GentrAIfication: Exploring the Association between Urban Greening and Gentrification

Tommaso Zambelli Franz, Felix Kesler, Ema Culibrk

October 22, 2025

Abstract

Green gentrification occurs when environmental improvements unintentionally lead to demographic and economic shifts, replacing lower-income residents with higher-income households. This report investigates the impact of urban greening on gentrification indicators in neighbourhoods across Amsterdam and develops an AI model to predict future gentrification risk. A sample of 439 neighborhoods between 2020 and 2024 was analysed, and a two-phase framework was employed: a quasi-experimental Difference-in-Differences (DiD) causal analysis and a machine learning-based predictive modeling approach. The results indicated that greening significantly affects neighborhood social composition, particularly attracting families with children, but its influence on economic indicators such as property values and income is minor or statistically insignificant in the short term. Only in areas with highly intensive greening a modest decline in low-income households was observed, suggesting that displacement risks mainly arise from large-scale projects. The feature importance report by a Random Forest Classifier model highlights that baseline economic vulnerabilities, such as average property values and median incomes, were far stronger predictors of gentrification risk than green space metrics. Forecasts identify 30 neighborhoods, particularly in Zuidoost, Nieuw-West, and Noord, as high risk for gentrification by 2028. Therefore, urban greening policies should balance environmental improvements with housing protections to prevent displacement. While greening makes neighborhoods more attractive and livable, it should be paired with measures like rent control and community involvement to ensure equity. Thus, high-quality green projects are best achieved through strong community participation and multifunctional green spaces.

1 Introduction

The term **green gentrification** refers to the process by which environmental improvements trigger socio-economic and demographic shifts, leading to the influx of higher-income residents and the displacement or marginalization of lower-income communities [1]. Urban greening initiatives, such as the creation of parks, green infrastructure, sustainable transportation, and ecological restoration projects, aim to improve environmental quality. Paradoxically, these green interventions lead to the displacement of vulnerable communities, an increase in property values, and demographic shifts [1].

Green gentrification was investigated across 28 cities in North America and Europe, finding a strong positive correlation between greening in the 1990s–2000s and gentrification occurring between 2000–2016 in the majority of the cities [2]. Additionally, Chen et al. show that proximity to urban parks and green spaces can lead to property value premiums of approximately 1.7 % to 9.3 %, depending on the specific model used and the property's distance to green space [3]. A similar correlation was observed in Amsterdam, as Bouwknegt and Rouwendal found that cutting down trees within 75 meters of a house led to a 1.19 % decrease in house prices, suggesting that the presence of green spaces contributes directly to higher property values [4]. Moreover, as property prices increase, low-income renters are more vulnerable to displacement, while homeowners are more likely to remain and benefit from capital gains and enhanced amenities [5].

Amsterdam's current greening policies focus on enhancing climate resilience and urban liveability, but may contribute to the displacement of lower-income residents. Policies such as *Agenda Groen* (2015–2018) allocated 20 million for 4 priorities by 2018: improve city parks, implement green roofs for biodiversity, more local greenery/play areas, and link green areas for better accessibility [6]. Long-term objectives to enhance public spaces, invest in natural assets, and reduce the impacts of climate change have also been implemented by *Structuurvisie 2040* [7]. In addition, while these policies are central to the city's sustainability agenda, they also risk contributing to socio-spatial inequality. Amsterdam's rent regulation relies on a points system (*Woningwaarderingsstelsel*, WWS), where scores <145 points implement strict housing rent caps, while scores >186 points allow for no contractual rent limits [8]. Environmental and structural improvements, such as green roofs and other amenities, can increase these scores, allowing landlords to charge higher rents. Hence, well-intentioned urban greening may indirectly amplify the gentrification process.

Furthermore, the effects of green gentrification don't always follow linear patterns, as multiple predictive factors may be present [9]. Research conducted by the U.S. Census Bureau utilized machine learning techniques to identify additional socioeconomic factors, including educational attainment, income levels, demographic composition, and housing tenure patterns. Additionally, environmental data such as air quality, climate change resilience features, green space coverage/proximity, and infrastructure development patterns all contribute to predicting gentrification risks. Lastly, building permits, rental prices, and property values were also investigated [9]. The paper found, using Random Forest Classification models achieved 72.5 % accuracy in forecasting gentrification compared to other models. Thus, machine learning techniques combined with real-time data integration can yield predictive insights in identifying areas at risk of green gentrification.

This being said, there are still gaps in research using AI models to both identify and predict green gentrification. Therefore, this report aims to investigate the relationship between urban greening and the displacement of lower-income residents, and develop models to predict the risk of green gentrification across Amsterdam's neighborhoods.

2 Methods

The primary objective of this study is to investigate the potential causal relationship between urban greening initiatives and socio-economic indicators of gentrification in Amsterdam. Building upon this causal inquiry, a secondary objective was to explore the feasibility of developing a predictive model to identify neighborhoods at future risk. To achieve these goals, a two-phase analytical framework was employed. The first phase utilizes a quasi-experimental Difference-in-Differences (DiD) model to estimate the causal impact of greening. The second phase involves the development and comparative evaluation of several machine learning classification models to construct a predictive "early-warning" system. This section details the comprehensive methodology, from data sourcing and integration to the specifics of each analytical model.

