Predicting Gentrification Across NYC Neighborhoods

Ahmad Nasrieh (an3951), Ananya Rajesh (ar7603), Claire Chen (hc3491), He Gao (hg2683), Liz Johnson (emj6930)

Abstract:

Gentrification has become a significant issue in New York City over the past few decades, leading to the displacement of lower-income, minority residents and the lack of affordable housing in neighborhoods accessible to public infrastructure and resources. Predicting gentrification and identifying which neighborhoods are at risk of it helps plan for the potential impact on low-income residents, the affordable housing crisis, and the preservation of minority group neighborhood dynamics. In this project, we propose a time series analysis of specific economic output variables, mainly median property sale values, and 311 complaint density, across Neighborhood Tabulation Areas (NTAs) in NYC. We isolate trends in these variables over a twelve-year period and adapt Group-Based Trajectory Model to cluster similar neighborhoods. We hypothesize the cluster output will return at least four distinct groups corresponding to (1) historically high-income, (2) historically low-income, (3) gentrified, and (4) currently or soon-to-be gentrifying neighborhoods. We will then use the historical time series data and clustering outputs to predict how gentrification will impact lower-income residents.

Introduction:

Gentrification is the renewal and revitalization process of lower-income neighborhoods due to an influx of middle-class residents and the displacement of the original, often poorer residents. Various economic and urban planning resources have studied gentrification, and it has been identified as a complex and multidimensional process involving a range of economic, social, and cultural factors. The first is economic growth driven by the influx of higher-income householders, as opposed to the upgrading of the economic status of current residents. The second is the displacement of lower-income residents in the neighborhood revitalization process. The third factor is the change in racial or ethnic distribution within the community.

Quantitative research on the topic of gentrification is limited. In this project, we propose an original methodology to predict when and where gentrification will impact lower-income residents by adapting the Group-Based Trajectory Model to cluster time series of various economic variables. By identifying which neighborhoods are at risk of gentrification, policymakers, and urban planners can take proactive steps to mitigate the negative impact of gentrification on low-income residents, the affordable housing crisis, homelessness, and the preservation of minority group neighborhood dynamics and local businesses. This research can inform urban planning decisions and housing policies that promote equitable development and social justice in NYC.

Literature Review:

The Urban Displacement Project[1]

The Urban Displacement Project defines nine types of neighborhood change to help communities better understand their trajectories and stabilize their resident population. This research is a collaboration between UC Berkeley and the NYU Center for Urban Science and Progress with sponsorships from the New York Local Initiatives Support Corporation (LISC) and other organizations. The project's methodology looks at demographic and housing changes in each neighborhood over 31 counties and compares them with the regional median. The interactive map provided by the Urban Displacement Project shows how these nine categories are spaced out across the city ranging from neighborhoods that are not losing low-income households to those that are experiencing super gentrification or exclusion. Some of the significant findings of this work are that displacement can take shape in different ways across the city and that some neighborhoods that do not seem like they are actively gentrifying at face value are experiencing a ripple effect of gentrification from other neighborhoods.

Furman Center[2]

The Furman Center also showcases how gentrification has impacted the city from 1990-2014. This research looks at rent affordability, demographic changes, and other indicators like the proximity to public housing developments. Further, the center categorizes neighborhoods into three categories, non-gentrifying neighborhoods, gentrifying neighborhoods, and higher-income neighborhoods. The research outlines some neighborhoods that have faced drastic implications of gentrification at higher average levels including Williamsburg/Greenpoint, Central Harlem, Lower East Side/Chinatown, Bushwick, East Harlem, Morningside Heights, and Bedford-Stuyvanstant. The Furman Center also notes some insightful inferences including that the average income rose by 14% only in the gentrifying neighborhoods within this timeframe when compared to the non-gentrifying and the higher-income neighborhoods.

