ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Data Science & Marketing Analytics

A Comparative Analysis of the Explanatory Power of Safety and Social Indices on Property Values in Rotterdam

Name student: Niels Stiebing

Student ID number: 544177

Supervisor: Dennis Fok

Second assessor: Vardan Avagyan

Date Final Version: 05/07/2024

Abstract

This research examines the explanatory value of social and safety indices, measured both objectively and subjectively, on property values in Rotterdam. The data in this study consist of neighborhood-level averages of 71 neighborhoods, measured every two years between 2014 and 2024. Utilizing the hedonic pricing model (HPM) as analytical framework, various linear regression models, fixed effects regression (FE) models, and geographically weighted regression (GWR) models are constructed to address the research question. The explanatory value of social and safety indices on property values is assessed by measuring the adjusted R^2. The results show that both social and safety indices possess significant explanatory power for the value of houses in both their objective and subjective measures. The findings also indicate spatial differences in the relationships between the various indices and property values. Globally, safety factors are more effective at explaining property values than social factors, while on a local level, social factors provide better explanations compared to safety factors. Overall, objective measures perform better in explaining housing values than subjective measures.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dennis Fok, for his support, guidance, and the insightful discussions throughout the course of my research. His feedback and suggestions greatly contributed to the quality of this thesis.

I would also like to thank my second assessor, Vardan Avagyan, for his time and effort.

Next, I would like to acknowledge the municipality of Rotterdam for providing me with the data for this research.

Finally, a special thanks to my parents, sister, girlfriend, and friends for their immense support throughout this thesis. Although they did not understand everything I researched, their encouragement and willingness to listen to my ideas and challenges have been amazing and have greatly helped me complete this thesis.

Contents

1. Introduction

1.1 Research Problem

A shortage of 390,000 houses. Sharply increased housing prices. Having to wait years for social housing or renting expensively in the private sector. Pretty much all political parties in the Netherlands agree: there are significant housing issues that need to be resolved. Voters also considered it an essential topic for the 2023 elections, where people aged 18 to 34 even recognized the housing market as their most important electoral theme (Mouissie & Kraniotis, 2023). Rabobank, the second biggest bank in the Netherlands, expects an increase of 6.2% in housing prices in 2024 and a further growth of 6.3% in 2025. On the demand side, this rise in market prices is caused by an increase in salaries and an expansion in loan capacity for potential buyers. On the supply side, it is caused by the limited, decreasing offer of houses (Groot & Vrieselaar, 2024).

Besides these factors, there are many more determinants of property values, including housing characteristics, like surface, number of rooms, or the presence of specific attributes (Zietz et al., 2007). Some demand-related aspects are demographic factors and trends, including population growth and urbanization, labor market conditions, and taxation. On the supply side, various elements influence housing prices as well (Nistor & Reianu, 2018).

Another example of a determinant of property values is neighborhood characteristics (e.g. Can, 1990; Sun et al., 2019). Much literature has shown the effect of various safety measurements on housing values (e.g. Gibbons, 2004; Kortas et al., 2022), while few papers discovered a correlation between social indicators and property values (e.g. Fu et al., 2016).

This paper aims to combine these findings and explore the association of neighborhood-level safety and social indices with housing values in Rotterdam. The research question to be answered is *"What is the explanatory value of safety and social indices, measured both objectively and subjectively, on property values in Rotterdam?"*

In this paper, I will use data from the 'Wijkprofiel,' which translates to 'District Profile.' Every two years, the municipality of Rotterdam publishes the 'Wijkprofiel,' which shows the status of its 14 districts and 71 neighborhoods in terms of social and physical conditions, as well as safety. Wijkprofiel offers both a subjective and an objective index to measure social conditions and safety. The subjective index is obtained through surveys, held every two years by the municipality. The objective index is determined by numbers and statistics, obtained from the municipal administration (Wijkprofiel Rotterdam, n.d.).

To be able to answer the main research question and gain a comprehensive understanding of the descriptive relationship between social and safety indices and property values in Rotterdam, I explore several subsidiary questions:

1) What is better in explaining the value of houses in Rotterdam, social or safety indices?
2) What is better in explaining the values of houses in Rotterdam, objective or subjective indices?
3) To what extent do subjective indices and objective indices complement each other in explaining property values in Rotterdam?
4) How do factors inside the social and safety indices relate to the value of houses in Rotterdam?
5) How do the associations between social and safety indices and property values vary across different neighborhoods in Rotterdam?

1.2 Practical Relevance

This research paper holds substantial relevance from a business perspective for several reasons, particularly in the real estate and urban development sectors:

Real Estate Investors: The real estate market often rewards those with superior information (Jaffe, 1980). Also, when homebuyers have varied levels of information, less informed buyers pay higher prices (Tu et al., 2016). If this study finds significant associations between social and safety indices and housing prices, investors could leverage these insights to gain a competitive advantage, enabling them to exploit market inefficiencies, potentially leading to higher returns on investment.

Homeowners: The residential real estate market is characterized by cyclical patterns and many uncertainties, making it hard to determine the optimal time to sell a property (Li et al., 2022). This study could enhance homeowners' understanding of their property's value over time, potentially improving the timing of sales. For instance, if a significant link between safety indices and property values is found, homeowners might postpone selling during high crime rates. They could potentially secure a higher sale price for their property by timing the sale for a more favorable safety climate.

Urban Planners and Policy Makers: Data and technology are increasingly pivotal in urban policy and design. However, genuinely listening to community needs and desires is also highly important (Economist Impact, 2024). This research incorporates data and technology on social and safety indices as well as subjective indices gathered through community surveys, which reflect the voice of the community. Together, this makes for a comprehensive approach to understanding the factors describing housing values, which are usually a reflection of the general state of a neighborhood, in Rotterdam. Although causal relationships cannot be inferred, recognizing these correlations may guide policy considerations and urban development plans aimed at enhancing neighborhood attractiveness.

1.3 Scientific Relevance

This research intersects urban planning, sociology, economics, and public policy, providing valuable interdisciplinary insights that can enrich each of these fields. Much research has been conducted to analyze property values and their determinants, including papers on the relationship of various social or safety elements with housing prices and analysis of associations between objective and subjective indices and property values. However, these fields of research have not been combined yet.

Safety elements literature: Many papers have analyzed the relationship between safety measurements, both objective and subjective, and property values. For example, Kortas et al. (2022) explored the relationship of different crime types with the prices of houses in Heerlen, another city in the Netherlands. They found that almost all the crime types they researched were negatively correlated with house prices. Pope (2008) analyzed the effect of fear of crime on housing prices. He concludes that fear of crime significantly decreases the price of a house. This research will fill the gap of comparing safety factors and social determinants in their relationship with the values of properties.

Social elements literature: Less research has been conducted on the relationship of property values with social elements than on the relationship with safety elements. Blair and Larsen (2010) studied whether neighborhoods where residents enjoy fulfilling social interactions with their neighbors tend to have higher housing prices than areas where people are less content with their neighborly relationships, and their research confirms this hypothesis. Whereas Uphoff et al. (2013) found that both objective and subjective social capital is positively associated with socioeconomic status. Property values can be seen as a socioeconomic characteristic but have not been used in that study. This research will fill the gap of combining objective and subjective social measurements in their relationship with the values of properties.

Objective and subjective literature: Not much research has been conducted on the comparison of subjective and objective indicators in explaining property values. Qiu et al. (2022) examined whether objective or subjective measures of street environment were more effective in describing property prices and whether these measurements are complementary or conflicting. They found that both measurements increase the explanatory value of the price of a house, with objective measurements explaining better than subjective measurements. They also conclude that both measurements are complementary. However, this point of research has not been combined with either social or safety measures. Qiu et al. included a safety variable, but this was purely a safety of street environment measurement, not a straightforward neighborhood-wide safety measure. This paper will fill this gap by analyzing whether objective or subjective measurements of social and safety indices are more effective in explaining property values and to what extent these measurements complement each other.

2. Literature Review

The literature review is divided into four different sections. Section 2.1 explains the hedonic pricing model (HPM), which is widely used to calculate the value of houses. Section 2.2 describes spatial differences in determining housing values, section 2.3 explores literature on safety and social factors in combination with property values, and Section 2.4 investigates the comparison of objective and subjective measurements in combination with property values and other fields. The concluding section, 2.5, denotes the expectations of this research.

2.1 Hedonic Pricing Model and Neighborhood-level Determinants

The HPM has evolved to become one of the most popular methods for assessing the value of various features. This model is one of the clearest ways to demonstrate how to determine consumers' willingness to pay (WTP) for various attributes of a property, including environmental features (Bishop et al., 2020).

Hedonic price functions provide an empirical overview of how prices correlate with the features of goods in markets with differentiated products (Pakes, 2003). It was initially introduced by Court (1939) and later revisited by Griliches (1961). Hedonic price functions address the issue of evaluating new goods. Both authors recognized that newer models of products typically possess improved features. Thus, price differences between new and older models should not be solely attributed to inflation. They proposed creating a model that relates prices to product characteristics over time. Rosen (1974) contributed to the hedonic framework by introducing it as an equilibrium model, providing insights into how prices of differentiated products can reflect consumer demand for specific product attributes.

The popularity of the hedonic framework in assessing property values results from its intuitive idea that is both economically logical and easily applicable in empirical research. The model reflects that buyers select properties based on specific housing features, such as the surface and the number of bedrooms and bathrooms, as well as location-based amenities like air quality, proximity to parks, and educational opportunities. Without market interference, these variations in amenities are naturally reflected in property prices. As buyers navigate these options, analyzing their purchasing choices helps reveal their WTP for improvements in these amenities (Bishop et al., 2020).

Literature has shown that neighborhood-level factors matter in determining the value of a property. Can (1990) utilized the HPM while adding spatial variances to this framework. He found the model to be more effective in explaining house price variations after the addition. This indicates that housing attributes influence property prices differently depending on the location. Sun et al. (2019) computed neighborhood-level blight indices, which refer to the condition where properties are abandoned, neglected, or poorly maintained, and included these in the HPM. Subsequently, they employed factor analysis and Shapley-Owen values to determine each variable's contribution to explaining variance in house values. They found that these neighborhood-level indices have a significant, negative effect on property values in a neighborhood.

2.2 Spatial Differences in Determining Property Values

In real estate research, it is widely recognized that hedonic prices can differ across various spatial divisions, including metropolitan areas, regions, and counties (Helbich et al., 2013). Goodman and Thibodeau (2003) showed that spatial disaggregation significantly enhances the accuracy of hedonic predictions, while Brady and Irwin (2011) encourage the usage of models that incorporate spatial heterogeneity to more accurately capture the variability in willingness to pay for environmental amenities.

Some specific examples of varying results across neighborhoods are Li et al. (2019) and Anderson and West (2006). Li et al. employed the random intercept multi-level regression (MLR) to determine whether there are spatial differences. They found that people who live in the inner-city area are prepared to pay more for improved service amenities. Anderson and West used local fixed effects to control for potential omitted spatial variables and included it in their econometric model inspired by the HPM. They showed that the value placed on proximity to open spaces is greater in densely populated neighborhoods, areas close to the central business district, high-income areas, regions with higher crime rates, and communities with many children. More related to this research is the paper of Tita et al. (2006). They examined whether variations in local and neighboring crime rates significantly influence housing prices while holding house and location characteristics constant. Their findings suggest that the average effects of crime rates on house prices can be inaccurate. They observed that crime impacts housing values differently across poor, middle-class, and wealthy neighborhoods.

2.3 Safety and Social Determinants of Property Values

Previous literature has shown that both actual safety and perceived safety play a role in the value of a house. For example, Buonanno et al. (2013) analyzed crime perception data from Barcelona, comparing it to the valuations of different districts. They constructed an HPM using panel methods with both district and year fixed effects. Their study revealed that an increase in perceived crime consistently lowers district valuations. Moreover, they discovered a significant negative correlation between crime perception and hedonic housing prices. Pope (2008) also found a significant relationship between a subjective safety measurement and property values. He created several HPMs, in which he added dummy variables for a sex offender moving into a neighborhood, as well as dummy variables to control for spatial and time differences. His study showed evidence of a significant drop in property values when a sex offender moves into a neighborhood. On the contrary, when this sex offender moves out of the neighborhood, housing prices immediately return to their original level.

Studies of the relationship of conducted crimes in a neighborhood with the property values in this neighborhood also show similar results, although they show varying results for different types of crimes. According to Gibbons (2004), crimes classified under the criminal damage category significantly reduce property prices. In contrast, burglaries do not appear to impact property prices. A possible explanation is that vandalism, graffiti, and other types of criminal damage can heighten community fear and are often seen as signs of overall community instability and neighborhood decline. Their research contained simple linear regression models based on the HPM and other regressions, which included spatial effects and instrumental variables. Instrumental variables were added since they assumed a problem in the context of the ordinary least squares (OLS) estimation, where there exists a correlation between the observed explanatory variables and the unobserved components of the price of a property, a problem known as endogeneity. This would lead to biased and inconsistent estimates.

Kortas et al. (2022) obtained a similar conclusion. This research considered four broad categories of crime data: total crime, total property crime, destruction and crimes against public order, and violent and sexual crimes. Home burglaries, typically grouped under property crimes, were analyzed separately to focus specifically on this variable. As the goal of their research was to assess spatial differences in relation to house prices, they decided not to use a linear regression, while also not using a geographically weighted regression (GWR), since they are not interested in proving the direction of the relationships with advanced regression methods. To obtain correlations, they combined the Getis-Ord G_i^* statistic and Global Moran's I statistic to compare regional averages with global averages. The study reveals that except for home burglaries, all types of crime generally negatively impact housing prices, showing consistent spatial patterns in relation to property values. Home burglaries do not exhibit the same trend, as the spatial relations between home burglaries and property values are complex, ranging from significantly negative to significantly positive.

Ihlanfeldt and Mayock (2010) constructed a panel method with time fixed effects. Change in price over the time span of a year for a property was used as the dependent variable, while independent variables included the time fixed effects, current change of various types of crimes, and four more lag variables of these types of crime. They concluded that only robbery and aggravated assault significantly affect neighborhood housing values among the seven crime categories examined in their study. The other crime categories studied were homicide, burglary, motor theft, larceny, and vandalism.

Social factors have been studied less in previous literature, even though there are many benefits that living in a neighborhood with high social capital offers (Blair & Larsen, 2010). Their research tested this by constructing a generalized least squares (GLS) regression model inspired by the HPM. Independent variables were various property characteristic control variables and the average neighborhood response to the question: "How satisfied are you with your neighbors?." The results show that residents' satisfaction with their neighbors is a significant determinant of housing values.

Property values perform well as a measurement of socioeconomic status (Coffee et al., 2013). Also, through the construction of various regression models, Han et al. (2014) found that family socio-economic status is significantly positively related to social capital. A regularly used social measurement is labor participation. Through linear regression and probit models, Johnson (2014) found a positive correlation between property values and female labor participation. Fu et al. (2016) conducted a similar method. On the contrary, he concluded that increased property value decreases the odds of a woman participating in the labor market. Sidenote is that Johnson conducted his research in the United States, and Fu et al. performed their analysis in China.

2.4 Objective and Subjective Measurements Comparison

While there is much literature on the influence of either subjective measurements or objective measurements on property values, there has not been much literature combining these two. Poor et al. (2001) explored and compared objective, scientific assessments of environmental quality with subjective evaluations from individuals in the context of HPMs. The subjective evaluations were obtained through a survey. They found that objective measurements significantly outperformed subjective measurements in predicting prices. Qiu et al. (2022) compared objective and subjective indicators of the street environment and their relationship with property values. Subjective measurements were obtained through two steps. First, surveys on street view images (SVI) were held, where people had to pick a preferred SVI out of two different ones. With these results, predictions were made for all streets through various machine learning methods. In the second step of their research, they assessed various linear regressions, which they tested for spatial autocorrelation using Moran's I statistic and Robust Lagrange Multiplier (LM) test. Afterward, they conducted spatial autoregressive combined (SAC) and GWR models to account for spatial effects. They concluded that objective measurements are better at describing property values but added that subjective factors are valuable, complementing the objective factors. Qiu et al. (2023) conducted research on the same variables and a very similar method; they only excluded the GWR model. In this research, subjectively measured qualities showed a stronger correlation with housing prices than objective measures. The overall predictive strength of subjective perceptions was nearly equivalent to the predictive strength of objective measures. Interesting to note, is that they found the roles of objective and subjective measures related to property values to be opposite.