2.1 Data Sourcing and Integration

A longitudinal panel dataset was constructed at the neighborhood (*buurt*) level for Amsterdam, capturing snapshots of the urban environment in 2020 and 2024. This required the integration of multiple heterogeneous data sources. A significant challenge was the harmonization of datasets that used different geographic coding systems (e.g., 2015 vs. 2022 *buurt* codes). This was resolved by creating a master mapping file which allowed all data to be linked to a consistent 2022 neighborhood identifier (`buurtcode_2022`), resulting in a final panel of 439 consistent neighborhoods. The data sources are summarized in Table 1.

Table 1: Overview of Data Sources

Data Category	**Source**	**Variables Used**	**Purpose in Analysis**
Green Space Data	OpenStreetMap (via Overpass API)	Historical geometric data: parks, green spaces.	To calculate neighborhood-level green metrics
Socio-Economic Data	Statistics Netherlands (CBS StatLine): *Kerncijfers Wijken en Buurten*	Demographics, income, education.	To characterize baseline conditions, measure change over time, and more.
Facilities Data	Onderzoek en Statistiek (O&S), City of Amsterdam	Number of amenities within 1km; distance to schools, train stations.	To serve as control variables.
Geographic Boundaries	Onderzoek en Statistiek (O&S), City of Amsterdam	Codenames and polygon geometries.	To provide geographic boundaries for spatial calculations.

The green space data was acquired via historical queries to the Overpass API, a tool for extracting data from OpenStreetMap. To manage API limitations, the administrative area of Amsterdam was divided into a grid, with separate queries run for each grid cell for the specified years. Socio-economic data was primarily sourced from CBS StatLine. Due to lags in the official publication of income statistics, it was necessary to use 2023 income data as a proxy for the 2024 time point, an approximation deemed acceptable for the purposes of this study.

2.2 Operationalizing Gentrification: A Data-Driven Indicator

The extant literature defines gentrification as a process involving the transformation of a vulnerable, low-value neighborhood into a high-value one through rapid, above-average socio-economic upgrading insert citation: link to your perplexity report. Established academic models, such as those developed by Freeman (2005) and Ellen & O'Regan (2011), consistently identify three core components: 1) a baseline of socio-economic disadvantage, 2) rapid economic appreciation in housing values or income, and 3) demographic upgrading, often measured by an influx of higher-educated residents.

Drawing directly from this theoretical framework, we translated these concepts into a specific, stringent, and computable three-part rule to create a binary `gentrified` flag (1 or 0) for each neighborhood. This indicator serves as the dependent variable in our predictive models.

- **1. Baseline vulnerability (2020):** A neighborhood was classified as "eligible" for gentrification only if it was unambiguously vulnerable at the start of the observation period. This was defined as having both an average property value (`woz_value`) AND an average household income (`income_household`) below the city-wide median in 2020. This strict `AND` condition ensures that already affluent areas are excluded.
- **2. Rapid upgrading (2020-2024):** A neighborhood was flagged as "upgrading" if it experienced a rate of change that was exceptional relative to the rest of the city. This was defined as its growth rate placing it in the top 25% (75th percentile) of all neighborhoods for at least two of the following three metrics:
  - Percentage growth in property value (`woz_growth_pct`).
  - Percentage growth in household income (`income_growth_pct`).
  - Absolute growth in the number of highly educated residents (`education_growth_abs`).
- **3. Final `gentrified` flag:** A neighborhood was assigned a `gentrified` flag of 1 only if it met both the "baseline vulnerability" and "rapid upgrading" criteria. This multi-stage definition provides a robust and theoretically grounded indicator of gentrification.

2.3 Data Pre-processing and Feature Engineering

A multi-step pre-processing pipeline was developed to transform the raw source files into a clean, analysis-ready panel dataset. This involved geospatial processing, data harmonization, and a sophisticated approach to handling missing data.

First, the raw JSON data from the Overpass API was processed into a structured `GeoDataFrame`. This involved parsing OSM element geometries, projecting coordinates from the WGS84 (EPSG:4326) system to the Dutch RD New (EPSG:28992) system for accurate area calculations, and computing neighborhood-level green metrics such as `green_coverage_percent` and `distance_to_nearest_green` for both 2020 and 2024.

Second, all datasets were cleaned and harmonized. This included translating Dutch headers to a consistent English schema, standardizing geographic codes, and filtering all sources to the common set of 439 valid Amsterdam *buurten*.

Third, to address missing values in socioeconomic data, a context-aware K-Nearest Neighbors (KNN) imputation strategy was employed. [1] This method was chosen over simpler techniques like mean or median imputation because it provides more realistic estimates by using values from the five most statistically similar neighborhoods. To prevent data leakage and improve accuracy, imputation was performed in separate, logical batches. A non-imputed version of the dataset was also retained for comparative analysis, which confirmed that the imputed data yielded more stable and statistically significant results.

