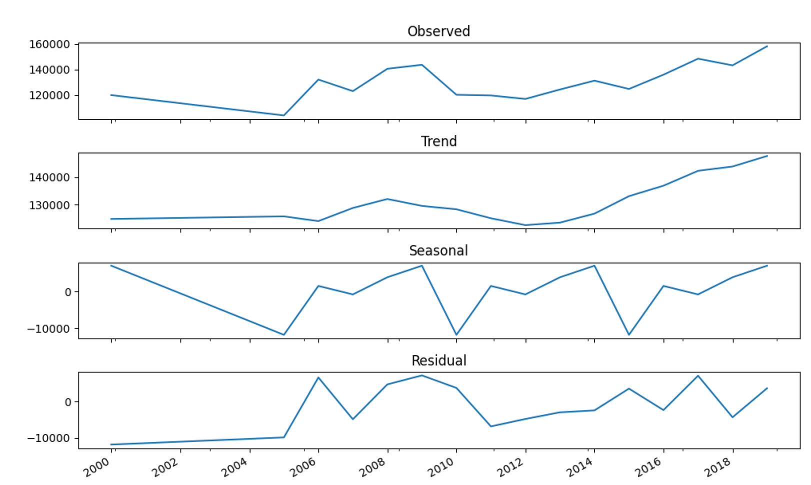

Figure 1: Decomposition of median income into trend, seasonal, and residual components

Time Series: Results

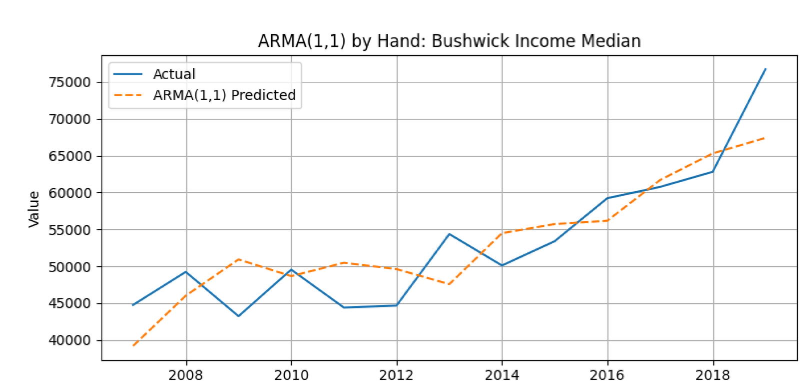

Figure 2: Median income with ARMA(1,1) fit

Neural Network: Methodology

- Flattened 3-dimensional data (15 years and 12 measures) into 180 features
- CNN architecture
  - 1-D Convolutional Layer
  - Leaky ReLU Activation Function
  - 1-D Max Pooling Layer
  - Fully Connected Layer
  - Sigmoid Activation Function
- Updated weights according to loss changes using numerical gradient estimation, minimizing binary cross-entropy loss

Neural Network: Results

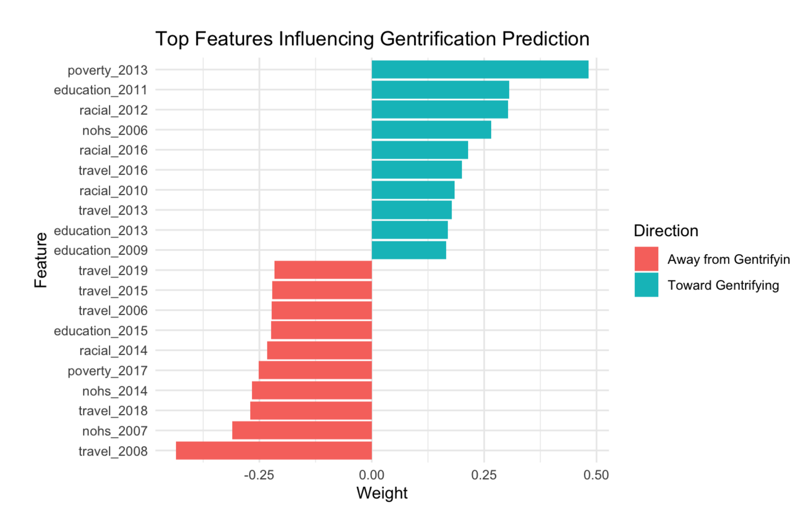

Figure 3: Largest Coefficient Features

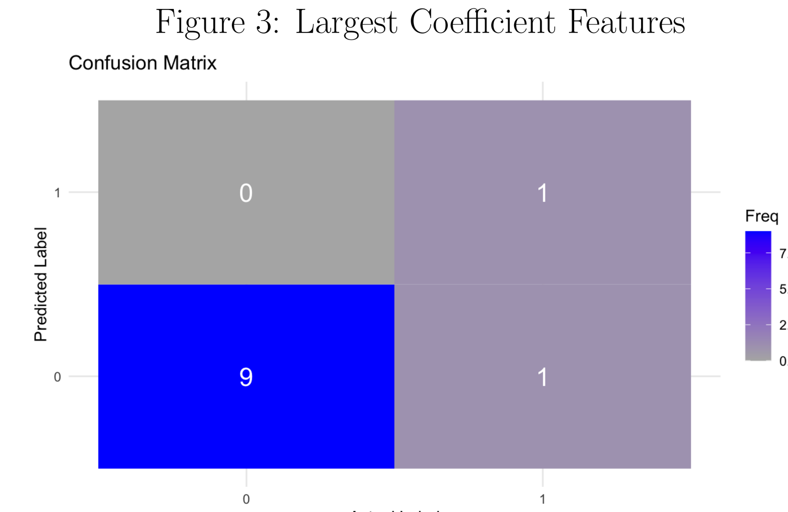

Figure 4: Predictions Matrix

Key Takeaways

- Median income rose steadily in the Upper West Side post 2012, a sign of gentrification
- ARMA(1,1) balances accuracy and simplicity for time series modeling
- We were able to achieve 90.91% accuracy for the CNN predictions of the gentrified label.
- The results are mixed in terms of which features have the strongest negative and positive weight to predict gentrification.
  - Education and Racial features were positive contributors for 6 of top 10 positively weighted features. This may indicate that improvements in elementary education, particularly in reading, and increase in racial diversity can be strong indicators that a neighborhood may become gentrified.
  - The travel feature was a negative contributor towards gentrification for 5 of the top 10 negatively weighted features. This may indicate that a higher average commute levels contributes to non-gentrification.