Band 109

Bastian Krämer

Enhancing Residential Property Valuation: A Journey through Machine Learning and Explainable Artificial Intelligence

International Real Estate Business School
Universität Regensburg

Schriften zu Immobilienökonomie und Immobilienrecht

Bastian Krämer

Enhancing Residential Property Valuation: A Journey through Machine Learning and Explainable Artificial Intelligence

Die Deutsche Bibliothek – CIP Einheitsaufnahme
Bastian Krämer
Enhancing Residential Property Valuation: A Journey through Machine Learning and Explainable Artificial Intelligence
Regensburg: Universitätsbibliothek Regensburg 2023
(Schriften zu Immobilienökonomie und Immobilienrecht; Bd. 109)
Zugl.: Regensburg, Univ. Regensburg, Diss., 2023
978-3-88246-496-2

978-3-88246-496-2

Verlag: Universitätsbibliothek Regensburg, Regensburg 2023
Zugleich: Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften, eingereicht an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg
Tag der mündlichen Prüfung: 14. Dezember 2023
Berichterstatter: Prof. Dr. Wolfgang Schäfers
Prof. Dr. Daniel Rösch

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivation and Background

Central to economic development, urban growth, and wealth creation, the real estate market serves as a key component in the global economy. As urban environments expand in response to surging population figures, the current real estate stock faces increasing pressure to evolve. At its core, the residential real estate market holds a substantial position, primarily attributed to its extensive financial value and significant impact on households worldwide (Nanda, 2019; ULI, 2020; IWF, 2023). This significance, however, is not only limited to individual homeowners or investors. It goes beyond that, impacting macroeconomic stability and consequently becoming an integral thread in the fabric of global economics (IWF, 2023). It is within this complex, multi-layered landscape that the importance of understanding market dynamics becomes evident. For homeowners, investors, urban planners, lenders, and policymakers alike, this understanding forms the basis of informed decision-making and prudent strategy development, making the residential real estate market a fertile ground for nuanced analysis and future prediction.

Diving into the core of residential real estate markets, property valuation stands out as a crucial element. It is instrumental in shaping the real estate landscape as it underpins transactional decisions, guides lending practices, and directs strategic planning as well as investment appraisals services (Schulz et al., 2014; RICS, 2021). The traditional property valuation process has historically heavily relied on the expertise and assessment of real estate appraisers. Despite the invaluable local knowledge and experience they offer, this approach presents significant challenges, such as the introduction of subjectivity and bias due to the heavy reliance on human judgment. Besides the fact that the human capacity to process extensive data is inevitably limited, appraisals by experts are also costly and time-consuming (Lausberg & Dust, 2017; Yeh & Hsu, 2018). At a time when the number of appraisers is declining, it is imperative to seek innovative solutions (Coyle, 2015).

Yet, a new paradigm has emerged that aims to overcome these limitations: machine learning-assisted property valuation. It marks a revolutionary shift in the real estate industry, with a particularly transformative impact on the residential sector (Royal Institution of Chartered Surveyors, 2017). This is mainly due to the higher degree of standardization and large amount of data available in this sector, which create a favorable environment for machine learning application. In particular, deep learning and tree-based ensemble learning approaches seem to be suitable for real estate valuation. As flexible and powerful models, these innovative methods learn from the data itself and provide many benefits over the rather traditional valuation process (see, e.g., Sangani et al., 2017; Mayer et al., 2019; Pace & Hayunga, 2020; Glumac et al., 2021). Machine learning algorithms can first and foremost process and analyze massive amounts of data. This allows them to detect complex patterns and relationships that human appraisers might overlook. Additionally, compared to how long it would take a human appraiser to manage such massive volumes of data, machine learning algorithms are much faster. Due to this, machine learning users may do large-scale real estate valuation and market analysis in real time or very close to it, providing accurate, reliable, and objective results (Royal Institution of Chartered Surveyors, 2017; Stang et al., 2022).

Despite the exceptional achievements of modern machine learning models in theory, the transition from these theoretical advancements to their practical implementation in the real estate industry has not kept pace and remains a challenge. To fully realize the benefits of machine learning-based solutions and sustainably enhance the property valuation process, this gap must be bridged. However, to achieve this, several key challenges and areas for improvement must be addressed.

First, these approaches have been criticized for their inherent opacity. The complex internal processes of these models, often referred to as 'black boxes', cannot straightforward be interpreted (see, e.g., Adadi & Berrada, 2018). This lack of transparency can form a significant barrier to the widespread adoption of machine learning-based valuation in the real estate industry, especially due to regulatory constraints (European Banking Authority, 2020). In response to this interpretability problem, a burgeoning field has emerged: explainable artificial intelligence. These approaches expose the functioning of machine learning algorithms, helping to make their operation more transparent. Explainable artificial intelligence makes it possible to understand and interpret the decision-making process within machine learning algorithms (Molnar, 2020). These insights can help mitigate the 'black box' problem and provide much-needed transparency in machine learning-based property valuation methods. However, within the field of real estate research, the application of explainable artificial intelligence has been limited, offering an exciting opportunity to deepen our understanding of algorithmic workings, property valuations, and general market dynamics.

Second, there is still untapped potential to improve the performance of machine learning-based property valuations by refining existing algorithms, optimizing model configurations, and actively exploring the integration of new data sources. Such improvements could increase valuation precision and lead to more robust and accurate results. These valuable enhancements can underline the ongoing evolution of machine learning in real estate and open avenues for future research that has the potential to make these valuations a more effective tool in the broader market (Glumac et al., 2021; Krämer et al, 2023).

Third, even though machine learning has the potential to significantly enhance the property valuation process, its applications have until today focused primarily on the property assessment. Other crucial areas have not yet been comprehensively integrated. For example, location analysis which significantly determines property values, is still largely unexplored and dependent on subjective human judgments. The complexity of the process is due to the need to consider multiple interacting factors such as accessibility to amenities, population and employment structure. The potential for machine learning and in particular complex models like deep neural networks and tree-based models in this context is substantial. They can model and capture the intricate relationships and interactions in location evaluation that may be missed by traditional econometric models. Explainable artificial intelligence approaches hold great potential to provide meaningful insights into location, its driving factors, and the broader market. Yet, there is currently a lack of initiatives that leverage advanced machine learning capabilities to automate and optimize the process of location quality analysis.

The objective of this thesis is to enhance residential property valuation by using modern machine learning algorithms in combination with explainable artificial intelligence approaches. More specifically, the goal is to deepen the understanding of market behavior, improve the performance of automated residential real estate valuations, and introduce a methodology to automatically evaluate both the location of a property and its influencing factors. This work extends the existing literature by building on established research approaches and showing how they can be refined to achieve optimal results. At the same time, it introduces entirely new methods and approaches to real estate research that future researchers can build upon. This work connects academic insights with practitioners and demonstrates how advanced methods can be applied to different areas of the real estate industry. These approaches not only help to achieve more efficient and objective results, but also enable a better understanding of property valuations.

1.2 Course of Analysis and Research Questions

This section delves into the course of analysis undertaken in this cumulative thesis. The three papers focus on residential property valuation through the application of machine learning and explainable artificial intelligence, providing insights into residential real estate markets and addressing significant challenges in the field.

Paper 1 | Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values

Paper 1 explores the need for transparent algorithms in evaluating residential real estate. For this purpose, the first paper employs a non-parametric machine learning method called eXtreme Gradient Boosting, in conjunction with two global post-hoc, model-agnostic explainable artificial intelligence techniques, namely Permutation Feature Importance and Accumulated Local Effects. These methods are used to overcome the "black box" image of modern machine learning algorithms and provide insights into the key drivers of residential real estate markets and their behavior. A dataset of 81,166 observations - 61,763 condominiums and 19,403 single-family homes - from the seven largest cities in Germany is used. The results show that for both, condominiums and single-family homes the same value determining features play an important role. However, the way in which these features affect market value partly differs significantly. In most cases, a non-linear relationship can be identified. In summary, our results highlight the importance of using purely data-driven approaches to facilitate the extraction of crucial market intelligence and update long-established rules regarding the determinants that influence residential real estate market values.

Research Questions

- Which characteristics are important for the market values of residential properties?
- To what extent are those features characterized by either linearity or non-linearity? Are there differences depending on different cities?
- Can any fundamental differences between condominiums and single-family homes be identified?

Paper 2 | Automated Valuation Models: Improving Model Performance by Choosing the Optimal Spatial Training Level

Paper 2 focuses on the optimization of the performance of Automated Valuation Models in the residential real estate sector. In this study, the effect of the choice of the spatial training level on the performance of different machine learning approaches is analyzed. For this purpose, a unique dataset of 1,212,546 observations across Germany is used. Four different machine learning algorithms are trained on four different spatial training levels in accordance with the European Union's Nomenclature of Territorial Units for Statistics. The results show that the choice of the spatial training level significantly influences model performance and that this influence varies widely. The more traditional approaches perform best when being trained on the smallest spatial level as they cannot adequately capture the assumed heterogeneity of the real estate markets. In contrast, for the two rather modern approaches, adding further observations increases their explanatory power and outweighs the effect of the heterogeneity. Our findings suggest that assumptions valid for applying traditional machine learning methods may not be appropriate for modern methods. They provide guidelines that can be used as a starting point for further research into the analysis of real estate markets using machine learning algorithms.

Research Questions

- Does the choice of spatial training level have a major impact on the model performance of machine learning algorithms?
- At which spatial training level do machine learning algorithms perform best? Can a difference be observed between rather traditional and modern machine learning approaches?
- Can the algorithms adequately capture the assumed heterogeneity of real estate markets?

Paper 3 | Changing the Location Game – Improving Location Analytics with the Help of Explainable AI

Paper 3 introduces a novel, purely data-driven approach for the analysis of real estate locations and their drivers. By combining a modern machine learning approach with a state-of-the-art local post-hoc, model agnostic explainable artificial intelligence method called Sharpley Additive Explanations, the newly developed SHAP location score represents an intuitive and flexible approach based on econometric modeling techniques and the assumptions of hedonic pricing theory that accounts for non-linearities and higher order interactions within the data. The approach can be applied post-hoc to any common machine learning method and can be flexibly adapted to meet specific needs. A dataset of 26,860 residential rental listing for the city of London from 2020 is used to show how an empirical implementation of the SHAP Location Score is possible and to demonstrate the results. The findings reveal that it can effectively assess the quality of a location and analyze the factors that determine it.

Research Questions

- What theoretical foundations are necessary to develop a machine learning-based framework for data-driven valuation of real estate locations?
- Can the integration of modern machine learning algorithms and explainable artificial intelligence techniques offer a purely data-driven methodology for assessing the quality of real estate locations?
- Are modern machine learning algorithms able to capture the interactions and non-linear relationships that define the quality of a location?

1.3 Co-Authors, Submissions and Conference Presentations

The following overview provides information about co-authors, journal submissions, publication status and conferences presentations.

Paper 1:	**Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values**

Authors:

Bastian Krämer, Dr. Cathrine Nagl, Moritz Stang, Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: Journal of Housing Research
Current status: accepted (11/01/2023) and pre-published online (14/02/2023)

Conference Presentations:
This paper was presented at:

- the 38th Annual Conference of the American Real Estate Society (ARES) in Bonita Springs, USA (2022)
- the 28th Annual Conference of the European Real Estate Society (ERES) in Milan, Italy (2022)
- the Doctoral Seminar of the Center of Finance of the University of Regensburg in Regensburg, Germany (2022)

Awards:
This paper was awarded the "Best PhD Paper Award 2022" at the 28th Annual Conference of the European Real Estate Society (ERES) in Milan, Italy (2022).

Paper 2:	**Automated Valuation Models: Improving Model Performance by Choosing the Optimal Spatial Training Level**

Authors:

Bastian Krämer, Moritz Stang, Vanja Doskoč, Prof. Dr. Wolfgang Schäfers, Prof. Dr. Tobias Friedrich

Submission Details:

Journal: Journal of Property Research
Current status: accepted (18/04/2023) and pre-published online (02/05/2023)

Conference Presentations:
This paper was presented at:

- the 39th Annual Conference of the American Real Estate Society (ARES) in San Antonio, USA (2023)
- the 29th Annual Conference of the European Real Estate Society (ERES) in London, UK (2023)

Paper 3: Changing the Location Game – Improving Location Analytics with the Help of Explainable AI

Authors:

Moritz Stang, Bastian Krämer, PD Dr. Marcelo Cajias, Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: Journal of Real Estate Research
Current status: accepted (05/09/2023) and pre-published online (30/09/2023)

Conference Presentations:
This paper was presented at:

- the 39th Annual Conference of the American Real Estate Society (ARES) in San Antonio, USA (2023)
- the 29th Annual Conference of the European Real Estate Society (ERES) in London, UK (2023)

Awards:
This paper was awarded the “Real Estate Market Analysis Award” at the 29th Annual Conference of the American Real Estate Society (ARES) in San Antonio, USA (2023).

1.4 References

Adadi, A. and Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access, 6*, 52138–52160.

Angrick S., Bals B., Hastrich N., Kleissl M., Schmidt J., Doskoč V., Molitor L., Friedrich T. & Katzmann M. (2022). *Towards explainable real estate valuation via evolutionary algorithms*. In Proceedings of the 22nd Proceedings of the Genetic and Evolutionary Computation Conference, New York, 1130–1138.

Bourassa, S.C., Cantoni, E. & Hoesli, M. (2010). Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions. *Journal of Real Estate Research, 32*(2), 139–159.

Coyle, M.L. (2015). *Future of Valuation*. Working RE magazine, Editor's note, San Diego.

European Banking Authority (2020). *Final Report: Guidelines on loan origination and monitoring.* European Banking Authority, Paris.

Glennon, D., Kiefer, H., & Mayock, T. (2018). Measurement error in residential property valuation: An application of forecast combination. *Journal of Housing Economics, 41*, 1-29.

Glumac, B., & Des Rosiers, F. (2021). Practice briefing–Automated valuation models (AVMs): their role, their advantages and their limitations. *Journal of Property Investment & Finance, 39*(5), 481-491.

IMF (2018). *Global Financial Stability Report*. International Monetary Fund.

Krämer B., Stang M., Doskoč V., Schäfers W. & Friedrich T. (2023). Automated valuation models: improving model performance by choosing the optimal spatial training level. *Journal of Property Research.*

Lausberg, C., & Dust, A. (2017). Reducing the appraisal bias in manual valuations with decision support systems. *Advances in Automated Valuation Modeling: AVM After the Non-Agency Mortgage Crisis*, 331-343.

Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*, *12*(1), 134–150.

Metzner, S., & Kindt, A. (2018). Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties. *International Journal of Housing Markets and Analysis*, *11*(1), 73–100.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

Nanda, A. (2019). *Residential real estate: urban & regional economic analysis*. Routledge.

Pace, R.K. & Hayunga, D. (2020). Examining the Information Content of Residuals from Hedonic and Spatial Models Using Trees and Forests. *The Journal of Real Estate Finance and Economics, 60*(1-2), 170–180.

Royal Institution of Chartered Surveyors (2017). *The Future of Valuations: The relevance of real estate valuations for institutional investors and banks – views from a European expert group*. RICS, London.

Royal Institution of Chartered Surveyors (2021). Automated valuation models: Roadmap for RICS members and stakeholders. *The Royal Institution of Chartered Surveyors*.

Sangani, D., Erickson, K. & Hasan, M.A. (2017). *Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting*. In IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Orlando.

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of clinical epidemiology, 98,* 146-151.

Schulz, R., Wersing, M., and Werwatz, A. (2014). Automated valuation modelling: A specification exercise. *Journal of Property Research*, *31*(2), 131–153.

Stang, M., Krämer, B., Nagl, C., & Schäfers, W. (2022). From human business to machine learning—methods for automating real estate appraisals and their practical implications. *Zeitschrift für Immobilienökonomie*, 1-28.

ULI (2020). *Promoting Housing Affordability*. Urban Land Institute.

Yeh, I. C., & Hsu, T. K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing, 65*, 260-271.

2 Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values

2.1 Abstract

A sound understanding of real estate markets is of economic importance and not simple, as properties are a heterogenous asset and no two are alike. Traditionally, parametric or semi-parametric and, thus, assumption-based hedonic pricing models are used to analyze real estate market fundamentals. These models are characterized by the fact that they require a-priori assumptions regarding their functional form. Usually, the true functional form is unknown and characterized by non-linearities and joint effects, which are hard to fully capture. Therefore, their results should be interpreted with caution. Applying the state-of-the art non-parametric machine learning XGBoost algorithm, in combination with the model-agnostic Accumulated Local Effects Plots, (ALE) enables us to overcome this problem. Using a dataset of 81,166 residential properties for the seven largest German cities, we show how ALE plots enable us to analyze the value-determining effects of several structural, locational and socio-economic hedonic features. Our findings lead to a deeper representation of real estate market fundamentals.

Keywords: Housing Markets, Machine Learning, Explainable AI, Feature Importance, ALE Plots

2.2 Introduction

Understanding real estate markets and its drivers is arguably one of the most important areas of real estate research. Compared to other asset classes, real estate is a heterogeneous asset and properties differ from one another in terms of their features. There is a large body of literature dealing with the factors which have a significant influence on the value or price of a property, subsumed under the term Hedonic Price Models. They are usually based on parametric and semi-parametric methods like the Ordinary Least Squares approach (see, e.g., Malpezzi, 2002; Sirmans et al., 2005; Schulz et al., 2014) or the Generalized Additive Models (see, e.g., Bourassa et al., 2007; Bourassa et al., 2010; Brunauer et al., 2010). In recent years, more advanced statistical and modern machine learning (ML) methods have attracted interest in the real estate community, as they are often less restrictive in terms of their model structure and thus are more flexible. Especially deep learning algorithms like Artificial Neural Networks (ANN), bagging techniques like random forest (RF) and boosting algorithms like the eXtreme Gradient Boosting (XGBoost) algorithm, seem better suited to real estate valuation.[1] In almost all cases, researchers conclude that machine learning techniques yield better predictions than standard linear models (see, e.g., Chun Lin & Mohan, 2011; Kok et al., 2017; Mayer et al., 2019). However, ML applications are usually criticized for their lack of transparency and are therefore often referred to as black boxes (see, e.g., Din et al., 2001; McCluskey et al., 2013). While parametric and semi-parametric applications are comprehensible to humans, the calculations of modern ML applications can only be understood with difficulty if at all. To overcome this problem, so-called eXplainable Artificial Intelligence (XAI) approaches have been developed. These approaches use model-agnostic frameworks to reveal the modes of operations of ML algorithms and thus help to make their mode of action more transparent.

In real estate, so far, XAI approaches have been explored only to a limited extent, but we believe they offer several benefits. First, they shed light on the mechanism underlying ML algorithms, thus overcoming their image of black boxes, and therefore increasing their acceptance in different regulated and unregulated areas within the real estate industry, for example in the mortgage lending industry. Second, XAI methods are able to support research in understanding the key drivers of real estate markets and their functional form

[1] Applications include Worzala et al. (1995), Din et al. (2001), Peterson & Flanagan (2009), McCluskey et al. (2013) and Chiarazzo et al. (2014) for neural networks. Antipov & Pokryshevskaya (2012), Bogin & Shui (2020) and Pace & Hayunga (2020) for random forests. Focusing on boosting-related methods, see van Wezel et al. (2005), Kagie & van Wezel (2007), Gu & Xu (2017), Sangani et al. (2017), Ho et al. (2021) and Stang et al. (2022).

by taking non-linearities and joint effects into consideration. These findings can among other things be used to validate or adapt the previously known understanding of econometric functional forms.

In order to further demonstrate and confirm these benefits empirically, we show in our study how XAI methods can be used to make the deep hidden patterns of residential real estate markets interpretable for human beings. Therefore, we use two different methods. First, we use Permutation Feature Importance (PFI), first introduced by Breiman (2001), to analyze which features actually influence the value of a property. Next, we use so-called Accumulated Effects Plots (ALE), established by Apley & Zhu (2020), to further make a statement about the effects themselves and whether non-linear relationships can be identified or not.[2] In particular, the former is used as a basis for the latter, to identify which variables have the greatest impact on property values. While PFI has been used – to a limited extent – for answering different real estate specific questions (see, e.g., Lorenz et al., 2022), we are – to the best of our knowledge – the first to use ALE plots to explore the determinants of residential property market values. Furthermore, the previous literature has mainly focused on the identification of non-linearities, but falls short on the interpretation of their economic implications. Accordingly, this paper not only focuses on identifying reliable and unbiased relations between features and residential property prices, but also discusses their economic implications.

To conduct our analysis, we use the modern ML algorithm eXtreme Gradient Boosting (XGBoost) and a unique dataset consisting of 81,166 residential properties for the seven largest cities of Germany. The dataset is from the years 2014 to 2020 and can be split into 61,763 condominiums and 19,403 single-family homes. We analyze the two groups separately in order to reveal differences between the two property subtypes, in addition to the general analysis of the value-determining features. We use the market value per square meter as our target variable. The underlying market values are based on appraiser valuations and are therefore verified by professional real estate appraisers.

Besides the general introduction of ALE plots in a real estate context, we contribute to the literature by addressing the following research questions:

I. Which characteristics are important for the market values of residential properties?

[2] In XAI research, Partial Dependence Plots (PDP) – proposed by Friedman (2001) – are one of the oldest and most widely used methods (see, e.g., Levantesi & Piscopo, 2020). However, PDP plots have been shown to produce biased results when features are correlated (Apley & Zhu, 2020). In real estate, many features have an intrinsic dependence that does not justify the use of PDP plots. In contrast, ALE plots do not have this disadvantage, and are therefore well suited to real estate market analysis.

II. To what extent are the features characterized by either linearity or non-linearity? Are there differences depending on different cities?

III. Are there fundamental differences between condominiums and single-family homes?

Our analyses reveal that the same value-determining features play an important role for both condominiums and single-family homes. However, there are fundamental differences within the two property types with regard to the shape of the individual ALE plots and thus the influence of the respective feature on the market value of a property. Furthermore, we identify non-linear relationships for the majority of features. Generalized rules of thumb such as "the larger the living area, the lower the market value per square meter" are refuted by our findings for condominiums, but can be confirmed for single-family homes. In summary, our results show how important it is for both real estate research and practice to conduct data-driven analyses with the help of modern ML and XAI approaches, in order to gain important market insights and, if necessary, to update long-established assumptions regarding the determinants of real estate market values. This is especially important for mortgage underwriters, valuation firms and regulatory authorities and, thus, of considerable interest to most of the real estate community.

2.3 Data

This study uses a dataset of 81,166 residential properties for the Top-7 cities of Germany. The data originate from the years 2014 to 2020. The Top-7 are the most important cities in Germany for the real estate industry and are: Berlin, Hamburg, Munich, Cologne, Frankfurt, Stuttgart and Dusseldorf. In comparison to other European countries such as England or France, the German real estate market is polycentric and not dominated by one large city. As Cajias & Freudenreich (2018) explain, analyzing the Top-7 cities leads to a "socially, culturally and economically well diversified overview of major urban areas all over Germany." As we are interested in analyzing differences between different subtypes of residential properties, the dataset is further split into two groups. The first consists of 61,763 condominiums and the second of 19,403 single-family homes. Table 2.1 shows how the individual observations are distributed among the seven cities.

Table 2.1: Observations per city and subtype

	Berlin	Hamburg	Munich	Cologne	Frankfurt	Stuttgart	Dusseldorf
Condominiums	15,166	5,703	12,743	13,189	5,559	4,108	5,295
Single-family Homes	6,545	3,555	1,408	4,933	1,140	1,008	814

The dataset is provided by a large German banking group and originates from their valuation department. The data was collected for traditional real estate lending to private customers and is used for the legally required valuation of the property serving as loan collateral. In principle, the banking group finances the entire range of properties available on the market across all cities. A potential selection bias (e.g., only financing a certain category of properties) can therefore be largely excluded. To avoid abuses, the valuations are not carried out by the lender itself, but by certified external third-party appraisers. The appraisals are carried out in accordance with the legal framework applicable in Germany and determined using the comparison approach.[3] The market value per square meter of the properties is used as the target variable. In addition to the dependent variable, a set of features defining the structural characteristics of the properties is used. All properties are georeferenced, making it possible to add a spatial gravity layer to account for spatial information. Features describing the location and neighborhood of the properties are added via Open Street Map and Acxiom.[4] The dataset is cleaned before being used to account for duplicates, incompleteness and erroneous data points.

In the area of structural characteristics, the construction year, living area, use of the property, condition and a quality grade were used for both apartments and single-family homes. Furthermore, the lot size, a variable describing whether the property has a basement or not and a feature outlining the subtype of the property, are used for the single-family homes. All these features were determined by professional appraisers in the context of their assessment process, which is why it can reasonably be assumed that these represent a detailed and truthful representation of the actual properties.

The use of the building describes the possible types of usage of the property, whereby the characteristics are either "Owner-occupied & Non-owner-occupied,"[5] "Owner-Occupied," or "Non-owner-occupied." Basically, the variable describes whether or not a property can be rented to a third-party. The quality of the property is measured in terms of a grade, on a scale ranging from 1 (very poor) to 5 (very good) and describes the general quality of the interior of a property (e.g., condition and quality of the bathrooms). The general condition of the property, on the other hand, refers to the exterior condition of the property (e.g., condition of the walls and facade) and is represented by a categorial variable with 5 different categories ranging from bad to very good. The features describing the subtype

[3] In the context of legally required real estate valuations in Germany, there are slight differences in the methodology used compared to the internationally common approaches. Detailed explanations can be found in Schnaidt & Sebastian (2012).