In other fields, comparisons between objective and subjective indicators have also been tested, such as in the health sector. Lee and Moudon (2006) used a survey to gather subjective measures and compared these with objective environmental measures to estimate the odds of walking for recreational or functional reasons. Through multinomial logit models, he found that there was poor agreement between subjective perceptions and objective indicators. Contrary, Nyunt et al. (2015) conclude that subjective and objective measures are complementary, each providing unique information. They obtained similar types of data as Lee and Moudon and constructed a regression model in which they used the adjusted R^2 as a performance metric to assess explanatory value.

2.5 Expectations

From the existing literature, expectations can be derived for each sub-question. Below, the sub-questions are listed, together with the expected answers.

1) What is better in explaining the value of houses in Rotterdam, social or safety indices?

Motivated by a larger amount of evidence of the relationship between housing values and safety factors than of that with social factors, I expect that safety measurements will be more effective in explaining property values in Rotterdam.

2) What is better in explaining the values of houses in Rotterdam, objective or subjective indices?

Most of the previous literature on the comparison of objective and subjective indicators and property values, has concluded that objective measures have stronger explanatory power. Therefore, I expect that objective indices are better at explaining housing values in Rotterdam.

3) To what extent do subjective indices and objective indices complement each other in explaining property values in Rotterdam?

The majority of literature comparing subjective and objective factors, stated that both types of measurement possess their own value, making it complementary. This results in my expectation that objective and subjective indices complement each other in explaining property values in Rotterdam.

4) How do factors inside the social and safety indices relate to the value of houses in Rotterdam?

Since previous literature has shown differences in significance for various measurements, I expect different relations between property values and various determinants for both social and safety indices. For instance, I expect an insignificant relation between burglaries and property values.

5) How do the associations between social and safety indices and property values vary across different neighborhoods in Rotterdam?

Existing literature gives reason to believe that the associations of social and safety indices for property values differ across different neighborhoods. Hence, the expectation that there will be various associations between social and safety indices and property values across different neighborhoods in Rotterdam.

3. Data

The data in this paper is obtained through the municipality of Rotterdam and originates from research called ‘Wijkprofiel,’ which translates to ‘District Profile.’ Section 3.1 explains what Wijkprofiel is and its purpose. Section 3.2 describes the property values used, while sections 3.3, 3.4, and 3.5 give more insight into the social index, the safety index, and the physical index, respectively. The summary statistics of the data used in this research can be found in Table 7 in the Appendix.

3.1 Wijkprofiel

Every two years, the municipality of Rotterdam publishes the ‘Wijkprofiel,’ which shows the status of its districts and neighborhoods in terms of social and physical conditions, as well as safety. Rotterdam is divided into 14 districts, and every district is split into various neighborhoods, resulting in 71 neighborhoods in total. The data from Wijkprofiel enables the city council and neighborhood councils to create neighborhood agreements in collaboration with partners, residents, and businesses (Wijkprofiel Rotterdam, n.d.). A neighborhood agreement is the basis for collaboration in the neighborhood. It consists of a plan that outlines what, according to the neighborhood, needs to be done in the coming period to improve the neighborhood. It also sets out agreements on how the neighborhood can participate in city-wide issues related to the neighborhood (Wijkraden, n.d.).

Wijkprofiel offers both a subjective and an objective index to measure social conditions, physical elements, and safety. The subjective index is obtained through surveys held every two years by the municipality. There are two surveys: one called "Wijkonderzoek" (Neighborhood Research) and the other called "Veiligheidsmonitor" (Safety Monitor). The municipality determines through a random sample which residents aged 15 and over will be invited to participate. If not enough participants have completed the survey in a particular neighborhood, the municipality tries to motivate people to participate through home visits (Privacy, 2023). The objective index is determined by numbers and statistics obtained from the municipal administration (Wijkprofiel Rotterdam, n.d.).

Wijkprofiel is published online and publicly available at both the city and district levels via wijkprofiel.rotterdam.nl. Through the municipality, I obtain this data on neighborhood-level for all years in which Wijkprofiel was conducted: 2014, 2016, 2018, 2020, 2022, and 2024. The data consist of the physical index, the safety index, and the social index. The scores for the themes are displayed as index scores, with the score for the city of Rotterdam in 2014 set at 100. All neighborhood scores are calculated in comparison to this baseline score. For example, if a neighborhood's safety index score in a particular year is higher than 100, it indicates that this neighborhood is considered safer in that year than Rotterdam was on average in 2014.

All indices are built up from various sub-indices, of which there is both a subjective and objective measurement. The average score of all objective sub-indices results in a total score for the objective index, and the subjective index is calculated in the same way. So, the scores of various indices in Wijkprofiel are based on both measurable facts and figures, as well as the perceptions of Rotterdam residents, each counting equally towards the final number representing the index score. Strikingly, there are sometimes big differences between the subjective and objective index scores, which will be shown in the following sub-sections.

3.2 Property values

The dependent variable in this research is the value of properties. To obtain a similar measurement across the dataset, I use the "waarde van onroerende zaken" (WOZ-waarde), which translates to "real estate value." Municipalities in the Netherlands determine the WOZ-waarde, or property value, through an evaluation process that calculates what the property would likely have cost if sold on January 1st of the previous year. The WOZ-waarde is based on actual sales in the neighborhood. The tax assessor reviews data about the land and the building and obtains information on one or more comparable properties that were sold around the reference date. Municipalities then use computer models that make comparisons based on various factors, such as the location and size of the properties. These valuations do not consider personal financial factors like mortgages or ground leases. (Ministerie van Algemene Zaken, 2023).

To control for a property's size, I use the property value per square meter. Since this is part of the physical index in Wijkprofiel, I obtain the neighborhood-level average property value per square meter for all 71 neighborhoods in Rotterdam. The data is obtained for six years: 2014, 2016, 2018, 2020, 2022, and 2024.

Figure 1 shows boxplots of the distribution of neighborhood-level average property values per square meter per year. The values decreased slightly in 2016, followed by a slight increase in 2018. Afterward, there was a significant rise in property values, which continued in the following years.

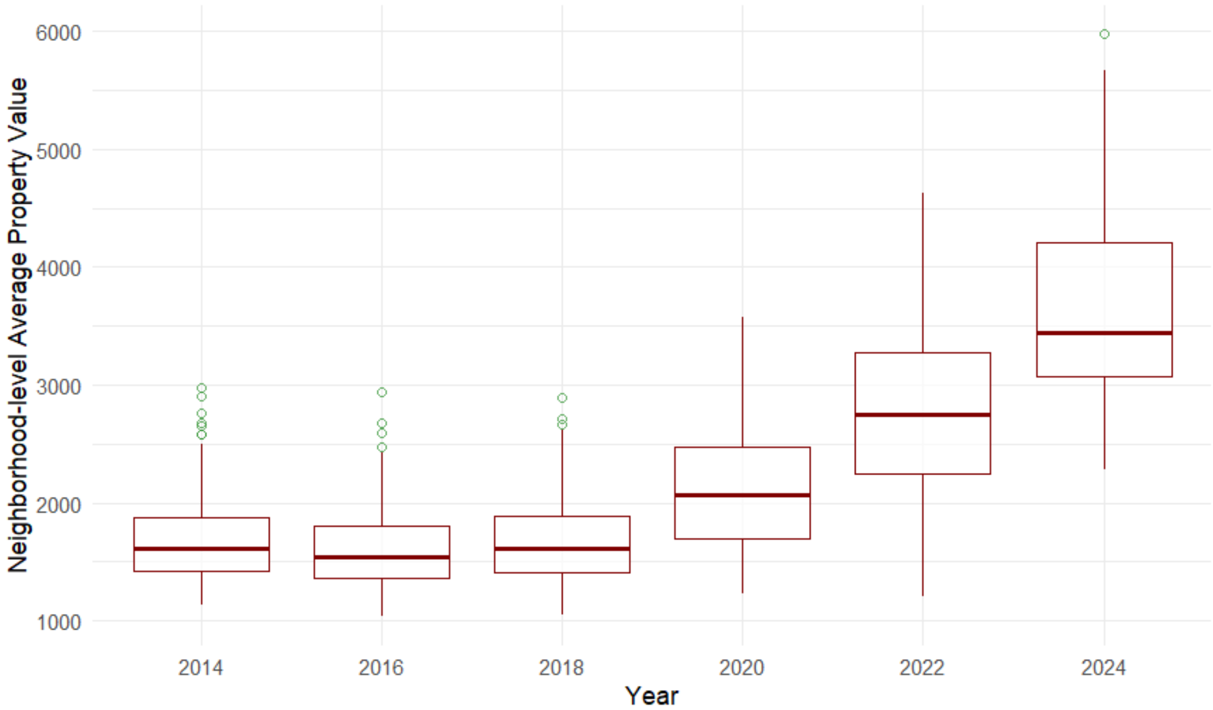

Figure 1 Distribution of Neighborhood-level Average Property Value

3.3 Social Index

The Social Index consists of five distinct sub-indices: judgment on quality of life, self-reliance, co-reliance, participation, and bonding.

Judgment on Quality of Life: This sub-index relies only on subjective assessments, measuring individual satisfaction with life. It captures personal evaluations through surveys about individuals' overall life satisfaction and subjective ratings of their life circumstances.

Self-Reliance: This dimension focuses on an individual's capacity to meet their own needs utilizing personal skills and resources independently. Objective measures include employment rates and cultural engagement, while subjective indicators explore perceptions of income sufficiency and feeling alone.

Co-Reliance: In contrast to self-reliance, co-reliance emphasizes mutual dependence and support among individuals or groups. Objectively, it is measured by factors like the proportion of individuals engaging in volunteer activities or assisting neighbors. Subjectively, it examines residents' perceptions of interactions with neighbors.

Participation: This sub-index involves participating in activities or processes where individuals or groups have a stake. Objectively, it includes participation in local planning and membership in organizations such as sports clubs. Subjectively, it assesses satisfaction with a person's level of engagement and experiences of discrimination.

Bonding: This sub-index captures the formation and quality of close interpersonal relationships and community ties. It considers objective factors such as residential mobility. Subjectively, it covers residents' feelings of happiness and belonging within the neighborhood.

Figure 2 visualizes the distribution of the neighborhood-level objective social index (left) and subjective social index (right) across different years, as shown through boxplots. Initially, the subjective scores exhibit a stronger increase than the objective scores over the first three years. The trend in the objective social index appears to stabilize after the initial three years. Contrary, the subjective social index started to decline, dropping to a lower point than in 2014. Moreover, it is noticeable that the subjective social index displays a broader range of values than the objective social index, indicating more variability in subjective assessments of social conditions.

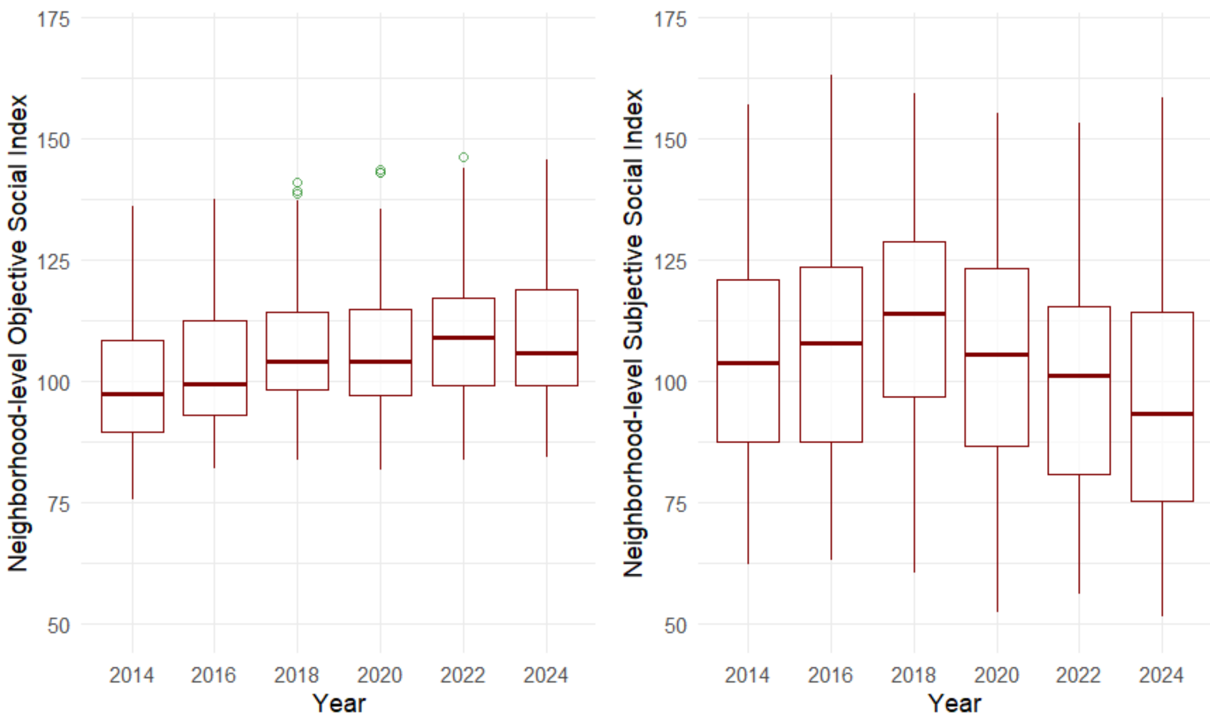

Figure 2 Distribution of Neighborhood-level Objective (left) and Subjective (right) Social Index

While Figure 2 provides an overview of the differences in distributions between the objective and subjective social indices, it does not offer insight into the variations within specific neighborhoods. To address this, Figure 3 presents the objective (left) and subjective (right) social index scores for five randomly selected neighborhoods in Rotterdam: Bergpolder, Carnisse, Hillegersberg-Zuid, Hoogvliet-Zuid, and Tussendijken. I do not show this plot for all neighborhoods, since this would result in a very unclear visualization due to the number of neighborhoods in this research. The results reveal significant differences between the objective and subjective scores, with these gaps varying in both directions. For example, the subjective score for Hillegersberg-Zuid in 2016 was almost 30 points higher than its objective counterpart, whereas in 2022, the subjective score for Carnisse was approximately 35 points lower than the objective measure.

Besides absolute differences between objective and subjective scores, significant differences are also visible in the trends of both scores. For example, in Bergpolder, every time the objective score increases, the subjective score decreases, and vice versa.

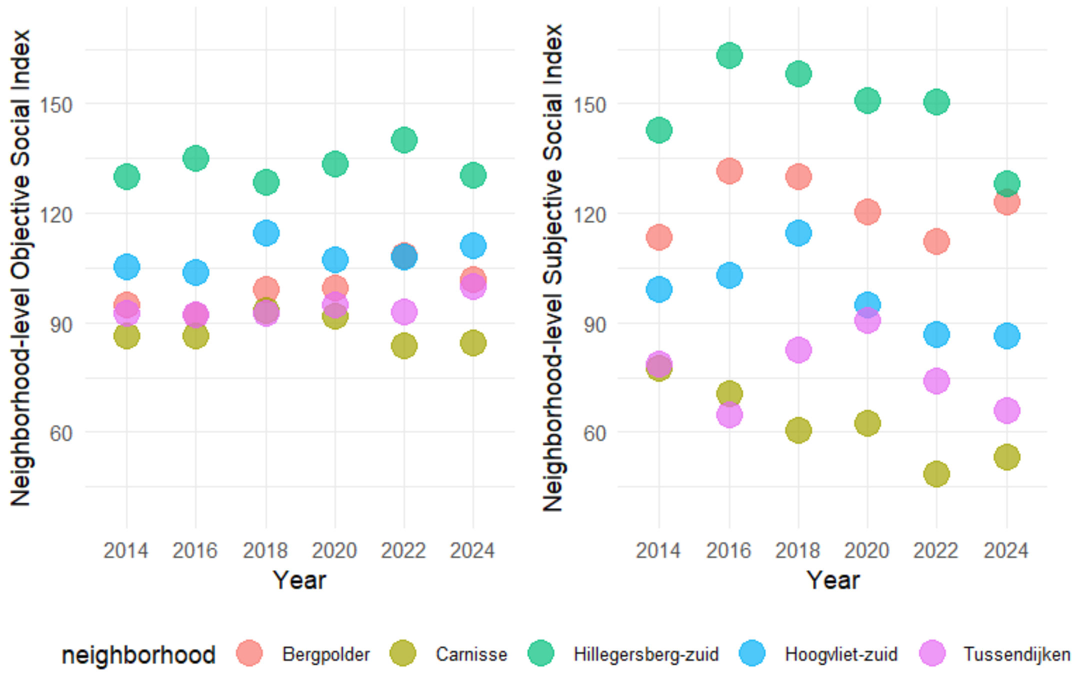

Figure 3 Observations of Objective (left) and Subjective (right) Neighborhood-level Social Index

3.4 Safety Index

The Safety Index is calculated using two primary components: the perception of safety and the prevalence of five distinct categories of crime: theft, violence, burglary, vandalism, and nuisance.