Finally, a number of features were engineered, including a composite `green_change` score for the DiD analysis and a suite of `_change` and `_growth_pct` variables used for defining the gentrification flag and profiling neighborhood transformations.

2.4 Analytical Strategy

2.4.1 Phase 1: Causal Analysis with Difference-in-Differences

The primary analytical goal was to estimate the causal impact of greening on gentrification indicators. For this, a Difference-in-Differences (DiD) model was employed. The DiD model is a quasi-experimental method that uses longitudinal data to estimate the causal effect of an intervention by comparing the change in an outcome over time between a treated group and an untreated control group. This approach was selected for our study to isolate the specific impact of the greening 'treatment' from broader, city-wide trends that affect all neighborhoods simultaneously. The "treatment" variable was defined by a composite `green_change` metric, capturing both the change in green coverage percentage and the change in distance to the nearest green space.

To ensure the robustness of the findings, a sensitivity analysis was conducted. The DiD regression was executed ten times, each time defining the "treated" group of neighborhoods with a different quantile cutoff for the `green_change` metric, ranging from the top 50% (median) to the top 5% (95th percentile). This allowed us to observe whether the estimated effect was consistent or only manifested at extreme levels of greening. The model was specified as a fixed-effects Ordinary Least Squares (OLS) regression [2] using the `statsmodels` library, with the key variable of interest being the interaction term `treated * post`. The model also included neighborhood fixed effects to control for time-invariant unobserved heterogeneity, and a set of time-varying control variables for amenities and proximity to services.

2.4.2 Phase 2: Predictive Modeling

Building upon the insights from the causal analysis, a secondary objective was to develop a machine learning model to function as an "early-warning system" for future gentrification risk. This required transforming the panel data into a wide format and framing the problem as a binary classification task: using the baseline characteristics of a neighborhood in 2020 ($\mathbf{X}$) to predict whether it would meet the criteria for the `gentrified` flag by 2024 ($\mathbf{Y}$).

The model selection process was systematic and comparative:

[1] Comparison pre-post imputed data included in Appendix B

[2] Included in Appendix A

1. **Initial Model Selection:** Based on its strong performance in a methodologically similar study by the U.S. Census Bureau [9], the Random Forest Classifier was selected as the initial model. An extensive hyperparameter search was conducted using `GridSearchCV`.

2. **Comparative "Bake-Off":** To ensure a robust model choice, a broader "scouting run" was then performed to compare the tuned Random Forest against four other diverse model architectures: Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and a Gradient Boosting Machine (LightGBM).

3. **Deep Dive and Champion Selection:** The top three contenders from the scouting run (KNN, SVM, and LightGBM) were subjected to a final, extensive hyperparameter tuning process, again using `GridSearchCV` with 5-fold cross-validation and optimizing for the F1-score to handle class imbalance. Based on its superior and most suitable performance characteristics for an early-warning system (notably, high recall), LightGBM was selected as the champion model.

The final step of the methodology involved training this champion LightGBM model on 100% of the historical data (2020 features predicting 2024 outcomes) and applying it to the current 2024 neighborhood data to generate a risk forecast for the period leading up to approximately 2028.

3 Results

This section presents the empirical findings from our two-phase analytical framework. We first detail the results of the causal inquiry, which used a Difference-in-Differences (DiD) model to estimate the potential impact of urban greening on various socio-economic and demographic indicators. Subsequently, we present the performance and diagnostic insights from the predictive modeling phase, where several machine learning algorithms were evaluated for their ability to forecast gentrification risk. All primary analyses were conducted on the fully imputed dataset, which proved to yield more stable and statistically significant results than the non-imputed version, a comparison of which can be found in the appendices.

3.1 Profile of Gentrified Neighborhoods

Our stringent, literature-informed definition identified **49 out of 439 consistent neighborhoods (11.2%)** [3] as having undergone a process consistent with gentrification between 2020 and 2024. A comparative analysis of these neighborhoods against their non-gentrified counterparts validates the logic of our flagging methodology, revealing a clear and plausible pattern of transformation.

Table 2 confirms that neighborhoods flagged as gentrified began the observation period from a position of significant socio-economic vulnerability. Their average property values and household incomes in 2020 were substantially lower than those of non-gentrified areas, validating the "baseline eligibility" criterion of our definition.