---

[1]Urban Displacement Project. (n.d.). New York Gentrification and Displacement. Retrieved May 3, 2023, from https://www.urbandisplacement.org/maps/new-york-gentrification-and-displacement/
[2]Stephanie Rosoff. (2016, September). Indicators of Neighborhood Change: Presentation to the Community Indicators Consortium 2016 Impact Summit. Retrieved from https://communityindicators.net/wp-content/uploads/2017/12/2016_Rosoff_Measuring_Gentrification_NYC-1.pdf

***Structure of 311 Service Requests as a Signature of Urban Location*[3]**

This research paper from the NYU Center for Urban Science and Progress examines how using spatiotemporal patterns from 311 data can offer insight into different neighborhood areas. The research utilizes classification to distinguish neighborhoods by 311 complaints as a means of grasping a better understanding of the socio-economic status. The literature found that 311 data can help predict changes in socio-economic characteristics including real estate prices. Our research aims to analyze 311 data and real estate values as well to explore the scope of gentrification in the city.

Data & Methodology:

Data

311 Service Requests from 2010 through the end of 2022.

- Provided by the NYC Open Data portal. We will leverage the Socrata API query language to extract data at a monthly cadence aggregated to the NTAs as defined by the New York City Department of City Planning. The New York City 311 Service Requests has three tiers of descriptions: agency name, complaint type, and descriptor, which together account for roughly 1950 unique request types. In order to make our analysis more efficient, we will look at the request frequency for the unique agency names which decreases the number of unique request types to 35.

Property Sales Data

- Available at monthly cadence across the neighborhoods of New York City through the StreetEasy Data Dashboard[4]. This dashboard contains the median sale price across their predefined neighborhoods beginning in January 2010 to the present day. This dataset is important in answering our hypothesis because:
  - Price appreciation and change in property values over time is a main indicator used to track gentrification. In gentrifying neighborhoods, property values tend to rise rapidly as more affluent residents move in, and demand for housing increases.
  - Changes in property sales patterns can also provide insight into broader neighborhood changes and trends that may signal gentrification.

Neighborhood Tabulation Areas (NTAs) Data

[3]Lingjing Wang, Cheng Qian, Philipp Kats, Constantine Kontokosta, & Stanislav Sobolevsky. (2017, October). Structure of 311 service requests as a signature of urban location. PLOS ONE, 12(10), e0186314. doi: 10.1371/journal.pone.0186314

[4] StreetEasy. (n.d.). Data dashboard. Retrieved May 2, 2023, from https://streeteasy.com/blog/data-dashboard/

- As defined by the New York City Department of City Planning (DCP), NTAs are medium-sized statistical geographies created by aggregating census tracts but are smaller than Community District Tabulation Areas. DCP acknowledges that NTA boundaries and their names roughly correspond to many neighborhoods recognized by NYC natives, however, NTAs are not intended to definitively represent neighborhoods[5]. It is this fact that made our analysis most difficult because the NTA boundaries do not correspond exactly with the boundaries of the StreetEasy neighborhoods. Out of the 195 unique NTAs, our analysis only contains 120 after excluding Staten Island, neighborhoods that do not overlap with StreetEasy's available data and neighborhoods with insufficient data.

Methodology

- First, we will construct a multivariate time series (MTS) for each NTA. We will use the time series of 311 complaints across the 35 unique agencies to predict the median property sales price from StreetEasy. We hope to observe neighborhood characteristics through the propensity to submit service requests of a certain category and hypothesize this can be used as an accurate predictor of other socioeconomic metrics at the neighborhood level based on previous research results[6].

- To identify neighborhoods at-risk of experiencing gentrification we will isolate the temporal trends of the sale price for each NTA by computing the first-order difference of the MTS predictions and by performing a cluster analysis to identify whether prices were generally increasing or decreasing. We will use the K-mode clustering algorithm with k=2 to divide the NTAs into two clusters. Cluster 1 indicates overall positive rate of change in property price over time, while cluster 0 indicates negative or constant growth rate of change. We expect neighborhoods in cluster 1 to be gentrified or gentrifying and cluster 0 would include more stable neighborhoods of historically high-income or low-income residents.

  For the purpose of this research, we will use the findings from this cluster analysis to identify price appreciation and neighborhood change which are both indicators of gentrification. It is important to note that we have not investigated any metrics to measure displacement and thus cannot definitively predict gentrification. We will look to supporting research and literature to compare our results.