Perception of Safety: This sub-index only incorporates subjective evaluations. It features an array of survey questions designed to obtain residents' feeling on the likelihood that they or someone in their household might fall victim to one of four specified crimes: burglary at their residence, physical abuse, pickpocketing (non-violent), and mugging (violent). Additionally, the survey examines whether residents adopt specific strategies to mitigate crime risk, such as avoiding particular areas within their neighborhood during nighttime.

Crime Categories: Each of the five mentioned crime types is further analyzed through its respective sub-index. These sub-indices objectively quantify the frequency of each crime type within the neighborhood, measured per 1,000 inhabitants. Subjectively, they explore the extent to which residents perceive these crimes as problematic within their neighborhood. This subjective component also includes residents' self-reports of victimization by these crimes.

Figure 4 shows the distribution of the neighborhood-level objective (left) and subjective (right) safety index across various years, visualized through boxplots. The objective index score demonstrates a slight upward trend in median scores, suggesting a constant increase in objective safety measures over the observed period. Conversely, the scores for the subjective index do not exhibit a clear trend, remaining relatively stable with only minor fluctuations in the median values over the years. The subjective safety index exhibits a wider variability, reflecting greater divergence in individual perceptions of safety. This broader range of values suggests more differences in how safety is subjectively assessed by respondents compared to the more consistent measurements of the objective safety index.

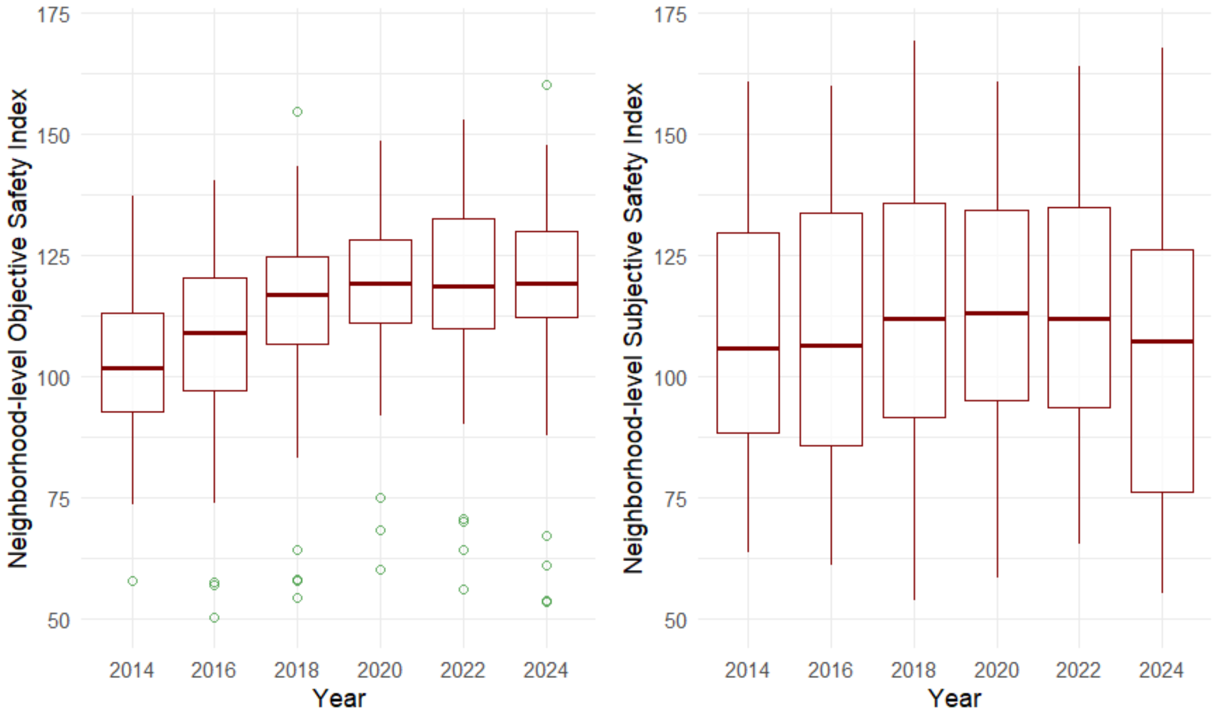

Figure 4 Distribution of Neighborhood-level Objective (left) and Subjective (right) Safety Index

Figure 5 illustrates the objective (left) and subjective (right) safety index scores for the same neighborhoods shown in Figure 4. Here too, significant contrasts are observed, with the differences pointing in both directions. For instance, in 2024, Hoogvliet-Zuid had the highest score for the objective safety index among these neighborhoods, but its subjective score was 30 points lower, placing it as the third highest. Conversely, in the same year, Hillegersberg-Zuid's subjective score was 27 points higher than its objective counterpart.

The trends of the objective and subjective safety index scores are not compatible in every neighborhood, although they are more compatible than those of the social index score. A not-so-similar trend can be spotted in Hoogvliet-Zuid, where the trends were the same from 2014-2020, but afterward, the objective score decreased and then increased. Meanwhile, the subjective score followed the opposite trend in these last two years.

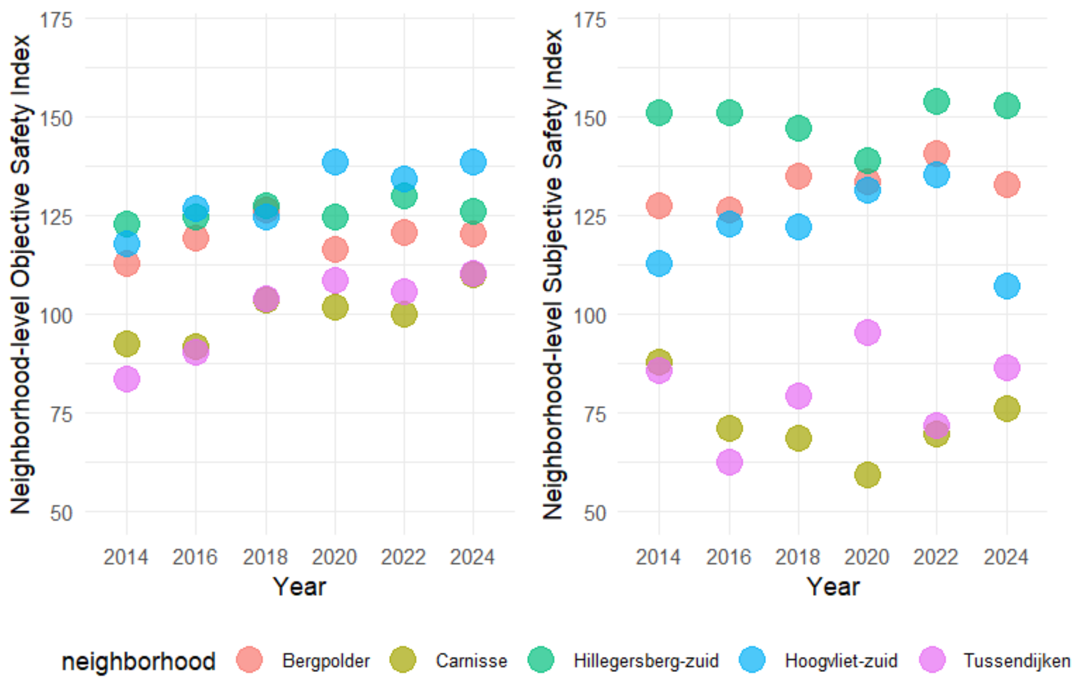

Figure 5 Observations of Objective (left) and Subjective (right) Neighborhood-level Safety Index

3.5 Physical Index

The Physical Index comprises five sub-categories: Living Experience, Housing, Public Space, Amenities, and Environment. Each category incorporates both objective and subjective components, except for the Living Experience, which is evaluated only through subjective measures.

Living Experience: This sub-index assesses individual satisfaction with their living situation by asking residents about their contentment and propensity to relocate if given the opportunity.

Housing: The objective component of this category is determined by analyzing data concerning the types and conditions of housing within a neighborhood. Conversely, the subjective aspect examines resident satisfaction with various housing characteristics.

Public Space: The objective evaluation of public spaces is based on statistics such as the occurrence of traffic accidents, street maintenance, measurements of the presence of garbage, and the quality of road surfaces. Subjectively, it encompasses assessments of street conditions, parks, trash management, and overall satisfaction with these elements.

Amenities: On the objective front, this sub-category measures the accessibility of stores, sports clubs, and schools within a reasonable distance. Subjectively, it reflects residents' satisfaction with the availability of these amenities in their neighborhood.

Environment: This sub-index objectively quantifies environmental quality through metrics like air quality and noise levels. The subjective component addresses residents' perceptions and concerns regarding noise, smell, and potential flooding due to various environmental factors.

Figure 6 displays the variation in the objective (left) and subjective (right) physical index across different years using boxplots. Notably, there appears to be a slight upward trend in the median values of the objective index over the observed period, suggesting a constant improvement in the objective measurements of physical conditions. The subjective physical index shows a broader range of values than the objective physical index, indicating significant variability in residents' perceptions of their physical environment. While the median values show some fluctuations, a declining trend is visible in the last three years, reflecting a decrease in subjective assessments of physical conditions.

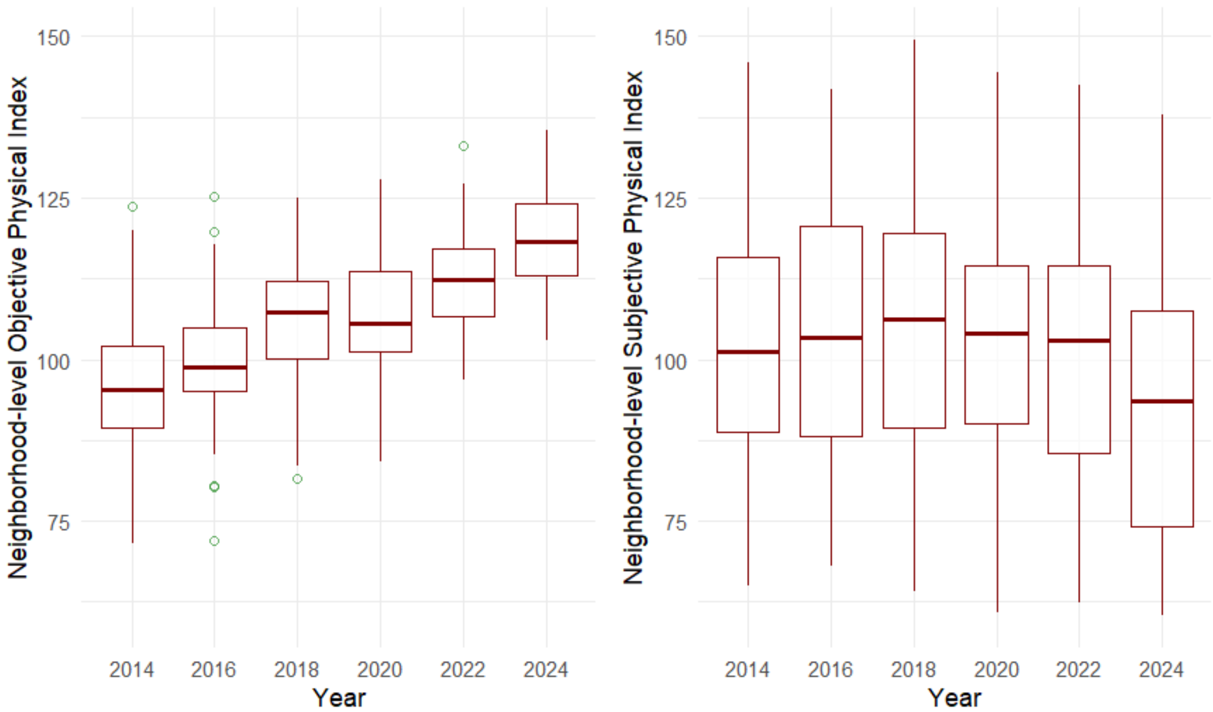

Figure 6 Distribution of Neighborhood-level Objective (left) and Subjective (right) Physical Index

Figure 7 provides an overview of the objective (left) and subjective (right) physical index scores for the five selected neighborhoods. The figure reveals significant discrepancies between the two measures. For instance, in 2024, the difference between the objective and subjective measures in Tussendijken is nearly 60 points, with the objective measure being higher. Conversely, in 2016, Bergpolder had a subjective score that was 15 points higher than the objective measure.

Significant differences can be spotted when analyzing the trends of the objective and subjective physical index scores. For instance, the objective index score for Carnisse increased every year, while its subjective counterpart only increased in 2022 and decreased in all other years.

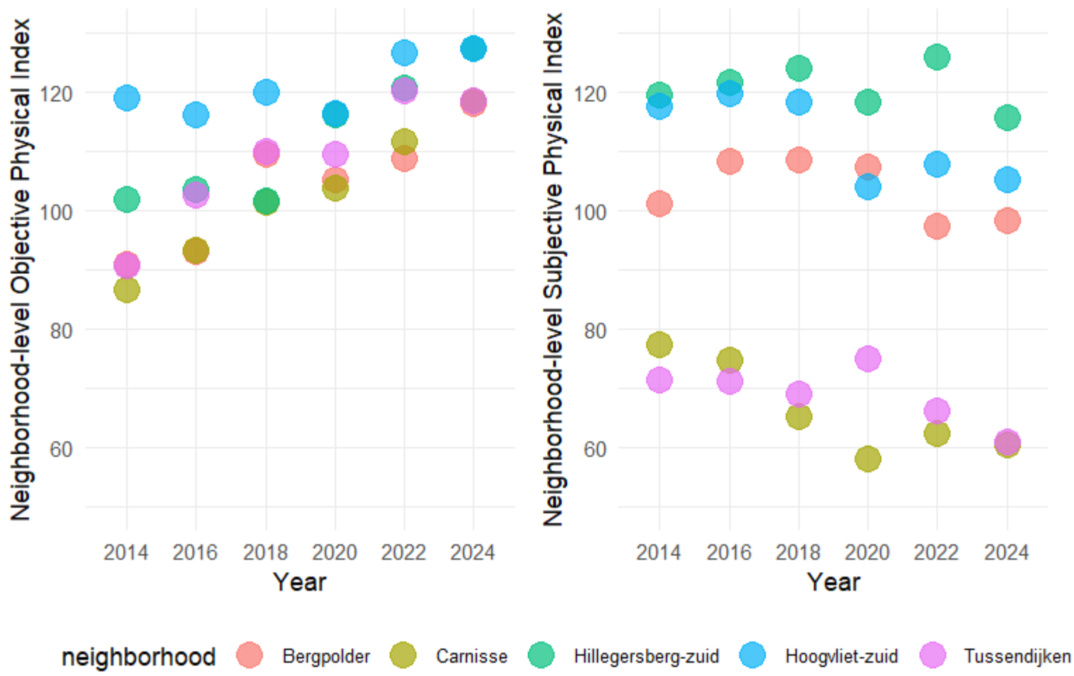

Figure 7 Observations of Objective (left) and Subjective (right) Neighborhood-level Physical Index

4. Methodology

This part describes the methods used in this research. Section 4.1 shows the use of the hedonic pricing model in this paper, section 4.2 describes how the analysis will take place, while section 4.2.1, 4.2.2 and 4.2.3 explain the different estimation methods which will be utilized in this research. These methods are linear regression, fixed effects regression and geographically weighted regression. Lastly, section 4.3 shows various metrics to assess model performance.

4.1 Hedonic Pricing Model

Since the HPM is widely used in housing prices literature, I will also perform my research inspired by this model. The simple HPM to be constructed using my data looks as follows:

$$(1)\ Property\ value_{it} = \beta_0 + \beta_1 * Housing\ objective_{it} + \beta_2 * Public\ Space\ objective_{it} + \beta_3 * Amenities\ objective_{it} + \beta_4 * Environment\ objective_{it} + \varepsilon.$$

Here, $Property\ value_{it}$ is the average property value per square meter of neighborhood i in year t. $Housing\ objective_{it}$, $Public\ Space\ objective_{it}$, $Amenities\ objective_{it}$ and $Environment\ objective_{it}$ are the index scores of corresponding indices of neighborhood i in year t. $\beta_0, \beta_1, \ldots, \beta_4$ represent the coefficients and ε is the error term.

These variables are the sub-indices of the physical index, holding information on structural, location, and neighborhood attributes. Since this research is conducted on neighborhood-level, they form a reasonable substitute for what had otherwise been structural, locational and neighborhood attributes of individual properties. Only objective measures are used, since the traditional HPM does not include subjective measures.