[3] Included in Appendix D

Table 2: Comparative Profile of Gentrified vs. Non-Gentrified Neighborhoods

Metric	Non-Gentrified	Gentrified
Section 1: Baseline Characteristics (2020)		
Average Property Value (€)	523.44	257.86
Average Household Income (€)	36.31	23.60
Share of Rental Housing (%)	63.47	81.94
Share of Low-Income Households (%)	47.78	65.31
Green Coverage (%)	9.26	11.25
Section 2: Transformation Metrics (Change from 2020-2024)		
Change in Property Value (€)	75.28	87.94
Change in Household Income (€)	7.13	9.31
Change in Population (nr. people)	46.27	245.20
Change in Low-Income Share (p.p.)	-0.49	-2.28
Change in High-Income Share (p.p.)	-0.24	0.98
Change in Green Coverage (p.p.)	0.73	-1.02

The transformation these neighborhoods underwent was markedly more intense. Gentrified areas experienced greater increases in property value and household income, a dramatically larger influx of population, and a rate of decline in the share of low-income households nearly five times that of non-gentrified areas. Geographically, these 49 neighborhoods were highly concentrated in districts undergoing significant urban renewal, primarily Zuidoost (31.8% of its neighborhoods gentrified), Nieuw-West (23.3%), and Noord (18.3%).

A notable and counter-intuitive finding from this profiling is the "green paradox": on average, gentrified neighborhoods saw a slight net decrease in green coverage (-1.02 percentage points), while non-gentrified areas saw a slight increase. This suggests that the relationship between greening and gentrification is not a simple linear one and may be confounded by other development pressures, necessitating the more nuanced causal modeling that follows.

3.2 The Causal Impact of Greening: A Difference-in-Differences Analysis

The Difference-in-Differences (DiD) sensitivity analysis, designed to isolate the causal impact of greening from general city-wide trends, revealed a nuanced and dual effect on neighborhood demographics. The statistical significance of these findings implies that the observed changes in the "treated" (greened) neighborhoods were above and beyond the average changes occurring across the entire city during the same period.

3.2.1 Statistically Significant Findings

Two primary causal effects were identified with statistical significance ($p < 0.05$).

1. **An Association Between Greening and an Influx of Families:** A consistent, positive causal effect of greening was observed on the population of children

aged 0-15 (`pop_age_0_15`). This effect was statistically significant across multiple moderate-to-high greening thresholds. At its most significant point (the 65th percentile of greening), the model estimates that treated neighborhoods saw an increase of approximately **14.9 children** relative to the control group (p=0.018). This suggests that greening initiatives may be successful in making neighborhoods more attractive to families.

2. **An Association Between Intense Greening and a Decline in Low-Income Households:** A statistically significant negative effect was found on the share of low-income households (`pct_bottom40`), but this was only observable at the most extreme levels of greening. At the 90th percentile cutoff, the model estimates that the most intensely greened neighborhoods experienced a relative **decrease of 2.4 percentage points** in their share of low-income households compared to other areas (p=0.036). This finding points towards potential displacement pressures specifically coinciding with the most intensive greening projects.

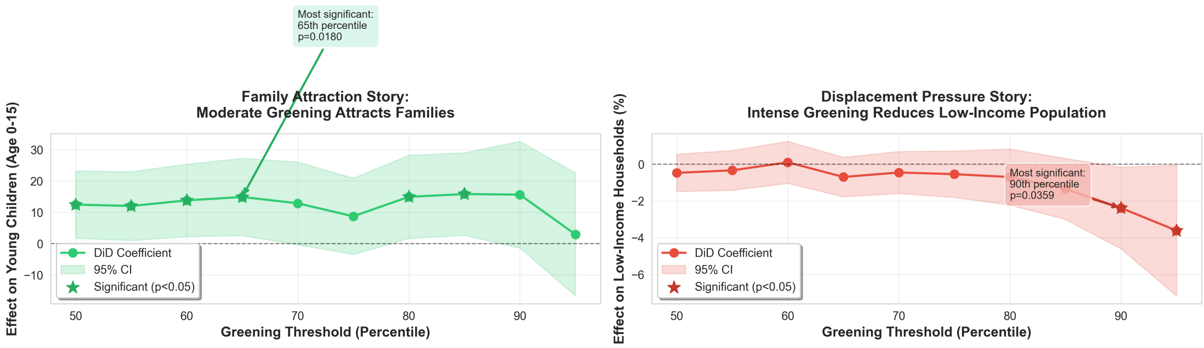

Figure 1: DiD Coefficient Plots for Key Outcomes. The left panel shows the positive effect of greening on the young children population, which becomes significant at several treatment thresholds. The right panel shows the negative effect on low-income households, which is only significant at the highest intensities of greening.

Figure 1 visually represents this dual effect, illustrating the different thresholds at which these phenomena become statistically significant.[4]

3.2.2 Null and Borderline Results

It is equally important to report that for most core economic indicators of gentrification, no statistically significant causal effect of greening was detected within our 2020-2024 timeframe. There was no significant effect on average property value (`woz_value`), household income (`income_household`), or the share of high-income households (`pct_top20`). A number of "borderline significant" results ($0.05 < p < 0.10$) were observed at the 90th percentile cutoff, including a decrease in renters and an increase in high-income residents, which may suggest the beginning of a trend but do not meet the standard threshold for statistical significance.

[4]Full regression tables are available in Appendix A.

3.3 Predictive Modeling of Gentrification Risk

The second phase of our analysis sought to develop a model to predict which neighborhoods were at risk of gentrification, using their 2020 baseline characteristics to predict the 2024 outcome.