[5] City of New York. (2020). 2020 Neighborhood Tabulation Areas (NTAs) tabular data. Retrieved May 2, 2023, from https://data.cityofnewyork.us/City-Government/2020-Neighborhood-Tabulation-Areas-NTAs-Tabular/9nt8-h7nd

[6] Lingjing Wang, Cheng Qian, Philipp Kats, Constantine Kontokosta, & Stanislav Sobolevsky. (2017, October). Structure of 311 service requests as a signature of urban location. PLOS ONE, 12(10), e0186314. doi: 10.1371/journal.pone.0186314

- Group-based trajectory modeling is designed to identify clusters of individuals who are following similar trajectories of a single indicator of interest, sale price as a function of 311 requests in this case. It recognizes groups of individuals that exhibit similar temporal progressions and calculates the influence of covariates on both group membership and trajectory shape[7]. By clustering neighborhoods based on predicted sale prices over time, we look to identify communities that have or are expected to experience rapid price growth and gentrification. We will use K-Means clustering to achieve our outputs and optimize the number of clusters by iterating through k-values and investigating SSE (Sum of Squared Errors) and Silhouette Scores.

Results:

Multivariate Time Series Analysis

A multivariate time series model was trained with the 311 complaints by city agency as dependent variables and the median property sale price as our independent variable. Below you will find the results of our MTS for a subset of NTAs, which showcases relatively accurate predicted values (orange) as compared to the actual property prices (blue). The predicted values follow the same trend as the actual values, indicating that the linear regression model is a reasonable fit for the data.

Overall, the results showed that the model was able to make accurate predictions for most of the NTAs. The majority of the r-square values ranged from 0.4 to 0.9, indicating that the model was able to capture a significant portion of the variation in the median sale price. However, there may be additional factors not included in the dataset that could influence median sale prices, and further analysis and refinement of the model may be necessary.

[7]Nagin, D. S. (n.d.). *Group-based trajectory modeling in clinical research*. Retrieved April 28, 2023, from http://www.contrib.andrew.cmu.edu/~bjones/refpdf/NaginOdgers2010.pdf