Since there has been an enormous increase in property values from 2018 to 2024, I expect that it will not be possible to explain this rise in prices through an improvement in attributes. To control for this inflation, I add the Dutch consumer price index (CPI) to the HPM. The CPI is calculated as an index score as well, with the value for 2014 set at 100. The model then looks like this:

$$(2)\ Property\ value_{it} = \beta_0 + \beta_1 * Housing\ objective_{it} + \beta_2 * Public\ Space\ objective_{it} + \beta_3 * Amenities\ objective_{it} + \beta_4 * Environment\ objective_{it} + \beta_5 * CPI_t + \varepsilon.$$

Here, $Property\ value_{it}$ is the average property value per square meter of neighborhood i in year t. $Housing\ objective_{it}$, $Public\ Space\ objective_{it}$, $Amenities\ objective_{it}$, and $Amenities\ objective_{it}$ are the index scores of corresponding indices of neighborhood i in year t. CPI_t is the Dutch consumer price index in year t, $\beta_0, \beta_1, \ldots, \beta_5$ represent the coefficients and ε is the error term.

Equation (2) suits as the baseline model for this research. I assume that an additive effect of CPI is unlikely. The CPI represents the percentage growth in inflation, which, when applied in a linear regression model, translates into a growth in euros. This transformation can be problematic because it ignores the fact that properties in different neighborhoods have different baseline values. Consequently, the same inflation rate has a larger effect on property values in expensive neighborhoods compared to less expensive ones. However, it seems like the most reasonable variable to account for the huge inflation. In 4.2.3 I explain GWR, which provides me with the information to assess the assumption of the effect of CPI not being additive. It also accounts for the effect possibly not being additive.

When adding all social and safety indices to the model, the total model looks as follows: (3) $Property\ value_{it} = \beta_0 + \beta_1 * Housing\ objective_{it} + \beta_2 * Public\ Space\ objective_{it} + \beta_3 * Amenities\ objective_{it} + \beta_4 * Environment\ objective_{it} + \beta_5 * CPI_t + \beta_6 * Judgement\ on\ Quality\ of\ Life_{it} + \beta_7 * Perception\ of\ Safety_{it} + \beta_8 * Selfreliance\ objective_{it} + \beta_9 * Selfreliance\ subjective_{it} + \beta_{10} * Coreliance\ objective_{it} + \beta_{11} * Coreliance\ subjective_{it} + \beta_{12} * Participation\ objective_{it} + \beta_{13} * Participation\ subjective_{it} + \beta_{14} * Bonding\ objective_{it} + \beta_{15} * Bonding\ subjective_{it} + \beta_{16} * Theft\ objective_{it} + \beta_{17} * Theft\ subjective_{it} + \beta_{18} * Violence\ objective_{it} + \beta_{19} * Violence\ subjective_{it} + \beta_{20} * Burglary\ objective_{it} + \beta_{21} * Burglary\ subjective_{it} + \beta_{22} * Vandalism\ objective_{it} + \beta_{23} * Vandalism\ subjective_{it} + \beta_{24} * Nuisance\ objective_{it} + \beta_{25} * Nuisance\ objective_{it} + \varepsilon.$

Here, $Property\ value_{it}$ is the average property value per square meter of neighborhood i in year t, all indices are the index scores of corresponding indices of neighborhood i in year t. CPI_t is the Dutch consumer price index in year t, $\beta_0, \beta_1, \dots, \beta_{25}$ represent the coefficients and ε is the error term.

Equation (3) consists of the baseline model, expanded by all safety and social indices, both objective and subjective. To check for multicollinearity, a variance inflation factor (VIF) test is conducted. If a variable has a VIF score higher than 10 will be removed, if multiple variables have a VIF score higher than this threshold, the variable with the highest VIF score will be removed. This process continues, until no variables surpass this threshold.

In the first VIF test, *Perception of safety* is the only variable that surpasses this threshold, meaning that it will be removed from this research. It is strongest correlated with *Nuisance – subjective, Self-reliance – objective* and *Violence – subjective* with correlation coefficients of 0.841, 0.837 and 0.825, respectively. After the removal of *Perception of safety,* another VIF test is conducted, in which none of the variables have a VIF score higher than 10.

4.2 Analysis

As an initial analysis, I construct 9 simple linear regressions. Linear regression models are explained in 4.2.1. Each of these linear regression models has the neighborhood-level average property value per square meter as the dependent variable. All different variables from the baseline model (Equation 2) will have their own linear regression in which they are the independent variable. There will also be distinct regressions with either social indices, safety indices, objective measures, or subjective measures as independent variables. This will be done to assess the explanatory value of each category by itself.

In the second step of the analysis using linear regression estimated by OLS, I start with the baseline HPM, as shown in Equation (2). I then expand this basic model into five different versions. The Social Model includes the baseline model and all objective and subjective social indices, while the Safety Model adds all objective and subjective safety indices to the baseline model. The Objective Model combines the baseline model with all objective measurements of social and safety conditions, and the Subjective model incorporates the subjective measurements of the social and safety indices into the baseline model. Finally, the Total Model combines all the different social and safety measurements with the baseline model into one model.

The second step of the linear regression analysis, as described in the above paragraph, will be repeated for both the FE regression method, and the GWR method. The FE regression method is explained in 4.2.2, while GWR is described in 4.2.3. The performance of all models will be measured using the metrics explained in section 4.3.

To evaluate the associations of specific indices with neighborhood-level average property values, the best-performing linear regression or FE model will be analyzed and interpreted based on the sign and magnitude of the coefficients. For this specific goal, the GWR model will not be utilized since these models contain varying coefficients per neighborhood, which means they are very hard to interpret at a city level. The GWR model will be utilized to assess the spatial differences of associations between neighborhoods.

4.2.1 Linear Regression Model

The initial analysis will be conducted using linear regression. Linear regression is a statistical method to estimate the relationship between a dependent variable and one or more independent variables. The goal is to find the line of best fit that minimizes the sum of the squared differences between the observed values and those predicted by the model, also known as errors. The method is picked for the initial analysis due to its simplicity, efficiency, and flexibility. It is very easy to construct a linear regression model, while they are also highly interpretable. The model equation looks as follows:

$$(4)\ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon.$$

Where y is the dependent variable, $x_1, x_2, \ldots, x_k$ are the dependent variables, $\beta_0, \beta_1, \ldots, \beta_k$ are the coefficients to be estimated, and ε is the error term.

In a linear regression model, the goal is to determine the line that best fits the data by minimizing the sum of squared residuals. The coefficients are estimated using OLS, where residuals represent the differences between the observed values y_i and the predicted values $\hat{y}_i$. Thus, the function to minimize looks as follows:

$$(5)\ Minimize\ \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

In this equation ε represents the error term, y_i is the observed value, and $\hat{y}_i$ is the predicted value of y_i.

When the sum of squared residuals is minimized through this process, the resulting coefficients $\beta_0, \beta_1, \ldots, \beta_k$ provide the best linear estimates to predict y_i.

Linear regression holds five key assumptions: Linearity of the relationship between the independent and dependent variables, independence of the residuals, homoscedasticity, normality, and no multicollinearity. When these assumptions hold, a linear regression model is effective.

However, in the case of property values, residuals are usually not independent (Gibbons, 2004). In cases where there are unobserved variables that vary across neighborhoods, but are constant over time, a linear regression model may not capture the relationship between the dependent and independent variables well. This can lead to inconsistent estimates due to omitted variable bias, which occurs when a relevant variable that is correlated with both the independent variable and the dependent variables is left out of a regression model.

4.2.2 Fixed Effects Regression Model

To address omitted variable bias arising from unobserved variables that vary across neighborhoods but are constant over time, I utilize a panel method, incorporating either fixed effects or random effects. Both methods account for unobserved heterogeneity, but they hold different assumptions.

The fixed effects approach controls for all time-invariant characteristics of the neighborhoods by including a separate intercept for each neighborhood. This means that any unobserved, constant factors that vary between neighborhoods are accounted for, effectively eliminating their biasing effect on the estimated coefficients of the independent variables. This method is particularly robust to omitted variable bias when these unobserved factors are correlated with the independent variables. For instance, if certain neighborhoods have intrinsic characteristics such as long-standing socioeconomic conditions or a cultural identity, that influence property values, fixed effects will control for these factors, increasing the likelihood that the estimates of the independent variables more accurately reflect their relationship with the dependent variable.

On the contrary, random effects models assume that the neighborhood-specific effects are random and uncorrelated with the independent variables. If this assumption holds, random effects are more efficient than fixed effects, as it allows for the inclusion of time-invariant variables in the analysis. Random effects models can provide more precise estimates by using both within-group and between-group variation. However, if the assumption that the unobserved effects are uncorrelated with the independent variables is violated, the estimates will be biased.

To determine whether to use fixed or random effects, a Hausman test is constructed. This test compares the fixed effects and random effects estimates to assess whether the random effects assumption holds. To test this, first a random effects (RE) model and a fixed effects (FE) model should be constructed. The formula of the Hausman test looks as follows:

$$(6)\ H = \left(\widehat{\beta_{RE}} - \widehat{\beta_{FE}}\right)^T \left[var\left(\widehat{\beta_{FE}}\right) - var\left(\widehat{\beta_{RE}}\right)\right]^+ \left(\widehat{\beta_{RE}} - \widehat{\beta_{FE}}\right) \sim \chi_v^2.$$

Here, $+$ stand for pseudo inverse, $\widehat{\beta_{FE}}$ refers to the fixed effects estimates, and $\widehat{\beta_{RE}}$ refers to the random effects estimates. The test statistic H follows a chi-squared distribution with the degrees of freedom v equal to the number of independent variables.

The null hypothesis of the Hausman test states that there is no correlation between the unique errors and the independent variables in the model. In this research, the null hypothesis was rejected for all considered models, indicating correlation between the unique errors and the independent variables. Thus, the FE model should be used.

Since the primary interest of this research lies in understanding variations within neighborhoods over time, I use fixed effects on the different neighborhoods. This approach allows me to control for all neighborhood-specific characteristics that do not change over time, effectively isolating the impact of the independent variables on property values within each neighborhood. The FE regression looks as follows:

$$(7)\ y_{nt} = \alpha_n + x'_{nt}\beta + \varepsilon_{nt}.$$

Here, y_{nt} represents the dependent variable for neighborhood n at time t, α_n captures the neighborhood-specific effect, x_{nt} is a vector of the independent variables in neighborhood n at time t with β as its coefficients, lastly, ε_{nt} represents the error term.

As the goal of this model is to eliminate neighborhood-invariant effects (α_n), the neighborhood-specific mean from each variable is subtracted. The equation then looks like this:

$$(8)\ (y_{nt} - \overline{y_n}) = (x_{nt} - \overline{x_n})'\beta + (\varepsilon_{nt} - \overline{\varepsilon_n}).$$

Here, $\overline{y_n}$ and $\overline{x_n}$ represent the means of y_{nt} and x_{nt} over time for neighborhood n, ε_{nt} is the error term, and β is the vector of coefficients for the FE regression.

Note that Equation (8) does not have a constant variable, or intercept. Due to the elimination of neighborhood-invariant effects, the global intercept is removed. Each neighborhood now has its own intercept value. Consequently, the intercept differs locally, but the coefficients of the variables are estimated as a global effect, affecting each neighborhood the same way.

By focusing on within-neighborhood changes over time, this methodology ensures that the estimates reflect temporal variations in property values while controlling for the constant characteristics of each neighborhood. This means that I am not primarily examining differences across neighborhoods, which are controlled for by the fixed effects.

By accounting for unobserved heterogeneities across neighborhoods, this method also controls for spatial heterogeneity to some extent. However, it is important to consider that the fixed effects of adjacent neighborhoods may be correlated, introducing spatial autocorrelation into the model. Spatial autocorrelation occurs when the unobserved factors influencing property values in one neighborhood are related to those in neighboring areas. Ignoring this spatial dependence can lead to biased and inefficient estimates.

4.2.3 Geographically Weighted Regression Model

To account for spatial autocorrelation and better capture the spatial relationships between neighborhoods, I employ Geographically Weighted Regression (GWR), first introduced by Brunsdon et al. (1996). Unlike traditional regression models that construct global parameter estimates, GWR allows for local parameter estimation, capturing spatial heterogeneity in the data. This method provides a more nuanced understanding of how spatial context influences property values.

GWR differs from creating separate regressions for all neighborhoods by also incorporating effects from nearby neighborhoods. An estimation per region is not convenient because it would ignore the spatial dependencies and interactions between neighborhoods, possibly leading to a less accurate analysis. By using a spatial kernel, GWR ensures that nearby neighborhoods have more influence on local parameter estimates than neighborhoods that are located further away, which results in realistic spatial processes and smooth transitions between areas. This approach leverages the spatial structure of the data to produce more reliable and contextually relevant estimates.

GWR also includes a bandwidth parameter to control for the spatial extent of influence, ensuring that only neighborhoods within a certain distance of a neighborhood can influence the outcome of the regression for that neighborhood. In this paper, distance is measured using Euclidean distance, which is the straight-line distance, between the geographical centers of the neighborhoods. [1] The formula of the GWR looks as follows:

$$(9)\ y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i.$$

In this equation, y_i represents the dependent variable of location i, which is a neighborhood in my research, x_{ik} is the k-th independent variable at location i, $\beta_k(u_i, v_i)$ represents the k-th coefficient at location (u_i, v_i), where (u_i, v_i) represents the coordinates of the center of neighborhood i. Lastly, ε_i denotes the error term.

The spatial kernel I utilize to assign weights to nearby neighborhoods is the Gaussian kernel. The Gaussian kernel is the most suitable spatial kernel for continuous variables as it provides a smooth transition of weights, ensuring that all points within the bandwidth influence the local parameter estimates, but closer points have more influence.

The optimal bandwidth, which determines the extent to which the influence of an observation decreases with distance, will be estimated through cross validation. Cross-validation is a technique used to measure the performance of a model by dividing the data into subsets, training the model on some subsets, and testing it on the remaining subsets to ensure it handles unseen data well. This method ensures the optimal bandwidth will be used in the GWR model. Since I construct various GWR models, the bandwidth will differ across models, to ensure each model has its optimal bandwidth incorporated.

In GWR, there is no need for fixed effects because GWR accounts for spatial heterogeneity by allowing the coefficients to vary by location. This flexibility addresses the issue of spatial autocorrelation without the need for fixed effects to control for unobserved, time-invariant characteristics. The local

[1] I obtain the coordinates of the centers of the neighborhoods through www.coordinatenbepalen.nl, which is a website that contains geographical coordinates.

parameter estimates provided by GWR offer a more detailed and spatially aware estimation of the relationships between variables.

In Section 4.1 I stated the assumption of an additive effect of *CPI* being unlikely. Since GWR allows for the examination of spatially varying relationships by estimating local regression coefficients for each neighborhood, the influence of CPI on property values can be different in each neighborhood. By capturing these local variations, GWR can reveal whether the impact of CPI is indeed non-uniform across different neighborhoods, and account for the differing baseline property values.

4.3 Performance Metrics

This part discusses various performance metrics. 4.6.1 explains the adjusted R^2, 4.6.2 describes the F-statistic for regression models, and lastly, 4.6.3 discusses Moran's I test statistic.

4.3.1 Adjusted R^2

After computing the regression models, performance will be assessed through various metrics. The most important metric to be considered in this research is adjusted R^2. Adjusted R^2 is used to evaluate the goodness-of-fit of a regression model while accounting for the number of predictors in the model. It adjusts the R^2 value based on the number of predictors and the sample size. R^2 is the proportion of the variance in the dependent variable that is predictable from the independent variables. It is calculated as follows:

$$(10)\ R^2 = 1 - \frac{RSS}{TSS}.$$

In this equation, RSS is the residual sum of squares, which can be seen as the variance in the outcome variable that was not explained by the model, and TSS is the total sum of squares, which is the total variance in the outcome variable.

Since adding more variables to a model will always improve the variance explained, it is better to use the adjusted R^2 when assessing the explanatory value of a combination of variables. The formula of the adjusted R^2 looks like this:

$$(11)\ Adjusted\ R^2 = 1 - \left(\frac{(1-R^2)(n-1)}{n-k-1}\right).$$

Here, n represents the number of observations, and k is the number of predictors in the model.

Since this research focuses on the explanatory value of social and safety indices, adjusted R^2 is the most important metric. It reflects how well the model explains variance in the outcome variable while controlling for the number of predictors. This indicates that when a model possesses a higher adjusted R^2, the values added contribute to the explanatory power of the model.