3.3.1 Model Comparison and Performance

A comparative analysis of five machine learning models was conducted to identify the best architecture for this classification task, which is complicated by significant class imbalance (11.2% positive class).

Contrary to the findings of the reference study by Yoo (2023), the Random Forest Classifier, even after extensive hyperparameter tuning, did not emerge as the top performer. Its performance was modest (F1-score of 0.45), indicating that the features driving gentrification in Amsterdam may interact differently than in the study's U.S. context.

After a rigorous "bake-off" and "deep dive" tuning process, the LightGBM model was selected as the champion. [5] Its primary strength was a high **recall of 80.0%** on the held-out test set, correctly identifying 8 out of 10 truly gentrified neighborhoods. For an early-warning system, this high recall (minimizing missed cases) was deemed more valuable than the higher precision of other models like K-Nearest Neighbors (which only had a recall of 50.0%). This performance, however, comes with the trade-off of a low precision of 42.1%, indicating a significant number of false positives. The overall performance of all models was limited, with no model achieving an F1-score above 0.55 on the test set, highlighting the inherent difficulty of predicting this complex social phenomenon with the available data.

3.3.2 Key Predictors of Gentrification Risk

An analysis of the feature importances from the Random Forest model provided crucial insights into the primary drivers of gentrification risk.

The results overwhelmingly show that the most powerful predictors are the baseline economic conditions of a neighborhood in 2020. The top two features — `woz_value_2020` and `income_recipient_2020`, which represent the average property value and the median income per income recipient in the buurt — collectively account for over 60% of the model's predictive power. This confirms that pre-existing economic vulnerability is the dominant precondition for gentrification.

While `total_green_area_m2_2020` and `dist_nearest_green_2020`, the two baseline green metrics, appear in the top 15 predictors, their role is secondary. Their combined importance is less than 4%, an order of magnitude smaller than that of the primary economic indicators. This suggests that while the presence of green space is a contributing factor, it is far from the most decisive one in predicting which vulnerable neighborhoods will gentrify.

[5] Predictive Model Diagnostics included in Appendix C

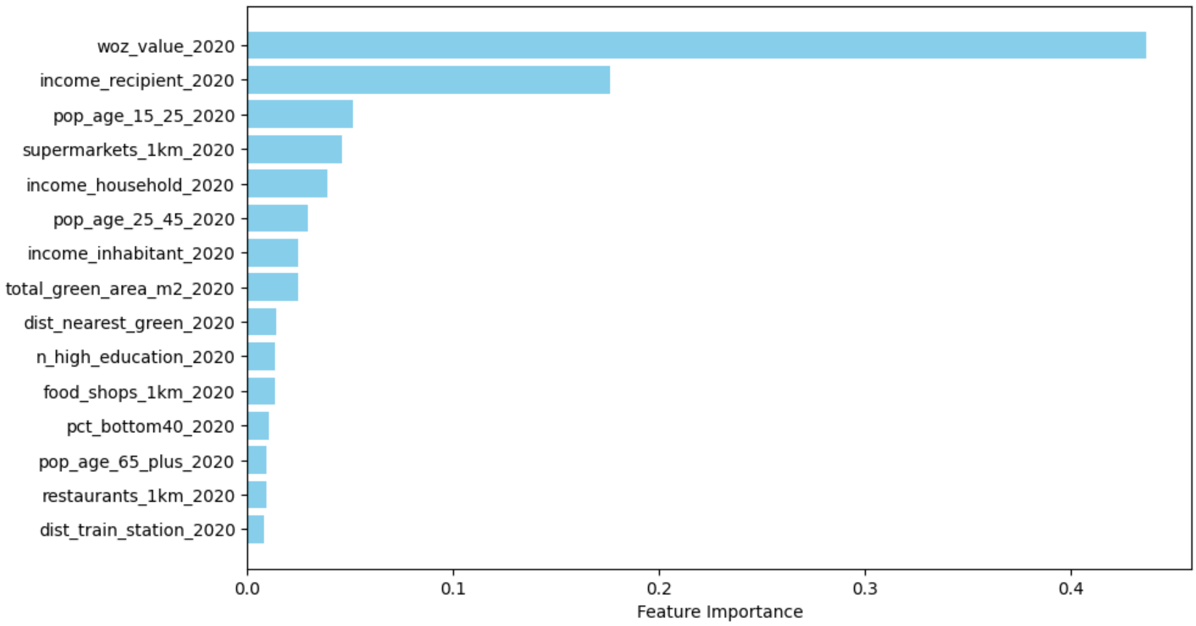

Figure 2: Top 15 Most Important Predictors of Gentrification Risk from the Random Forest Model.

3.4 Future Risk Forecast for Amsterdam (2028)

The final step involved training the champion LightGBM model on the full 2020-2024 historical dataset and applying it to the current 2024 data to forecast future risk. The model predicted that **30 neighborhoods** are at high risk of gentrifying by approximately 2028, with a heavy concentration in **Zuidoost, Nieuw-West, and Noord**. Crucially, this forecast identifies several "newly at-risk" neighborhoods that did not meet our gentrification criteria by 2024 but now exhibit the key risk characteristics learned by the model.