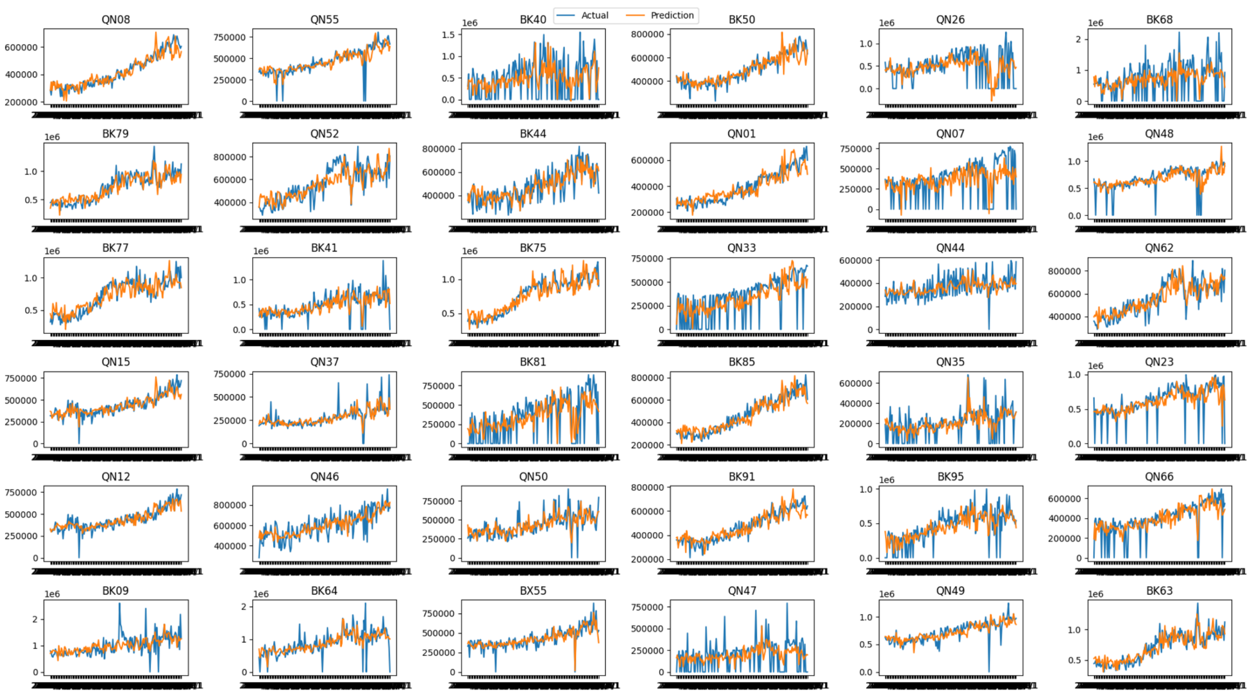


Figure 1.1: Predicted Values vs. Actual Values

Utilizing MTS Trends to Identify At-Risk Neighborhoods

We isolated the temporal trends for each NTA by computing the first-order difference of the MTS prediction and conducted K-mode clustering with k=2 to visualize the temporal property trends across the city.

When observing these clusters spatially, we can see that communities with historically high-income residents have negative or constant property value growth. This can be seen in areas like Downtown Brooklyn, Brooklyn Heights, Cobble Hill, Park Slope, Upper West Side, Hell's Kitchen, and parts of the Upper East Side. Unfortunately, due to the amount of data lost when connecting NTAs to the StreetEasy neighborhoods, a lot of low-income communities in the Bronx are excluded from this map but we hypothesize that historically low-income communities would be included in this cluster.

Our findings show that wealthier neighborhoods have lower or stable property price growth, while areas like Bushwick, Williamsburg, Greenpoint, and Long Island City have experienced an increase in property prices in the past twelve years. This is similar to what was found in the Furman Centers Gentrification map. However, we think it is important to mention that the areas of Bedford-Stuyvanstant and Crown Heights, which have been labeled as gentrifying in Furman's analysis, are labeled as negative or stagnant neighborhoods in our results.