4.3.2 F-statistic

The F-statistic is used to assess a model's significance. The null hypothesis is that all regression coefficients are equal to 0, indicating that the model has no explanatory power. The alternative hypothesis states that at least one of the coefficients is not equal to 0, suggesting that the model has explanatory power. The F-statistic is calculated as follows:

$$(12)\ F = \frac{(\frac{RSS}{k})}{(\frac{ESS}{n-k-1})}.$$

In this formula, RSS is the residual sum of squares, and ESS is the error sum of squares. n represents the number of observations, and k equals the number of predictors in the model.

After computing the F-statistic, a corresponding p-value is computed. The null hypothesis will be rejected if the p-value is less than the significance level. In this research, the F-statistic is not used to compare models, it is simply used as a metric to assess whether a model is significant.

4.3.3 Moran's I Test on Residuals

Moran's I test is a test to assess spatial autocorrelation in a model. The null hypothesis states that there is no spatial autocorrelation, meaning that the variables are randomly distributed in space. The alternative hypothesis is that he spatial distribution of high and low values in the dataset is more clustered than would be expected under the assumption of random spatial processes, indicating that spatial autocorrelation is present in the model. In this analysis, Moran's I is applied to the residuals from the regression model to check for spatial autocorrelation in the model's errors. The Moran's I statistic is calculated as follows:

$$(13)\ I = \frac{N \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^{N} (x_i - \bar{x})^2}.$$

Here N denotes the number of spatial units, x_i represent the value of the variable at location i, $\bar{x}$ is the mean value of the variable, w_{ij} contains the spatial weight between locations i and j based on distance, and lastly, S_0 is the sum of all spatial weights: $S_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$.

The spatial weights, w_{ij}, are calculated using the k-nearest neighbors method. This method identifies the nearest k neighbors for each neighborhood based on the coordinates of the center of each neighborhood. The optimal number of nearest neighbors is determined by comparing the results of Moran's I test across different k values. The optimal k is chosen by identifying the point where Moran's I values stabilize. In this research, the optimal k is found to be 72. Since the dataset includes all neighborhoods observed over 6 different years, a k of 72 effectively means that each neighborhood's spatial context is defined by its 12 nearest neighbors. After selecting the optimal k, the weights are standardized using a row-standardization method, ensuring that the sum of weights for each neighborhood equals 1. This standardization facilitates a more accurate comparison of influence across neighborhoods by normalizing the impact of each neighboring area.

After computing the Moran's I statistic, a corresponding p-value is computed. The null hypothesis will be rejected if the p-value is less than the significance level. If this is the case, it indicates that a model accounting for spatial differences should be used.

5. Results

The results section is structured as follows: section 5.1 shows the results of the initial linear regression analysis, section 5.2 discusses the results of the second step of linear regression analysis, section 5.3 examines the results of the FE models, and section 5.4 displays the results of the GWR models. Section 5.5 discusses differences across neighborhoods in Rotterdam, and the final section, 5.6, interprets the coefficients of the indices.

5.1 Comparison of Model Metrics Across Attribute Categories

Table 1 shows the performance of the initial analysis containing 9 linear regression models with property values as the outcome variable and various attributes as independent variables.

Table 1 Comparison of Model Metrics Across Attribute Categories

Linear Regression	Housing	Public Space	Amenities	Environment	CPI
Adjusted R^2	0.395***	0.018***	-0.001	0.126***	0.608***
Moran's I	0.129***	0.080***	0.085***	0.176***	0.246***
N	1	1	1	1	1

Linear Regression	Social Attributes	Safety Attributes	Objective Attributes	Subjective Attributes
Adjusted R^2	0.493***	0.367***	0.454***	0.343***
Moran's I	0.027***	0.031***	0.033***	0.021***
N	9	10	9	10

Notes: p-value *<0.1, ** p < 0.05, *** p < 0.01; For the adjusted R^2, the hypothesis tested is whether the regression is significant, determined using the F-statistic; For Moran's I test on residuals, the hypothesis tested is whether Moran's I statistic equals 0; N refers to the number of explanatory variables.

As can be seen, except for *amenities*, all attributes are distinctly significant at the 1% level explaining property values. This verifies the assumption that these attributes possess significant explanatory power for housing values. This explanatory power, based on adjusted R^2, can be classified as follows: *CPI* (0.608) > *Social Attributes* (0.493) > *Objective Attributes* (0.454) > *Housing* (0.395) > *Safety Attributes* (0.370) > *Subjective Attributes* (0.342) > *Environment* (0.126) > *Public Space* (0.018) > *Amenities* (-0.001). A side note on the explanatory value is that *Housing, Public Space, Amenities, Environment,* and *CPI* all consist of only one independent variable. At the same time, the other regression models contain multiple attributes as independent variables.

Furthermore, Moran's I test on the residuals confirms that the spatial distribution of high and low values in the dataset is more clustered than would be expected if the underlying spatial processes were random. This encourages the use of a model that accounts for spatial autocorrelation to improve accuracy and robustness in the analysis.

Higher values of Moran's I test statistic indicate stronger autocorrelation in the residuals, suggesting the following order of attributes with the strongest unexplained spatial autocorrelation: *CPI* (0.246) > *Environment* (0.176) > *Housing* (0.129) > *Amenities* (0.085) > *Public Space* (0.080) > *Objective Attributes* (0.033) > *Safety Attributes* (0.031) > *Social Attributes* (0.027) > *Subjective Attributes* (0.021). This order indicates that social and subjective attributes explain most of the spatial autocorrelation in the data, resulting in lower Moran's I values for their residuals. In contrast, attributes such as *Housing, Public Space, Amenities, Environment,* and *CPI* do not explain as much spatial autocorrelation, leading to higher Moran's I values and indicating stronger unexplained spatial patterns in their residuals. A sidenote here is that their linear regressions contain only one explanatory variable, whereas the other models include multiple explanatory variables. Models with multiple explanatory variables can usually explain more of the spatial processes, resulting in lower spatial autocorrelation in the residuals.

5.2 Analysis of Linear Regression Models

The next step in the analysis is the construction of the baseline linear regression model and the expanded models. As stated in Section 4.2, the explanatory variables in the baseline model are *Housing – objective, Public Space – objective, Amenities – objective, Environment – objective* and *CPI.* The Social Model includes the baseline model along with all objective and subjective social indices. The Safety Model adds all objective and subjective safety indices to the baseline model. The Objective Model combines the baseline model with all objective measurements of social and safety conditions, while the Subjective Model incorporates the subjective measurements of the social and safety indices into the baseline model. Finally, the Total Model integrates all the different social and safety measurements with the baseline model into one model. The performance of these models can be seen in Table 2.

Table 2 Performance of Linear Regression Models

Linear Regression	Baseline	Social	Safety	Objective	Subjective	Total
Adjusted R^2	0.743***	0.853***	0.840***	0.864***	0.847***	0.881***
Moran's I	0.161***	0.056***	0.071***	0.086***	0.039***	0.043***
N	5	14	15	14	15	24

Notes: p-value *<0.1, ** p < 0.05, *** p < 0.01; For the adjusted R^2, the hypothesis tested is whether the regression is significant, determined using the F-statistic; For Moran's I test on residuals, the hypothesis tested is whether Moran's I statistic equals 0; N refers to the number of explanatory variables.

As can be seen, the baseline model possesses an adjusted R^2 of 0.743. All expanded models have a higher adjusted R^2, confirming their explanatory power in determining property values. The order of explaining power is as follows: *Total* (0.881) > *Objective* (0.864) > *Social* (0.853) > *Subjective* (0.847) > *Safety* (0.84) > *Baseline* (0.743). These findings imply that social indices are better at explaining property values than safety indices and that objective measurements have greater explanatory value than subjective measurements. Besides, it also suggests that both measurements complement each other since the combination results in the highest variance explained.

Additionally, the significant values for Moran's I test again suggest using a model that accounts for spatial autocorrelation. Here, the values of Moran's I test statistic suggest the following order of models with highest spatial autocorrelation in the residuals: *Baseline* (0.161) > *Objective* (0.086) > *Safety* (0.071) > *Social* (0.056) > *Total* (0.043) > *Subjective* (0.039). The baseline model exhibiting the strongest spatial autocorrelation in the residuals is consistent with the results in Section 5.1. The assumption that including more independent variables captures more of the spatial processes is also largely supported by these findings.

5.3 Analysis of Fixed Effects Regression Models

After obtaining the results for the different linear regression models, FE models are computed for the same combinations of variables, with fixed effects on the neighborhood level. The fixed effects model controls for all time-invariant characteristics of the neighborhoods, thus isolating the within-neighborhood (over time) variation. This allows the model to focus on how changes within a neighborhood over time are related to the independent variables. In contrast, a simple linear regression model does not distinguish between within-neighborhood and between-neighborhood variations, potentially mixing up the two types of effects. The results can be found in Table 3.

Table 3 Performance of FE Models

FE	Baseline	Social	Safety	Objective	Subjective	Total
Adjusted R^2	0.873***	0.886***	0.890***	0.893***	0.883***	0.897***
Moran's I	-0.014	-0.014	-0.014	-0.014	-0.014	-0.014
N	5	14	15	14	15	24

Notes: p-value *<0.1, ** p < 0.05, *** p < 0.01; For the adjusted R^2, the hypothesis tested is whether the regression is significant, determined using the F-statistic; For Moran's I test on residuals, the hypothesis tested is whether Moran's I statistic equals 0; N refers to the number of explanatory variables.

Like the outcomes of the linear regression models, the variance explained, after controlling for the number of variables added, improves when social and safety factors are included in the FE model. This again confirms the importance of safety and social indices in describing housing values. The adjusted

R^2 of all models is higher compared to their linear regression model counterparts, indicating that the FE regression model is a better fit for this data than a simple linear regression. By effectively accounting for unique characteristics of each neighborhood that do not change over time, it provides a clearer picture of how the variables explain housing values over time within the same neighborhood.

However, the order of explanatory value changes with the different method used. The ranking is now: *Total* (0.897) > *Objective* (0.893) > *Safety* (0.890) > *Social* (0.886) > *Subjective* (0.883) > *Baseline* (0.873). This indicates that safety indices are more effective in explaining housing values than social indices. The order between objective and subjective measurements remains unchanged, while they also remain complementary since the FE model containing all measures yields the highest adjusted R^2.

Notably, the insignificant values for Moran's I test suggest that accounting for neighborhood-level fixed effects has successfully captured much of the spatial patterns in the residuals. However, despite the reduction in spatial autocorrelation, it is still beneficial to utilize a method which allows spatial differences to occur, to further refine the model. To address this, I utilize GWR, which allows for local variations in the relationships between the independent variables and the dependent variable by estimating separate coefficients for each location. This approach will help ensure that any remaining spatial correlation is adequately addressed, thereby improving the robustness and accuracy of the analysis.

5.4 Analysis of Geographically Weighted Regression Models

A GWR model is estimated for the same sets of previously used variables. Although GWR constructs neighborhood-specific linear regressions instead of a global model, an overall adjusted R^2 can still be calculated using the same formula as for linear regression and fixed effects models. This involves calculating the residual sum of squares (RSS) and total sum of squares (TSS) across all observations and adjusting for the number of predictors and observations. However, it is important to note that this measure should be interpreted cautiously due to the local nature of GWR. This time, Moran's I is not included in the table, since this model already accounts for spatial heterogeneity. The performance of these models is shown in Table 4.

Table 4 Performance of GWR Models

GWR	Baseline	Social	Safety	Objective	Subjective	Total
Adjusted R^2	0.914***	0.925***	0.908***	0.938***	0.863***	0.904***
N	5	14	15	14	15	24

Notes: p-value *<0.1, ** p < 0.05, *** p < 0.01; For the adjusted R^2, the hypothesis tested is whether the regression is significant, determined using the F-statistic; For Moran's I test on residuals, the hypothesis tested is whether Moran's I statistic equals 0; N refers to the number of explanatory variables.

The GWR models present varying results in the explained variance after accounting for the number of variables included. In the linear regression models and FE models, the addition of variables to the baseline models consistently results in a higher adjusted R^2, whereas the model with all variables possesses the strongest explanatory power. In GWR models, only the inclusion of either social attributes or objective indices results in higher adjusted R^2 values compared to the baseline model, whereas adding other variables does not increase the adjusted R^2.

The explanatory value ranking using GWR models is as follows: *Objective* (0.938) > *Social* (0.914) > *Baseline* (0.925) > *Safety* (0.908) > *Total* (0.904) > *Subjective* (0.863). Consistent with the FE models, these results indicate that objective measurements are more effective in explaining housing values than subjective measurements. However, the combination of objective and subjective indices does not appear to be very complementary in this context, as the adjusted R^2 decreases when subjective measurements are added to the objective measurements, which can be seen in the *Total* column. Furthermore, in the GWR model, social indices possess greater explanatory power than safety indices, which is the opposite of the result of FE models.

While adjusted R^2 provides a global measure of model fit, indicating how well the model explains the variability in the data across the entire study area; it does not capture local variations in the model's performance. The GWR model generates separate regression equations for each neighborhood, allowing for spatial heterogeneity in the relationships between the dependent and independent variables. Since the strength of GWR lies in capturing these spatial differences, I also consider local R^2 values, to gain insight into how well the model explains variability over time within each neighborhood, reflecting the localized fit of the model.

Local R^2 values are calculated for each neighborhood based on the fit of the GWR model within that specific area. Specifically, for each neighborhood, the R^2 value is determined by comparing the observed and predicted values of the dependent variable using the neighborhood-specific regression coefficients. These local R^2 values vary across neighborhoods, indicating areas where the model performs better or worse.

Figure 8 shows the local R^2 values of each neighborhood of the GWR model with all variables included. The grey neighborhoods are neighborhoods without any residents, for example, because they are part of the port. These neighborhoods are not included in the research. In the Appendix, Figure 14 shows the map of Rotterdam with the names of the neighborhoods included in this paper. As can be seen, the model shows varying performance across neighborhoods, but R^2 does not reach a level lower than 0.90 in any neighborhood. Strikingly, in Dorp/Rijnpoort, Rozenburg and Strand en Duin, a local R^2 of

1.00 is reached, which implies that the GWR model including all variables explains all the variance in the dependent variable, neighborhood-level average property value per square meter.

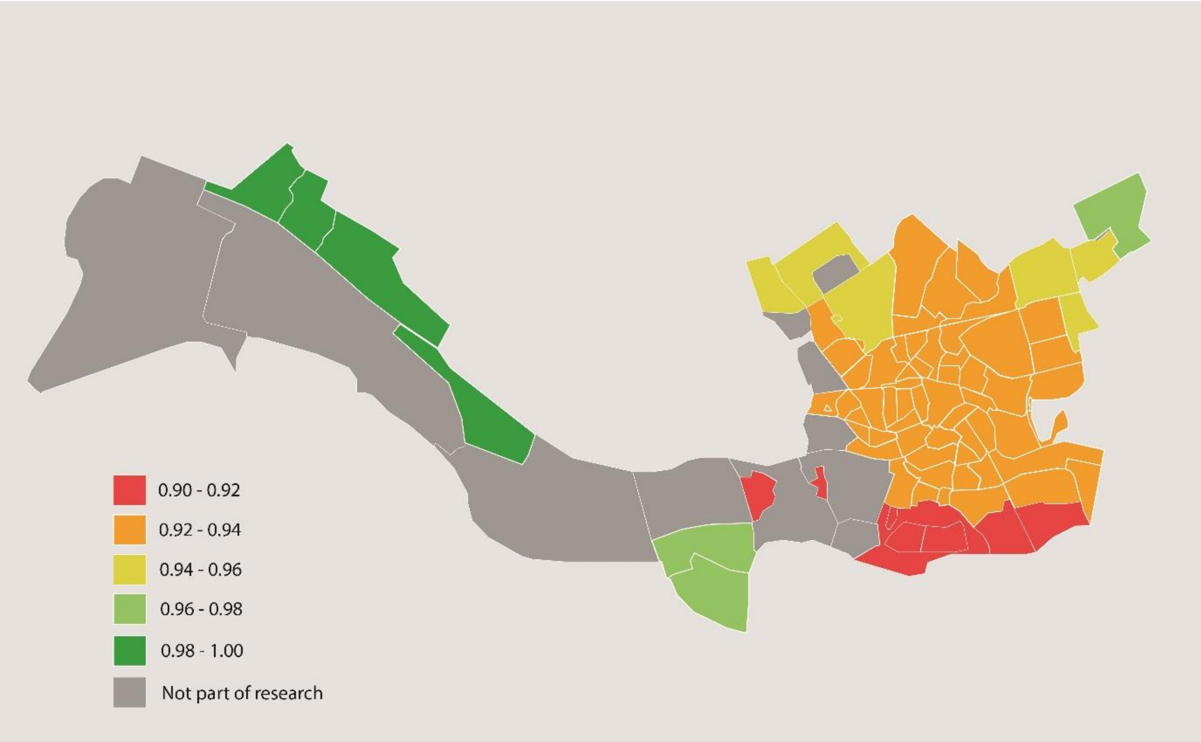

Figure 8 Local R^2 Across Neighborhoods

However, the perfect R^2 values in some neighborhoods could indicate potential overfitting. Overfitting occurs when a model is excessively complex, capturing noise in the data as if it were a true underlying pattern, which may result from having many variables but not enough observations. Given the high number of variables relative to the number of years and neighborhoods, this is a concern here. Overfitting can lead to an excellent fit in the considered data but poor generalizability to new, unseen data, making it inaccurate for predictions or conclusions.