4 Discussion

This forward-looking risk profile, which identified 30 high-risk neighborhoods for the period up to 2028 — including newly flagged areas like *Gein 4* and *De Kleine Wereld* — represents the most actionable output of this research, providing a targeted list for potential policy interventions designed to mitigate the negative social impacts of urban renewal.

The preceding analysis of the relationship between urban greening and gentrification indicators reveals a complex and nuanced pattern: the principal short-term effect of greening was observed on the social structure of neighborhoods, rather than consistently predicting immediate economic shifts. While moderate levels of greening coincided with an increase in families with children in a neighborhood, the only other statistically significant effect ($p = 0.036$) was a modest decline in low-income households observed only at highly intensive greening projects. Other classical gentrification indicators, such as property values or average income, showed no significant association with greening over the examined timeframe. Overall, these results suggest that the primary impact of greening is on the social composition, with only minor or delayed effects on economic indicators.

The insight that greener neighborhoods tend to attract more families with children is consistent with earlier studies conducted in the US showing that parks and other publicly accessible recreation spaces increase perceived quality of life and attractiveness for young families [10, 11]. According to previous literature, this demographic shift can be seen as a subtle early indicator in gentrifying neighborhoods, preceding the more dramatic effects of gentrification, such as increased rents and house prices [12]. Interesting insights can also be drawn from the decline in low-income households at intense levels of greening, as it points to a threshold effect. This indicates that displacement is most likely to occur as a result of large-scale greening projects. This finding is especially pertinent to the context of our study, as a large-scale green redevelopment project is currently underway in the Gaasplas in Amsterdam Zuid-Oost [13]. Four out of the ten neighborhoods most at risk of gentrification until 2028, according to our model, are situated geographically adjacent to the park, making it an especially interesting perspective for policy approaches.

Further policy implications arising from the results are that, while greening can improve neighborhood quality of life, it is paramount to pair such projects with housing measures to counter long-term displacement pressures on the most socio-economically vulnerable residents. This is essential, as ensuring equitable greening with appropriate safeguards, such as rent and legal protection, community participation, or even non-profit oriented housing organization, allows it to attain its objective of improving the quality of life across all socio-economic strata. This prevents a well-intentioned urban planning intervention from devolving into a further, admittedly indirect, pressure point on those citizens it is meant to serve. In addition, one must also focus on the quality of the greening interventions themselves. This can include multi-faceted community-oriented greening through involving residents in planning, including free or low-cost leisure options or, for example, increasing green-space functionality through accessible food parks.

References

1. Dell'Ovo M, Datola G, Maiullari D, Oppio A, and Schretzenmayr M. Green Gentrification: A Literature Review of Trends, Challenges, and Research Opportunities. *Computational Science and Its Applications – ICCSA 2025 Workshops*. Ed. by Gervasi O, Murgante B, Garau C, Karaca Y, Lago MF, Scorza F, and Braga AC. Vol. 15893. Lecture Notes in Computer Science. Cham: Springer, 2026 :222–33. DOI: 10.1007/978-3-031-97645-2_15
2. Anguelovski I, Connolly JJT, Cole H, Garcia-Lamarca M, Triguero-Mas M, Baró F, Martin N, Conesa D, Shokry G, Pérez del Pulgar C, Ramos LA, Matheney A, Gallez E, Oscilowicz E, Máñez JL, Sarzo B, Beltrán MA, and Minaya JM. Green gentrification in European and North American cities. Nat Commun 2022; 13:3734. Available from: https://www.nature.com/articles/s41467-022-31572-1
3. Chen K, Lin H, You S, and Han Y. Review of the impact of urban parks and green spaces on residence prices in the environmental health context. Front Public Health 2022; 10:933589. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9490231/

4. Bouwknegt L and Rouwendal J. The effect of urban trees on house prices: evidence from cut-down trees in Amsterdam. Discussion Paper Series 23-059/VIII. Tinbergen Institute, 2023. Available from: https://research.vu.nl/ws/portalfiles/portal/259221423/23059.pdf
5. Wang W. Environmental Gentrification. PhD thesis. Duke University, 2020. Available from: https://hdl.handle.net/10161/21022
6. Wijten G, Van der Veur W, Koerselman P, Bosch N, Timmermans G, Kaljee H, Eilander P, and Vries H de. Agenda Groen 2015–2018. Ed. by Wijten G. Amsterdam, 2015. Available from: https://openresearch.amsterdam/nl/page/53874/agenda-groen-2015---2018
7. Netherlands Commission for Environmental Assessment (NCEA). Structure vision Amsterdam 2040 and SEA. Strategic Environmental Assessment for the Structure Vision Amsterdam 2040. Report. 2023. Available from: https://api.commissiemer.nl/docs/mer/diversen/os_structurevisionamsterdam.pdf
8. Amsterdam Life Homes. A complete guide to understanding Amsterdam's rent control laws. 2024. Available from: https://amsterdamlifehomes.com/blog/a-complete-guide-to-understanding-amsterdam-s-rent-control-laws [Accessed on: 2024 Oct 22]
9. Yoo J. Identifying gentrification using machine learning. SEHSD Working Paper 2023-15. U.S. Census Bureau, 2023. Available from: https://www.census.gov/content/dam/Census/library/working-papers/2023/demo/sehsd-wp2023-15.pdf
10. Wolch JR, Byrne J, and Newell JP. Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. Landsc Urban Plan 2014; 125:234–44
11. Rigolon A. Parks and young people: An environmental justice study of park proximity, acreage, and quality in Denver, Colorado. Landsc Urban Plan 2017; 165:73–83
12. Freeman L. Displacement or Succession?: Residential Mobility in Gentrifying Neighborhoods. Urban Aff Rev 2005; 40:463–91
13. Gemeente Amsterdam. Gaasplerplas: vernieuwen park. Bouwprojecten en verkeersprojecten. 2025. Available from: https://www.amsterdam.nl/projecten/gaasperplas/ [Accessed on: 2025 Oct 21]