[4] Acxiom is an American data provider for international data. Further information can be found at: https://www.acxiom.com/.

[5] Applies if the property is both partly owner-occupied and partly non-owner-occupied (e.g., single-family home with attached rental unit).

of the single-family homes are binary, and state whether it's a detached or non-detached house.

Features representing the micro-location of a property are the latitude and longitude and four different micro scores. The micro scores of a location are calculated via a gravity model and reflect the accessibility as the proximity to selected everyday destinations for each category. A gravity model is a common method for approximating the accessibility of a location and is based on the assumption that nearby destinations play a greater role in everyday life than more distant ones (Handy & Clifton, 2001). The scores can range from 0 to 100 points, and the higher the score, the better the accessibility of the location. The relevant points-of-interest (POIs) are selected on the basis of the findings of Powe et al. (1995), Metzner & Kindt (2018), Yang et al. (2018), Nobis & Kuhnimhof (2018) and Huang & Dall'erba (2021) and are provided in Table 2.2. The scores are mainly used to reduce dimensionality and complexity. A more detailed description of the construction of the micro scores can be found in Appendix I.

Table 2.2: Features of the micro scores of a location

Points-of-Interests	Category	Description
University	Education & Work	University campus: an institute of higher education
School	Education & Work	Place for education
Kindergarten	Education & Work	Facility for early childhood care
CBD	Education & Work	Center of the next city
Supermarket	Local Supply	Supermarket – a large store with groceries
Marketplace	Local Supply	A marketplace where goods are traded daily or weekly
Chemist	Local Supply	Shop focused on selling articles of personal hygiene, cosmetics, and household cleaning products
Bakery	Local Supply	Place for fresh bakery goods
ATM	Local Supply	ATM or cash point
Hospital	Local Supply	Facility providing in-patient medical treatment
Doctors	Local Supply	Doctor's practice / surgery
Pharmacy	Local Supply	Shop where a pharmacist sells medications
Restaurant	Leisure & Food	Facility to go out to eat
Café	Leisure & Food	Place that offers casual meals and beverages
Park	Leisure & Food	A park, usually urban (municipal)
Fitness Centre	Leisure & Food	Fitness Centre, health club or gym
Movie Theater	Leisure & Food	Place where films are shown
Theater	Leisure & Food	Theatre or opera house where live performances occur
Shopping Mall	Leisure & Food	Shopping Centre– multiple stores under one roof
Department Store	Leisure & Food	Single large store selling a large variety of goods
Subway Station	Transportation	City passenger rail service
Tram Station	Transportation	City passenger rail service
Railway Station	Transportation	Railway passenger-only station.
Bus Stop	Transportation	Bus stops of the local bus lines.
E-Charging Station	Transportation	Charging facility for electric vehicles

Notes: The descriptions of the selected Points-of-Interests is based on the explanations of Open Street Map.[6]

[6] See https://wiki.openstreetmap.org/wiki/Map_features.

The macro-location is considered by means of the features "Unemployment Rate" and "Time-On-Market." These two features have been used frequently in other studies (see, e.g., Cheng & Fung, 2015). Both variables are available at ZIP Code level. The "Unemployment Rate" measures the percentage of workers in the labor force who do not currently have a job, but are actively looking for work, and is used as a proxy for the social status of the local inhabitants. The feature "Time-On-Market" is used as a proxy for liquidity and is defined as the average number of days properties are advertised on the market within a certain ZIP Code.

To capture a temporal trend and seasonality, the year and quarter of the valuation are included. There are no correlations of concern within the dataset, so that all variables can be integrated accordingly.[7]

Table 2.3 summarizes the descriptive statistics of the features used for condominiums and Table 2.4 those for single-family homes.[8]

Table 2.3: Condominium - Descriptive statistics

Variable	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Market value per square meter	Euro	3,691.20	3,254.55	1,911.97	1,8384.40	216.96
Living area	Sqm	72.19	69.00	28.34	203.57	15.0
Longitude	Coordinate	10.06	10.00	2.52	13.64	6.70
Latitude	Coordinate	50.80	50.95	1.79	53.68	48.07
Micro score - education and work	Percentage	94.59	97.88	7.58	99.89	0.00
Micro score - shopping	Percentage	88.84	92.78	11.04	99.29	0.00
Micro score - leisure	Percentage	98.84	99.64	3.35	99.98	0.00
Micro score - public transport	Percentage	64.22	67.86	19.22	97.90	0.00
Unemployment ratio	Percentage	6.27	5.60	4.42	26.89	0.04
Year of construction	Year	1974	1973	33.80	2020	1900
Year of valuation	Year	2016	2017	2.02	2020	2014
Quarter of valuation	Integer	2.45	2.00	1.12	4.00	1.00
Quality grade	Integer	3.19	3.00	0.52	5.0	1.00
Time on market	Weeks	10.50	9.90	4.23	60.7	2.80
Condition very good	Binary	0.18	0.00	0.38	1.00	0.00
Condition good	Binary	0.39	0.00	0.48	1.00	0.00
Condition middle	Binary	0.45	0.00	0.50	1.00	0.00
Condition moderate	Binary	0.01	0.00	0.11	1.00	0.00
Condition bad	Binary	0.00	0.00	0.04	1.00	0.00
Owner-occupied & Non-owner-occupied	Binary	0.10	0.00	0.31	1.00	0.00
Owner-occupied	Binary	0.44	0.00	0.50	1.00	0.00
Non-owner-occupied	Binary	0.46	0.00	0.50	1.00	0.00

[7] The correlation matrix is available on request.
[8] The individual summary statistics for each city are available on request.

Table 2.4: Single-family homes - Descriptive statistics

Variable	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Market value per square meter	Euro	3,064.06	2,693.19	1,538.35	2,2781.21	199.44
Living area	Sqm	133.68	126.43	42.19	402.00	73.07
Lot size	Sqm	467.92	396.00	296.63	1500.00	1.00
Longitude	Coordinate	10.25	10.03	2.64	13.75	6.70
Latitude	Coordinate	51.60	52.40	1.59	53.71	47.58
Micro score - education and work	Percentage	85.59	88.29	11.87	99.83	0.00
Micro score - shopping	Percentage	75.20	79.49	15.99	98.88	0.00
Micro score - leisure	Percentage	95.49	98.27	8.75	99.98	0.00
Micro score - public transport	Percentage	43.28	42.78	16.48	95.37	0.00
Unemployment ratio	Percentage	8.34	9.44	4.33	26.89	0.08
Year of construction	Year	1974	1977	30.18	2020	1900
Year of valuation	Year	2016	2016	1.97	2020	2014
Quarter of valuation	Integer	2.44	2.00	1.11	4.00	1.00
Quality grade	Integer	3.15	3.00	0.50	5.00	1.00
Time on market	Weeks	11.30	10.20	3.71	60.70	3.70
Condition very good	Binary	0.15	0.00	0.36	1.00	0.00
Condition good	Binary	0.42	0.00	0.49	1.00	0.00
Condition middle	Binary	0.41	0.00	0.49	1.00	0.00
Condition moderate	Binary	0.02	0.00	0.13	1.00	0.00
Condition bad	Binary	0.00	0.00	0.04	1.00	0.00
Basement	Binary	0.19	0.00	0.39	1.00	0.00
No basement	Binary	0.81	1.00	0.39	1.00	0.00
Owner-occupied & Non-owner-occupied	Binary	0.17	0.00	0.37	1.00	0.00
Owner-occupied	Binary	0.74	1.00	0.44	1.00	0.00
Non-owner-occupied	Binary	0.09	0.00	0.28	1.00	0.00
Detached house	Binary	0.41	0.00	0.49	1.00	0.00
Non-detached house	Binary	0.59	1.00	0.49	1.00	0.00

2.4 Methodology

2.4.1 Extreme Gradient Boosting – XGBoost

The XGBoost algorithm is chosen as our underlying ML model, since it yielded reasonable results in several research articles (see, e.g., Truong et al., 2020). Especially Stang et al. (2022) showed that the XGBoost achieved the best results for estimating real estate market values in Germany. Therefore, the XGBoost ensures a good model-fit and enables a post-hoc analysis of the results and the application of the PFI and the ALE plots. The XGBoost is a treebased ensemble learning algorithm. Tree-based ensemble learning algorithms combine the results of several decision trees to produce better predictive performance than utilizing a single decision tree. The basic idea behind ensemble learning algorithms is that individual so-called weak learners (e.g., single decision trees) can be combined with each other and thus a strong learner can be achieved. These algorithms were developed to overcome the bias and variance problems associated with single decision trees. In the case of the XGBoost, the following technical expression can be stated:

$$h(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{M} u_m h_m(\boldsymbol{y}|\boldsymbol{x})\,. \tag{1}$$

where h is the strong learner, h_m are the individual weak learners and u_m is used to weight the weak learners. M is the number of single trees, $\boldsymbol{x}$ is the full features space and $\boldsymbol{y}$ the response variable. Boosting is a type of ensemble learning in which the weak learners h_m are trained sequentially. Starting with one tree, the following models learn from the previous errors. Gradient boosting uses the so-called gradient decent algorithm by adding new trees to minimize the loss of the model. The eXtreme Gradient Boosting is a computationally effective and highly efficient version of gradient boosting. The advantage of XGBoost is that it can recognize very complex patterns within large amount of data. For more information about tree-based methods, ensemble learning and gradient boosting, the interested reader is recommended to Hastie et al. (2001).

2.4.2 Testing Concept

In order to evaluate the XGBoost, five-fold cross validation is used.[9] To obtain the overall performance, we use the set of evaluation metrics presented in Table 2.5. The selected metrics are applied continuously, to evaluate the results of hedonic and machine learning approaches (see, e.g., Mayer et al., 2019).

Table 2.5: Evaluation metrics

Error	Formula	Description
Mean Absolute Percentage Error (MAPE)	$MAPE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert$	Mean of all absolute percentage errors. A lower MAPE signals higher overall prediction accuracy in percent.
Median Absolute Percentage Error (MdAPE)	$MdAPE(y,\hat{y}) = median(\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert)$	Median of all absolute percentage errors. A lower MdAPE denotes a higher precision in percent without being sensitive to outliers.
Error buckets (PE(x))	$PE(x) = 100\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert < x$	Percentage of predictions where the relative deviation is less than x%, with x being 10 and 20. A larger PE(x) signals a lower variation in the predictions.
R^2	$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$	Coefficient of determination. A high R^2 is an indication of better goodness of fit of the model.

[9] K-fold cross validation is a method to test how good the predictive power of a statistical model is. It randomly splits the dataset into k equal-sized folds (= blocks). One fold is used to test the model, the remaining folds are used for training. This process is performed k times, so that each fold is used once as test data. At the end, the cross-validation error is calculated by averaging the errors of the individual test folds.

2.4.3 Permutation Feature Importance

Permutation feature importance (PFI) is a so-called post-hoc global model-agnostic technique. This term describes the fact that this technique is applied after the actual training of an algorithm (= post-hoc), the results are determined based on all available training observations (= global) and can be applied to different algorithms (= model-agnostic). Starting with the results of a trained valuation algorithm, the PFI can be used to detect the influence of individual features on the target variable of a model. The basic idea behind this technique is that if the values of a given feature are permutated and this particular feature is important for predicting the target variable, then the loss function should increase. The greater the increase, the more important the feature. The result of the PFI analysis is a ranking that indicates the relative importance of all individual features with regard to the ML model. The PFI analysis provides a simple and intuitive visual representation of which factors are important for the algorithm and which play only a subordinate role. We use the PFI to first identify the most important features before analyzing them in more detail afterwards. One main advantage of the PFI is that it can be applied to all machine learning models. Initially introduced only for random forest by Breiman (2001), Fisher et al. (2019) developed this method further to be applicable to all models and called it 'model reliance'. As explained, a feature is considered as important for the final prediction of an ML model, if the prediction error increases after its entries are permuted. Therefore, the permutation feature importance of feature j can be defined as:

$$PFI_j = E\left(L\left(\hat{f}\left(\boldsymbol{x}_j, \boldsymbol{x}_{-j}\right), \boldsymbol{y}\right)\right) - E(L\left(\hat{f}(\boldsymbol{x}, \boldsymbol{y})\right), \tag{2}$$

where L denotes a chosen loss function, $\hat{f}$ refers to a fitted supervised machine learning model, $\boldsymbol{x}_j$ and $\boldsymbol{x}_{-j}$ are the permuted variable j and its complementary set of features. Furthermore, $\boldsymbol{x}$ defines the full features space and $\boldsymbol{y}$ the response variable. In this paper, the Mean Absolute Percentage Error is used as a loss function. For every feature j, the permutation feature importance is computed 100 times, each time randomly permuting its entries. To obtain the final PFI of j, hundred permutation feature importances are averaged.

2.4.4 Accumulated Local Effects

To analyze how a single feature influences the prediction of an ML model on average, so-called global model-agnostic feature effect methods can be used. These methods allow a precise visual representation of the patterns learned by the ML model and tell us how a feature interacts with a target variable. There are different XAI approaches that have been developed to perform such an analysis. To the best of our knowledge, the field of real estate research focused on using Partial Dependence Plots (PDPs) (see, e.g., Lorenz et al., 2022). PDPs are calculated by varying each feature over all observed values (marginal distribution), while holding all other features constant and re-predicting the target variable. The basic idea behind this is that, by varying the inputs and then calculating the outputs, it is possible to analyze how the influence of a feature develops along its actual distribution. By using the marginal distribution, a function is created that is only dependent on the feature of interest. By plotting the average prediction as a function of the respective feature values, the relationships learned by the ML model can be identified and visualized. A detailed description of PDPs can be found in Friedman (2001).

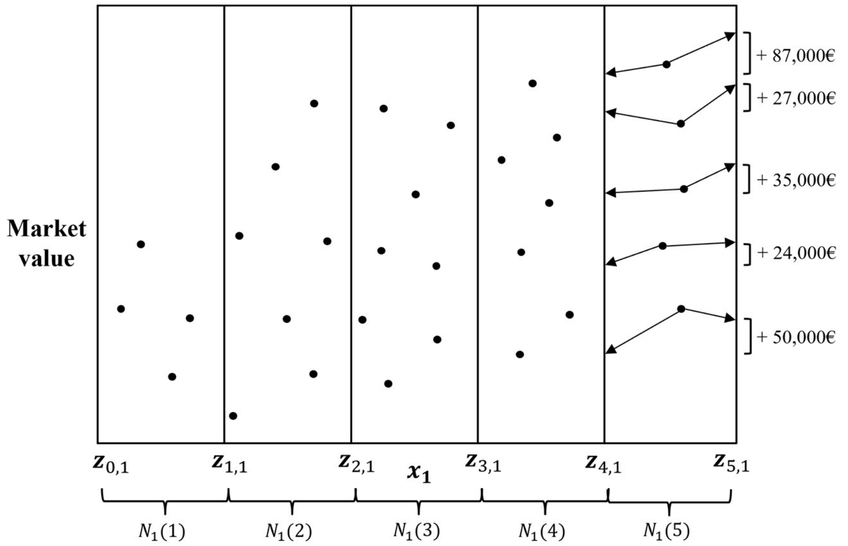
However, the way PDPs are calculated suffers from disadvantages. First, PDPs require a lot of computing time. Second, PDPs assume the relationship between features to be independent, otherwise the PDPs are severely biased. Yet, in the case of real estate, many features are inherently correlated. For example, the living area and the number of rooms are intrinsically interdependent. In the calculation of the PDPs, one would incorporate unrealistic data pairs, such as a house with 400 square meters and 1 room and or a house with 40 square meters and 10 rooms (Molnar, 2020). Therefore, the results of the PDPs should be interpreted with caution.

To overcome this disadvantages, Apley & Zhu (2020) developed Accumulated Local Effects (ALE) plots which attempt to answer the same question as PDPs, namely how features interact with the target variable. But unlike PDPs, ALE plots compute differences in predictions by varying the features value of interest only with closely related data instances (conditional distribution) instead of using the marginal distribution. Accordingly, this time small bins are created for the feature of interest and variations of the feature are made only by means of the upper and lower bounds of these bins. A simple visual representation regarding the logic of ALE plots can be found in Figure 2.1. The figure represents a simple one-dimensional case in which feature x_1 serves as our feature of interest and a total of five bins $N_1(1), \dots, N_1(5)$ are used to separate the dataset.

The first step of computing the ALE plots is to calculate the so-called Local Effect (LE) of each bin. The LE tells us what effect the feature of interest has on the target variable within a selected bin. For example, in order to build the LE of feature x_1 within the bin $N_1(5)$, we first use the upper bound $z_{5,1}$ for all observations of x_1 within the bin $N_1(5)$ to calculate the prediction of the already trained ML model (while holding the values of all other features constant) and then repeat the same process using the lower bound $z_{4,1}$.

Figure 2.1: Logic of ALE plots (Adapted from Galkin et al., 2020)

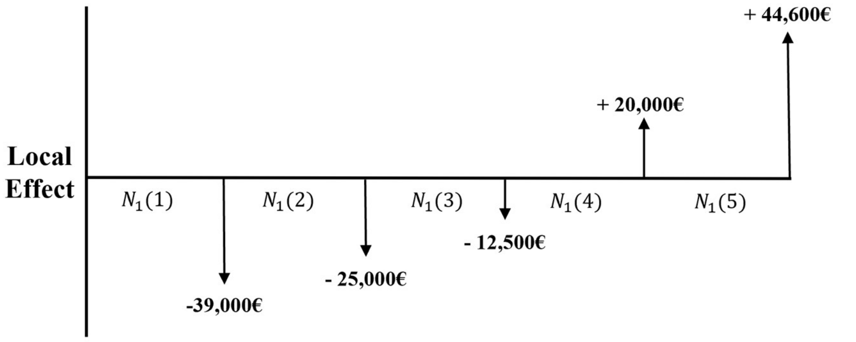
Step 1:



Step 2:

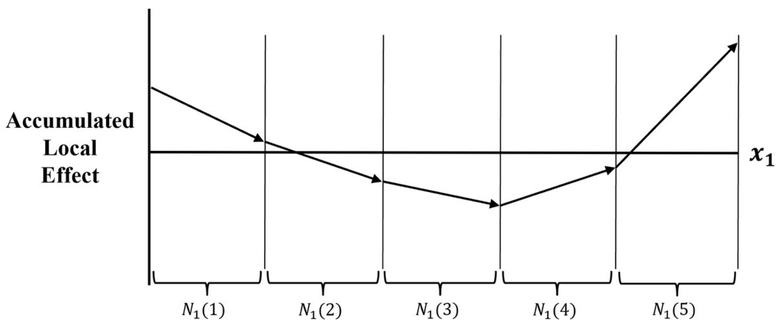
$$LE_{N_1(5)} = \frac{(87,000€ + 27,000€ + 35,000€ + 24,000€ + 50,000€)}{5} = +44,600€$$

Step 3:



Step 4:

$$ALE_{N_1(1)} = LE_{N_1(1)}$$
$$ALE_{N_1(2)} = LE_{N_1(1)} + LE_{N_1(2)}$$
$$\ldots$$
$$ALE_{N_1(5)} = LE_{N_1(1)} + LE_{N_1(2)} + LE_{N_1(3),} + LE_{N_1(4),} + LE_{N_1(5)}$$



In a next step, for all observations of x_1 within the bin $N_1(5)$, the predictions obtained with the lower bound $z_{4,1}$ are subtracted from those with the upper bound $z_{5,1}$ to calculate the change in prediction within the range $N_1(5)$. In the example shown in Figure 3.1, these range from +24,000 € to +87,000 €. Once these steps have been completed for all observations within the bin $N_1(5)$, we calculate the average of all changes to obtain the LE of x_1 within the bin $N_1(5)$, in our example €44,600.

For the final construction of the ALE plots, the individual LEs are accumulated and centered on the basis of the average prediction of all observations. Regarding the accumulation, the LE of the first bin is taken first and then expanded by that of the subsequent bin. This process is repeated until the last bin, and results in a line plot showing the ALEs. The centering of the individual ALEs ensures that the final ALE plot can always be interpreted in such a way that the line shown represents the effect of a feature as a function of the average prediction of the ML model over the full feature space. Therefore, the ALE plots have to be interpreted slightly differently to the PDPs. While PDPs show the average prediction depending on the feature values, ALEs tell us how changes in a feature influence the prediction on average compared to the average prediction. For interested readers, a more technical explanation of how ALE plots work can be found in Appendix II.

In summary, the ALE plots provide us with the opportunity to analyze the features identified in advance by the PFI analysis in more detail and to examine how the inherited relationship is learned by the ML model. By displaying the results via plots, they can be easily and intuitively interpreted. In contrast to the PDPs that have been predominantly used so far, the results of the ALE plots can be assumed to show more accurate and representative results due to the sole use of realistic data pairs. For further information about the ALE plots, we recommend reading Apley & Zhu (2020).

2.5 Results

An extra XGBoost model was trained for each of the seven cities. Furthermore, different algorithms were trained for the condominiums and the single-family homes. To ensure that the machine learning models provide reliable results and can therefore be used as underlying models for the post-hoc XAI analysis, the models were tested by five-fold cross validation on five different evaluation metrics, namely the MAPE, MdAPE, PE(10), PE(20) and R^2. Table 2.6 shows the average results of the evaluation metrics across all cities. To make a statement about the quality of the results, we estimated a basic OLS regression for each city and can thus benchmark the results.[10]

Table 2.6: Results XGBoost for all Top-7 cities

	XGBoost		OLS	
Metrics	**Condominiums**	**Single-family homes**	**Condominiums**	**Single-family homes**
MAPE	0.1253	0.1441	0.1986	0.1834
MdAPE	0.0829	0.0988	0.1467	0.1314
PE(10)	0.5698	0.5073	0.3654	0.3999
PE(20)	0.8247	0.7861	0.6396	0.6833
R^2	0.8117	0.7133	0.6230	0.5623

Our results indicate strong and robust model performance across all Top-7 cities for the XGBoost. For all metrics, the XGBoost yields a better result than the OLS, which could be expected from the literature. The MAPE of the XGBoost, relative to the OLS, is 37% lower for condominiums and 21% lower for single-family homes. The results for each city can be seen in Table 2.9, Appendix III. It is evident that the XGBoost outperforms the basic OLS regression for each city with respect to each evaluation metric used. These differences can already serve as a first indicator of non-linearities, joint effects and higher order interactions within the data. The trained XGBoost algorithms are therefore well suited to a post-hoc analysis of the results and the application of the PFI and the ALE plots.[11]

2.5.1 Results Permutation Feature Importance (PFI)

In a first step, we use the PFI to determine which variables are important for predicting the market value per square meter. The PFI provides a highly compressed, global insight into the machine learning model's behavior. The PFI is easy to interpret and also takes into account interactions within the individual features, as described by Molnar (2020). The PFI ranks all features used in the model according to their influence on the dependent variable. Therefore, in our case, the higher the ranking of a feature, the greater its influence on the market value per square meter of a property. In Appendix IV, the detailed results of the feature importance analysis can be found. The focus of the PFI analysis is to identify the most important features, which are then used for further investigation using the ALE plots.

[10] Overall, the magnitude of the improvement over the OLS benchmark may be inflated as it ignores the well-established literature on functional forms of the variables. However, we follow this path for three reasons. First, we want to point out that non-linearities and interactions can imply large performance differences. Second, this baseline OLS will also be used as a benchmark for the ALE plots in the next chapter to emphasize the non-linearity of the data as much as possible and third, by using the ALE plots we want to show how non-parametric ML can help choose suitable functional forms for parametric and semi-parametric models.

[11] To obtain the results of the permutation feature importance and accumulated local effects the scikit-learn (https://scikit-learn.org/stable/modules/permutation_importance.html) and PyALE (https://pypi.org/project /PyALE/) packages are used.

Therefore, the five most influenceable features for condominiums and single-family homes for each city can be seen in Tables 2.7 and 2.8.