5.5 Differences Across Neighborhoods

A GWR model is constructed to assess the varying associations between the various indices and property values across neighborhoods. The regression results of the GWR model with all variables are shown in Table 5. The actual interpretation of the coefficients of the indices will be performed in Section 5.6, this section focuses on the spatial differences in the relationships between neighborhood-level average property values and social and safety indices.

Table 5 Regression Results of GWR Model Total

	GWR Model total			
	Min	Median	Mean	Max

(Intercept)	-7967.747	-7173.190	-6872.835	-2524.743
Housing - objective	-8.924	13.383	11.275	14.816
Public Space - objective	-2.258	-0.531	0.146	7.124
Amenities - objective	-3.331	0.968	0.818	7.734
Environment - objective	-1.612	3.461	3.208	4.426
CPI	36.244	71.502	68.676	76.988
Self-reliance - subjective	-1.650	3.756	4.056	7.669
Co-reliance - subjective	-10.414	-4.622	-4.639	5.746
Bonding - subjective	-4.447	2.815	3.774	7.916
Participation - subjective	-8.505	-2.351	-2.701	-0.333
Burglary - subjective	-0.980	1.363	1.810	4.400
Vandalism - subjective	-3.157	-0.093	0.055	2.929
Nuisance - subjective	-3.204	-1.098	-0.017	3.515
Violence - subjective	-3.069	-0.550	0.062	3.052
Theft - subjective	-4.471	-1.048	-1.080	3.911
Self-reliance - objective	0.743	8.773	8.966	16.542
Co-reliance - objective	-6.230	-1.881	-1.809	3.370
Bonding - objective	-4.956	-0.510	-0.568	6.096
Participation - objective	-3.349	1.715	1.438	6.144
Burglary - objective	-8.097	-6.137	-6.011	-1.525
Vandalism - objective	-9.018	-3.351	-3.522	-0.665
Nuisance - objective	-2.503	-0.044	0.148	3.123
Violence - objective	-4.116	1.326	1.848	8.709
Theft - objective	-8.816	-0.952	-1.625	3.198
Judgement on quality of life	-3.610	-0.050	0.046	4.689

Notes: Different regressions were constructed for all neighborhoods in the GWR model. This table shows the minimum, median, and maximum coefficients across neighborhoods, as well as the mean coefficient across the neighborhoods.

The regression results confirm significant spatial differences in associations between the social and safety measurements and property values since there are – sometimes big – differences between the minimum and maximum values of the coefficients. Most coefficients have a negative minimum value and a positive maximum value, underscoring the varying relationships across neighborhoods. Only *CPI* and *Self-reliance – objective* have a positive coefficient in all neighborhoods, while *Participation – subjective, Burglary – objective* and *Vandalism – objective* have a negative coefficient in all neighborhoods. These negative coefficients are surprising, for example, this would mean that if there are more burglaries in a neighborhood, the predicted average property value in this neighborhood would decrease. The other variables can all have either a positive or negative coefficient, depending on the neighborhood. Further interpretation of the coefficients of variables will be done in Section 5.6.

To visualize these differences, the varying coefficients of three variables are shown on a map of Rotterdam. Figure 9 shows the coefficients of *Self-reliance – objective*; Figure 10 shows those of *Theft – subjective*, while Figure 11 maps out the coefficients of *Violence – objective*.

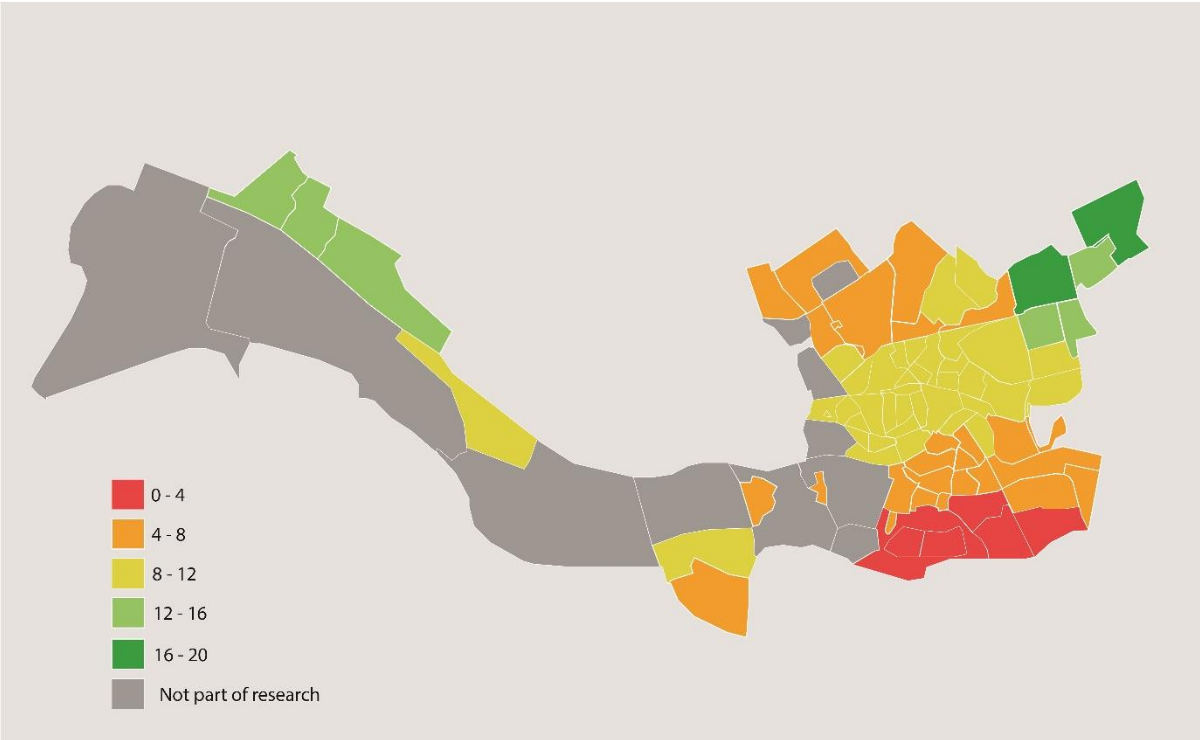

Figure 9 Self-reliance - objective Coefficients Across Rotterdam

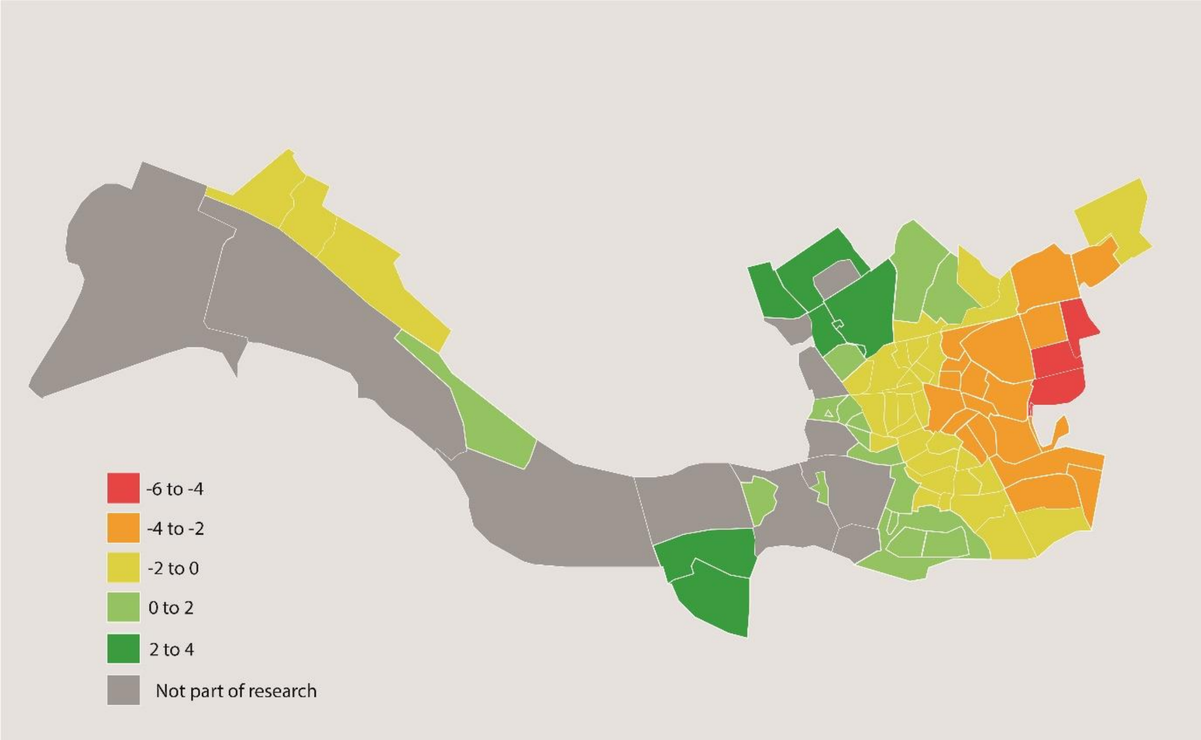

Figure 10 Theft - subjective Coefficients Across Rotterdam

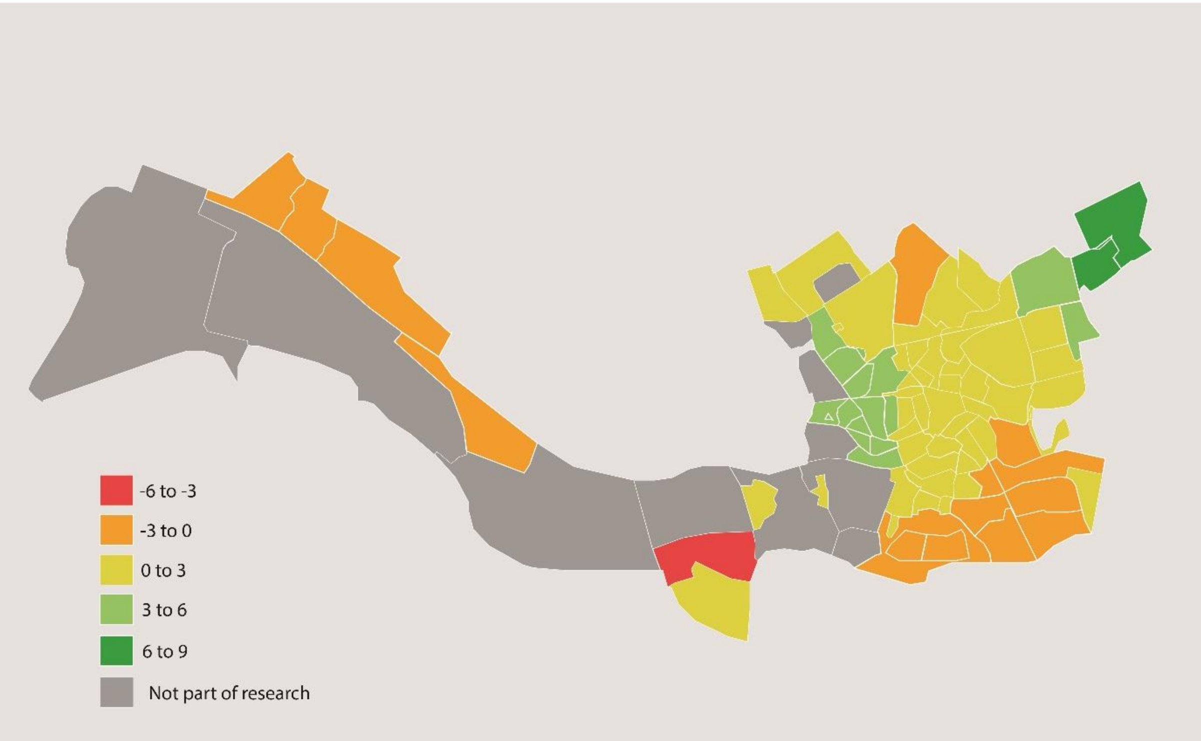

Figure 11 Violence - objective Coefficients Across Rotterdam

As stated in the methodology, GWR can assess whether the effect of CPI is additive, and account for the effect not being additive. I assume that the same inflation rate has a larger effect on property values in expensive neighborhoods compared to less expensive ones. Surprisingly, this is not the case everywhere. The coefficients of *CPI* across neighborhoods are visualized in Figure 12. For comparison, Figure 13 visualizes the mean of the neighborhood-level average property values per square meter over the period 2014-2024. The legend displays the values in euros.

In Figure 12, a large part of central Rotterdam is dark green, showing high *CPI* coefficients, which matches with the expensive houses in this area. Interestingly, even in the surrounding neighborhoods where houses are cheaper, the *CPI* coefficient is still high. However, the neighborhoods on the left side of the graph have the lowest *CPI* coefficients, even though their average property values are not the lowest. Additionally, the northern neighborhoods of Rotterdam have high average property values, but this is not completely reflected in their *CPI* coefficients.

This difference between property values and *CPI* coefficients can be attributed to variations in the relative increase in neighborhood-level average property values. The neighborhoods with lower *CPI* coefficients are also those where the relative increase in average property values has been relatively low.