A DiD Regression Tables (Full Results)

Table 3: Difference-in-Differences (DiD) Regression Results for Key Gentrification Indicators (2020–2024)

Independent Variables	Dependent Variables: Δ Population Age 0-15 (Treated at 65% Δ Green)	Dependent Variables: Δ Share Low-Income (Treated at 90% Δ Green)
DiD Interaction Term (Greening Impact)		
Treated × Post (Standard Error)	14.893 **	−2.380 *
Time Dummies and Controls		
Treated (Baseline Diff.)	243.289 ***	−9.751 **
Post (Time Trend)	−14.494 ***	−1.033 *
supermarkets_1km	4.712 ***	−0.074
food_shops_1km	0.772 **	0.003
restaurants_1km	−0.452 **	0.001
dist_primary_school	−14.975	3.192 *
urbanicity	2.136	−0.016
Summary Statistics		
N (Neighborhood-Years)	878	878
R^2 (Overall)	0.989	0.972
R^2 (Adjusted)	0.977	0.944
Neighborhood Fixed Effects	Yes	Yes
Covariance Type	HC1 Robust	HC1 Robust

*p < 0.05; **p < 0.01; ***p < 0.001. All models include full set of amenity controls and fixed effects.

B DiD Sensitivity: pre-post KNN Imputation

Table 4: DiD Sensitivity: Comparison of Imputed vs. Non-Imputed Data

Outcome Variable	**Imputed Data (KNN)**		**Non-Imputed Data**	
	Lowest p-value	**Sig. Cutoffs ($p < 0.05$)**	**Lowest p-value**	**Sig. Cutoffs ($p < 0.05$)**
Primary Effects				
Δ Pop Age 0-15	**0.018** (at 65%)	**6/10**	0.016 (at 65%)	7/10
Δ Share Low-Income	**0.036** (at 90%)	**2/10**	0.140 (at 65%)	0/10
Economic Indicators				
Δ Property Value (`woz`)	0.109 (at 90%)	0/10	**0.040** (at 90%)	**1/10**
Δ Household Income	0.062 (at 70%)	0/10	0.027 (at 70%)	**2/10**
Housing/Tenure Indicators				
Δ Share Top 20% Income	0.060 (at 90%)	0/10	0.037 (at 65%)	**3/10**
Δ Share Rental Housing	0.063 (at 85%)	0/10	0.231 (at 85%)	0/10

Note: The imputed data shows significant displacement (Δ Share Low-Income) where the non-imputed data did not. The non-imputed data shows isolated significance for some economic indicators (e.g., Δ Property Value), which is likely spurious due to sample size fluctuations caused by missing data.

C Predictive Model Diagnostics

Table 5: Comparative Performance of Predictive Models (Ranked by F1-Score on Test Data)

Model	**Best F1-Score**	**Best Hyperparameters**
LightGBM (Champion)	**0.538**	{`learning_rate`: 0.01, `max_depth`: 10, `n_estimators`: 200, `num_leaves`: 20}
Support Vector Machine	0.520	{`classifier__C`: 1, `classifier__gamma`: 'scale', `classifier__kernel`: 'poly'}
K-Nearest Neighbors	0.421	{`classifier__n_neighbors`: 3, `classifier__p`: 2, `classifier__weights`: 'uniform'}

Note: The LightGBM model was selected for its balance of precision and recall (specifically high recall, crucial for an early warning system), evidenced by the highest F1-score achieved after rigorous grid search cross-validation.

D Gentrified Neighborhood Profile and List

The following tables provide the detailed characteristics and full list of the 49 neighborhoods (11.2% of the total) that met the strict three-part gentrification criteria between 2020 and 2024 (Section 2.2).