Table 2.7: Top-5 features per city – Condominiums

	Berlin	Hamburg	Munich	Cologne	Frankfurt	Stuttgart	Dusseldorf
Top-1 Feature	Year of valuation	Year of valuation	Year of valuation	Year of valuation	Year of construction	Year of valuation	Year of construction
Top-2 Feature	Year of construction	Unemploy-ment ratio	Year of construction	Year of construction	Year of valuation	Year of construction	Year of Valuation
Top-3 Feature	Unemploy-ment ratio	Year of construction	Longitude	Unemploy-ment ratio	Longitude	Unemploy-ment ratio	Longitude
Top-4 Feature	Longitude	Longitude	Living area	Longitude	Unemploy-ment ratio	Living area	Latitude
Top-5 Feature	Latitude	Latitude	Unemploy-ment ratio	Latitude	Latitude	Longitude	Living area

Table 2.8: Top-5 features per city – Single-family homes

	Berlin	Hamburg	Munich	Cologne	Frankfurt	Stuttgart	Dusseldorf
Top-1 Feature	Year of valuation	Year of valuation	Year of valuation	Year of valuation	Year of valuation	Year of valuation	Lot size
Top-2 Feature	Year of construction	Lot size	Lot size	Living area	Lot size	Lot size	Year of valuation
Top-3 Feature	Lot size	Unemploy-ment ratio	Living area	Lot size	Living area	Living area	Year of construction
Top-4 Feature	Living area	Living area	Year of construction	Unemploy-ment ratio	Longitude	Year of construction	Living area
Top-5 Feature	Longitude	Year of construction	Latitude	Year of construction	Year of construction	Unemploy-ment ratio	Unemploy-ment ratio

In summary, the results of the PFI analysis show the expected results and are also in line with the findings of previous studies, e.g., Cajias (2018) and Lorenz et al. (2022). Thus, it turns out that even with modern ML models, similar to more traditional parametric and semi-parametric approaches, the usual features seem to play a predominant role. The results show that the same features are important for both condominiums and single-family homes. The valuation year is by far the most important. For condominiums, it is always at the top of the list except in Dusseldorf and Frankfurt, and for single-family homes it is also always the highest ranked feature except in Dusseldorf. It shows that the market values of the properties are mainly influenced by the market phase and thus the general market trends. In addition to the valuation year, it can be seen that structural, location-related and socio-economic features have an important influence on the market value of the properties. These findings are also in line with Dubin (1988) and Sirmans et al. (2005). Not only structural characteristics, but also the location and the economic or social environment of a property are decisive for the composition of market values in hedonic pricing models. In the case of condominiums, the year of construction and the living area are the most important factors in terms of property characteristics. In the case of single-family homes, lot size is added to these features. This is also in line with the general findings of other studies (see, e.g., Fan et al., 2006). In our case, the location of the properties is represented by the latitude and longitude. It turns out that depending on the location, the values for otherwise identical properties are different. Socio-economically, we find that the unemployment ratio seems to play a partially important role.

2.5.2 Results Accumulated Local Effects Plots (ALE)

To analyze the identified features in more detail, we use ALE plots to take a closer look at how the individual effects work and what economic insights can therefore be drawn. ALE plots describe the main effect of a feature at a certain point in comparison to the average predicted value. Compared to traditional hedonic price functions, they can capture non-linearities independently, without the need for a priori manual specification. This enables us to visualize a more realistic representation of the actual market fundamentals. Our findings are therefore beneficial for all real estate market players, to reach more targeted and, in particular, data-supported decisions. In addition, the results of the ALE plots can also be used to specify the functional form for each feature for other parametric and semi-parametric models and thus improve the performance of these models as well. Therefore, our results are also important for applications where machine learning is not the method of choice. Researchers and practitioners can use the approximated relations and implement them into their parametric or semi-parametric models.[12] Accordingly, the following results are interesting for a broad audience in real estate research and practice. The ALE plot is centered and the mean effect of the features is zero. Therefore, the y-axis of the ALE plot can be interpreted as the main effect of the independent variable at a certain point, in comparison to the average predicted value. It shows wheter a feature changes, how much it affects the prediction on average compared to the average prediction. The ALE algorithm divides the feature space into intervals containing the same number of data points, whereby feature intervals with a greater observation density are chosen to be smaller than intervals with a low density. In our case, the maximum number of intervals is set to 250. In order to check whether there are non-linearities within the data, we show within the ALE plots, in addition to the effect identified by the XGBoost, the results of a basic OLS as a benchmark. The results of the OLS are shown by means of

[12] An example of how an OLS can be optimized by using the results of the ALEs can be seen in Appendix V.

a red line and show the main difference between previous hedonic pricing approaches and to the results of our analysis. We are thus able to show which effects are covered well by parametric models and which effects must be manually specified.

Starting with the structural characteristics, Figure 2.2 shows the ALE plots for the year of construction for condominiums, and Figure 2.3 for single-family homes. In contrast to the red OLS line, it is obvious that the trend is not linear. The effects are approximately the same across all cities. Comparing the graphs of condominiums with those of single-family homes, we see that the effects are essentially the same. It is notable that the negative trend for middle-aged properties is more pronounced for the former, suggesting that the year of construction has a generally greater influence on condominiums. Traditionally, the effect of the year of construction is described and incorporated as u-shaped (see, e.g., Mayer et al., 2019), as values for new buildings and for old buildings are generally higher than those for middle-aged properties. This effect can be described, for example, by the increased renovation rate for old properties and the generally higher quality of new buildings. While the u-shape transformation seems reasonable in a parametric context, the effects shown in Figures 2.2 and 2.3 indicate that this transformation cannot be supported here. Hence, this transformation can only be seen as a rough approximation of the true underlying relationship. While we also see the effect that middle aged properties tend to have a lower market value than the average valuation, the increase is much more significant for properties with newer construction years than for older buildings. In particular, for properties built between 2010 and 2020, we see that the increase is already almost exponential. This trend reflects the current high demand for new buildings in major German cities. Due to the lack of supply, people are currently willing to pay significantly higher prices for properties on the market. Another explanation lies in the sharp rise of construction costs. The construction of a property has become significantly more expensive in Germany's metropolitan regions in recent years, which is also ultimately reflected in market values.

Figure 2.2: Condominiums – Year of construction

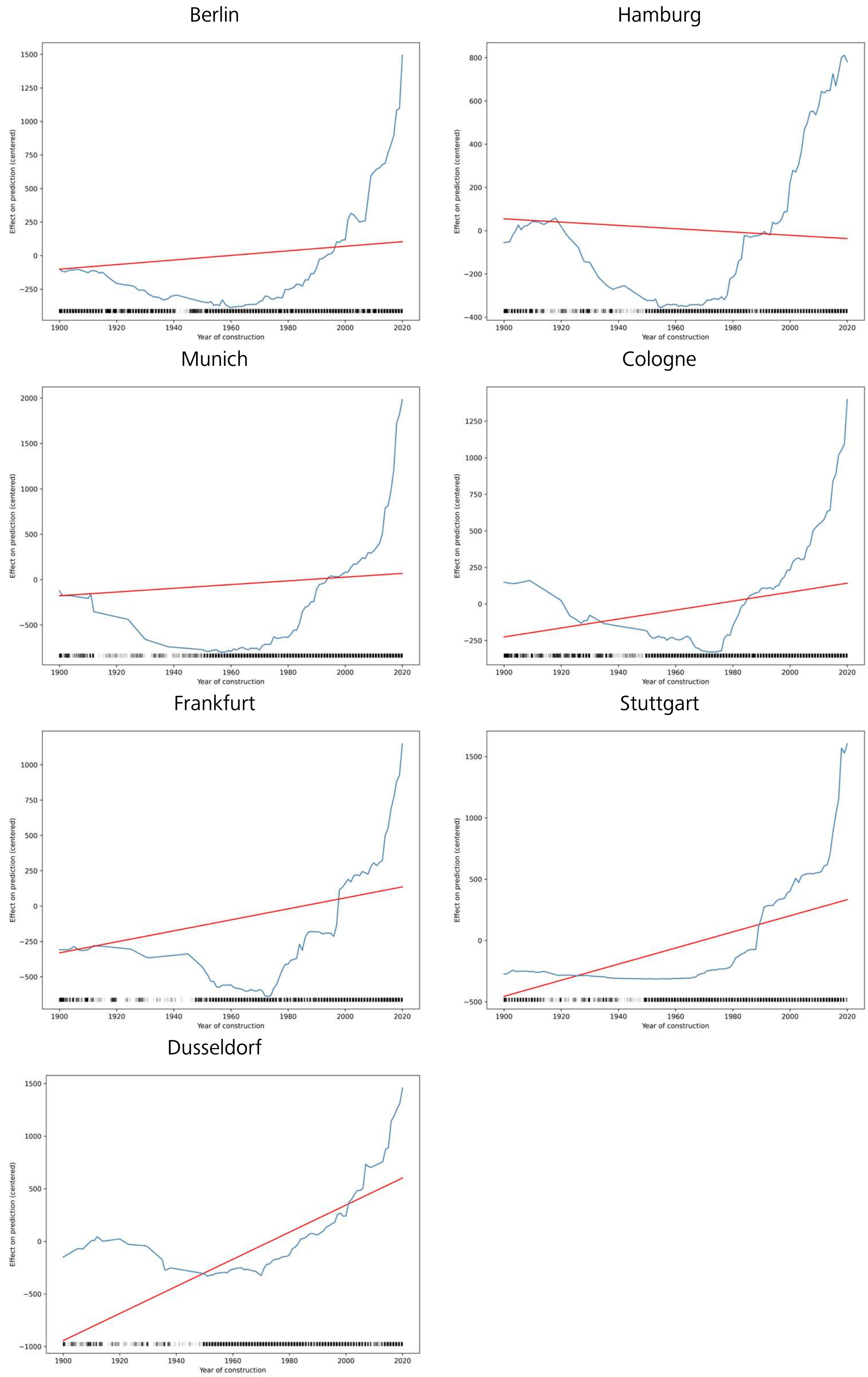
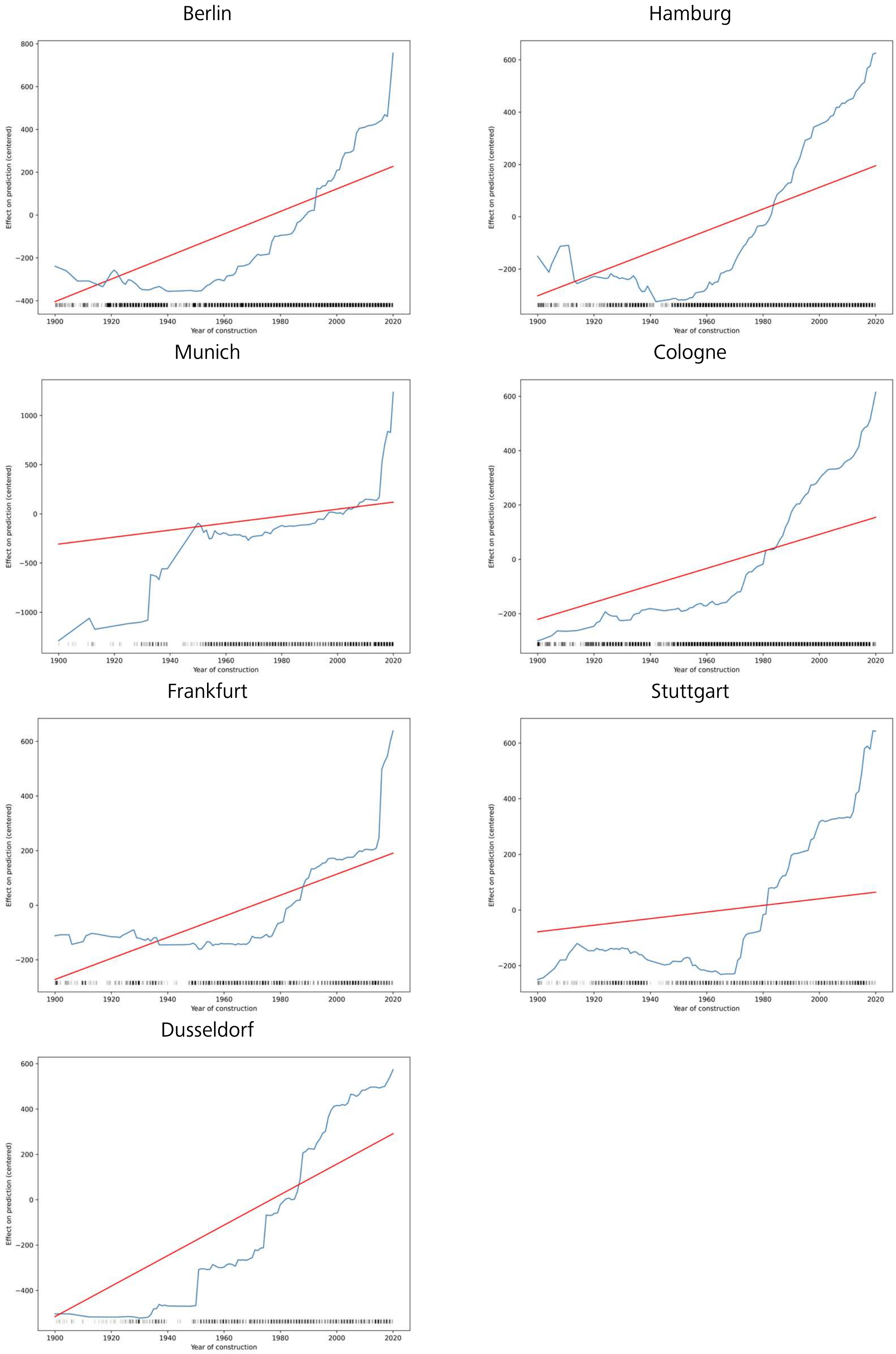


Figure 2.3: Single-family homes – Year of construction



The effect of the living area on market values of condominiums and single-family homes can be seen in Figures 2.4 and 2.5. Focusing on the condominiums, for all Top-7 cities, a clear non-linear relationship is identified with the XGBoost, which would be very difficult to represent by parametric or semi-parametric models. As no recurring pattern is evident, the effects seem to differ in each city. The findings show that there is no generally applicable rule for the analyzed cities and that a well-known rule of thumb in the real estate industry "the larger the area, the lower the market value per square meter" (see, e.g., the results of Wittowsky et al. (2020)) does not hold for condominiums. The ALE plots clearly indicate that there are different patterns regarding the market values within the cities. For example, a high demand for small apartments in the cities of Berlin, Frankfurt, Hamburg, Munich, and Stuttgart is evident. In the cities of Dusseldorf and Cologne, on the other hand, this is not as pronounced. These results offer important implications for the real estate industry. The ALE plots support the analysis of which type of apartment sizes are in demand in which region and what prices can be achieved. Currently, such decisions are often still made on the basis of personal experience or purely descriptive market statistics. The combination of machine learning and ALE plots, on the other hand, enables an empirically valid and data-driven analysis. With respect to the derivation of a generally valid functional form for parametric and semi-parametric models, the results show that this is only possible to a limited extent and that, therefore, in our case an individual function should ideally be selected in each case at the regional level. In contrast to condominiums, the effect of living area on the market value per square meter of single-family homes is homogeneous across all seven cities. Furthermore, the effect is almost linear and can be mapped by the basic OLS to a large extent. Overall, the effect shows a negative trend, which can most likely be attributed to the marginal cost effect. Major components of the costs of a single-family house (e.g., land area, development costs, etc.) are fixed to a certain extent and increase only noticeably as the size of the living area increases. In the case of larger houses, these costs are distributed over the additional square meters and lead to the negative trend shown in Figure 2.5. What also stands out is that the effect for single-family homes is larger on average, which indicates that the size of the property, compared to condominiums, is more important for houses.

Figure 2.4: Condominiums – Living area

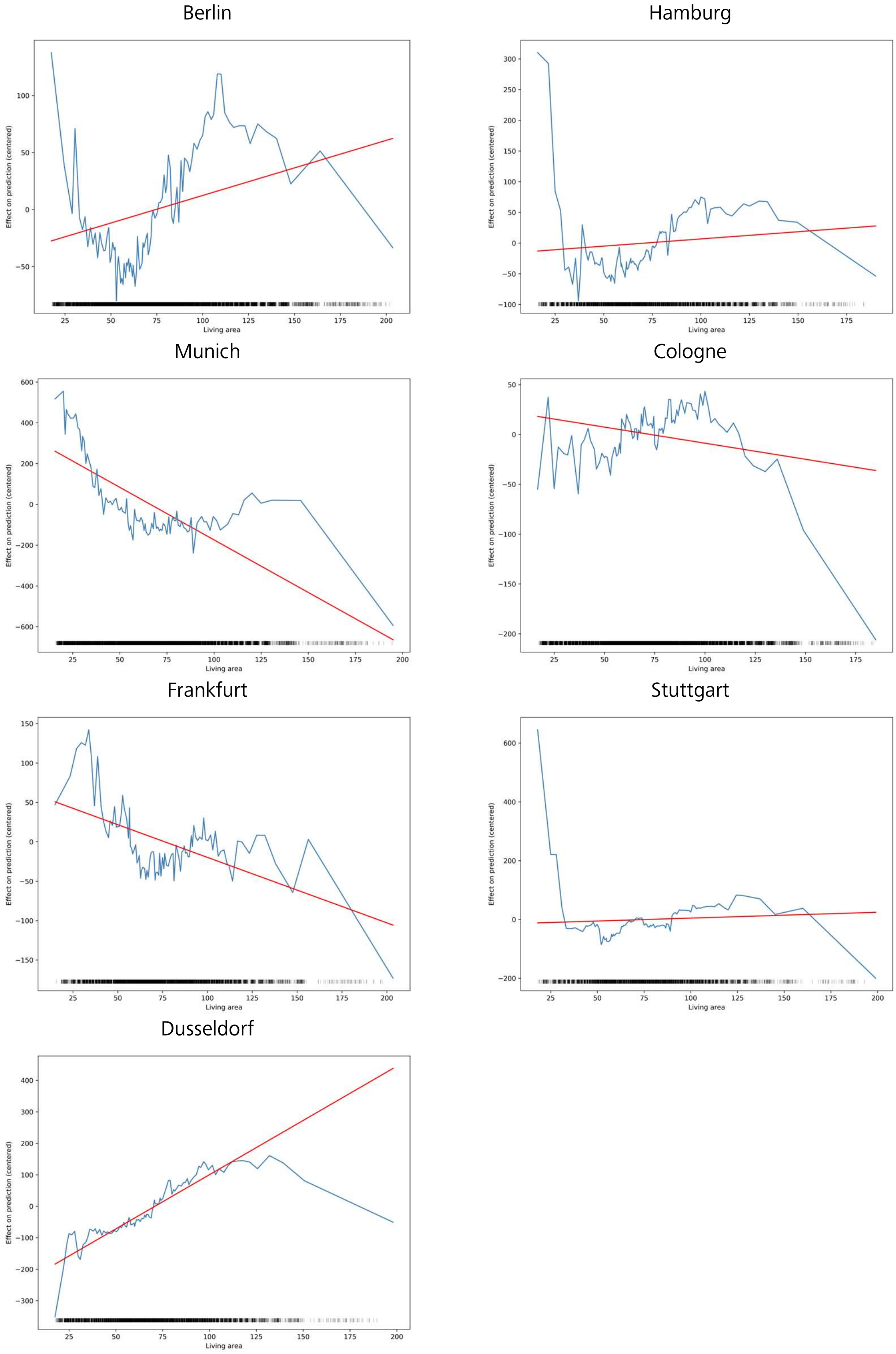


Figure 2.5: Single-family homes – Living area

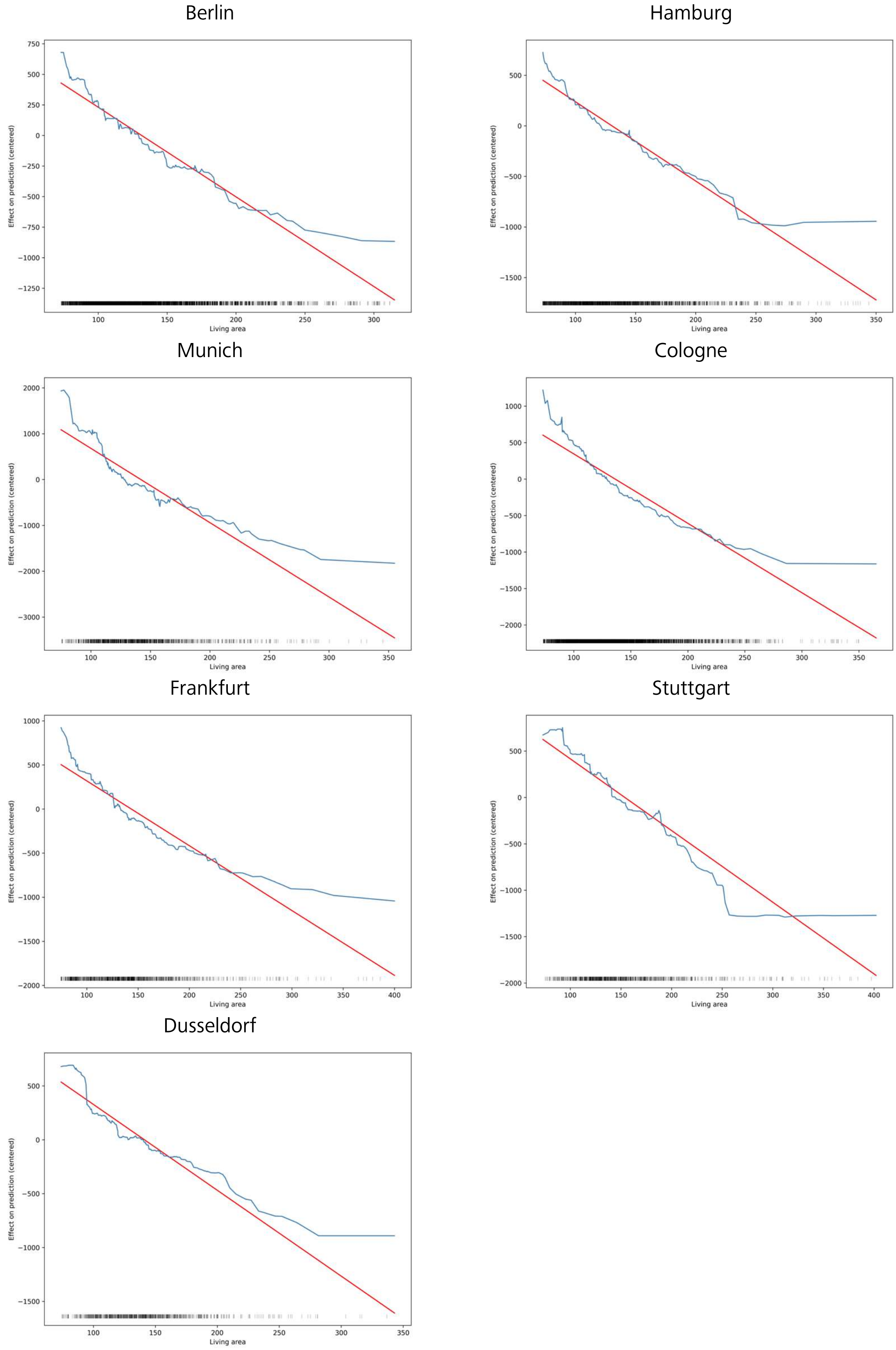


Figure 2.6 highlights the effect of lot size on the market value of single-family homes. In general, the effect is reasonably consistent across all Top-7 cities. The larger the lot size, the higher the market value per square meter of living area. This generally indicates that there is a higher demand for larger plots of land and that market values are rising due to a lack of supply, although a decreasing marginal utility can be seen for very large plots in all cities. While the basic OLS also assumes such a progression, the results of the XGBoost are more granular and thus more accurately reflect the actual effect. In general, it can be seen that a log transformation of the lot size appears to be useful with respect to a manual specification in parametric and semi-parametric approaches. This can also be adopted as generally valid for all cities analyzed in our case. Differences between cities can be seen mainly in the strength of the effect. In Munich and Stuttgart, for example, the lot size seems to play a more important role than in Berlin or Hamburg.

As the results of the PFI show, the year of valuation was by far the most important feature. The effect of the valuation year is shown in Figure 2.7 for condominiums, and in Figure 2.8 for single-family houses. Since this variable is discrete, the next lower and next higher values are used as interval limits. The bars represent the size of the sample in each year, and the number is summarized with a second y-axis on the right of the plot. The red line again represents the results of a basic OLS. The results show that in all Top-7 cities, market values have risen sharply and constantly over the observation period. In principle, the OLS and XGBoost curves are relatively similar. However, it is apparent that the XGBoost identifies a stronger price increase for the last three years. In general, the demand for both condominiums and single-family homes has risen sharply in German metropolitan areas. Since supply is inelastic because of long development periods, this increase in demand leads to a dynamic rise in prices (Belke & Keil, 2018). Our findings show that this price increase also affects the market values of the properties. Property values have risen over the past few years, irrespective of their structural and locational characteristics. This decoupling effect can be seen as quite critical, as the generally strongly rising prices can lead to speculation, which was also observable on the U.S. residential real estate markets before the Global Financial Crisis (Martin, 2011). In combination with a longer time series and other important macroeconomic features, the ALE plots could be used to conduct a more in-depth analysis and thus analyze key developments and drivers of real estate price bubbles. We consider this to be a promising and interesting area of research that should be pursued further.