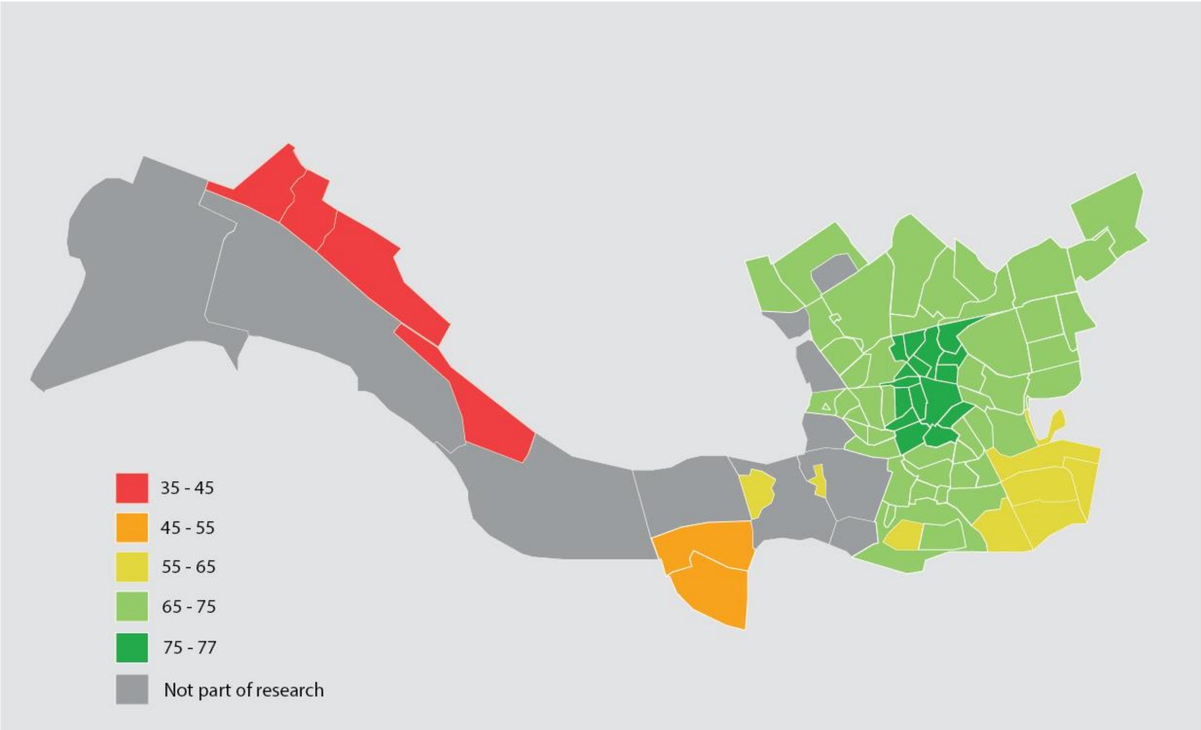

Figure 12 CPI Coefficients Across Rotterdam

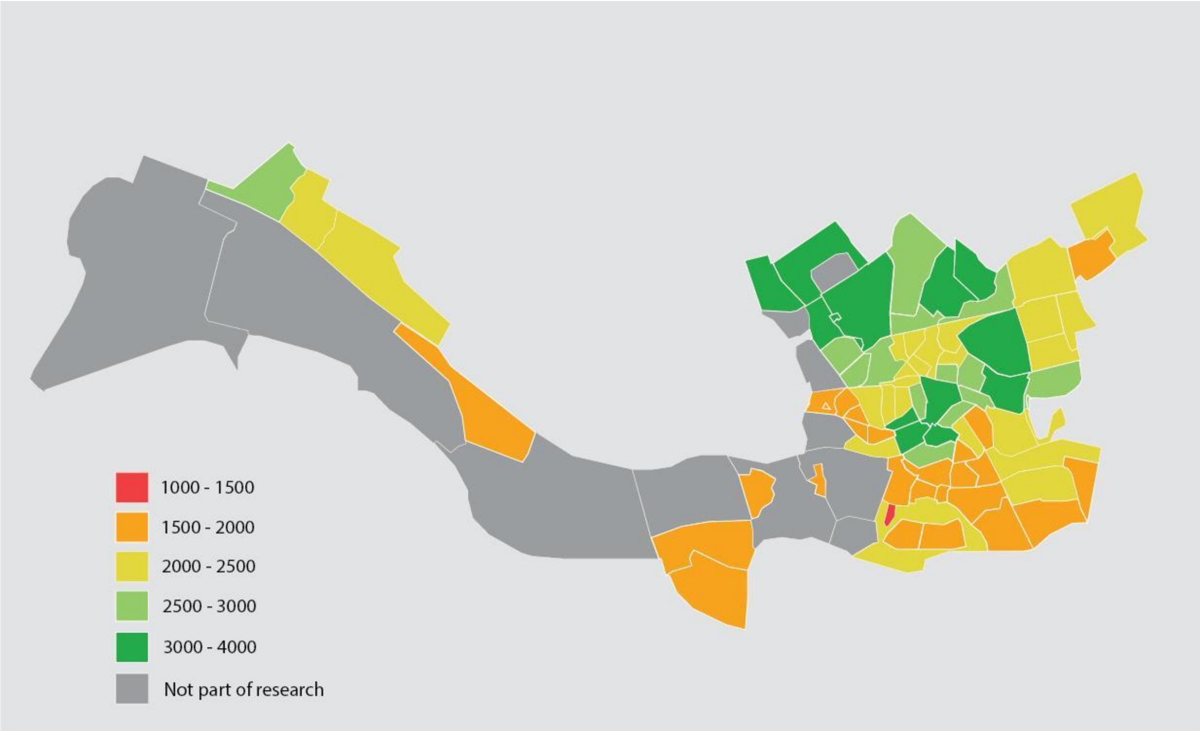

Figure 13 Mean Neighborhood-level Average Property Value Across Rotterdam

5.6 Interpretation of Regression Results

To interpret the explanatory value of specific indicators for property values, I decide to analyze the FE model that includes all variables. Out of all the linear regression and FE models, the FE model including all indices is the model that reaches the highest adjusted R^2, and it is also best suited for comparison, as all variables are included in this regression.

Including fixed effects controls for all time-invariant characteristics of neighborhoods, isolating the within-neighborhood variation. This allows the model to focus on changes within each neighborhood over time, rather than differences between neighborhoods. Fixed effects account for unique, constant factors such as historical significance and cultural identity, ensuring more accurate estimates of the variables' relationships with property values. This makes the FE model reliable for interpreting the explanatory value of specific indicators.

When comparing this to the mean results of the GWR model, it is important to note that the GWR model provides localized estimates that vary across neighborhoods, capturing spatial heterogeneity. The mean results of the GWR model offer a summary of these local estimates, reflecting the average relationship between variables and property values across different neighborhoods. However, an average might be inaccurate if some neighborhoods have very high or low coefficients.

While the FE model focuses on within-neighborhood changes over time, the GWR model highlights spatial variations and local patterns. Therefore, the FE model is better suited for understanding temporal dynamics within neighborhoods, whereas the GWR model is more appropriate for capturing spatial diversity and identifying areas with distinct relationships between the variables and property values. The results of the FE model can be found in Table 6.

Table 6 Regression Results of FE Model Total

	Fe Model Total	
	Coefficients	Standard error
Housing - objective	8.823***	(2.008)
Public Space - objective	-1.363	(0.863)
Amenities - objective	8.500**	(3.318)
Environment - objective	4.684***	(1.146)
CPI	63.679***	(3.498)
Self-reliance – subjective	1.764	(1.585)
Co-reliance – subjective	-0.075	(2.128)
Bonding - subjective	-2.906	(1.887)
Participation - subjective	-2.552***	(0.834)
Burglary - subjective	0.847	(0.857)
Vandalism - subjective	-1.149	(1.067)
Nuisance - subjective	-0.020	(1.055)
Violence - subjective	-0.353	(0.926)
Theft - subjective	0.492	(1.141)
Self-reliance - objective	11.503***	(3.805)
Co-reliance - objective	-0.733	(1.208)
Bonding - objective	1.004	(2.072)
Participation - objective	-3.607	(2.386)
Burglary - objective	-3.907***	(0.893)
Vandalism - objective	-4.047***	(1.039)
Nuisance - objective	0.630	(1.327)
Violence - objective	2.092	(1.661)
Theft - objective	0.572	(1.634)
Judgment on quality of life	0.380	(0.792)
Observations	426	
R^2	0.920	

Adjusted R^2	0.897
F Statistic	158.228***

Notes: p-value *<0.1, ** p < 0.05, *** p < 0.01; The hypothesis tested is whether the coefficient equals 0; For the F statistic, the hypothesis tested is whether the regression is significant.

As can be seen, among the variables analyzed, only seven exhibit coefficients significant at the 1% level. These are *Housing—objective, Environment—objective, CPI, Participation—subjective, Self-reliance—objective, Burglary—objective,* and *Vandalism—objective.* Additionally, the coefficient of *Amenities—objective* is significant at the 5% level, while the remaining sixteen variables are insignificant.

Since this research assesses the explanatory value of social and safety indices, interpretations will focus on their coefficients. *Self-reliance—objective* has the largest coefficient (11.503), and notably, it is also the only positive significant coefficient. This is consistent with the GWR model in Table 5, in which *Self-reliance—objective* is the only variable that has a positive coefficient across all neighborhoods, besides *CPI.*

The other significant variables, *Vandalism—objective, Burglary—objective,* and *Participation—subjective,* have coefficients of -4.047, -3.907, and -2.552, respectively. Again, this is consistent with the GWR model in Table 5, as these variables are the only variables that exhibit a negative coefficient across all neighborhoods. Given that many indicators of these three indices are typically positively correlated with property values in a neighborhood, and the indices themselves are positively correlated with property values in the data being utilized in this research, the regression result is unexpected. An empirically supported explanation for these significant negative coefficients has not been found, though potential reasons will be mentioned in the discussion, Section 7.3.

Given that many coefficients in the FE model containing all variables are not significant and some significant coefficients are in a surprising direction, I construct a new model from this FE model. Using an iterative approach, I refine the model by systematically removing the least significant variable. The process involves fitting the FE model including all safety and social indices, and then identifying the variable with the highest p-value. If this p-value exceeds a threshold of 5%, I remove the variable from the model. I repeat this procedure, refitting the model after each removal, until all remaining variables are significant at the 5% level. This iterative approach ensures that the final model includes only variables that have a statistically significant relationship with property values. The results of this model can be found in Table 7.

Table 7 Regression Results of Significant FE Model

	Significant Fe Model	
	Coefficients	Standard error
Housing - objective	8.851***	(1.747)
Amenities - objective	8.227**	(3.275)
Environment - objective	4.635***	(0.899)
CPI	64.433***	(2.388)
Participation - subjective	-3.060***	(0.734)
Self-reliance - objective	14.043***	(3.436)
Participation - objective	-4.401**	(1.891)
Burglary - objective	-3.427***	(0.774)
Vandalism - objective	-4.867***	(0.949)
Observations	426	
R^2	0.916	
Adjusted R^2	0.897	
F Statistic	421.697***	

Notes: p-value *<0.1, ** p < 0.05, *** p < 0.01; the hypothesis tested is whether the coefficient equals 0; For the F statistic, the hypothesis tested is whether the regression is significant.

As can be observed, the variables that are significant in Table 6 are also included in the Significant FE Model. Notably, only one variable that does not exhibit a significant coefficient in FE Model Total is included in Table 7, which is *Participation – objective*. The only social or safety index in this model with a positive coefficient is *Self-reliance – objective*. Given that the signs of the coefficients remain consistent with those in Table 6, and their magnitudes show minimal changes, the interpretation will be based on FE Model Total.

In FE Model Total, the order for social indices, from strong positive to strong negative coefficient, is as follows: *Self-reliance – objective* (11.503) > *Self-reliance – subjective* (1.764) > *Bonding – objective* (1.004) > *Judgement on quality of life* (0.380) > *Co-reliance – subjective* (-0.075) > *Co-reliance - objective* (-0.733) > *Participation - subjective* (-2.552) > *Bonding – subjective* (-2.906) > *Participation – objective* (-3.607). This indicates that an increase in the index score would lead to a higher predicted value of neighborhood-level average property value for only four out of nine variables.

The ranking of coefficients of safety indices from positive to negative contribution is as follows: *Violence - objective* (2.092) > *Burglary - subjective* (0.847) > *Nuisance - objective* (0.630) > *Theft - objective* (0.572) > *Theft - subjective* (0.492) > *Nuisance - subjective* (-0.020) > *Violence - subjective* (-

0.353) > *Vandalism - subjective* (-1.149) > *Burglary - objective* (-3.907) > *Vandalism - objective* (-4.047). This suggests that for half of the safety sub-indices, an increase in the index score indicates a higher predicted neighborhood-level average property value, while an increase in the index score of the other determinants would predict a lower neighborhood-level average housing value.

The total value of each sub-index is obtained by summing the coefficients for the objective and subjective measurements. For Judgment on quality of life, the coefficient remains unchanged as it only consists of a subjective measurement. The aggregate scores are ranked as follows: *Self-reliance* (13.267) > *Violence* (1.739) > *Theft* (1.064) > *Nuisance* (0.610) > *Judgement on quality of life* (0.380) > *Co-reliance* (-0.808) > *Bonding* (-1.902) > *Burglary* (-3.060) > *Vandalism* (-5.196) > *Participation* (-6.159). The significances are tested by examining the combined effect of both subjective and objective measurements for each indicator. The analysis involves checking whether the combined influence of these measurements on the dependent variable is statistically different from 0. Out of these values, *Self-reliance, Burglary,* and *Vandalism* are significant at the 1% level, while *Participation* is significant at the 5% level. This indicates that *Self-reliance* is the strongest predictor, as its coefficient is of the largest magnitude. Moreover, only half of the aggregate coefficients are positive. *Participation* emerges as the strongest negative predictor.

The sum of coefficients for all social indices totals 4.778, whereas the aggregated sum for all safety indices is -4.843. The significance is again tested by checking whether the combined influence of either social or safety factors is statistically different from 0. Both sums of coefficients are insignificant, even at the 10% level. The sum of coefficients for all social indices has a p-value of 0.250, and the sum of coefficients for all safety indices has a p-value of 0.100.

6. Conclusion

In this paper, I examined the explanatory value of social and safety indices, measured both objectively and subjectively, on property values in Rotterdam. The research was conducted on the neighborhood level; the social indices, the safety indices, and the property level were all measured as the average of a neighborhood. Neighborhood-level average WOZ-waarde per square meter was used as the value of a property. The WOZ-waarde was used to ensure uniformity in the measurement of property values, and this value is measured per square meter to control for the size of a house. A hedonic pricing model (HPM) was constructed, including information about the houses, public space, amenities, and the environment in a neighborhood. To account for the huge inflation, consumer price index was added to the HPM. Next, to assess the explanatory value of the various indices, extended models of the baseline HPM were created, each containing a specific set of variables. Specifically, one model incorporated social indices, another included safety indices, a third comprised all objective measures, a fourth encompassed all subjective measures, and finally, a comprehensive model was created that included all variables. All these models were constructed three times, each with a different estimation method. First using linear regression, afterward through a fixed effects regression (FE), and lastly, using the geographically weighted regression (GWR) method. All models were compared based on adjusted R^2, to measure the variance explained while controlling for the number of predictors in the model. Several conclusions were obtained from this research. Since the FE and GWR models outperformed all linear regression models, conclusions are based on the results of the FE model and GWR model.

The main research question in this study was: *"What is the explanatory value of safety and social indices, measured both objective and subjective, on property values in Rotterdam?".* To begin with, all models that included social and safety factors exhibited a higher adjusted R^2 compared to the baseline HPM on the global level, underscoring their significance in explaining property values. Consequently, I conclude that social and safety indices, measured both objective and subjective, possess significant explanatory power.

First, the analysis sought to determine whether social indices or safety indices better explain the value of houses in Rotterdam. The findings indicate a context-dependent conclusion: Globally, safety indices provide a better explanation, whereas, when accounting for spatial differences, social indices possess more explanatory power. This is based on the safety indices outperforming the social indices in the global (FE) model. At the same time, this order was reversed in the GWR model, which models the local relationships between the predictors and property values.

Secondly, I conclude that objective measures of both indices possess stronger explanatory value than subjective measures since the models that included objective measures outperformed those that included subjective measures in all methods. To what extent these measures complement each other in explaining property values is, again, context dependent. Globally, they are highly complementary, as the FE model containing all indices reached the highest adjusted R^2. In contrast, in the GWR model, the objective indices reach a higher adjusted R^2 without the addition of subjective indices. This suggests that, when accounting for spatial variability, subjective measures do not contribute enough additional explanatory value to increase the explained variance, considering the number of variables added.

Next, the GWR model has shown significant differences in associations between social and safety indices and property values across different neighborhoods in Rotterdam. The GWR model with all objective indices obtained the highest adjusted R^2 out of all models included in this research, with a value of 0.938. In fact, all GWR models showed better performance than the FE models, except for the model that included subjective indices. This indicates that accounting for spatial differences increases the explanatory power of social and safety indices. In the GWR model with all variables, besides *Self-reliance – objective, Participation – subjective, Burglary – objective,* and *Vandalism – objective,* all indices could have either a positive or negative coefficient, depending on the neighborhood. This again confirms spatial differences in the relationship between the safety and social indices and property values. For example, *Violence – objective* only exhibits a positive coefficient in the center, north, and west of Rotterdam, suggesting that fewer cases of violence in these neighborhoods are associated with higher property values. Conversely, in the south and far west of Rotterdam, the coefficient for *Violence – objective* is negative, indicating that fewer cases of violence in these areas would result in a lower predicted neighborhood-level average property value.

Lastly, after analyzing the FE model containing all indices, I conclude that big differences exist between the relations of various factors and property values. The strongest positive predictor of property values is *Self-reliance – objective*. This is quite logical since this contains information on, among others, employment, educational certificates, and debt. However, it also contains less intuitive factors, such as contact with neighbors, or information on events visited. More surprisingly, the significant negative predictors were *Burglary – objective, Vandalism – objective,* and *Participation – subjective*. These three indices hold factors that are usually positively correlated with the value of properties in a neighborhood. For example, this result suggests that when there are more burglaries or acts of vandalism in a neighborhood, the predicted neighborhood-level average property value would decrease. *Participation – subjective* includes measures on themes like satisfaction with neighborhood participation or whether a person has felt discriminated in their neighborhood. So, this result also indicates that when more people in a neighborhood have experienced discrimination, this is a positive predictor for the neighborhood-level average property value. A possible explanation for these unexpected results is given in Section 7.3.

7. Discussion

The discussion is divided into three parts. Section 7.1 summarizes the practical implications of this research, section 7.2 denotes the scientifical contribution, and 7.3 discusses limitations of this research while also making suggestions for future research.

7.1 Practical Implications

This paper highlights the value that social and safety indices possess in determining the value of a property. These insights can inform strategic decisions across various sectors. By understanding and leveraging the explanatory power of social and safety indices, stakeholders can better navigate the complexities of the real estate market and contribute to the development of more desirable and valuable neighborhoods. Since causal relationships were not examined nor stated in this research, implications should act as guidance rather than as direct prescriptions for policy or investment decisions.

This research found that social indices were positive predictors of neighborhood-level average property values. Real estate investors, armed with these insights, can identify undervalued neighborhoods based on social indices. By strategically targeting neighborhoods with low social indices and investing in social infrastructure, investors can enhance neighborhood attractiveness and achieve higher returns.