Table 6: Geographic Distribution of Gentrified Neighborhoods by District

District (*Stadsdeel*)	**Non-Gentrified**	**Gentrified**	**Total**	**% Gentrified**
Zuidoost	30	14	44	31.8
Nieuw-West	46	14	60	23.3
Noord	49	11	60	18.3
West	54	6	60	10.0
Oost	63	3	66	4.5
Zuid	75	1	76	1.3
Centrum	66	0	66	0.0
Weesp	1	0	1	0.0
Westpoort	6	0	6	0.0

Table 7: Full List of 49 Gentrified Neighborhoods (2020–2024)

Buurt Name	District	WOZ 2020 (€k)	Income 2020 (€k)	WOZ Δ	Income Δ	Pop. Δ
Sloterdijk Stationskwartier	Nieuw-West	135.0	20.0	155.00	6.46	460
Spaarndammerbuurt-Midden	West	375.0	28.6	102.00	6.40	160
Borgerbuurt	West	349.0	26.2	93.00	9.50	-85
Vivaldi	Zuid	179.0	20.2	136.00	37.70	870
Gibraltarbuurt	West	327.0	27.0	49.00	8.30	-85
Robert Scottbuurt-West	West	250.0	24.7	81.00	8.10	90
Laan van Spartaan	West	334.0	26.3	91.00	6.50	530
Kolenkitbuurt-Noord	West	302.0	25.0	131.00	7.80	900
Havenkwartier IJburg	Oost	394.0	30.6	56.00	16.00	315
De Eenhoorn	Oost	266.0	22.8	88.00	13.50	1220
Betondorp	Oost	345.0	24.9	92.00	9.55	-220
Bloemenbuurt-Zuid	Noord	322.0	23.1	98.00	6.40	-85
Vogelbuurt-Zuid	Noord	330.0	24.5	88.00	6.30	-130
Tuindorp Nieuwendam-Oost	Noord	322.0	25.6	118.00	8.22	-85
Blauwe Zand	Noord	304.0	25.6	111.00	9.94	20
Terrasdorp	Noord	289.0	25.2	85.00	7.60	-70
Circus/Kermisbuurt	Noord	304.0	30.3	81.00	4.90	-10
Markengouw-Zuid	Noord	303.0	25.5	85.00	6.70	180
Loenermark	Noord	262.0	25.8	82.00	4.70	595
Banne-Noordoost	Noord	301.0	25.7	80.00	5.90	-80
NDSM terrein	Noord	167.0	23.2	160.00	18.80	1990
Elzenhagen-Zuid	Noord	125.0	19.9	86.00	12.90	805
Noordoever Sloterplas	Nieuw-West	310.0	29.1	112.00	1.40	55
Wildeman	Nieuw-West	222.0	21.8	55.00	9.30	-80
Osdorp-Zuidoost	Nieuw-West	294.0	25.6	92.00	5.50	155
Reimerswaal	Nieuw-West	215.0	22.8	113.00	8.60	-70
Dijkgraafpleinbuurt	Nieuw-West	252.0	25.4	71.00	4.80	80
Middelveldsche Akerpolder	Nieuw-West	355.0	28.9	102.00	11.00	75
Rembrandtpark-Noord	Nieuw-West	352.0	24.6	158.00	8.70	75
Johan Jongkindbuurt	Nieuw-West	262.0	20.1	22.00	13.20	300
Riekerhaven	Nieuw-West	118.0	17.1	182.91	16.90	-160
Schipluidenbuurt	Nieuw-West	197.0	22.5	91.00	11.50	450
Belgiëplein e.o.	Nieuw-West	334.0	30.3	71.00	11.50	70
Louis Chrispijnbuurt	Nieuw-West	306.0	24.3	84.00	6.30	50
Medisch Centrum Slotervaart	Nieuw-West	282.0	13.6	67.00	17.00	50
Amstel III deel A/B-Zuid	Zuidoost	101.0	14.6	80.00	7.77	2230
Venserpolder-Oost	Zuidoost	229.0	23.8	70.00	4.80	-20
D-buurt	Zuidoost	48.0	13.0	8.00	14.20	400
Amsterdamse Poort	Zuidoost	207.0	20.3	43.00	6.90	920
Hoptille	Zuidoost	167.0	17.6	55.00	8.10	55
E-buurt	Zuidoost	372.0	28.5	136.00	2.00	170
Bijlmermuseum-Noord	Zuidoost	222.0	20.5	59.00	6.30	-80
Kortvoort	Zuidoost	194.0	20.9	72.00	9.60	255
Kelbergen	Zuidoost	251.0	24.1	91.00	6.40	10
G-buurt-Oost	Zuidoost	262.0	30.1	101.00	3.20	5
G-buurt-Noord	Zuidoost	151.0	19.2	70.00	11.30	-140
Holendrecht-West	Zuidoost	215.0	20.2	57.00	9.30	-45
Gein 1	Zuidoost	214.0	23.4	34.00	9.40	-40
Gein 2	Zuidoost	218.0	23.6	64.00	9.20	-40

E Data and Code Repository

For full reproducibility, the data construction, processing notebooks, and final analysis scripts are publicly available at the following repository:

`https://github.com/tommasozf/amsterdam-green-gentrification`