Figure 2.6: Single-family homes – Lot size

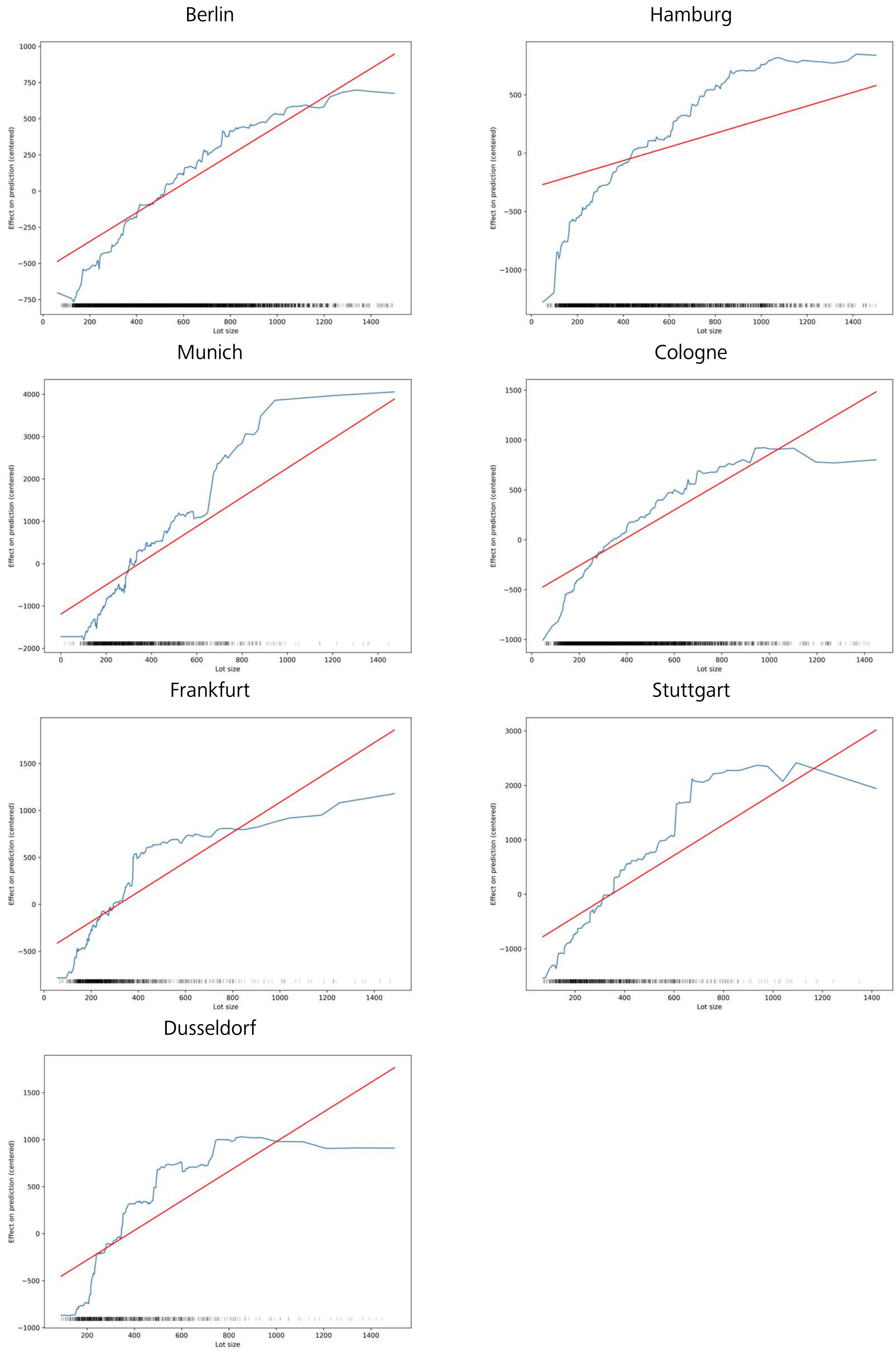


Figure 2.7: Condominiums – Year of valuation

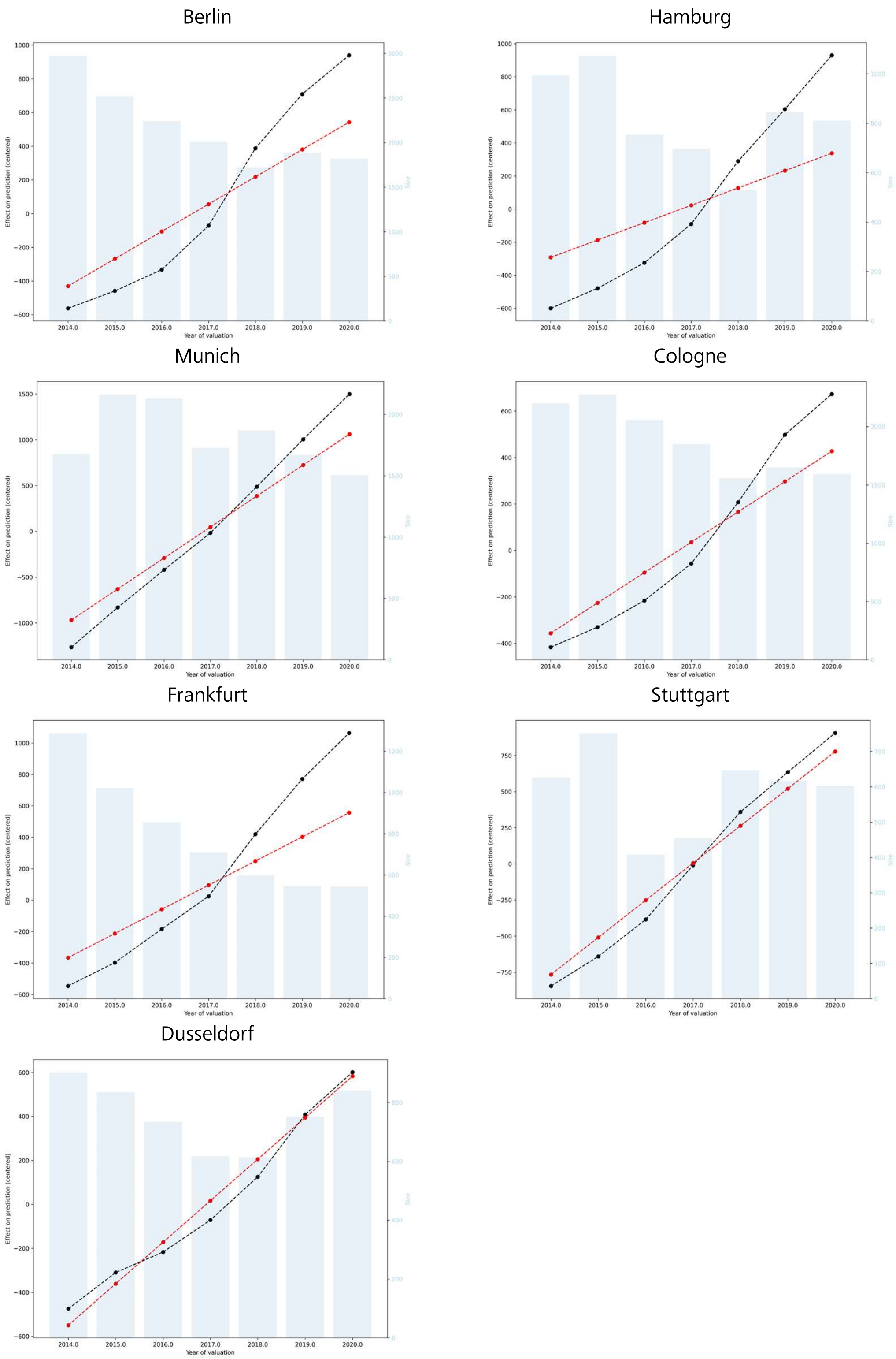
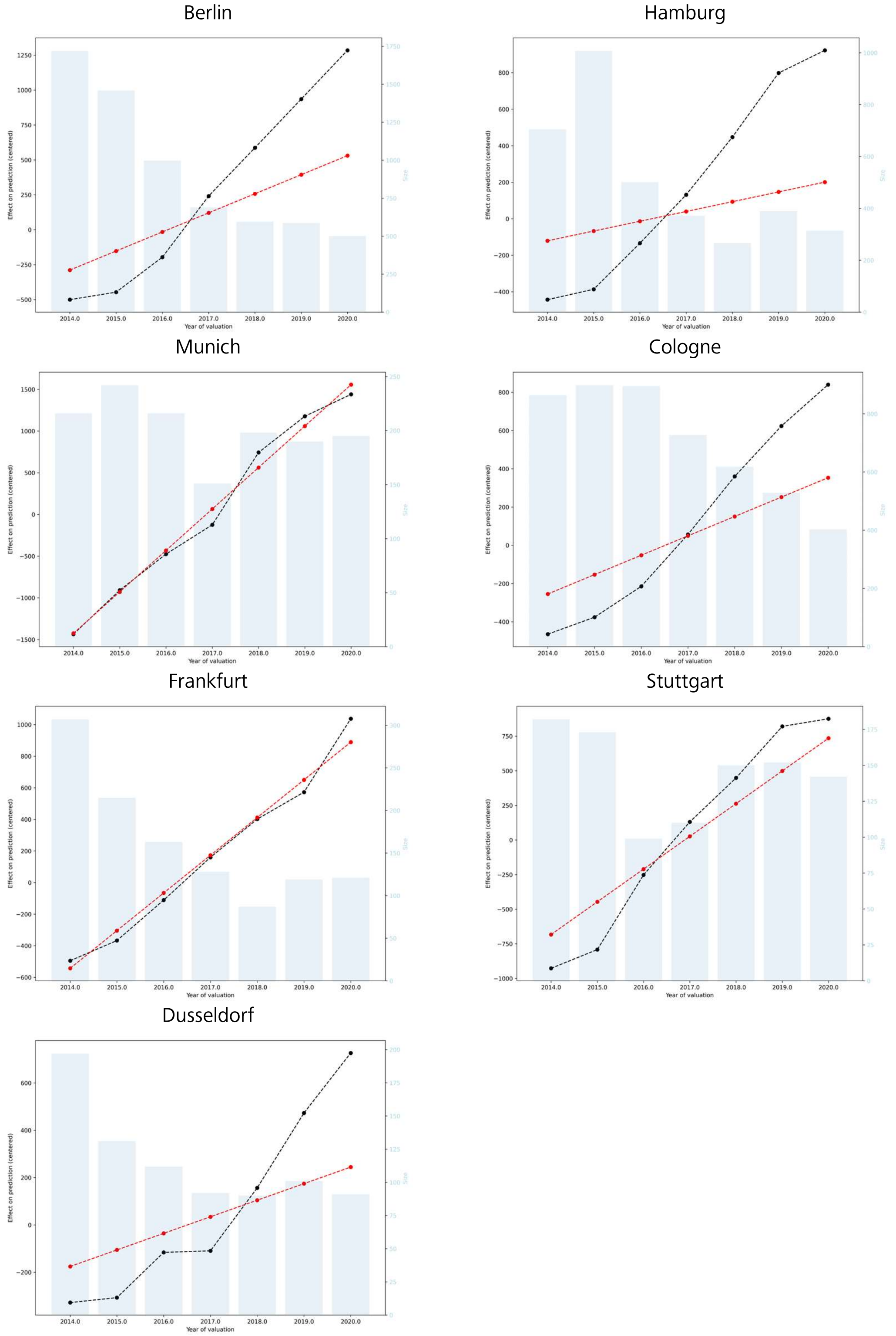


Figure 2.8: Single-family homes – Year of valuation



In addition to the temporal and structural characteristics, the PFI analysis shows that the location-related features of latitude and longitude play a significant role in predicting the market values per square meter for both condominiums and single-family homes. The effect of the latitude for condominiums is shown in Figure 2.9 and the effect of the longitude in Figure 2.10. In both figures, one can see that prices are rising sharply in certain regions – the city centers. An exception to this is the latitude of the city of Stuttgart. This is due to the unique location of Stuttgart, in a valley with a lot of industry in central locations and thus has its own geographical characteristics. Comparing the ALE plots of the XGBoost and the basic OLS, it can be clearly stated that a simple OLS cannot reproduce these non-linear locational effects. It is also not possible to derive a universally valid functional form, which is why in our case a specification at the local level is also recommended. Looking at the impact of latitude and longitude on single-family houses in Figures 2.11 and 2.12, a clear difference can be seen. Not only is the effect much less pronounced, but the expensive regions are no longer in the center of the cities, which is not surprising, since the houses in these cities are located in the suburbs. In summary, the ALE plots of latitude and longitude can help to identify promising locations within the cities.[13]

Finally, we consider the impact of the unemployment ratio on market values. Figure 2.13 shows the impact on condominiums and Figure 2.14 on single-family homes. Overall, the ALE plots of the XGBoost seem reasonable and are in line with the findings of other studies (see, e.g., Grum & Govekar, 2016). In all cities, the XGBoost identifies a downward trend in market values the higher the unemployment rate. Comparing the results of the XGBoost with the baseline OLS, one can see a large difference between the two graphs. In this context, the focus is on Stuttgart in particular, where positive effect of the OLS is assumed. Once again, the OLS is not able to capture the effects in a granular and comprehensive way. Decisions made on the basis of this flawed assumption can have far-reaching consequences and should be avoided.[14]

[13] However, these results have to be interpreted with caution. There are several location-based features in our dataset. Besides latitude and longitude, there are the four micro-scores, which also describe the surrounding location of the properties. Furthermore, there are three socioeconomic variables in the dataset, which are available at the ZIP code level and thus could also be seen as a proxy for location. To obtain the overall effect of the location on the price, these individual effects would have to be aggregated.

[14] However, it bears repeating at this point that the results should also be interpreted with caution, as the unemployment rate can also serve as a simple proxy for the location of a property, due to its availability at the ZIP code level.