The sub-index *Self-reliance – objective* is the strongest positive predictor of neighborhood-level average property values. Homeowners can benefit from this insight by improving factors related to this sub-index. Logically, it is hard to improve the percentage of people with a higher education certificate or decrease the unemployment rate. Still, *Self-reliance - objective* also contains information on factors that are more feasible to improve. Factors such as social contact between neighbors, other people in the neighborhood, or friends can be improved. This enhancement is also possible for other determinants like visiting a sports club or attending religious or cultural events. By fostering these improvements, for example, through homeowner associations, they can potentially increase the value of their properties.

Urban planners and policymakers can create targeted strategies to improve neighborhood conditions by focusing on enhancing social infrastructure and addressing safety concerns where necessary. This research, utilizing GWR, identifies precisely which factors would increase property values in specific neighborhoods. GWR shows the sub-indices that form the strongest predictors per neighborhood, indicating what should be improved to create value in this neighborhood, and making the neighborhood better, which is a goal for the municipality.

For instance, in the far west of Rotterdam, *Self-reliance – objective* has a very high coefficient, while the coefficient of *Violence – objective* is negative. Indicating that, to create value, the policies should be aimed at improving factors inside the *Self-reliance objective* index, such as employment ratio, educational attainment, and social contact between neighbors. At the same time, improving safety to decrease the number of acts violence would not necessarily result in a higher predicted neighborhood-level average property value in these neighborhoods.

7.2 Scientific Contributions

Extensive research has been conducted on property values and their determinants, often focusing on either safety factors or social indices. However, studies examining the relationship between property values and a combination of objective and subjective measures are less common. This research fills multiple gaps in the literature by exploring these multifaceted relationships.

Before this study, the combined impact of safety and social indices on housing values had not been thoroughly investigated. This research concludes that, on a global scale, safety attributes are more effective in explaining property values. At the same time, social indicators provide better explanations at the local level when spatial differences are considered. This finding highlights the importance of context and scale in understanding the determinants of property values.

Few studies have incorporated objective and subjective safety measures in relation to property values, and, to my knowledge, none have done so for social factors, let alone in combination with safety indices. This research concludes that objective measures of safety and social indices possess greater explanatory value for housing values than their subjective counterparts. This underscores the reliability of objective data in real estate analysis while also acknowledging the supplementary role of subjective perceptions.

While some research has examined the combination of objective and subjective data for safety concerning property values, they have yet to address to what extent these measures complement each other. This research reveals that objective and subjective measures are highly complementary at the global level, enhancing the explanatory power of the models. However, subjective measures do not add much explanatory value when accounting for spatial heterogeneity, indicating that the relationship between objective and subjective data can vary significantly based on the analytical approach.

7.3 Limitations & Further Research

This research possesses several limitations regarding the data, the external validity, the methods, and the interpretation.

First, all data in this study consist of neighborhood-level averages. While neighborhood-level averages provide a useful representation of individual data, significant variations within neighborhoods could change the outcomes of this research. For instance, in the traditional HPM, structural attributes typically include building characteristics such as the number of bedrooms, the number of bathrooms, and the presence of a yard. In this research, *Housing – objective* was used as a structural attribute, which includes factors like the percentage of small one-person houses, the percentage of well-maintained houses, and the average number of days between the listing and sale of a house. Although this is a reasonable substitute at the neighborhood level, individual property data might yield different results. Therefore, future research should consider using data on individual properties to achieve more precise outcomes.

Also, the large number explanatory variables compared to the number of observations in this study increases the risk of overfitting. Overfitting happens when a model becomes too complex and starts to capture random noise instead of just the actual patterns in the data. This can result in coefficients that look important in this dataset but do not hold up when applied to new data. While an iterative process was used to remove insignificant variables, future research could use other methods to prevent overfitting, such as regularization techniques like LASSO or Ridge regression. Additionally and even better, increasing the sample size by extending the study over a longer period of time can help make the model more reliable and generalizable.

Second, this research was conducted in Rotterdam, the Netherlands, a city with unique characteristics. There are no guarantees that social and safety indices will have the same explanatory value in other cities. To obtain more robust and generalizable results, it is recommended to extend this research to multiple cities across various countries and continents, incorporating diverse city types.

Third, the research identified spatial heterogeneity in the relationship between property values and social and safety indices. To account for this, a GWR model was constructed, outperforming the global regression methods. However, other spatially explicit models could potentially handle the varying relationships between neighborhoods even better. Future research should incorporate multiple spatially explicit models to investigate whether alternative techniques perform even better.

Lastly, this study assessed the explanatory value of social and safety indices, measured both objectively and subjectively, on housing values. While clear results were obtained regarding their explanatory power, interpreting the magnitude and direction of various indices' predictive values is challenging. For example, the study considered objective and subjective indices separately, making it difficult to estimate the combined effect of simultaneous changes in both. Future research could address this by including interaction terms in the analysis, providing a more nuanced understanding of how different indices interact and contribute to property value changes.

The unexpected significant negative coefficients for indices like *Participation - objective, Burglary - objective,* and *Vandalism - objective* could be due to endogeneity. Endogeneity occurs when an explanatory variable is correlated with the error term, possibly due to omitted variables, reverse causality, or simultaneity. I suspect that endogeneity, caused by omitted variable bias, might be influencing the results in this study, especially considering that all three mentioned variables are positively correlated with the neighborhood-level average property values. There may be other important factors influencing property values that are not included in the model. If these omitted variables are correlated with the social and safety indices and property values, they could bias the estimated coefficients. In the case of burglaries and vandalism, the presence of police officers, or neighborhood-specific interventions could influence both the incidence of crime and property values, leading to biased estimates if not properly controlled for. Although neighborhood fixed effects were used in the model to account for time-invariant characteristics of neighborhoods, this approach may not fully capture all relevant time-varying factors, suggesting that omitted variable bias could still be affecting the results.

Addressing endogeneity by employing instrumental variables could provide a more accurate estimate of the true effect of these indices on property values. An instrumental variable approach would involve finding instruments that are correlated with the participation, burglary, and vandalism indices but uncorrelated with the error term, helping to isolate the descriptive effect.

References

Anderson, S. T., & West, S. E. (2006). Open space, residential property values, and spatial context. *Regional Science And Urban Economics, 36(6),* 773–789.

Bishop, K. C., Kuminoff, N. V., Banzhaf, H. S., Boyle, K. J., Von Gravenitz, K., Pope, J. C., Smith, V. K., & Timmins, C. D. (2020). Best Practices for Using Hedonic pricing Value Models to Measure Willingness to Pay for Environmental Quality. *Review Of Environmental Economics And Policy, 14(2),* 260–281.

Blair, J. P., & Larsen, J. E. (2010). Satisfaction with neighbors and neighborhood housing prices. *Journal Of Place Management And Development, 3(3),* 194–204.

Brady, M., & Irwin, E. (2011). Accounting for Spatial Effects in Economic Models of Land Use: Recent Developments and Challenges Ahead. *Environmental & Resource Economics, 48(3)*, 487–509.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis, 28(4),* 281–298.

Buonanno, P., Montolio, D., & Raya-Vílchez, J. M. (2012). Housing prices and crime perception. *Empirical Economics, 45(1),* 305–321.

Can, A. (1990). The Measurement of Neighborhood Dynamics in Urban House Prices. *Economic Geography, 66(3),* 254-272.

Coffee, N. T., Lockwood, T., Hugo, G., Paquet, C., Howard, N. J., & Daniel, M. (2013). Relative residential property value as a socio-economic status indicator for health research. *International Journal Of Health Geographics, 12(1),* 22.

Court, A. (1939). Hedonic Price Indexes with Automotive Examples, *The dynamics of automobile demand,* 99-117.

Fu, S., Liao, Y., & Zhang, J. (2016). The effect of housing wealth on labor force participation: Evidence from China. *Journal Of Housing Economics, 33,* 59–69.

Gibbons, S. (2004). The Costs of Urban Property Crime. *The Economic Journal, 114(499),* 441–463.

Goodman, A. C., & Thibodeau, T. G. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal Of Housing Economics, 12(3),* 181–201.

Griliches, Z. (1961). Hedonic Price Indexes for Automobiles: An Econometric of Quality Change. *National Bureau of Economic Research,* 173–196.

Groot, S., & Vrieselaar, N. (2024, march 12). Kwartaalbericht Woningmarkt: hoge prijsstijgingen in het verschiet. *Rabobank*. https://www.rabobank.nl/kennis/d011418749-kwartaalbericht-woningmarkt-hoge-prijsstijgingen-in-het-verschiet

Han, J., Chu, X., Song, H., & Li, Y. (2014). Social Capital, Socioeconomic Status and Self-efficacy. *Applied Economics And Finance*, 2*(1).*

Helbich, M., Brunauer, W., Vaz, E., & Nijkamp, P. (2013). Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria. *Urban Studies, 51(2),* 390–411.

Ihlanfeldt, K., & Mayock, T. (2010). Panel data estimates of the effects of different types of crime on housing prices. *Regional Science And Urban Economics*, *40(2–3),* 161–172.

Jaffe, A. J. (1980). Real Estate as if Management Mattered. *California Management Review, 22(3),* 92–96.

Johnson, W. R. (2014). House prices and female labor force participation. *Journal Of Urban Economics*, 82, 1–11.

Kortas, F., Grigoriev, A., & Piccillo, G. (2022). Exploring multi-scale variability in hotspot mapping: A case study on housing prices and crime occurrences in Heerlen. *Cities, 128,* 103814.

Li, H., Wei, Y. D., Wu, Y., & Tian, G. (2019). Analyzing housing prices in Shanghai with open data: Amenity, accessibility and urban structure. *Cities, 91,* 165–179.

Li, L., Bao, H. X. H., & Chau, K. W. (2022). On the Strategic Timing of Sales by Real Estate Developers: To Wait or To Presell? *Journal Of Real Estate Finance And Economics, 66(1),* 169–196.

Lee, C., & Moudon, A. V. (2006). Correlates of Walking for Transportation or Recreation Purposes. *Journal Of Physical Activity & Health, 3(s1),* 77–98.

Ministerie van Algemene Zaken. (2023, 20 april). Hoe bepalen gemeenten de WOZ-waarde? *Rijksoverheid.* https://www.rijksoverheid.nl/onderwerpen/waardering-onroerende-zaken-woz/vraag-en-antwoord/woz-waarde-bepalen

Mouissie, S., & Kraniotis, L. (2023, november 11). Schreeuwend tekort aan woningen en hoge huizenprijzen: hoe is het zo gekomen? *NOS*. https://nos.nl/collectie/13960/artikel/2497415-schreeuwend-tekort-aan-woningen-en-hoge-huizenprijzen-hoe-is-het-zo-gekomen

Nistor, A. P., & Reianu, D. (2018). Determinants of housing prices: evidence from Ontario cities, 2001-2011. *International Journal Of Housing Markets And Analysis (Print), 11(3),* 541–556.

Nyunt, M. S. Z., Shuvo, F. K., Eng, J. Y., Yap, K. B., Scherer, S., Hee, L. M., Chan, S. P., & Ng, T. P. (2015). Objective and subjective measures of neighborhood environment (NE): relationships with transportation physical activity among older persons. *The International Journal Of Behavioural Nutrition And Physical Activity, 12(1).*

Pakes, A. (2003). A Reconsideration of Hedonic Price Indexes with an Application to PC's. *The American Economic Review, 93(5),* 1578–1596.

Poor, P. J., Boyle, K. J., Taylor, L. O., & Bouchard, R. (2001). Objective versus Subjective Measures of Water Clarity in Hedonic pricing Value Models. *Land Economics, 77(4),* 482–493.

Pope, J. C. (2008b). Fear of Crime and Housing Prices: Household Reactions to Sex Offender Registries. *Journal Of Urban Economics, 64(3),* 601–604.

Privacy. (2023, March 3). Wijkprofiel Rotterdam. https://wijkprofiel.rotterdam.nl/nl/2024/privacy

Qiu, W., Zhang, Z., Liu, X., Li, W., Li, X., Xu, X., & Huang, X. (2022). Subjective or objective measures of street environment, which are more effective in explaining housing prices? *Landscape And Urban Planning, 221,* 104358.

Qiu, W., Li, W., Liu, X., Zhang, Z., Li, X., & Huang, X. (2023). Subjective and objective measures of streetscape perceptions: Relationships with property value in Shanghai. *Cities, 132,* 104037.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal Of Political Economy, 82(1),* 34–55.

Sun, W., Huang, Y., Spahr, R. W., Sunderman, M. A., & Sun, M. (2019). Neighborhood Blight Indices, Impacts on Property Values and Blight Resolution Alternatives. *The Journal Of Real Estate Research, 41(4),* 555–604.

The Economist Newspaper. (2024, March 25). A new urbanism: Building more liveable and resilient cities. *Economist Impact.* https://impact.economist.com/perspectives/infrastructure-cities/new-urbanism-building-more-liveable-and-resilient-cities

Tita, G. E., Petras, T. L., & Greenbaum, R. T. (2006). Crime and Residential Choice: A Neighborhood Level Analysis of the Impact of Crime on Housing Prices. *Journal Of Quantitative Criminology, 22(4),* 299–317.

Tu, Y., Li, P., & Qiu, L. (2016). Housing search and housing choice in urban China. *Urban Studies, 54(8),* 1851–1866.

Uphoff, E. P., Pickett, K. E., Cabieses, B., Small, N., & Wright, J. (2013). A systematic review of the relationships between social capital and socioeconomic inequalities in health: a contribution to understanding the psychosocial pathway of health inequalities. *International Journal For Equity in Health, 12(1),* 54.

Wijkprofiel Rotterdam. (n.d.). https://wijkprofiel.rotterdam.nl/nl/2022/rotterdam

Wijkraden. (n.d.). Gemeente Rotterdam. https://www.rotterdam.nl/wijkraden

Zietz, J., Zietz, E. N., & Sirmans, G. S. (2007). Determinants of House Prices: A Quantile Regression Approach. *Journal Of Real Estate Finance And Economics, 37(4),* 317–333.

Appendix

Table 5 Summary Statistics

Variable	N	Mean	SD	Min	Max
Safety Index	426	111.514	21.156	60.423	164.691
Safety Index - subjective	426	110.632	27.607	53.691	176.510
Safety Index - objective	426	112.397	20.688	37.901	160.258
Perception of safety	426	112.575	40.550	33.453	179.069
Theft - subjective	426	109.324	25.666	48.247	182.685
Violence - subjective	426	109.785	31.404	31.428	188.468
Burglary - subjective	426	125.613	34.103	19.815	196.220
Vandalism - subjective	426	106.552	29.247	37.182	179.770
Nuisance - subjective	426	99.942	38.800	24.820	197.220
Theft - objective	426	110.852	25.787	20.592	160.522
Violence - objective	426	114.485	24.276	14.883	157.765
Burglary - objective	426	129.055	28.057	8.828	186.973
Vandalism - objective	426	106.028	25.413	40.242	162.019
Nuisance - objective	426	101.565	31.512	9.358	179.588
Social Index	426	105.742	18.827	66.087	149.347
Social Index - subjective	426	104.710	25.642	48.537	163.092
Social Index - objective	426	106.774	15.145	75.534	146.312
Judgment on quality of life	426	104.354	41.587	19.597	180.019
Self-reliance - subjective	426	107.283	28.809	41.061	176.125
Co-reliance - subjective	426	102.448	20.119	47.433	173.258
Participation - subjective	426	92.466	32.386	19.046	170.607
Bonding - subjective	426	116.998	26.488	54.372	174.160
Self-reliance - objective	426	104.193	21.868	45.744	157.607

Co-reliance - objective	426	114.637	20.994	68.341	176.550
Participation - objective	426	110.562	21.282	64.586	165.986
Bonding - objective	426	97.704	28.181	16.605	156.241
Living Experience	426	109.161	40.469	19.260	186.454
Housing - subjective	426	103.806	34.698	27.295	178.168
Public Space - subjective	426	95.831	23.279	37.463	145.590
Amenities - subjective	426	99.716	17.652	41.097	133.953
Environment - subjective	426	95.938	20.968	44.572	153.790
Housing - objective	426	111.287	19.852	57.888	157.054
Public Space - objective	426	106.214	19.540	3.525	160.832
Amenities - objective	426	95.305	18.175	40.008	129.796
Environment - objective	426	115.281	32.211	18.951	185.533
Physical Index	426	103.956	13.552	74.811	136.587
Physical Index - subjective	426	100.890	21.699	51.557	149.495
Physical Index - objective	426	107.022	11.714	71.482	135.490
Property Value	426	2290.059	928.047	1039.147	5975.670
CPI	426	108.710	9.217	100.000	127.160

Notes: This table shows the summary statistics of all variables on the neighborhood-level.

Figure 14 Map Containing Names of Neighborhoods