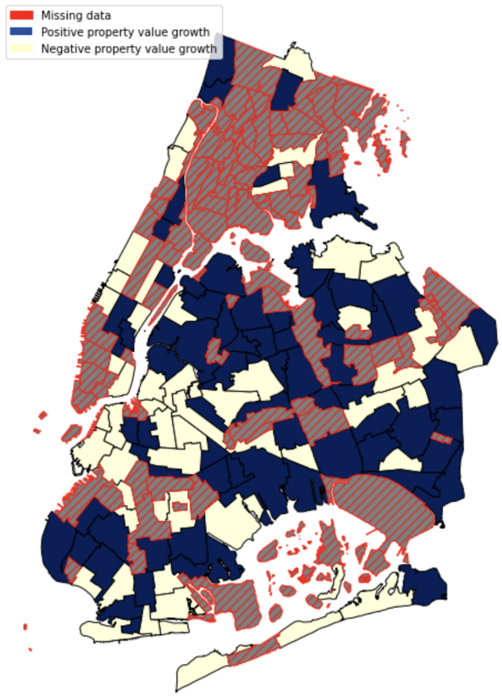


Figure 2.1: Our Gentrification Map

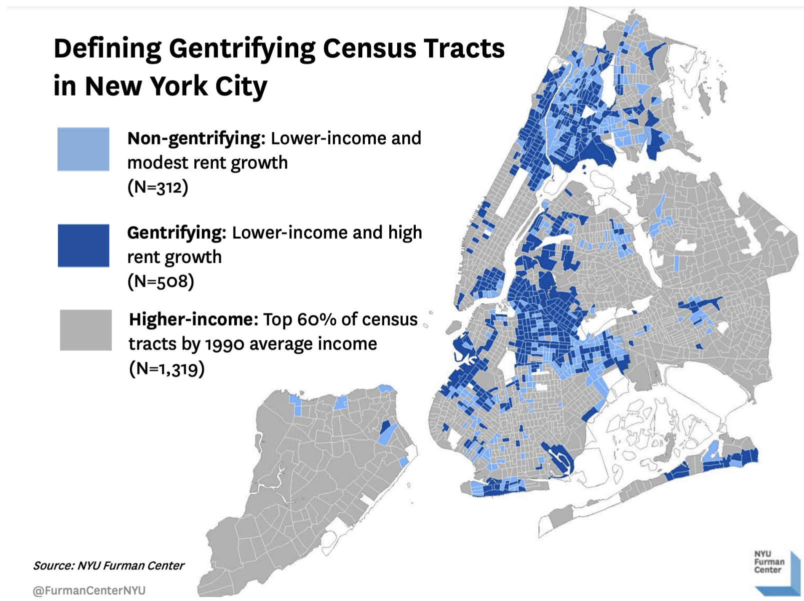


Figure 2.2: Furman Center Gentrification Map

Group-Based Trajectory Model

After conducting k-means clustering on the MTS predictions we found four to be our optimal cluster number. See the figure below for the plot of the silhouette score (left) and SSE (right) as a function of cluster number. We chose four instead of three as the silhouette score remained the same but the error continued to decrease. This result also matched our original hypothesis that the cluster output will return at least four distinct groups corresponding to (1) historically high-income, (2) historically low-income, (3) gentrified, and (4) currently or in the near future gentrifying neighborhoods.

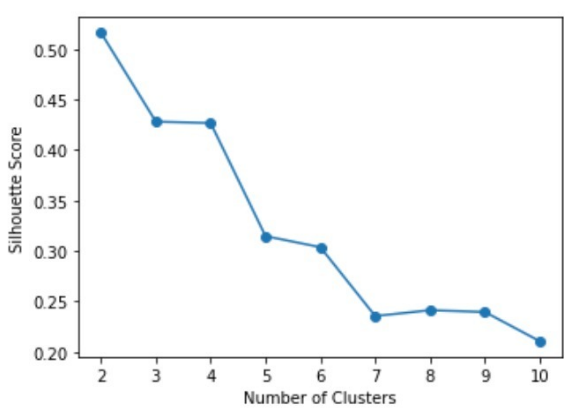


Figure 3.1: Silhouette Score

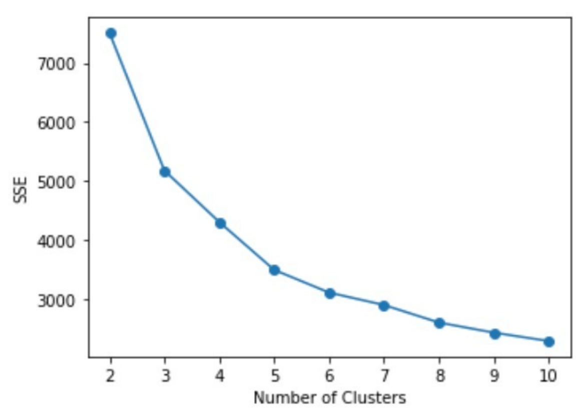
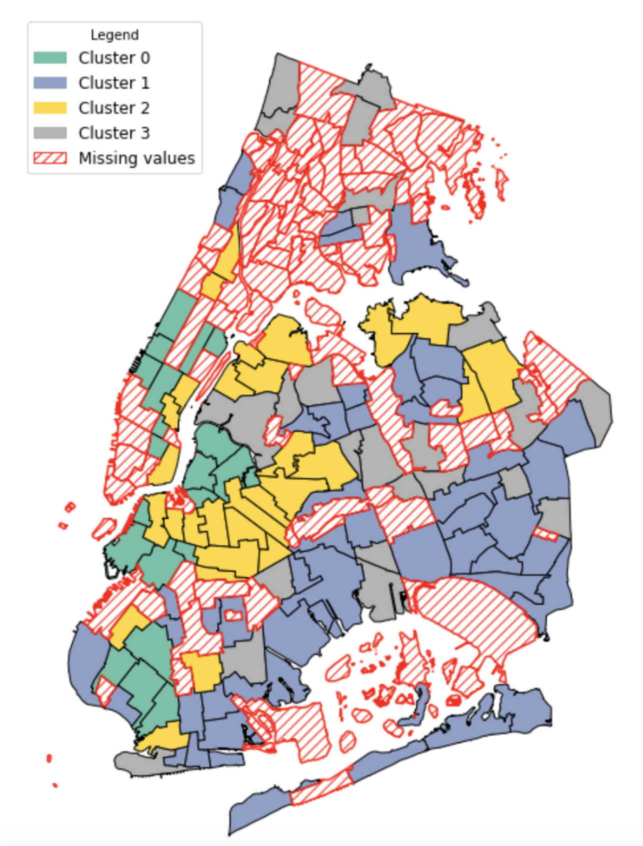


Figure 3.2: SSE

Our results (left), when compared to those of the Urban Displacement Project's Gentrification and Displacement Map (right), seem consistent. We see that cluster 0 and cluster 1 in our results correspond closely to the Stable Exclusion typology. Cluster 2 corresponds closely to At Risk of Gentrification and Ongoing Gentrification typology, mainly the areas of Bedford-Stuyvesant, Crown Heights, and Bushwick in Brooklyn, Astoria in Queens as well as East Village and the Lower East Side of Manhattan. Cluster 3 overlaps with areas of low-income typology. For example, some neighborhoods in East New York, the Bronx, and Flushing are all labeled cluster 3 in our results which overlaps with the Urban Displacement Project's results.