Figure 2.9: Condominiums – Latitude

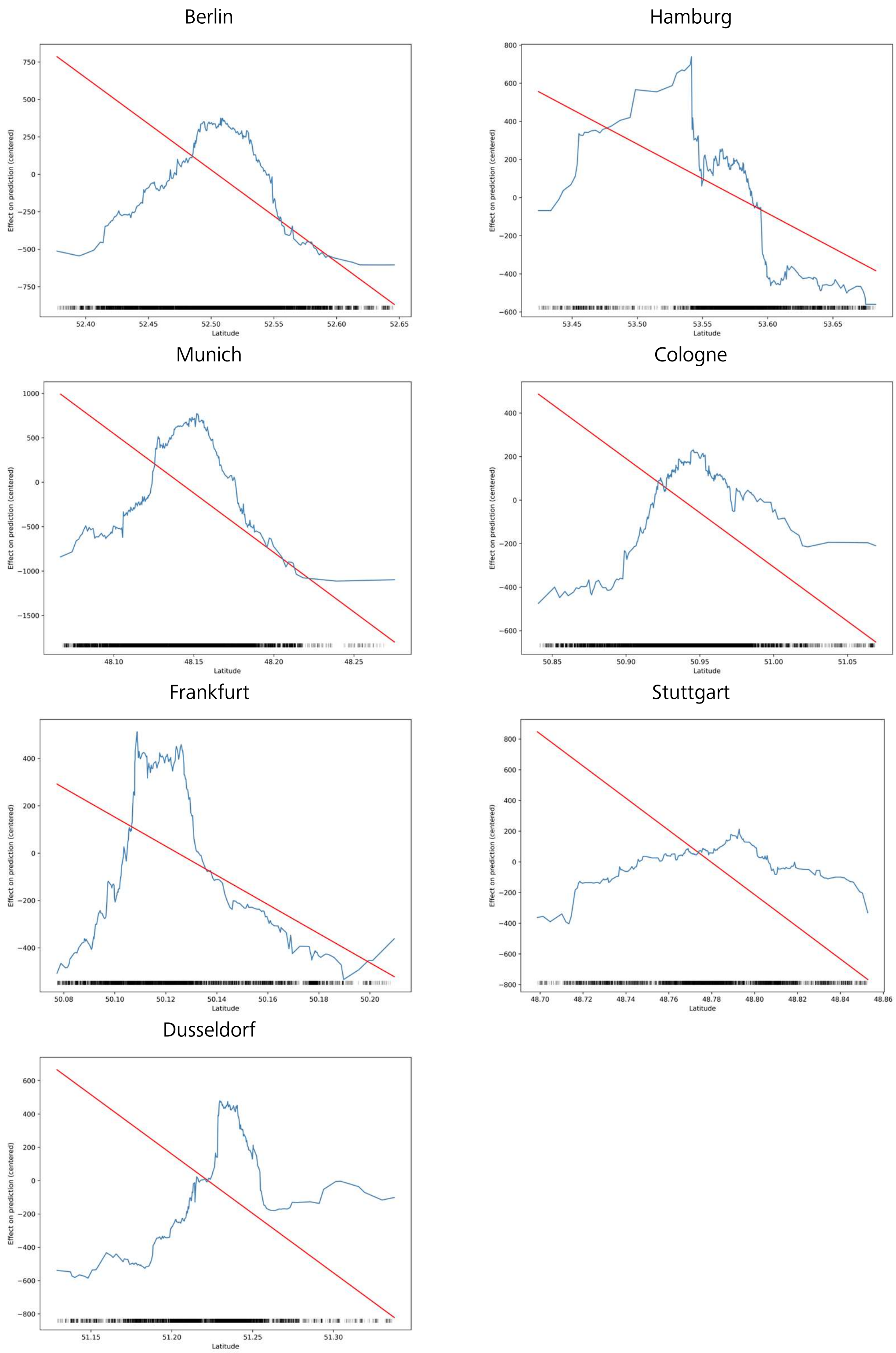


Figure 2.10: Condominiums – Longitude

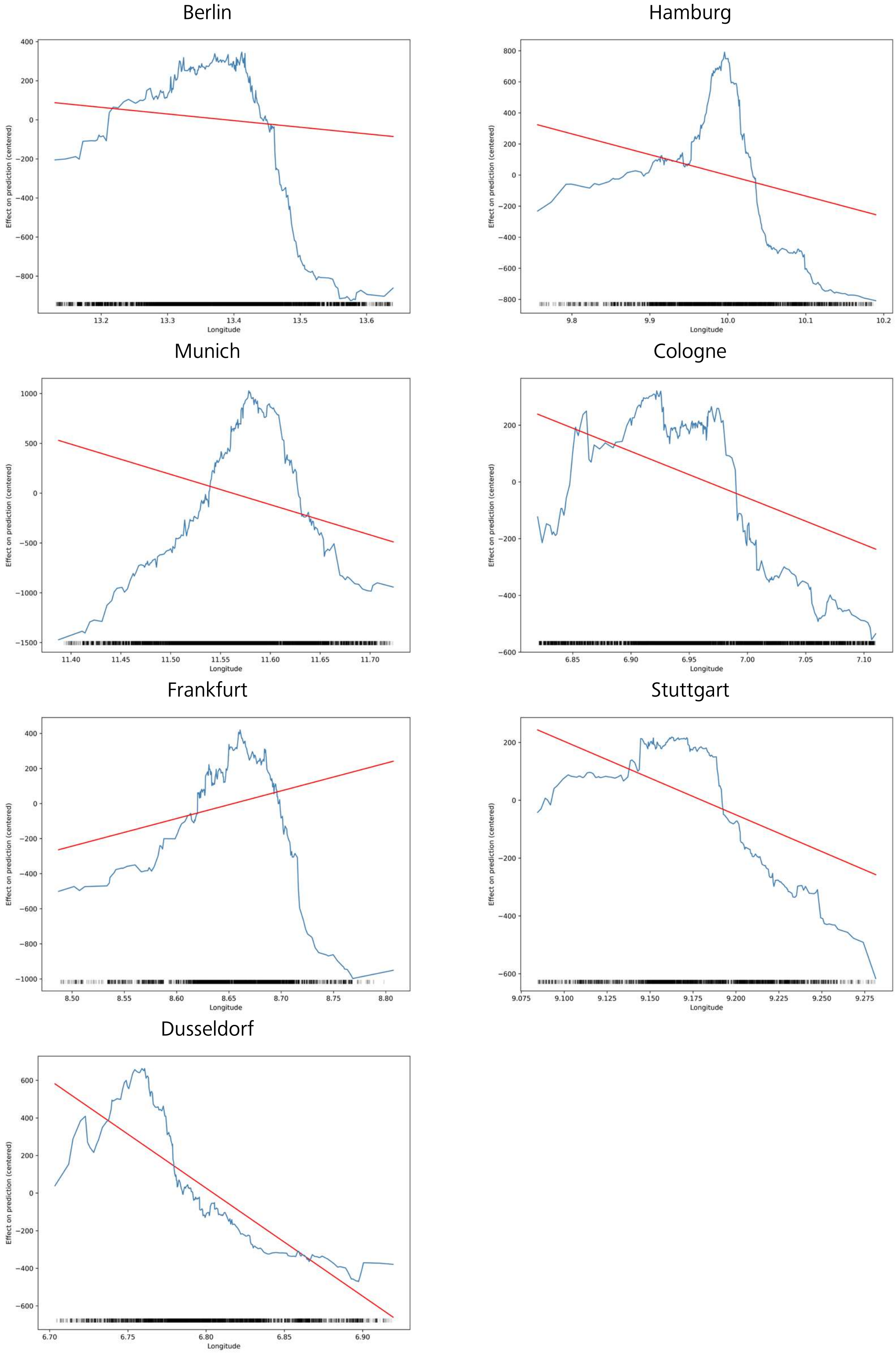


Figure 2.11: Single-family homes – Latitude

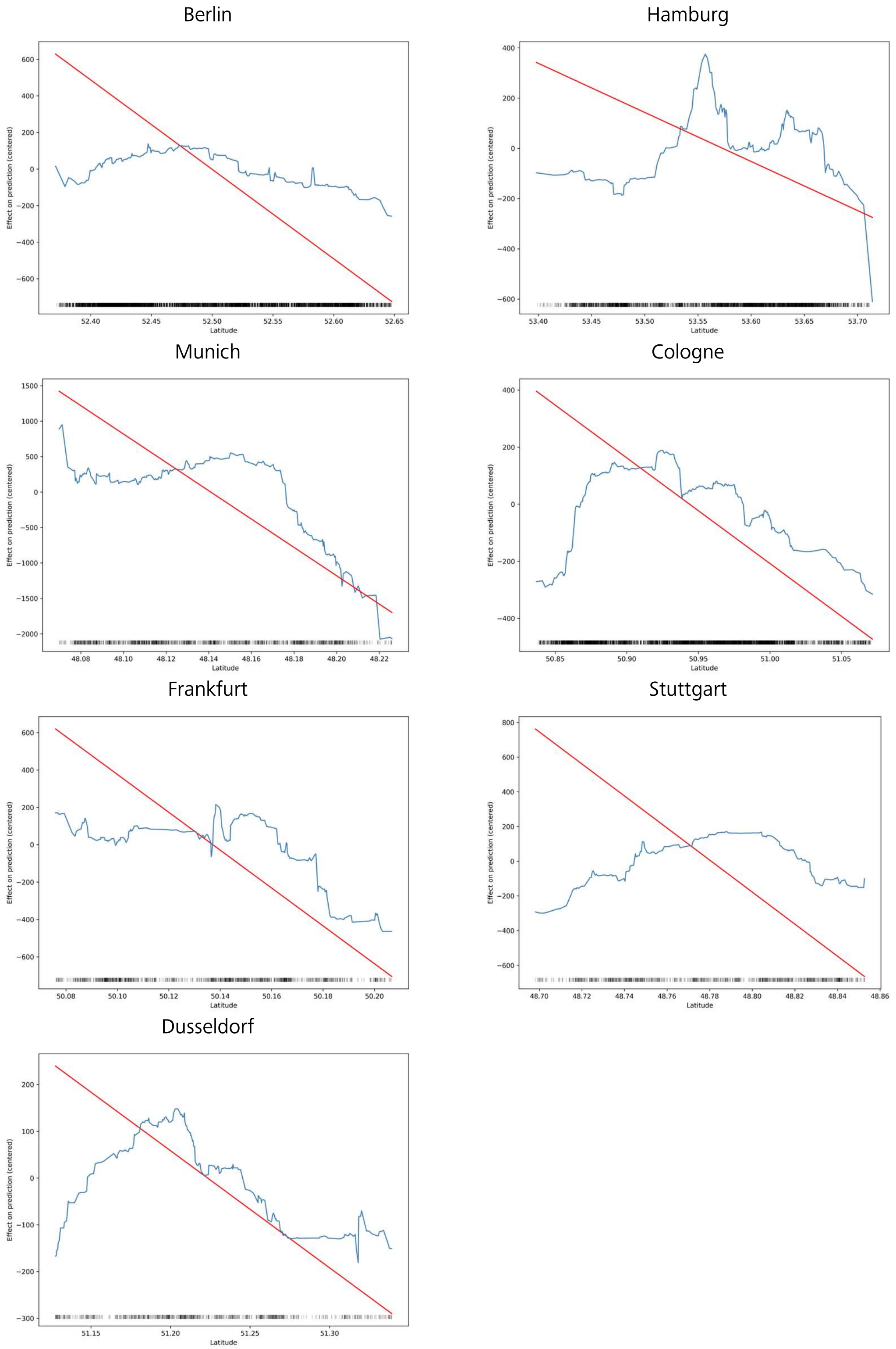


Figure 2.12: Single-family homes – Longitude

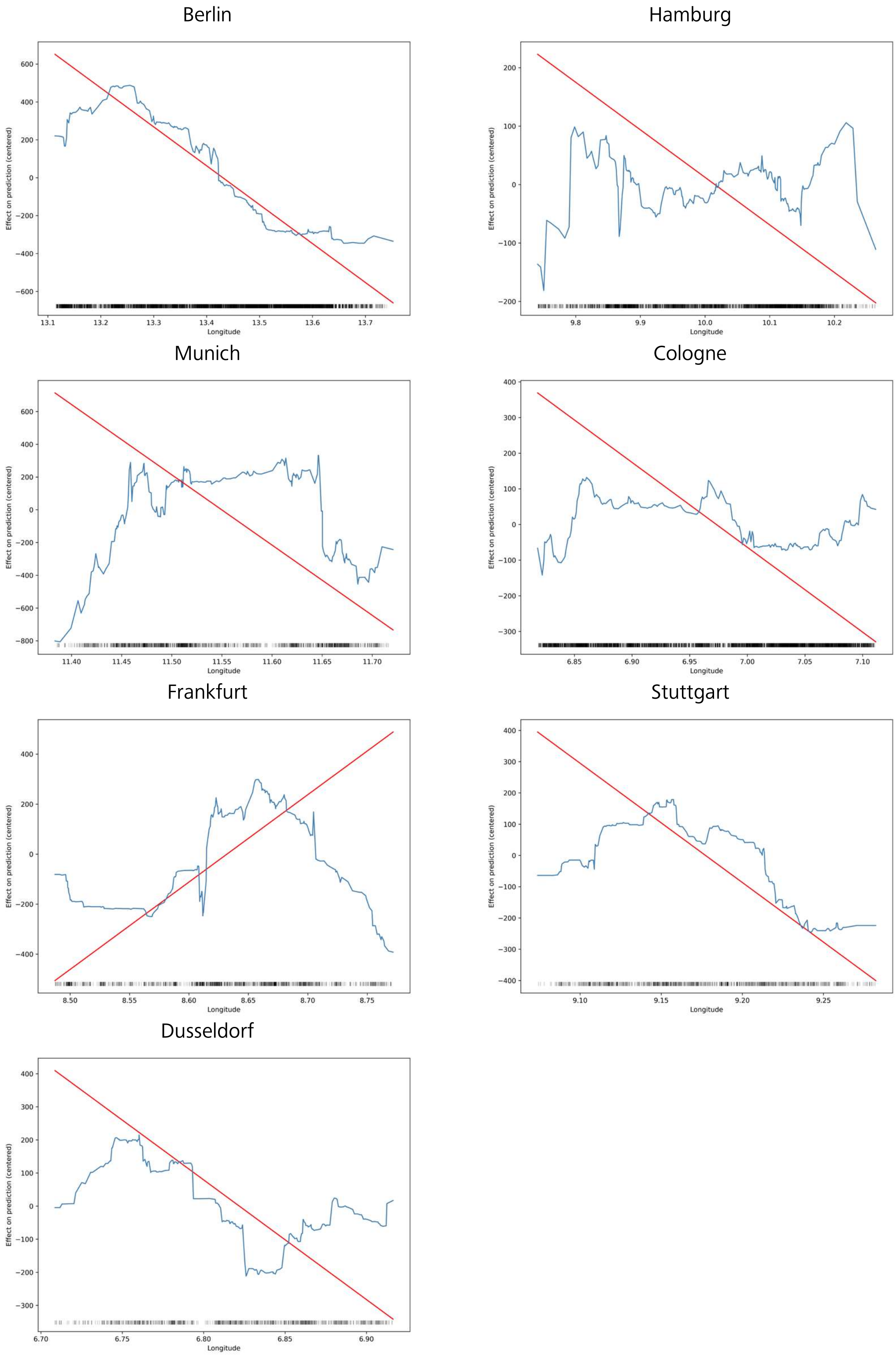


Figure 2.13: Condominiums – Unemployment ratio

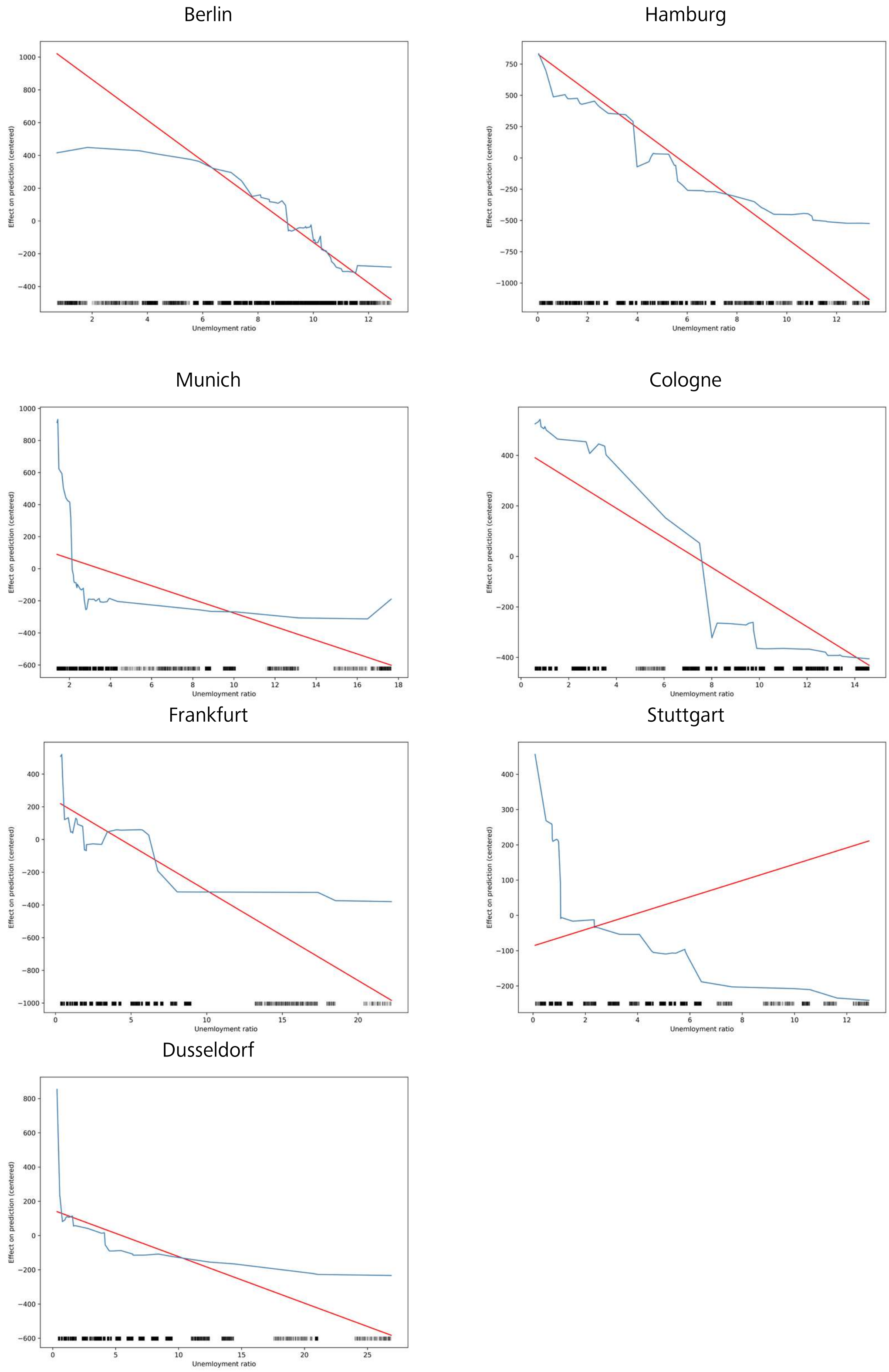
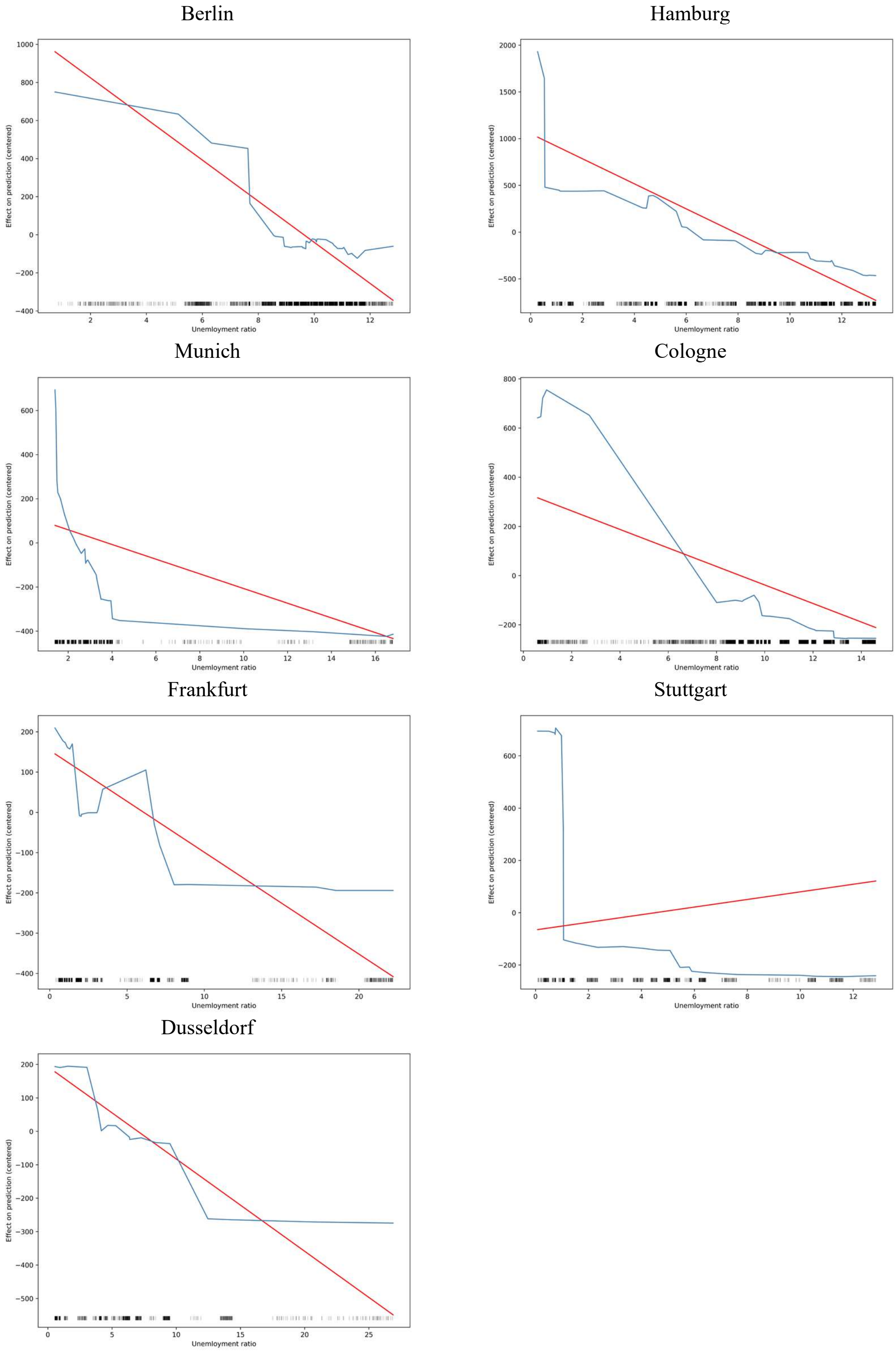


Figure 2.14: Single-family homes – Unemployment ratio



2.6 Conclusion

This study is intended to introduce eXplainable Artificial Intelligence (XAI) in a real estate context, and updates the existing literature with the application of Accumulated Local Effects plots (ALE). Compared to the Partial Dependence Plots (PDP), which are commonly used in real estate research to date, ALE plots can also handle correlated features and are therefore more suitable for real-estate-related problems. We use a dataset consisting of 61,763 condominiums and 19,403 single-family homes for the Top-7 cities of Germany to empirically apply our analysis. We are interested in identifying the most value-determining features of the two property subtypes, and then analyzing them separately with respect to existing non-linearities. We apply Permutation Feature Importance (PFI) to identify the most important features, and ALE plots to visualize their individual effects. As an underlying machine learning (ML) model, we use the XGBoost algorithm for a hedonic estimation of the properties' market values.

The PFI analysis shows that the same features play a predominant role for both condominiums and single-family homes. The valuation year is by far the most important feature. In addition to the valuation year, a mix of structural, location-related and socio-economic features influence the market value of the properties. Among the structural characteristics, the most important are the year of construction, the living area and the lot size. In terms of location features, the latitude and longitude are decisive in terms of market values. Socio-economically, we find that the unemployment ratio seems to play a partially important role.

To the best of our knowledge, we are the first to use ALE plots for visualizing individual effects on the market value, and we see that both non-linear and linear effects can be observed. In terms of year of construction, our results show that for both condominiums and single-family homes, the u-shaped transformation traditionally used for Hedonic Price Models (HPMs) is not evident. Properties with newer construction years are valued much higher than is the case for older buildings. We can confirm that properties with middle age tend to have lower market values. The results for living area among condominiums are particularly interesting. The ALE plots show no clear trend here for the cities studied, but that this effect varies greatly and is clearly non-linear. For single-family homes, on the other hand, a linear trend can be observed for all cities. An approximately linear trend is also evident for the year of valuation feature. The analysis of the ALE plots of latitude and longitude shows that market values within the city can vary greatly, depending on the particular location. Our results show that market values for condominiums tend to be highest in the centers of cities, whereas values for single-family homes tend to be highest outside city centers. Both findings appear intuitive and are in line with the prevailing opinion within the real estate industry. The effect of the unemployment rate is also clearly non-linear and different across the cities analyzed. In general, however, the presumed negative influence is evident.

In summary, the ALE plots provide a deeper understanding of the fundamentals of real estate markets and either empirically confirm long-established rules of thumb or, as in the case of living area for condominiums, challenge them. Our results show that linear relationships indeed occur in the housing market. Here, parametric estimates can also provide valuable results. However, the analysis of the features year of construction, living area, lot size, latitude, longitude and unemployment ratio reveal non-linear effects. Therefore, non-parametric ML approaches seem to be the right choice. The ALE plots offer a way to represent these effects in a well-founded way and thus make an important contribution to the housing market literature. Moreover, the effects seen in the ALE plots can be used to optimize parametric and semi-parametric models in order to achieve a higher predictive performance.

Model-agnostic methods are still a rather young field of research, but will play a major role in the acceptance of ML methods in the future, as they allow us to look into the "black box "of ML approaches and are thus an important tool in deciphering them. The tradeoff between explainability and model performance can thus be mitigated in the long run. However, further research is still needed before widespread use is possible. With regard to future work, for example, a more in-depth analysis of neighborhood and environmental features constitutes a promising field of application. While the focus of this study was rather on the analysis of structural characteristics, the analysis of neighborhood and environmental features in the context of property prices has become an important part of real estate research and can, in our opinion, also benefit from the advantageous properties of XAI approaches and especially those of ALE plots.

2.7 Appendix

2.7.1 Appendix I – Micro Score

Our gravity model can be described using an activity function $f(A_p)$ and a distance function $f(D_{i,p})$:

$$A_{i,p} = \sum f(A_p) f(D_{i,p}). \tag{3}$$

$A_{i,p} \in [0{,}100]$ denotes the accessibility of point i for the POI p, whereby the activity function $f(A_p)$ specifies the relative importance of POI p, with $f(A_p) \in [0{,}1]$. $f(D_{i,p})$ measuring the travel time from point i to the POI p by using a non-symmetric sigmoidal distance function. The travel time was obtained for the selected POIs via Open Street Map, and normalized using the following function:

$$L(x) = \frac{K}{(1 + Qe^{0.5x})^{\frac{1}{v}}}, \tag{4}$$

where $K, Q \in \mathbb{R}$ and $v \in \mathbb{R}^+$ are defined for all possible distances $x \in \mathbb{R}$. Furthermore, we have:

$$\begin{gathered} K = (1 + Q)^{1+v}, \\ Q = v \cdot \exp(B \cdot x^*), \\ v = \frac{\exp(B \cdot x^*) - 1}{\ln(y_i) - 1}, \end{gathered} \tag{5}$$

where x^* denotes a feature specific point of inflection and y^* is 0.5.

2.7.2 Appendix II – Accumulated Local Effects

Accumulated local effects (ALE), developed by Apley & Zhu (2020), is a feature effect approach that shows how a feature influences a prediction on average. The technical explanation of how this can be achieved is given in the following section.

Assume that $\boldsymbol{x}$ is the full feature space containing d variables and $\boldsymbol{y}$ the response variable. $\hat{f}$ is a fitted supervised machine learning model that is differentiable and uses $\boldsymbol{x}$ to predict $\boldsymbol{y}$. Define $\boldsymbol{x}_j$ as the feature of interest and $\boldsymbol{x}_{-j}$ the complement set of features, $j \in \{1, \dots d\}$. X_j represents the j^{th} feature as a random variable. Then, the ALE main effect of $\boldsymbol{x}_j$ can be calculated:

$$\hat{f}_{j,ALE} = \int_{z_{0,j}}^{x_j} E\left[\frac{\partial \hat{f}(X_j, X_{-j})}{\partial X_j} \mid X_j = z_j\right] dz_j - constant, \tag{6}$$

with $z_{0,j}$ being a lower bound of X_j. Usually, $z_{0,j}$ is defined as $\min\{\boldsymbol{x}_j\}$. The expected value E is computed conditional on the representation of $\boldsymbol{x}_j$ and over the marginal distribution of $\boldsymbol{x}_{-j}$. The constant is subtracted to center the plot.

Since not every machine learning model is differentiable, Apley & Zhu (2020) introduced a way to estimate the ALE for any supervised machine learning model. Therefore, the value range of the j^{th} feature is divided into K intervals $N_j(k)$, $\{N_j(k) = (z_{k-1,j}, z_{k,j}]: k = 1,2, \dots K\}$, where $z_{j,k}$ refers to the upper and $z_{k-1,j}$ the lower boundary of interval k. Furthermore, x^* is a specific value of $\boldsymbol{x}_j$ and $k_j(x^*)$ denotes the index of the interval x^* belongs to. $n_j(k)$ is the number of observations in each interval k and $\boldsymbol{x}_{i,-j}$ represents the observations of the remaining features, $i \in \{1, 2, \dots N\}$.

Before we can compute the main effect $\hat{f}_{j,ALE}$, the uncentered ALE $\hat{g}_{j,ALE}$ of the j^{th} feature has to be calculated for every $x^* \in (z_{0,j}, z_{K,j}]$, where $z_{0,j}$ is just below the minimum observation of $\{x_{i,j}: i = 1, \dots n\}$ and $z_{K,j}$ is the maximum observation of $\{x_{i,j}: i = 1, \dots n\}$:

$$\hat{g}_{j,AL}(x^*) = \sum_{k=1}^{k_j(x^*)} \frac{1}{n_j(k)} \sum_{\{i: x_{i,j} \in N_j(k)\}} \left[\hat{f}(z_{k,j}, \boldsymbol{x}_{i,-j}) - \hat{f}(z_{k-1,j}, \boldsymbol{x}_{i,-j})\right]. \tag{7}$$

Now the ALE main effect estimator can be computed by subtracting an estimate of $E[g_{j,ALE}(X_j)]$:

$$\begin{aligned} \hat{f}_{j,ALE}(x^*) &= \hat{g}_{j,ALE}(x^*) - \frac{1}{n}\sum_{i=1}^{n} \hat{g}_{j,ALE}(x_{i,j}) \\ &= \hat{g}_{j,ALE}(x^*) - \frac{1}{n}\sum_{k=1}^{K} n_j(k) \cdot \hat{g}_{j,ALE}(z_{k,j}). \end{aligned} \tag{8}$$

2.7.3 Appendix III – Evaluation Metrics at City Level

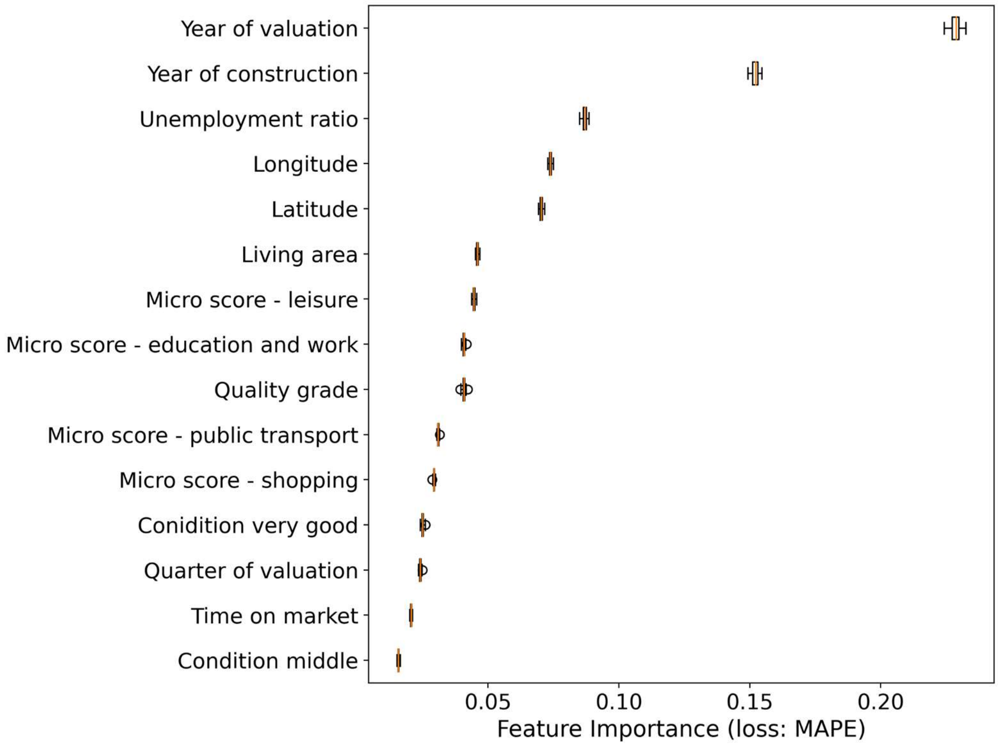
Table 2.9: XGBoost results at city level

	XGBoost		OLS	
Metrics	**Condominiums**	**Single-family homes**	**Condominiums**	**Single-family homes**
		Berlin		
MAPE	0.1431	0.1415	0.2311	0.1837
MdAPE	0.0984	0.0988	0.1741	0.1342
PE(10)	0.5059	0.5062	0.3077	0.3927
PE(20)	0.7752	0.7910	0.5593	0.6698
R^2	0.8052	0.7544	0.6192	0.6074
		Hamburg		
MAPE	0.1291	0.1505	0.1990	0.2039
MdAPE	0.0806	0.1047	0.1455	0.1466
PE(10)	0.5697	0.4805	0.3721	0.3595
PE(20)	0.8199	0.7710	0.6369	0.6450
R^2	0.7936	0.7245	0.6123	0.5288
		Munich		
MAPE	0.1051	0.1735	0.1718	0.2016
MdAPE	0.0618	0.0981	0.1233	0.1251
PE(10)	0.6559	0.5099	0.4163	0.4154
PE(20)	0.8772	0.7670	0.7104	0.7095
R^2	0.8079	0.6381	0.5734	0.5264
		Cologne		
MAPE	0.1278	0.1232	0.2008	0.1587
MdAPE	0.0865	0.0878	0.1469	0.1180
PE(10)	0.5577	0.5530	0.3609	0.4381
PE(20)	0.8267	0.8220	0.6398	0.7217
R^2	0.8388	0.7256	0.6820	0.5672
		Frankfurt		
MAPE	0.1121	0.1571	0.2124	0.1866
MdAPE	0.0738	0.1041	0.1571	0.1222
PE(10)	0.6080	0.4816	0.3360	0.4351
PE(20)	0.8512	0.7640	0.6050	0.7070
R^2	0.8312	0.6639	0.6061	0.5431
		Stuttgart		
MAPE	0.1160	0.1513	0.1449	0.1859
MdAPE	0.0824	0.1033	0.1082	0.1337
PE(10)	0.5772	0.4851	0.4698	0.3720
PE(20)	0.8508	0.7778	0.7607	0.6845
R^2	0.8002	0.6869	0.6855	0.5518
		Dusseldorf		
MAPE	0.1334	0.1848	0.1907	0.2019
MdAPE	0.0930	0.1270	0.1438	0.1450
PE(10)	0.5296	0.4158	0.3622	0.3624
PE(20)	0.7921	0.6695	0.6440	0.6474
R^2	0.7793	0.4909	0.5867	0.4188

2.7.4 Appendix IV – Permutation Feature Importance

To get a better understanding on how to interpret the PFI results, Figure 2.15 shows the PFI plot including the 14 most important features for condominiums in Munich.

Figure 2.15: Condominiums – PFI feature importance – Berlin



The features are ranked on the y-axis from most important to least important. The x-axis provides information of how much the average prediction accuracy changes when the values of the features are permutated 100 times. The individual feature importance ratios in Figures 2.15 can be interpreted like box plots. The orange line in the middle represents the median values and the ends of the bars the 25% and 75% quantiles. Furthermore the 1.5x interquartile range (IQR) and individual outliers are shown.

The year of valuation is seen to have the highest impact on valuation accuracy for condominiums in Berlin. By randomly permuting the year of valuation 100 times, the models MAPE increases by 23.0% for condominiums, while the other features are kept constant. The second most important feature with an increase of 15.3% in the MAPE is the year of construction for condominiums, followed by the unemployment ratio.

The remaining PFI plots in Figures 2.16 and 2.17 can be interpreted in the same way.

Figure 2.16: Feature importance – Condominiums

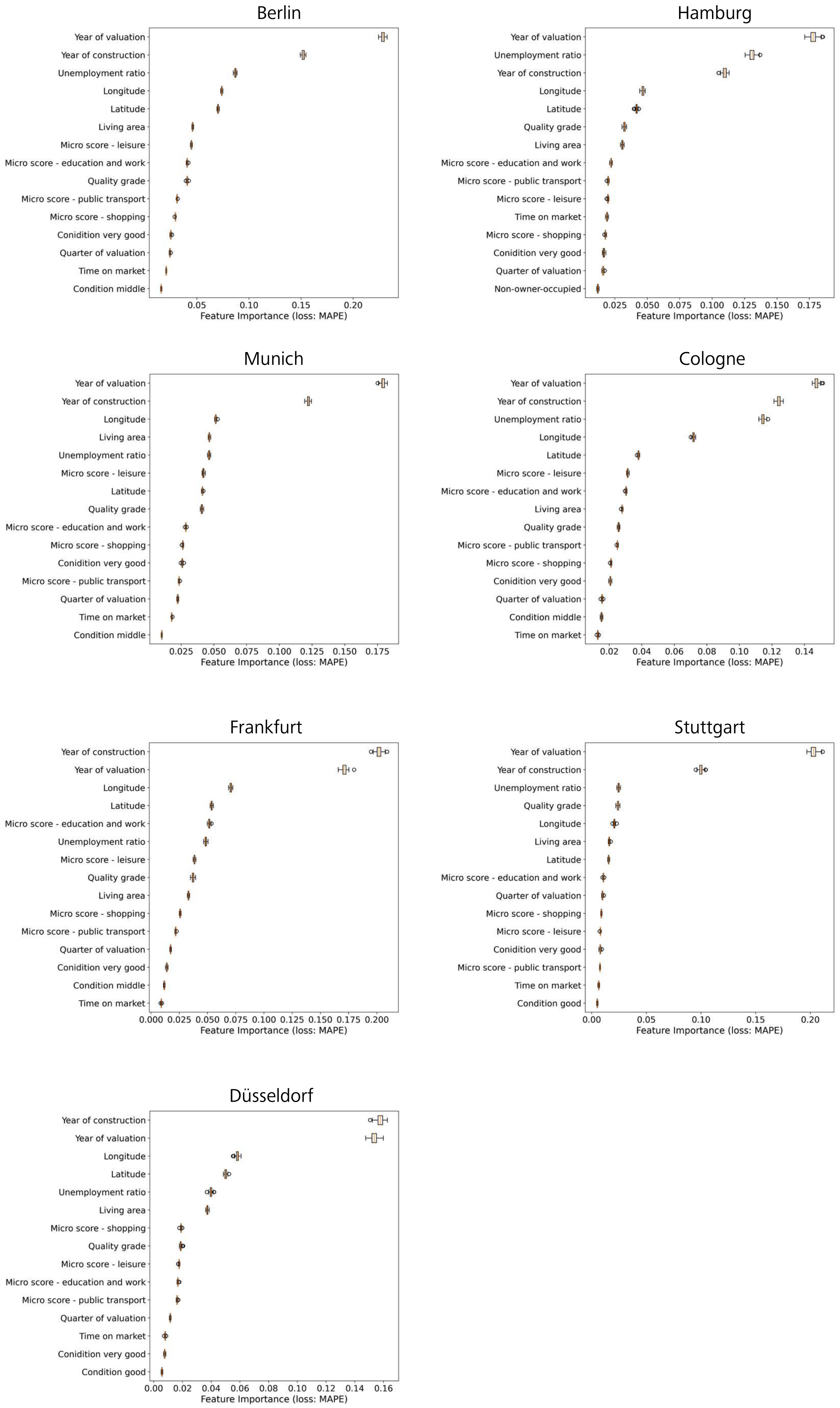
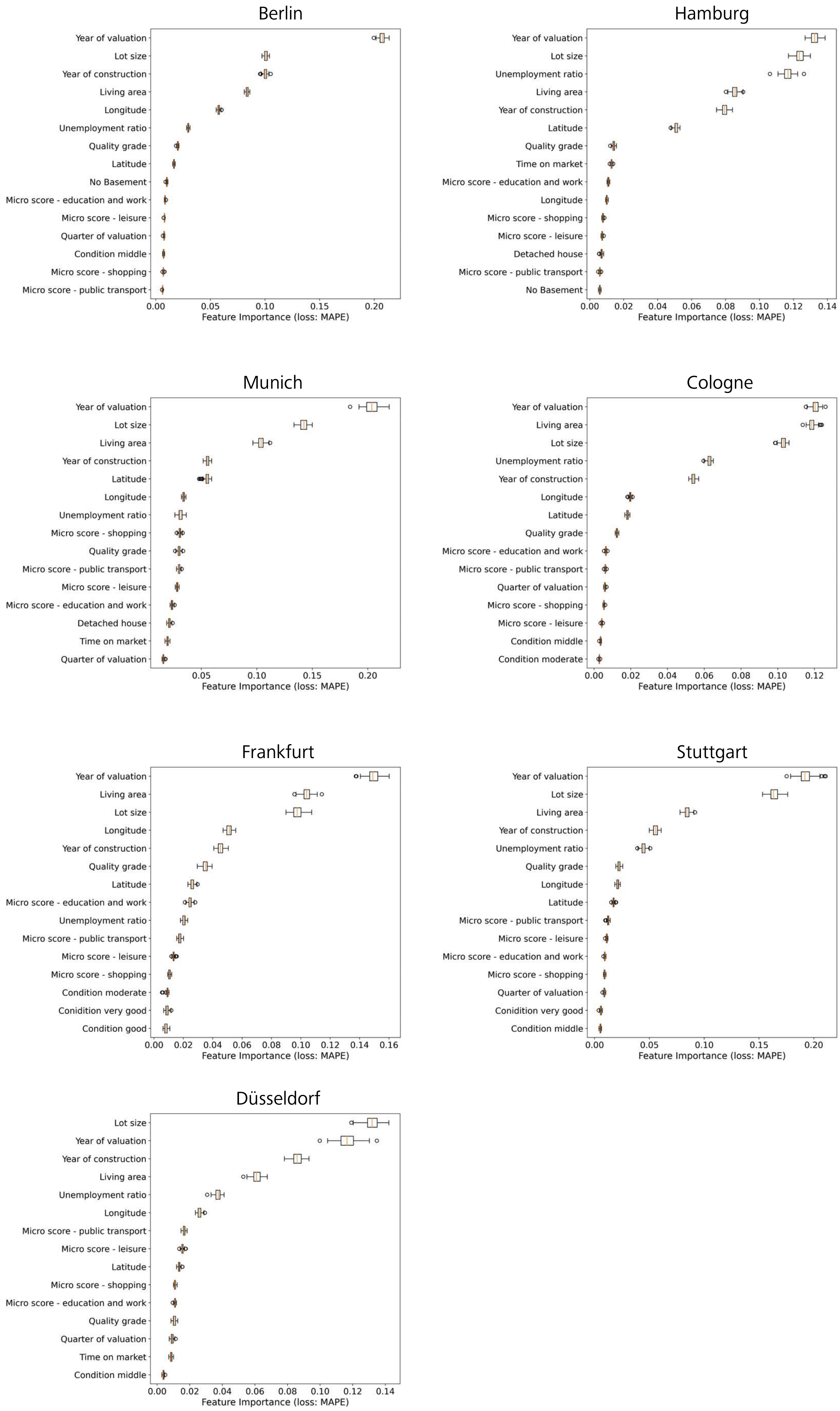


Figure 2.17: Feature importance – Single-family homes



2.7.5 Appendix V – Using ALE Plots to Optimize an Ordinary Least Squares Regression

Below, it is shown how, for example, the results of the ALEs can be used to optimize the benchmark OLS used in this paper to achieve higher predictive performance. To do this, as the first step, the ALE plots for each city for condominiums and single-family homes have to be analyzed and relevant inflections of the function have to be found. Then, the feature space of the OLS is split according to those functions in order to better adapt the data to the non-linearities found. Table 2.10 shows the exact feature splits. If we take the year of construction of single-family houses in Berlin as an example, this means that instead of one feature with a feature space of 1900-2020, we now have three features, each with feature spaces of 1900-1953, 1954-2016, and 2016-2020.

Table 2.10: Feature space splits

	Condominiums	**Single-family homes**
Berlin		
Year of construction	$x \leq 1960$ $1960 < x \leq 2008$ $2008 < x$	$x \leq 1953$ $1953 < x \leq 2016$ $2016 < x$
Year of valuation	$x \leq 2016$ $2016 < x$	$x \leq 2016$ $2016 < x$
Latitude	$x \leq 52.52$ $52.52 < x$	$x \leq 52.59$ $52.59 < x$
Longitude	$x \leq 13.415$ $13.415 < x \leq 13.75$ $13.75 < x$	$x \leq 13.25$ $13.25 < x \leq 13.51$ $13.51 < x$
Unemployment ratio	$x \leq 7.7$ $7.7 < x \leq 11$ $11 < x$	$x \leq 7.68$ $7.68 < x \leq 8.01$ $8.01 < x$
Living area	$x \leq 62.5$ $62.5 < x \leq 110$ $110 < x$	$x \leq 150$ $150 < x \leq 200$ $200 < x$
Lot size	-	$x \leq 500$ $500 < x$
Hamburg		
Year of construction	$x \leq 1918$ $1918 < x \leq 1955$ $1955 < x \leq 1977$ $1977 < x$	$x \leq 1941$ $1941 < x \leq 1968$ $1968 < x$
Year of valuation	$x \leq 2016$ $2016 < x$	$x \leq 2015$ $2015 < x \leq 2018$ $2018 < x$
Latitude	$x \leq 53.545$ $53.545 < x \leq 53.60$ $53.60 < x$	$x \leq 53.515$ $53.515 < x \leq 53.555$ $53.555 < x \leq 53.675$ $53.765 < x$
Longitude	$x \leq 9.975$ $9.975 < x \leq 10$ $10 < x \leq 10.05$ $10.05 < x$	$x \leq 9.8$ $9.8 < x \leq 10.16$ $10.16 < x$
Unemployment ratio	$x \leq 6$ $6 < x$	$x \leq 0.8$ $0.8 < x$
Living area	$x \leq 35$ $35 < x \leq 100$ $100 < x$	$x \leq 150$ $150 < x \leq 235$ $235 < x$
Lot size	-	$x \leq 400$ $400 < x \leq 400$ $800 < x$

	Munich	
Year of construction	$x \leq 1952$ $1952 < x \leq 1980$ $1980 < x \leq 2009$ $2009 < x$	$x \leq 1950$ $1950 < x \leq 2015$ $2015 < x$
Year of valuation	$x \leq 2017$ $2017 < x$	$x \leq 2017$ $2017 < x$
Latitude	$x \leq 48.15$ $48.15 < x \leq 48.21$ $48.21 < x$	$x \leq 48.165$ $48.165 < x$
Longitude	$x \leq 11.58$ $11.58 < x \leq 11.66$ $11.66 < x$	$x \leq 11.465$ $11.465 < x \leq 11.647$ $11.647 < x$
Unemployment ratio	$x \leq 3$ $3 < x$	$x \leq 4$ $4 < x$
Living area	$x \leq 60$ $60 < x \leq 150$ $150 < x$	$x \leq 135$ $135 < x \leq 225$ $225 < x$
Lot size	-	$x \leq 510$ $510 < x \leq 930$ $930 < x$
	Cologne	
Year of construction	$x \leq 1977$ $1977 < x \leq 2007$ $2007 < x$	$x \leq 1970$ $1970 < x \leq 2000$ $2000 < x$
Year of valuation	$x \leq 2017$ $2017 < x$	$x \leq 2016$ $2016 < x$
Latitude	$x \leq 50.895$ $50.895 < x \leq 50.947$ $50.947 < x \leq 51.02$ $51.02 < x$	$x \leq 50.875$ $50.875 < x \leq 50.935$ $50.935 < x$
Longitude	$x \leq 6.93$ $6.93 < x \leq 6.98$ $9.98 < x \leq 7.015$ $7.015 < x$	$x \leq 6.865$ $6.865 < x \leq 6.995$ $6.995 < x$
Unemployment ratio	$x \leq 8$ $8 < x$	$x \leq 8$ $8 < x$
Living area	$x \leq 100$ $100 < x \leq 138.5$ $138.5 < x$	$x \leq 140$ $140 < x \leq 250$ $250 < x$
Lot size	-	$x \leq 300$ $300 < x \leq 950$ $950 < x$
	Frankfurt	
Year of construction	$x \leq 1945$ $1945 < x \leq 1975$ $1975 < x \leq 2009$ $2009 < x$	$x \leq 1980$ $1980 < x \leq 2015$ $2015 < x$
Year of valuation	$x \leq 2017$ $2017 < x$	$x \leq 2017$ $2017 < x$
Latitude	$x \leq 50.112$ $50.112 < x \leq 50.132$ $50.132 < x \leq 50.138$ $50.143 < x$	$x \leq 50.16$ $50.16 < x$
Longitude	$x \leq 8.68$ $8.68 < x \leq 8.74$ $8.74 < x$	$x \leq 8.57$ $8.57 < x \leq 8.66$ $8.66 < x$
Unemployment ratio	$x \leq 9$ $9 < x$	$x \leq 2$ $2 < x \leq 8$ $8 < x$
Living area	$x \leq 55$ $55 < x \leq 150$ $150 < x$	$x \leq 150$ $150 < x \leq 250$ $250 < x$
Lot size	-	$x \leq 400$ $400 < x$

	Stuttgart	
Year of construction	$x \leq 1967$ $1967 < x \leq 1988$ $1988 < x \leq 2009$ $2009 < x$	$x \leq 1970$ $1970 < x \leq 2009$ $2009 < x$
Year of valuation	$x \leq 2016$ $2016 < x$	$x \leq 2015$ $2015 < x$
Latitude	$x \leq 48.75$ $48.75 < x \leq 48.75$ $48.79 < x$	$x \leq 48.81$ $48.81 < x$
Longitude	$x \leq 9.15$ $9.15 < x \leq 9.19$ $9.19 < x$	$x \leq 9.165$ $9.165 < x$
Unemployment ratio	$x \leq 1$ $1 < x$	$x \leq 1.5$ $1.5 < x$
Living area	$x \leq 30$ $30 < x \leq 125$ $125 < x$	$x \leq 150$ $250 < x$
Lot size	-	$x \leq 680$ $680 < x$
	Dusseldorf	
Year of construction	$x \leq 1971$ $1971 < x \leq 2009$ $2009 < x$	$x \leq 1952$ $1952 < x$
Year of valuation	$x \leq 2016$ $2016 < x$	$x \leq 2017$ $2017 < x$
Latitude	$x \leq 51.20$ $51.20 < x$	$x \leq 51.21$ $51.21 < x$
Longitude	$x \leq 6.76$ $6.76 < x \leq 6.785$ $6.785 < x$	$x \leq 6.765$ $6.765 < x$
Unemployment ratio	$x \leq 1.3$ $1.3 < x$	$x \leq 12.5$ $12.5 < x$
Living area	$x \leq 117$ $117 < x$	$x \leq 175$ $175 < x$
Lot size	-	$x \leq 790$ $790 < x$

To test wheter the splits improved the model performance, we again used five-fold cross validation with the same evaluation metrics as before. Table 2.11 shows the overall results of the optimized OLS with the basic OLS as a benchmark. We can see that the feature splits clearly improved the OLS regarding all evaluation metrics for both condominiums and single-family homes.

Table 2.11: Optimized OLS results for all Top-7 cities

	$\mathbf{OLS_{optimized}}$		**OLS**	
Metrics	**Condominiums**	**Single-family homes**	**Condominiums**	**Single-family homes**
MAPE	0.1841	0.1705	0.1986	0.1834
MdAPE	0.1345	0.1243	0.1467	0.1314
PE(10)	0.3934	0.4149	0.3654	0.3999
PE(20)	0.6780	0.7072	0.6396	0.6833
R^2	0.6626	0.6148	0.6230	0.5623

The results on a city level are shown in Table 2.12. The optimized OLS outperforms the basic OLS for all cities regarding all evaluation metrics.

Table 2.12: Optimized OLS results at city level

	$OLS_{optimized}$		OLS	
Metrics	**Condominiums**	**Single-family homes**	**Condominiums**	**Single-family homes**
		Berlin		
MAPE	0.2216	0.1723	0.2311	0.1837
MdAPE	0.1659	0.1268	0.1741	0.1342
PE(10)	0.3233	0.4138	0.3077	0.3927
PE(20)	0.5816	0.6956	0.5593	0.6698
R^2	0.6462	0.6512	0.6192	0.6074
		Hamburg		
MAPE	0.1823	0.1824	0.1990	0.2039
MdAPE	0.1279	0.1301	0.1455	0.1466
PE(10)	0.4037	0.3911	0.3721	0.3595
PE(20)	0.6910	0.6879	0.6369	0.6450
R^2	0.6644	0.6512	0.6123	0.5288
		Munich		
MAPE	0.1633	0.1813	0.1718	0.2016
MdAPE	0.1171	0.1213	0.1233	0.1251
PE(10)	0.4388	0.4173	0.4163	0.4154
PE(20)	0.7350	0.7171	0.7104	0.7095
R^2	0.5921	0.6003	0.5734	0.5264
		Cologne		
MAPE	0.1839	0.1509	0.2008	0.1587
MdAPE	0.1304	0.1140	0.1469	0.1180
PE(10)	0.3924	0.4444	0.3609	0.4381
PE(20)	0.6863	0.7418	0.6398	0.7217
R^2	0.7260	0.6068	0.6820	0.5672
		Frankfurt		
MAPE	0.1810	0.1796	0.2124	0.1822
MdAPE	0.1344	0.1237	0.1571	0.1247
PE(10)	0.3876	0.4210	0.3360	0.4188
PE(20)	0.6761	0.6951	0.6050	0.6886
R^2	0.6835	0.5683	0.6061	0.5601
		Stuttgart		
MAPE	0.1310	0.1748	0.1449	0.1859
MdAPE	0.0971	0.1317	0.1082	0.1337
PE(10)	0.5120	0.3863	0.4698	0.3720
PE(20)	0.8060	0.6949	0.7607	0.6845
R^2	0.7430	0.6038	0.6855	0.5518
		Dusseldorf		
MAPE	0.1734	0.1994	0.1907	0.2019
MdAPE	0.1325	0.1420	0.1438	0.1450
PE(10)	0.3914	0.3476	0.3622	0.3624
PE(20)	0.6862	0.6698	0.6440	0.6474
R^2	0.6344	0.4442	0.5867	0.4188

2.8 References

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, *39*(2), 1772–1778.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(4), 1059–1086.

Belke, A., & Keil, J. (2018). Fundamental Determinants of Real Estate Prices: A Panel Study of German Regions. *International Advances in Economic Research*, *24*(1), 25–45.

Bogin, A. N., & Shui, J. (2020). Appraisal Accuracy and Automated Valuation Models in Rural Areas. *The Journal of Real Estate Finance and Economics*, *60*(1-2), 40–52.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial Dependence, Housing Submarkets, and House Price Prediction. *The Journal of Real Estate Finance and Economics*, *35*(2), 143–160.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2010). Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions. *Journal of Real Estate Research*, *32*(2), 139–159.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Brunauer, W. A., Lang, S., Wechselberger, P., & Bienert, S. (2010). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *The Journal of Real Estate Finance and Economics*, *41*(4), 390–411.

Cajias, M. (2018). Is there room for another hedonic model? *Journal of European Real Estate Research*, *11*(2), 204–245.

Cajias, M., & Freudenreich, P. (2018). Exploring the determinants of liquidity with big data – market heterogeneity in German markets. *Journal of Property Investment & Finance*, *36*(1), 3–18.

Cheng, A. C. S., & Fung, M. K. (2015). Determinants of Hong Kong's Housing Prices. *Journal of Economics, Business and Management*, *3*(3), 352–355.

Chiarazzo, V., Caggiani, L., Marinelli, M., & Ottomanelli, M. (2014). A Neural Network based Model for Real Estate Price Estimation Considering Environmental Quality of Property Location. *Transportation Research Procedia*, *3*, 810–817.

Chun Lin, C., & Mohan, S. B. (2011). Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis*, *4*(3), 224–243.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental Variables and Real Estate Prices. *Urban Studies*, *38*(11), 1989–2000.

Dubin, R. A. (1988). Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms. *The Review of Economics and Statistics*, *70*(3), 466.

Fan, G.-Z., Ong, S. E., & Koh, H. C. (2006). Determinants of House Price: A Decision Tree Approach. *Urban Studies*, *43*(12), 2301–2315.

Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research*, *20*(177),1–81.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189-1232.

Galkin, F., Aliper, A., Putin, E., Kuznetsov, I., Gladyshev, V. N., & Zhavoronkov, A. (2020). Human Gut Microbiome Aging Clock Based on Taxonomic Profiling and Deep Learning. iScience, *23*(6), 101199.

Grum, B., & Govekar, D. K. (2016). Influence of Macroeconomic Factors on Prices of Real Estate in Various Cultural Environments: Case of Slovenia, Greece, France, Poland and Norway. *Procedia Economics and Finance*, *39*, 597–604.

Gu, G., & Xu, B. (2017). Housing Market Hedonic Price Study Based on Boosting Regression Tree. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *21*(6), 1040–1047.

Handy, S. L., & Clifton, K. J. (2001). Evaluating Neighborhood Accessibility: Possibilities and Practicalities. *Journal of Transportation and Statistics*, *4*(2-3), 67-78.

Hastie, T., Friedman, J. & Tibshirani, R. (2001). *The Elements of Statistical Learning.* New York: Springer.

Ho, W. K., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, *38*(1), 48–70.

Huang, Y., & Dall'erba, S. (2021). Does Proximity to School Still Matter Once Access to Your Preferred School Zone Has Already Been Secured? *The Journal of Real Estate Finance and Economics*, *62*(4), 548–577.

Kagie, M., & van Wezel, M. (2007). Hedonic price models and indices based on boosting applied to the Dutch housing market. *Intelligent Systems in Accounting, Finance & Management*, *15*(3-4), 85–106.

Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal of Portfolio Management*, *43*(6), 202–211.

Levantesi, S., & Piscopo, G. (2020). The Importance of Economic Variables on London Real Estate Market: A Random Forest Approach. *Risks*, *8*(4), 112.

Lorenz, F., Willwersch, J., Cajias, M., & Fuerst, F. (2022). Interpretable machine learning for real estate market analysis. Journal of *Real Estate Economics*, Forthcoming.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. *Housing Economics and Public Policy, 1*, 67–89.

Martin, R. (2011). The local geographies of the financial crisis: From the housing bubble to economic recession and beyond. *Journal of Economic Geography*, *11*(4), 587–618.

Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*, *12*(1), 134–150.

McCluskey, W. J., McCord, M., Davis, P. T., Haran, M., & McIlhatton, D. (2013). Prediction accuracy in mass appraisal: A comparison of modern approaches. *Journal of Property Research*, *30*(4), 239–265.

Metzner, S., & Kindt, A. (2018). Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties. *International Journal of Housing Markets and Analysis*, *11*(1), 73–100.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

Nobis, C. & Kuhnimhof, T. (2018). *Mobilität in Deutschland – MiD: Ergebnisbericht*. Bundesministerium für Verkehr und digitale Infrastruktur, Bonn.

Pace, R. K., & Hayunga, D. (2020). Examining the Information Content of Residuals from Hedonic and Spatial Models Using Trees and Forests. *The Journal of Real Estate Finance and Economics*, *60*(1-2), 170–180.

Peterson, S., & Flanagan, A. (2009). Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research*, *31*(2), 147–164.

Powe, N. A., Garrod, G. D., & Willis, K. G. (1995). Valuation of urban amenities using an hedonic price model. *Journal of Property Research*, *12*(2), 137–147.

Sangani, D., Erickson, K. & Hasan, M.A. (2017). *Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting*. In IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Orlando.

Schnaidt, T., & Sebastian, S. (2012). German valuation: Review of methods and legal framework. *Journal of Property Investment & Finance*, *30*(2), 145–158.

Schulz, R., Wersing, M., & Werwatz, A. (2014). Automated valuation modelling: a specification exercise. *Journal of Property Research*, *31*(2), 131–153.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, *13*(1), 1–44.

Stang, M., Krämer, B., Nagl, C., & Schäfers, W. (2022). From human business to machine learning—Methods for automating real estate appraisals and their practical implications. *Zeitschrift Für Immobilienökonomie*, Advanced online publication.

Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, *174*, 433–442.

van Wezel, M., Kagie, M., & Potharst, R. (2005). Boosting the accuracy of hedonic pricing models. *Report / Econometric Institute, Erasmus University Rotterdam* (EI 2005-50).

Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. (2020). Residential housing prices: Impact of housing characteristics, accessibility and neighbouring apartments – a case study of Dortmund, Germany. *Urban, Planning and Transport Research*, *8*(1), 44–70.

Worzala, E., Lenk, M., & Silva, A. (1995). An Exploration of Neural Networks and Its Application to Real Estate Valuation. *Journal of Real Estate Research*, *10*(2), 185–201.