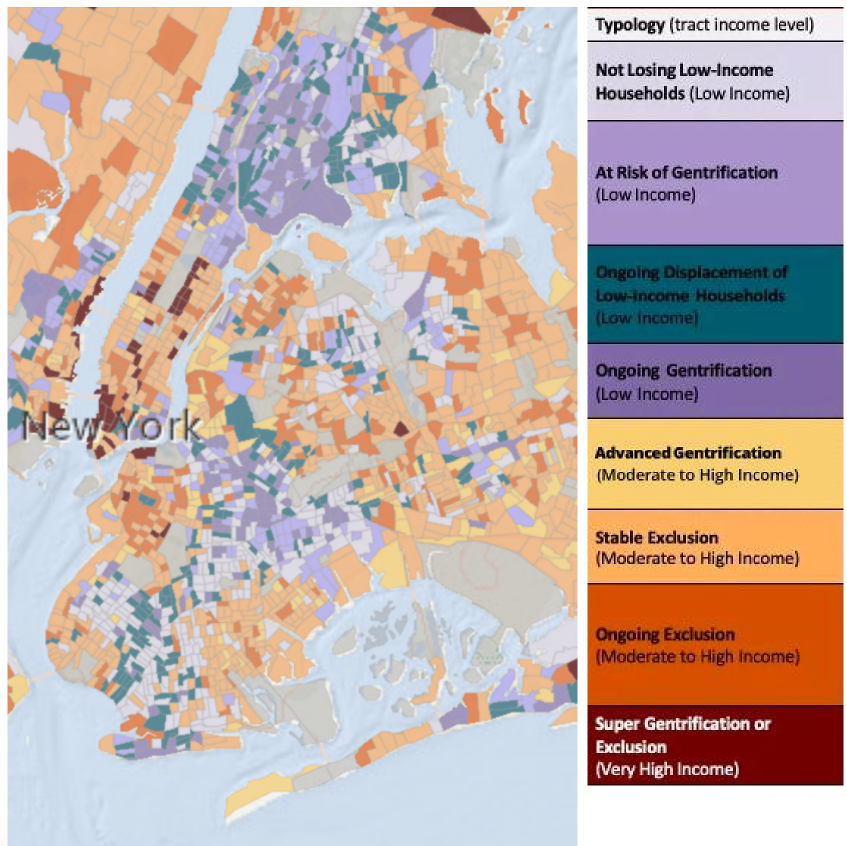


Figure 4.1: Group-Based Trajectory Model Results

Figure 4.2: New York Gentrification and Displacement Map

Discussion & Conclusion:

Our research project presents a methodology that could potentially predict when and where gentrification will impact lower-income residents in NYC by utilizing data analysis methods including time series analysis and group-based trajectory Modeling. More accurate and significant results could have been possible across more community areas, however, issues were encountered due to the lack of granular reporting from ACS and Census in both a spatial and temporal manner.

It is also important to note that the literature can differ in terms of how gentrification is defined and quantified. Although property prices comprise a large portion of discussions around the topic, it is important to use supplemental data sources like 311, income, and racial/ethnic distribution data to grasp a better understanding of the scope of the problem.. Gentrification remains a very complex issue that is not only impacted by various factors within each neighborhood area but also factors in surrounding neighborhoods as well.

However, with the conditions outlined above, the findings of this research project can provide more insight into the topic and how using 311 data can offer a different lens to the literature that currently exists. Our research can help inform urban planning decisions and housing policies that promote equitable development and social justice in NYC. The methodology and findings can be adapted to other cities facing similar gentrification challenges as well, making it a valuable contribution to the literature. Further research in this area can focus on the impact of gentrification on local businesses, the displacement of long-time residents, and the affordability of housing in gentrified areas.