Yang, J., Bao, Y., Zhang, Y., Li, X., & Ge, Q. (2018). Impact of Accessibility on Housing Prices in Dalian City of China Based on a Geographically Weighted Regression Model. *Chinese Geographical Science*, *28*(3), 505–515.

3 Automated Valuation Models: Improving Model Performance by Choosing the Optimal Spatial Training Level

3.1 Abstract

The academic community has discussed using Automated Valuation Models (AVMs) in the context of traditional real estate valuations and their performance for several decades. Most studies focus on finding the best method for estimating property values. One aspect that has not yet to be studied scientifically is the appropriate choice of the spatial training level. The published research on AVMs usually deals with a manually defined region and fails to test the methods used on different spatial levels. Our research aims to investigate the impact of training AVM algorithms at different spatial levels regarding valuation accuracy. We use a dataset with 1.2 million residential properties from Germany and test four methods: Ordinary Least Squares, Generalized Additive Models, eXtreme Gradient Boosting and Deep Neural Network. Our results show that the right choice of spatial training level can significantly impact the model performance, and that this impact varies across the different methods.

Keywords – Automated Valuation Models (AVMs), Machine Learning, Spatial Training Level, Model Performance, Valuation Accuracy

3.2 Introduction

The academic community has discussed using Automated Valuation Models (AVMs) in the context of traditional real estate valuations and their performance for several decades, and practitioners are also now increasingly scrutinizing it. Most studies focus on the comparison of different statistical methods. Accordingly, a large body of literature compares traditional hedonic models with more modern machine learning (ML) approaches or approaches from spatial econometrics (see, e.g., Pace & Hayunga, 2020). These studies aim to identify which method is best suited for estimating real estate values or prices.

Besides the method selection, AVMs can be optimised in many other areas. For example, the selection, cleaning and preparation of data play an important role for the overall performance of the AVM. Another aspect is the choice of spatial level to train the selected methods. This is decisive for determining which data are ultimately included in the estimation of the AVM and, thus, what information is used or ignored. Thanks to georeferencing, models can, in principle, be trained at any level. For example, a model can be trained at the city level, the associated commuter belt, or even nationwide. However, this aspect has received little to no attention from the academic community until now.

The published research on AVMs usually deals only with a manually defined region and fails to test the methods used on different spatial levels. One reason for this might be that historically, the availability of suitable real estate data[15] for academic purposes has been limited. Therefore, analyses could only be conducted in the limited area where the data was available. However, data availability has improved massively in recent years, so this has become less of a factor (Mortgage Bankers Association, 2019). In the meantime, there are providers of real estate-related data in almost every country, which centrally force a collection of existing data and make them available for further analysis. Another reason could be the usually assumed heterogeneity of real estate markets. Traditionally, real estate markets are believed to have a certain regionality, meaning that data from other diverging regions would not provide additional explanatory power. However, the fundamental question arises as to whether this heterogeneity is generally present or whether there are not also basic characteristics that apply consistently to all markets. If this is the case, achieving a higher degree of valuation accuracy may be possible by adding further data from different markets.

[15] To avoid a structural break within the dataset, the data should ideally come from one source or have been collected according to the same criteria.

Therefore, it raises the question of whether considering different spatial levels for training AVMs could be an important and undervalued factor in enhancing their valuation accuracy. Our research aims to answer this question and investigate the influence of training statistical models used for AVMs on different spatial levels.

For this purpose, we compare four different methods trained on four differing spatial levels each and compare the overall performance of the models. Our objective is not primarily a comparison of the methods used, but a specific comparison within the individual methods concerning their performance on different spatial levels. We are interested in whether different methods deliver different results and whether any patterns emerge.

The methods selected represent a collection of regularly used ones in academic studies related to AVMs. In addition to parametric Ordinary Least Squares (OLS) regressions, we analyse semi-parametric Generalised Additive Models (GAM), eXtreme Gradient Boosting (XGBoost) algorithms and Deep Neural Networks (DNN) from the field of modern ML. Our analysis is based on a dataset of 1.2 million residential properties across Germany provided by professional real estate appraisers. The four spatial levels are based on the NUTS nomenclature of the European Union. The NUTS (Nomenclature of territorial units for statistics) classification is a hierarchical system for dividing up the economic territory of the EU and the UK. There are four different subdivision levels, called NUTS-0, NUTS-1, NUTS-2 and NUTS-3, which we use to train our models on a country, state, cross-regional, and county level, respectively[16].

Our research has theoretical and practical implications that collectively help improve AVMs' valuation accuracy. Our findings show that the right choice of spatial training level can significantly influence the model performance of different AVM algorithms, and that this influence varies considerably, depending on the type of method. The results indicate that for parametric and semi-parametric approaches, choosing a relatively small training level is advisable. This shows that the trained OLS and GAM cannot draw additional explanatory power from observations outside a particular region. The results for the two modern ML algorithms are quite different. We observe that they can gain more explanatory power by adding further observations, and that this effect outweighs local heterogeneity. Therefore, we recommend, choosing a generally higher training level for modern ML algorithms.

The contributions of our paper are manifold. First and foremost, our findings provide further evidence that when it comes to applying more traditional versus modern ML methods, fundamental differences in their application should be considered to achieve the

[16] The models could not be analysed at an even smaller spatial level because of data availability.

best model performance. Our findings indicate that assumptions valid for applying traditional ML methods may not be suitable for modern methods.

This provides real estate researchers and practitioners with new guidelines for using different AVM algorithms, which can help improve the performance of their valuation results. Additionally, our findings also shed light on the question of whether real estate markets are characterised by high local heterogeneity. The results of our OLS and GAM models study suggest significant heterogeneity in local real estate markets. Still, the results of the XGBoost and DNN indicate that there are overall patterns that apply to all real estate markets. In summary, our paper provides a new set of guidelines that can be used to answer various real estate-related questions more accurately. These new guidelines are a starting point for further research into the analysis of real estate markets using modern ML algorithms.

3.3 Literature Review

AVMs are computer-based applications that use various statistical and algorithmic approaches to assess the value or price of a property in an automated manner. They can be a cost-effective and rapid alternative to traditional valuation procedures (Schulz et al., 2014). AVMs emerged mainly from research on hedonic price models (HPM). HPMs were developed to estimate the effects of individual characteristics, so-called marginal prices, of a good on its value or price. By aggregating these marginal prices, the overall value of a good can subsequently be calculated (Chau & Chin, 2003). HPMs were first brought into a real estate context by Lancaster (1966) and Rosen (1974). As Malpezzi (2003) and Sirmans et al. (2005) show, a diverse and dynamic field of research has emerged since then, addressing a wide variety of real-estate-specific issues.

To improve the quality of automated real estate appraisals, the research community's focus in recent years has been almost exclusively on finding the best-fitting method. For this purpose, many approaches were either newly designed, or adapted and applied from other areas. The applied methods cover the full bandwidth of statistical methods and can be classified as parametric, semi-parametric or non-parametric approaches. Regarding parametric approaches, the most common multiple linear regression (MLR) models are applied and tested. Schulz et al. (2014), for example, use a flexible parametric hedonic regression introduced by Bunke et al. (1999) to measure the potential predictive performance of an AVM applied to the housing market of Berlin in Germany. Other examples of parametric approaches can be found at Tse (2002), R. Kelly Pace & LeSage (2004), Páez et al. (2008), Bourassa et al. (2010), Osland (2010) and Zurada et al. (2011).

Semi-parametric approaches can come in a variety of different forms. An often-used semi-parametric approach is the GAM, first introduced by Hastie & Tibshirani (1986). In contrast to traditional MLR models, the GAM can automatically control for non-linear relations between the dependent and independent variables. An early and prominent application within a real estate context is the study of Pace (1998). The author applies a GAM to a dataset for residential properties in Memphis (Tennessee) and finds that the GAM can outperform parametric and polynomial methods in terms of predictive behaviour.

A more recent example of the GAM can be found in Dąbrowski & Adamczyk (2010). Non-parametric approaches are a category of methods which do not need an a-priori specified functional form regarding the predictor. Instead, the form is learned from the information derived from the data itself. Given this flexibility, non-parametric approaches usually account for non-linearities and interactions within datasets and outperform parametric and semi-parametric approaches (Stang et al., 2022).

Prominent examples of non-parametric approaches include modern machine learning methods like Support Vector Machines, Artificial Neural Networks or Tree-Based Models. A real-estate-specific application of ML methods can be found in Mayer et al. (2019). The authors apply three commonly used basic techniques of modern ML (Random Forrest Regression, Gradient Boosting and Neural Networks) and compare their performance against some more traditional parametric approaches. Their findings show that the non-parametric methods can outperform stricter parametric approaches. Other real-estate-specific applications of non-parametric modelling techniques can be found in Chun Lin & Mohan (2011), Yoo et al. (2012), Antipov & Pokryshevskaya (2012), W. J. McCluskey et al. (2013), Kok et al. (2017), and Yilmazer & Kocaman (2020).

Another aspect with regard to the optimization of the valuation accuracy of AVMs is, besides the method selection, the choice of spatial level for training the models. The level at which the models are trained implies for which data, and thus ultimately also which information is considered in the context of the valuation and which is not. This could have a strong influence on the results of the models and is therefore a factor that should not be neglected. AVM-related studies currently always focus on a predefined region. The region to which the analyses are limited is in most cases the city level or the immediate surroundings of a city. Yao et al. (2018), for example, focus on the city level of Shenzhen (China), and W. McCluskey et al. (2012) choose the Lisburn District Council area around Belfast (North Ireland) to test their hypotheses. Other authors go a step further and conduct their analysis at the city district level. Baldominos et al. (2018), for example, focused on the Salamanca district of Madrid (Spain), Hong et al. (2020) run their analysis for the Gangnam district of Soul (South Korea), and Yilmazer & Kocaman (2020) run their model at the Mamak district of Ankara (Turkey). However, none of the authors investigates whether the chosen level is also the best one for training the models.

To the best of our knowledge, no study currently that deals with the optimal spatial level for training AVMs. Therefore, we aim to close this gap in the literature and determine the influence of the choice of spatial level on the quality of statistical valuation results. In particular, we are interested in whether this influence is the same for different types of methods (parametric, semi-parametric, non-parametric) or whether there are fundamental differences. In our analysis, we calculate the valuation accuracy of four different statistical methods (OLS, GAM, XGBoost, DNN), each trained at four different spatial levels, and compare their results subsequently.

3.4 Data

We base our analysis on a dataset consisting of 1,212,546 residential properties. These observations are distributed across Germany and were collected between 2014 and 2020. The dataset originates from the valuation department of one of Germany's largest mortgage lenders. Table 3.1 shows the distribution of the data over the observation period.

Table 3.1: Observations per year

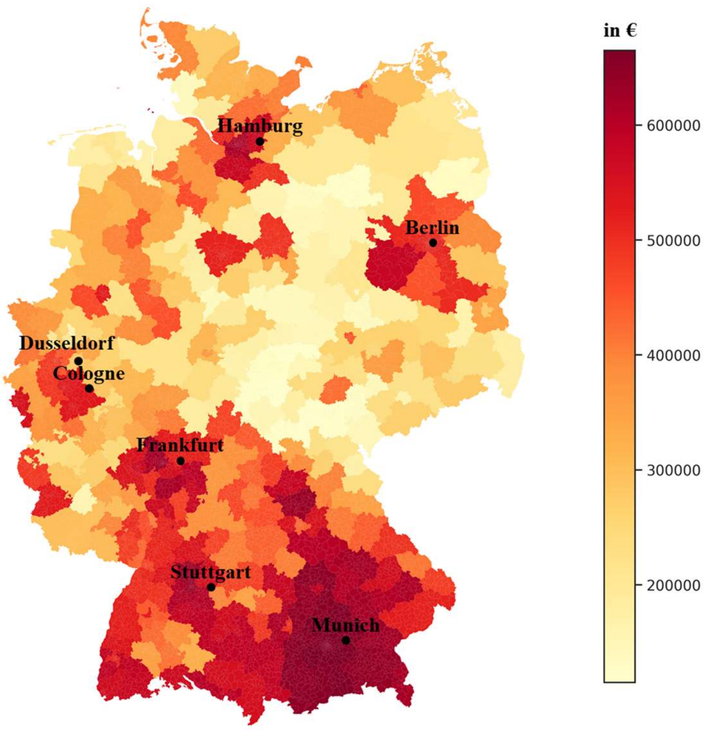
	2014	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
n	196318	196403	176238	163365	165106	165996	149120
(%)	0.1619	0.1620	0.1453	0.1347	0.1362	0.1369	0.1230

Notes: This table reports the number of observations available for each year. Over the years, the trend is slightly downward. Especially in 2020, the number of observations is lower, due to the COVID restrictions prevailing at that time. Due to the contact restrictions in place, on-site visits by appraisers were limited.

The data are actual valuation data collected by professional appraisers. We use the assessed market value as our target variable. An overview of the average market values across Germany is provided in Figure 3.1. The areas with the highest market values are in the Top-7[17] cities and commuter belts. Furthermore, the market values are by far the highest in the south of Germany and tend to be lower in the east.

[17] The Top-7 are the most important cities in Germany, namely Berlin, Munich, Hamburg, Frankfurt, Cologne, Dusseldorf and Stuttgart. Their importance is based on their market size and market activity. They can be seen as the most liquid and dynamic real estate markets in Germany.

Figure 3.1: Average market value per district



Notes: This figure shows the average market values per NUTS-3 district. The average was calculated using all available observations within the individual districts. The highest market values are near the major metropolitan regions and in the south of Germany. The substantial discrepancy between the west and east of Germany is striking. The market values observed here are also consistent with other studies (see, e.g., Just & Maennig, 2012), so it can be assumed that the observations used are representative.

As hedonic characteristics, we use a set of features describing the properties' structural characteristics, the micro-location and the macro-location. In addition, the year and quarter of the valuation are used to capture a temporal trend and seasonality. An overview of all the features used and their univariate distribution can be seen in Table 3.2.[18] Before being used the dataset was cleaned to account for duplicates, incompleteness, and erroneous data points. There are no correlations of concern within the dataset so that all variables be integrated accordingly.[19]

[18] Table 3.9 in Appendix I explains the individual variables.
[19] The correlation matrix is available on request.

Table 3.2: Descriptive statistics

Variable	Unit	Mean	Median	Standard Deviation	Maximum	Minimum
Market value	Integer	228157.10	200000.00	141717.54	3860000.00	20100.00
Modernization year	Integer	1989.10	1988.00	17.19	2020.00	1950.00
Year of construction	Integer	1978.48	1981.00	29.77	2023.00	1900.00
Year of valuation	Integer	2016.82	2017.00	2.00	2020.00	2014.00
Quarter of valuation	Integer	2.45	2.00	1.12	4.00	1.00
Quality grade	Integer	3.12	3.00	0.51	5.00	1.00
Living area	Float	120.31	114.68	51.69	440.00	15.00
Lot size	Float	436.48	323.00	541.66	10000.00	0.00
Longitude	Float	9.25	8.94	1.90	19.25	5.87
Latitude	Float	50.62	50.74	1.85	55.02	47.40
Micro score	Float	72.73	74.20	14.44	99.85	0.00
Unemployment ratio	Float	4.96	4.17	2.89	26.89	0.04
Time on market	Float	12.27	10.90	4.80	106.00	0.20
Basement condominium	Binary	0.38	0.00	0.48	1.00	0.00
No basement	Binary	0.19	0.00	0.39	1.00	0.00
Basement	Binary	0.44	0.00	0.50	1.00	0.00
Owner-occupied & Non-owner-occupied	Binary	0.09	0.00	0.29	1.00	0.00
Owner-occupied	Binary	0.70	1.00	0.46	1.00	0.00
Non-owner-occupied	Binary	0.21	0.00	0.41	1.00	0.00
Object subtype condominium	Binary	0.38	0.00	0.48	1.00	0.00
Object subtype detached house	Binary	0.42	0.00	0.49	1.00	0.00
Object subtype no detached house	Binary	0.20	0.00	0.40	1.00	0.00
Condition good	Binary	0.38	0.00	0.49	1.00	0.00
Condition disastrous	Binary	0.00	0.00	0.02	1.00	0.00
Condition middle	Binary	0.45	0.00	0.50	1.00	0.00
Condition moderate	Binary	0.02	0.00	0.14	1.00	0.00
Condition bad	Binary	0.00	0.00	0.05	1.00	0.00
Conidition very good	Binary	0.15	0.00	0.36	1.00	0.00
Regiotype agglo commuter belt	Binary	0.15	0.00	0.36	1.00	0.00
Regiotype agglo cbd	Binary	0.13	0.00	0.34	1.00	0.00
Regiotype agglo middle order centre	Binary	0.13	0.00	0.34	1.00	0.00
Regiotype agglo upper order centre	Binary	0.04	0.00	0.19	1.00	0.00
Regiotype rural commuter belt	Binary	0.15	0.00	0.36	1.00	0.00
Regiotype rural middle order centre	Binary	0.07	0.00	0.26	1.00	0.00
Regiotype rural upper order centre	Binary	0.01	0.00	0.07	1.00	0.00
Regiotype urban commuter belt	Binary	0.15	0.00	0.36	1.00	0.00
Regiotype urban middle order centre	Binary	0.10	0.00	0.29	1.00	0.00
Regiotype urban upper order centre	Binary	0.07	0.00	0.26	1.00	0.00
NUTS-1	String	-	-	-	-	-
NUTS-2	String	-	-	-	-	-
NUTS-3	String	-	-	-	-	-

Notes: This table reports the descriptive statistics of the dataset. Polytomous variables are one-hot encoded to binary variables to account for the requirements of modern machine learning methods. For the rather traditional methods – OLS and GAM – these polytomous variables are dummy encoded. The numbers were determined on the basis of all available observations. Overall, both structural features and location-describing features were used for model estimation. The selection of the parameters was in accordance with other publications in the AVM literature (see e.g., Metzner & Kindt, 2018). The parameter "market value" is the dependent variable in the model estimation.

Features describing the properties' structural characteristics include the property's subtype, year of construction, modernisation year, living area, lot size, use of the property, quality grade, condition and variable denoting whether the property has a basement or not. The subtype of a property can be either a 'Condominium', 'Detached house' or 'Not a detached house.' The year of modernisation represents when the last major refurbishment took place. The use of the building describes the possible uses, whereby the characteristics are either 'Owner-occupied & Non-owner-occupied',[20] 'Owner-Occupied' or 'Non-owner-occupied'. The variable describes whether a property can be rented to a third party. The quality of the property is measured via a grade on a scale ranging from 1 (very poor) to 5 (very good). The general condition of the property is represented by a categorical variable with five different categories ranging from disastrous to very good.[21] The variable 'Basement condominium' measures whether an apartment has an extra cellar compartment or not, whereas the 'Basement' and 'No Basement' variables are only valid for detached and non-detached houses.

The features describing the micro-location of the properties are the latitude and longitude, the different regiotypes and the micro score. The regiotype is provided by Acxiom[22] and clusters Germany into ten different area types. In general, Acxiom defines four different spatial types: 'Central-Business-District', 'Agglomeration Area, 'Urban Area' and 'Rural Area'. The last three can be divided further into three sub-categories each ('Upper Centres', 'Middle Centres' and 'Commuter Belt'). All addresses in Germany can be allocated to one of the ten area types. The individual area types are determined according to the respective settlement structure and population density within the municipality and its surrounding area. The micro score of a location is calculated via a gravity model and reflects the accessibility in the sense of proximity to selected everyday destinations. A more detailed description of the construction of the micro score of a location can be found in Appendix II. In addition, the two socio-economic variables, 'unemployment ratio' and 'Time-on-Market', are included to represent the properties' macro-location. All are available at the ZIP code level.

The spatial breakdown of our dataset is based on the NUTS nomenclature of the European Union and is done by creating three new features, namely 'NUTS-1', 'NUTS-2' and 'NUTS-3'. The NUTS system was introduced by the European Union and is monitored by Eurostat.

[20] Applies if the property is both partly owner-occupied and partly non-owner-occupied (e.g., single-family home with an attached rental unit).

[21] The assessment of the two variables, 'condition' and 'quality grade', was performed by professional appraisers during the property inspection process.

[22] Acxiom is an American provider of international macroeconomic and microeconomic data. Further information can be found at: https://www.acxiom.com/.

The system is used to provide a standardised system of territorial reference for the EU to make it easier to collect, compare and analyse statistics across different regions and countries. The NUTS system is, in general, divided into four levels, each with increasing geographical detail:

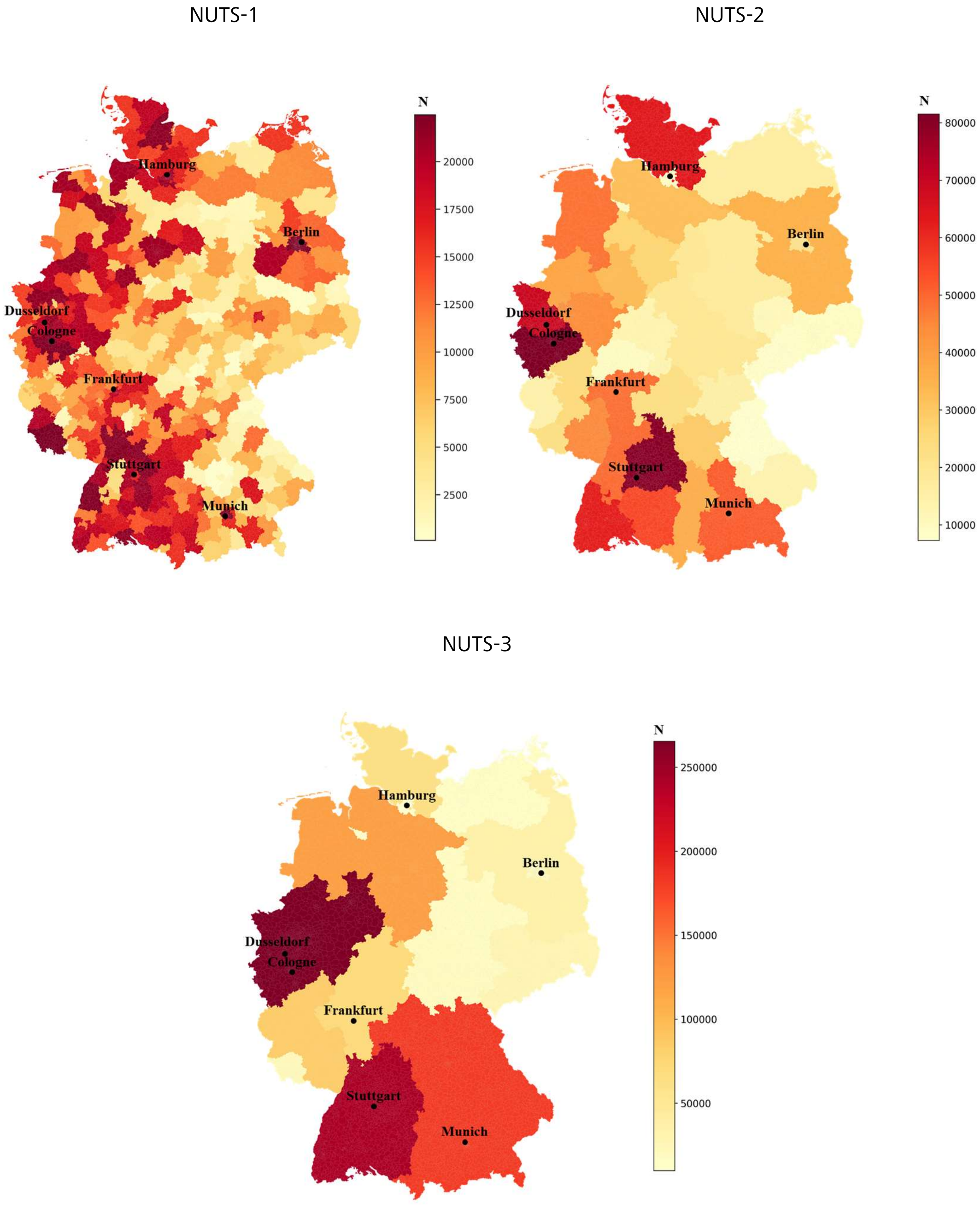
- NUTS-0: This level consists of larger regions that are typically based on the country or a group of countries
- NUTS-1: This level consists of major socio-economic regions within a NUTS-0 region
- NUTS-2: This level consists of basic regions for the application of regional policies within a NUTS-1 region
- NUTS-3: This level consists of small regions for specific diagnoses within a NUTS-2 region

The NUTS nomenclature is used for a wide range of purposes, including monitoring the progress of the EU's cohesion policy, as well as for other economic, social, and environmental statistics.[23]

Germany can generally be divided into a single NUTS-0, 16 NUTS-1, 38 NUTS-2 and 401 NUTS-3 regions. Since only a few observations were available in some NUTS-3 regions, we combined these regions and ended up with 327 NUTS-3 regions for our analysis. Figure 3.2 provides an overview of the different NUTS regions and the number of observations available for the specific regions. Analysing the NUTS-3 level shows most observations are located around the most significant German metropolitan areas like Berlin, Hamburg and Munich. In addition, the NUTS-2 and NUTS-1 levels indicate that a difference can be observed between west and east Germany, with the east tending to have fewer observations. This is consistent with the widely diverging population figures between these regions. Just and Schäfer (2017) provide a comprehensive introduction to the structure of the German regions. Just & Maennig (2012) give a more detailed overview of the German real estate markets.

[23]Further information about the NUTS nomenclature can be found at https://ec.europa.eu/eurostat/web/nuts/background

Figure 3.2: Number of observations per NUTS region



Notes: This figure highlights the observations available for the individual NUTS regions. Fewer observations are available in the eastern part of Germany. This can be explained by the generally lower market activity in these regions. Structurally, these regions are primarily rural and characterised by high out-migration and vacancies. Therefore, the data distribution is not a dataset-specific distortion but a representative reflection of the German residential real estate market.

3.5 Methodology

3.5.1 Ordinary Least Squares Regression – OLS

The first method applied is an Ordinary Least Squares Regression (OLS). The main advantage of the OLS is that it is easy to understand and interpret. Therefore, it is the most commonly used machine learning method and is often considered a benchmark. The aim of an OLS is to explain a dependent variable $\boldsymbol{y}$, with independent variables $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$, a-priori unknown parameters $\beta_0, \beta_1, \dots, \beta_k$ and an error term $\boldsymbol{\varepsilon}$:

$$\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{x}_1 + \cdots + \beta_k \boldsymbol{x}_k + \boldsymbol{\varepsilon}, \tag{9}$$

for all observations with

$$\boldsymbol{\mu} = E[\boldsymbol{y}] = \beta_0 + \beta_1 \boldsymbol{x}_1 + \cdots + \beta_k \boldsymbol{x}_k. \tag{10}$$

Thereby, the relationship between the dependent and independent variables is assumed to be linear in parameters, and the error terms ε are considered independent and to have a constant variance. For further information, we recommend Fahrmeir et al. (2013).

Several optimisations were performed to account for locational differences and to achieve the best model performance, including backward stepwise regression, interaction terms and variable transformations.

3.5.2 Generalized Additive Model – GAM

Our second method is a Generalized Additive Model (GAM). It is a further development of the OLS and is based essentially on the concept of the Generalized Linear Model. A monotonic link function $g(.)$ is used to model the relationship between the expected value $\boldsymbol{\mu}$ of the dependent variable $\boldsymbol{y}$ and the independent variables $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$. The main advantage of the GAM over the OLS is that unspecified, non-parametric smoothing functions s_j, $j \in \{1, \dots, k\}$, of the covariates can be included in the model:

$$g(\boldsymbol{\mu}) = \beta_o + s_1(\boldsymbol{x}_1) + \cdots + s_k(\boldsymbol{x}_k). \tag{11}$$

For a more extensive description of the GAM, we recommend Wood (2017).

Again, multiple model optimisations were carried out. In addition to the methods mentioned above, different penalised spline types like cubic and thin plane splines were considered. As in the OLS, these optimisations were implemented manually.

3.5.3 Extreme Gradient Boosting – XGBoost

The third method, an Extreme Gradient Boosting (XGBoost) algorithm, is a tree-based ensemble learning method. Ensemble learning algorithms train many weak learners h_m, in our case, single decision trees, and combine them to form one strong learner h:

$$h(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{M} u_m h_m(\boldsymbol{y}|\boldsymbol{x}), \tag{12}$$

with u_m being used to weight the weak learners. M denotes the number of single trees, $\boldsymbol{x}$ is the features space and $\boldsymbol{y}$ the response variable. In boosting, the weak learners h_m are trained sequentially. The algorithm starts with one model and uses the errors made to improve the subsequent trees. In Gradient boosting, the gradient descent algorithm is used to add new trees to minimise the loss of the model. The eXtreme Gradient Boosting is a computationally effective and highly efficient version of Gradient Boosting. The advantage of XGBoost is that it can recognise very complex patterns within a large amount of data. However, it is unclear from the model structure why a certain result occurs. The eXtreme Gradient Boosting is a computationally effective and highly efficient version of Gradient Boosting. The XGBoost can automatically detect complex non-linearities or higher-order interactions within a large dataset, with fewer manual optimisations than the OLS and GAM. Hastie et al. (2001) provide a detailed description of tree-based methods, ensemble learning and gradient boosting.

3.5.4 Deep Neural Network - DNN

Lastly, we consider deep neural networks (DNN), a popular and performant machine learning technique. DNNs are designed from biological neural networks (Pham, 1970), like the human brain, and consist of multiple layers, which are typically densely connected. Each layer consists of numerous neurons, each processing the weighted output of all (hence the term dense) neurons of the previous layer, combined with a bias value, and applies a so-called activation function onto this linear combination. To capture this formally, let y be a neuron in the current layer, and let n be the number of neurons in the previous layer. For $i \in \{1, \ldots, n\}$, let z_i be the output of the i-th neuron in the previous layer and let w_i be the according weight. Furthermore, let f be the activation function of the current neuron and b the bias term. Then, the output of the neuron is

$$f\left(b + \sum_{i=1}^{n} z_i \, w_i\right) \tag{13}$$

A DNN then consists of multiple such neurons and layers.

To train a DNN for a specific task and data, the weights and biases are adapted. The data is passed through the DNN in batches in a forward-propagation step. A prediction is calculated, for each datum in a batch, and the predictions are evaluated regarding loss function. The weights and biases are then adjusted using gradient descent to minimise the loss function. After all the data is passed through the DNN once, we say one epoch has passed. After many epochs, the DNN is trained, and predictions for a new object can be obtained by passing the object through the DNN again.

Finding the right architecture of a DNN for the task at hand is an essential yet tedious task. We use the hyperparameter optimisation framework Optuna (Akiba et al., 2019) to find suitable architectures for each region. In particular, we allow Optuna to choose the number of layers, the number of neurons per layer and the activation function per layer. Furthermore, we allow Optuna to choose the dropout rate per layer, which controls how many neurons per layer are activated.

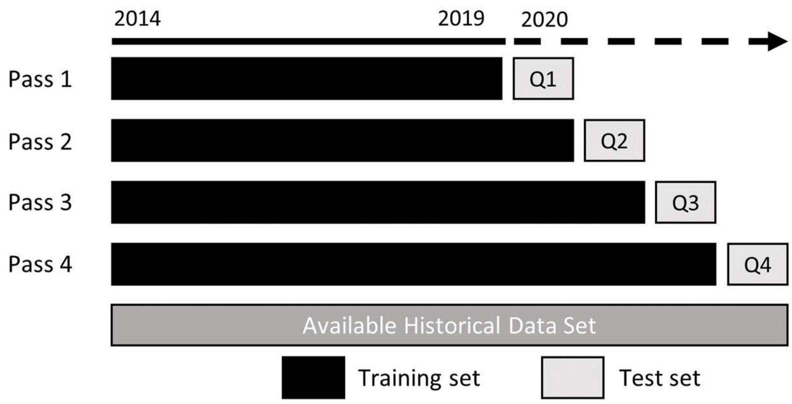
The advantages of deep neural networks are that they are very flexible and adapt automatically to all data. Therefore, they can capture complex non-linearities and higher-order interactions by themselves. Besides that, compared to other modern machine learning approaches, deep neural networks require less computation power to produce reliable results. For more information about DNNs, see Goodfellow et al. (2016).

3.5.5 Testing Concept

An extending window approach is implemented according to Mayer et al. (2019) to evaluate the predictive performance of the models. Figure 3.3 illustrates the testing concept. The first iteration divides the dataset into a training set with observations from Q1/2014 to Q4/2019 and a test set from Q1/2020. In the following steps, the data of the tested quarter is added to the training set, and the models are retrained and tested on data from the next quarter. The advantages of this approach are that all algorithms are tested on unseen data and thus produce unbiased, robust results. Furthermore, the testing approach provides a realistic testing scenario. Table 3.3 presents the number of training and test observations for each iteration.[24]

[24] The train/test split was selected to include as much data as possible in the training set, as some NUTS-3 regions have limited data. If an alternative split were employed, it would result in inconsistent results for these regions. Our study focuses on a nationwide comparison, rather than individual metropolitan regions, which typically have a large and dense amount of data. This approach is consistent with other studies that compare algorithms on a national scale, such as Stang et al. (2022).

Figure 3.3: Extending window approach



Notes: This figure visualizes the applied extending window approach-testing strategy. The strategy is the right choice for the purposes of this study, as it best reflects the test procedure of conventional AVM providers and thus provides a strong reference to reality. AVM providers usually update their models on a quarterly basis as well. The results obtained in this way therefore represent an extract that is in all probability also achievable in a real-life situation.

Table 3.3: Training and test observations

Data split	Q1	Q2	Q3	Q4
Training	1,063,426	1,106,866	1,141,612	1,180,741
Test	43,440	34,746	39,129	31,805

Notes: This table shows the number of training and test observations over the four quarters of 2020. The number of training data increases over the quarters by the number of test data from the previous quarter. With regard to the test data, it can be seen in particular that fewer observations are available in Q2 and Q4. This can be attributed to COVID restrictions which made it difficult to conduct assessment visits, especially shortly after the pandemic outbreak (Q2) and during the winter (Q4).

3.5.6 Evaluation Metrics

We compute the Mean Absolute Percentage Error (MAPE) and the Median Absolute Percentage Error (MdAPE) as accuracy measures for each model. Unlike Mayer et al. (2019), we use the relative rather than the absolute error measures to better compare the different spatial levels. To obtain an overall picture of the strength and weaknesses of the algorithms, we additionally provide the proportion of predictions within 10 and 20 per cent (PE(x)) following Cajias et al. (2019) and Stang et al. (2022). A detailed description of all metrics can be found in Table 3.4.

Table 3.4: Evaluation metrics

Error	Formula	Description
Mean Absolute Percentage Error (MAPE)	$MAPE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|$	Mean of all absolute percentage errors. A lower MAPE signals higher overall prediction accuracy in percent.
Median Absolute Percentage Error (MdAPE)	$MdAPE(y,\hat{y}) = median(\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|)$	Median of all absolute percentage errors. A lower MdAPE denotes a higher precision in percent without being sensitive to outliers.
Error buckets (PE(x))	$PE(x) = 100\left\|\frac{y_i - \hat{y}_i}{y_i}\right\| < x$	Percentage of predictions where the relative deviation is less than x%, with x being 10 and 20. A larger PE(x) signals a lower variation in the predictions.

Notes: This table reports the evaluation metrics used to determine the valuation accuracy of the different algorithms. All four metrics are regularly used to assess the quality of AVMS. The choice of several metrics in total allows a more differentiated statement to be made than would be the case with just one metric.

3.6 Results

Our study aims to find out whether the choice of spatial level for training statistical models has an influence on their performance, whether this influence is the same for all methods, or whether there are differences between more traditional and modern ML methods. In contrast to other publications, the main focus is not on which method performs best overall but on an intra-method comparison to determine which spatial level seems best suited for which method. This enables finding out whether the assumed local heterogeneity of real estate markets is also reflected in the results of the valuation methods or whether greater valuation accuracy can be achieved by adding further observations from other submarkets. For this purpose, two traditional approaches (OLS & GAM) as well as two modern ML approaches (XGBoost & DNN) are each trained for different spatial levels (NUTS-0, NUTS-1, NUTS-2, NUTS-3).

Below, we show the results for all four methods. To achieve comparability and to be able to make a valid statement, we evaluate the results on an aggregated level. For this purpose, we first provide a table for each method that shows the individual evaluation metrics for the four spatial levels of all test observations. For the metrics in the 'NUTS-3' row, for example, all test data is predicted with the different models calculated at the NUTS-3 level. Finally, the metrics are calculated for the nationwide aggregated residuals. For the other three levels, the procedure is then the same.

Furthermore, four maps are shown for each method. The maps are a cartographic representation of the results of the MAPE on a NUTS-3 level from the tables presented earlier. The representation allows for more detailed interpretations concerning regional performance. For example, it allows us to determine whether the results differ across different regions and whether general data availability plays a role.

3.6.1 Results of the ordinary least squares regression

The OLS results presented in Table 3.5 yield a clear pattern: The smaller the spatial level, the better the sperformance. Regarding the MAPE, the NUTS-3 models, which divide Germany into 327 submarkets, are more than three percentage points better than the NUTS-0 model, which considers Germany one overall market. In relative terms, this represents a performance increase of 18.0%. The PE-ratio also shows that the NUTS-3 models are far superior to the NUTS-0 model.

Table 3.5: OLS – model prediction errors 2020 throughout Germany

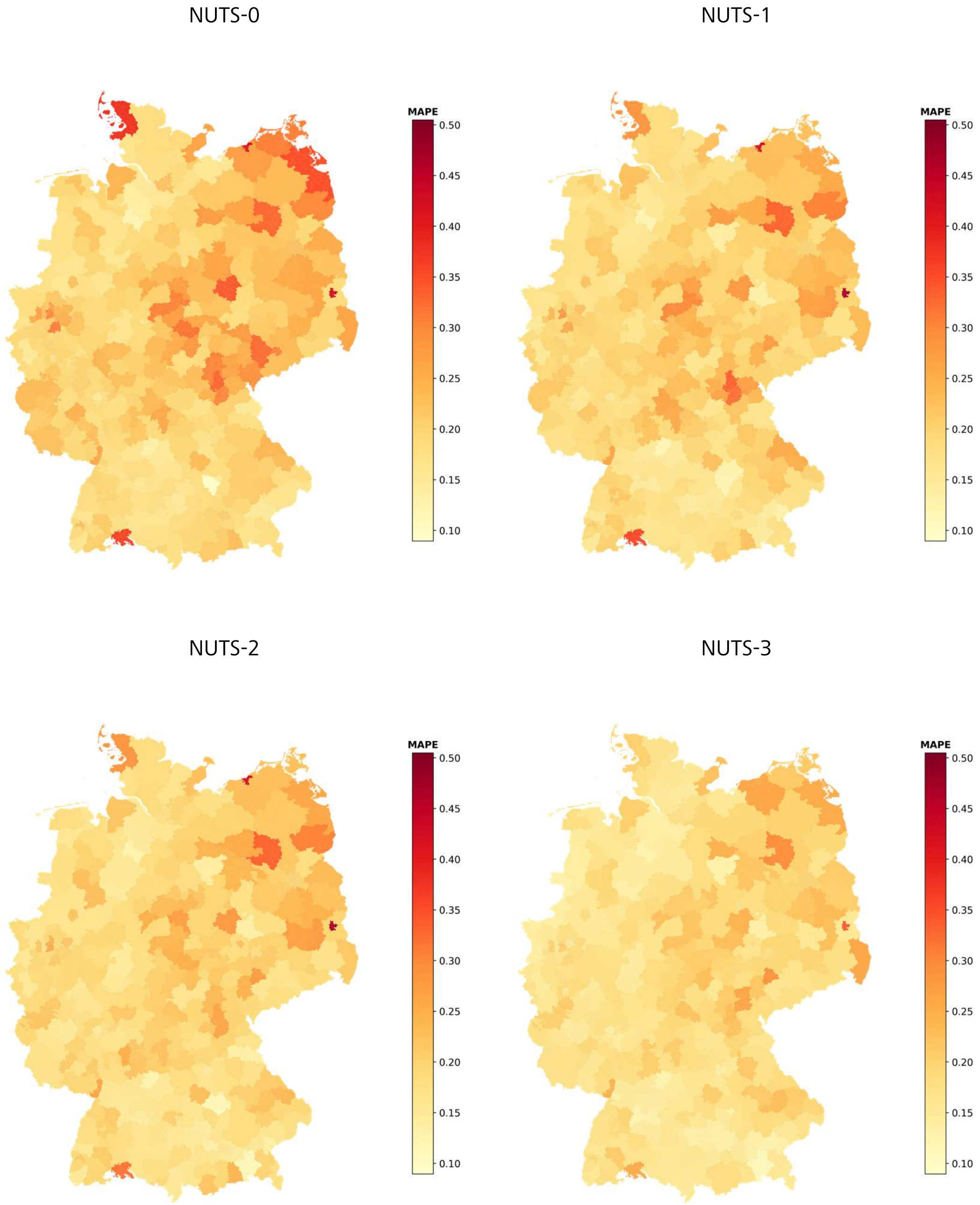
Models	MAPE	MdAPE	PE(10)	PE(20)
$OLS_{NUTS\text{-}0}$	0.2023	0.1521	0.3423	0.6236
$OLS_{NUTS\text{-}1}$	0.1914	0.1454	0.3577	0.6473
$OLS_{NUTS\text{-}2}$	0.1852	0.1407	0.3688	0.6612
$OLS_{NUTS\text{-}3}$	0.1714	0.1294	0.3985	0.7004

Notes: This table reports the model prediction errors for the OLS. The results are evident across all metrics and show that model performance improves with a decreasing spatial training level. This result confirms the correctness of the proceeding that in parametric approaches, a data selection that is as granular as possible must be conducted in each case.

The cartographic representation in Figure 3.4 illustrates the results from Table 3.5. It can be seen that the lower the spatial training level, the better the MAPE for each region. The maps further show that the increased performance at the aggregate level can be attributed to improved performance in the eastern parts of Germany. In addition, the German North Sea Island group around Sylt stands out on the top left of the maps. Here, it can be seen that the performance in the NUTS-3 models is much better than in the NUTS-0 model. Very distinct peculiarities characterise the real estate market on Sylt and the surrounding islands. Residential properties are traded there only at top prices, and there is a strong dependency between the property's specific location and its value.

In summary, the OLS can only capture local effects of the German residential real estate market when trained on a small spatial level. Therefore, it is advisable to use the smallest possible spatial level, in our case NUTS-3, for training the OLS. These results also make sense in theory since the OLS is generalising in its structure and, therefore, can hardly (or not at all) take into account the local characteristics of individual regions if training is done on a global level. For the NUTS-0 model, the coefficients of the OLS are smoothed by too many individual and inconsistent regional effects, leading to a significant deterioration in performance. In the case of an OLS, it should always be ensured that only regional data are used to determine the coefficients and, ideally, that different submarkets are delimited from one another in advance.

Figure 3.4: MAPE of the different OLS models



Notes: This figure visualises the MAPE of the four different OLS models. The maps show the average absolute percentage error obtained when applying the individual models within a given region. For a granular representation, the 327 NUTS-3 regions were selected as the corresponding levels of representation. The representation of the scale is chosen so that the minimum and maximum are the largest and smallest errors, respectively, of all four methods.

3.6.2 Results of the generalized additive model

The results for the GAM, shown in Table 3.6, are also clear and similar to those for the OLS. The more granular the spatial level for training, the better the estimation accuracy. This is true for all four evaluation metrics used. This time, the MAPE at the NUTS-3 level is 23.8% better than the NUTS-0 model. If we look at the general performance of the GAM and compare it with the results of the OLS, we see that the GAM is generally able to correctly estimate the market values of the properties better. The use of non-linear functions, which characterises the GAM, results in a performance boost. However, it is interesting to note that this effect only comes into play at a granular level. While the relative difference between the MAPEs of the NUTS-0 models of the OLS and the GAM is only 2.6%, it increases continuously and amounts to 7.7% at the level of the NUTS-3 models.

Table 3.6: GAM – model prediction errors 2020 throughout Germany

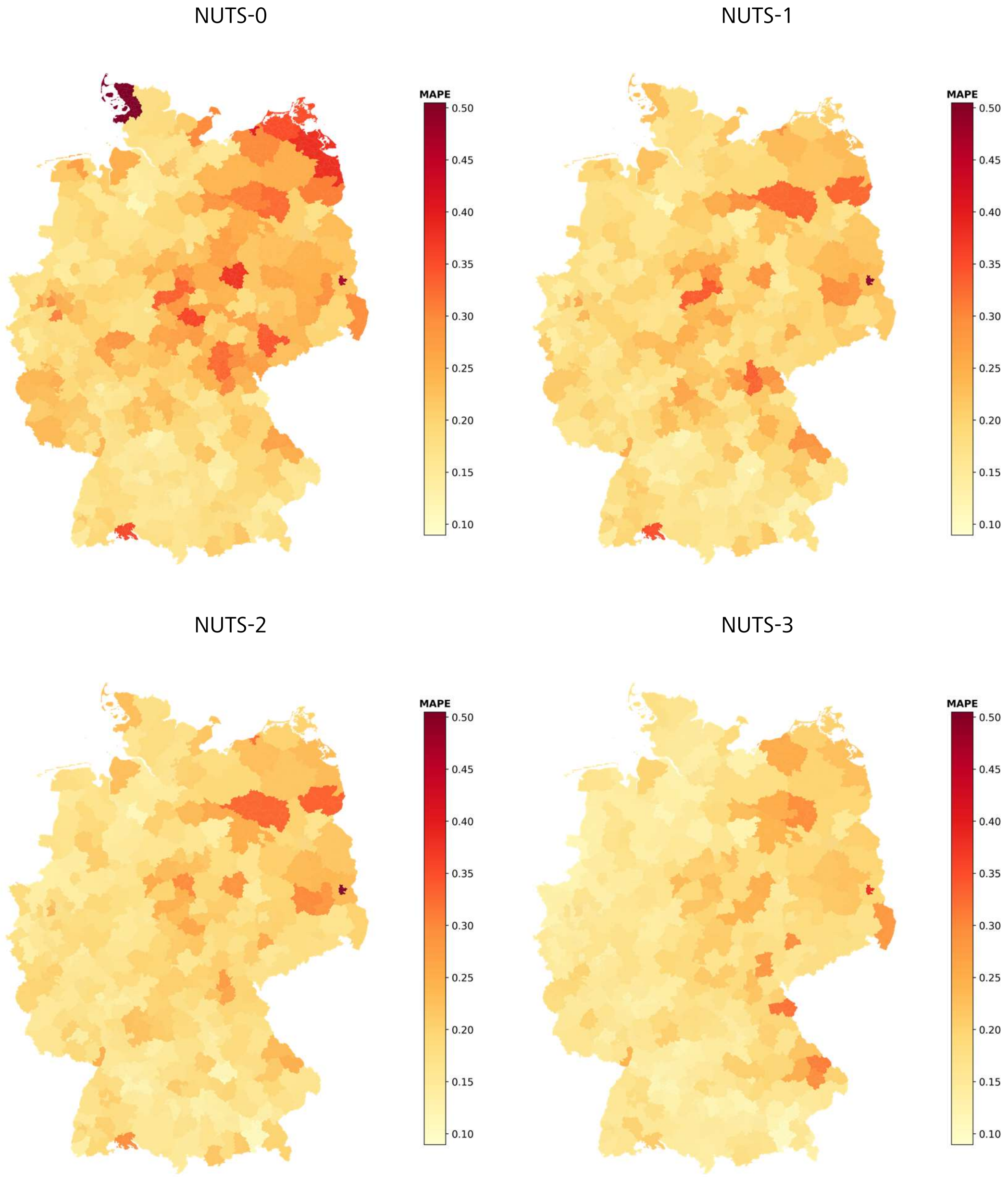
Models	MAPE	MdAPE	PE(10)	PE(20)
GAM_{NUTS-0}	0.1971	0.1423	0.3641	0.6504
GAM_{NUTS-1}	0.1832	0.1339	0.3852	0.6800
GAM_{NUTS-2}	0.1734	0.1273	0.4044	0.7028
GAM_{NUTS-3}	0.1592	0.1160	0.4398	0.7426

Notes: This table reports the model prediction errors for the GAM. The results are also clear across all metrics and similar to the results of the OLS. They show that model performance improves with a decreasing spatial training level. Again, the implication is that the smallest spatial level should be chosen to achieve the best model performance.

The cartographic representation in Figure 3.5 shows the same picture as the OLS. Once again, it is noticeable that the estimation accuracy in the eastern part of Germany can be improved by implementing the method on a granular level. Furthermore, the group of islands around Sylt stands out again. It implies that the smaller the spatial level for training the model, the better the performance.

In summary, the feedback for the GAM is the same as that for the OLS. On a higher spatial level, the GAM does not capture the complexity and heterogeneity of the individual residential real estate markets in a single model as accurately as on a granular level. Therefore, when using a GAM for estimating residential property values, the smallest possible level should be used for training.

Figure 3.5: MAPE of the different GAM models



Notes: This figure depicts the MAPE of the four different GAM models. The maps show the average absolute percentage error obtained when applying the individual models within a given region. For a granular representation, the 327 NUTS-3 regions were selected as the corresponding levels of representation. The representation of the scale is chosen so that the minimum and maximum are the largest and smallest errors, respectively, of all four methods.

3.6.3 Results of the extreme gradient boosting

Compared to the first two methods, the results of the XGBoost yield a different picture. The evaluation metrics from Table 3.7 show that the performance is similar on all four NUTS levels, and the greatest accuracy is achieved this time on the NUTS-1 level and not, as with the OLS and the GAM, on the NUTS-3 level. This is interesting because, as shown in the literature review, most academic studies on ML algorithms focus on the NUTS-3 level. This spatial level yields the worst performance in our case. Relative to the NUTS-1 level, the NUTS-3 level based on MAPE is 2.9% worse regarding valuation accuracy. Although the differences between the individual metrics are only minor in absolute values, if these are considered in relative terms, then a small performance boost is shown by the correct choice of the spatial level.

Table 3.7: XGBoost – model prediction errors 2020 throughout Germany

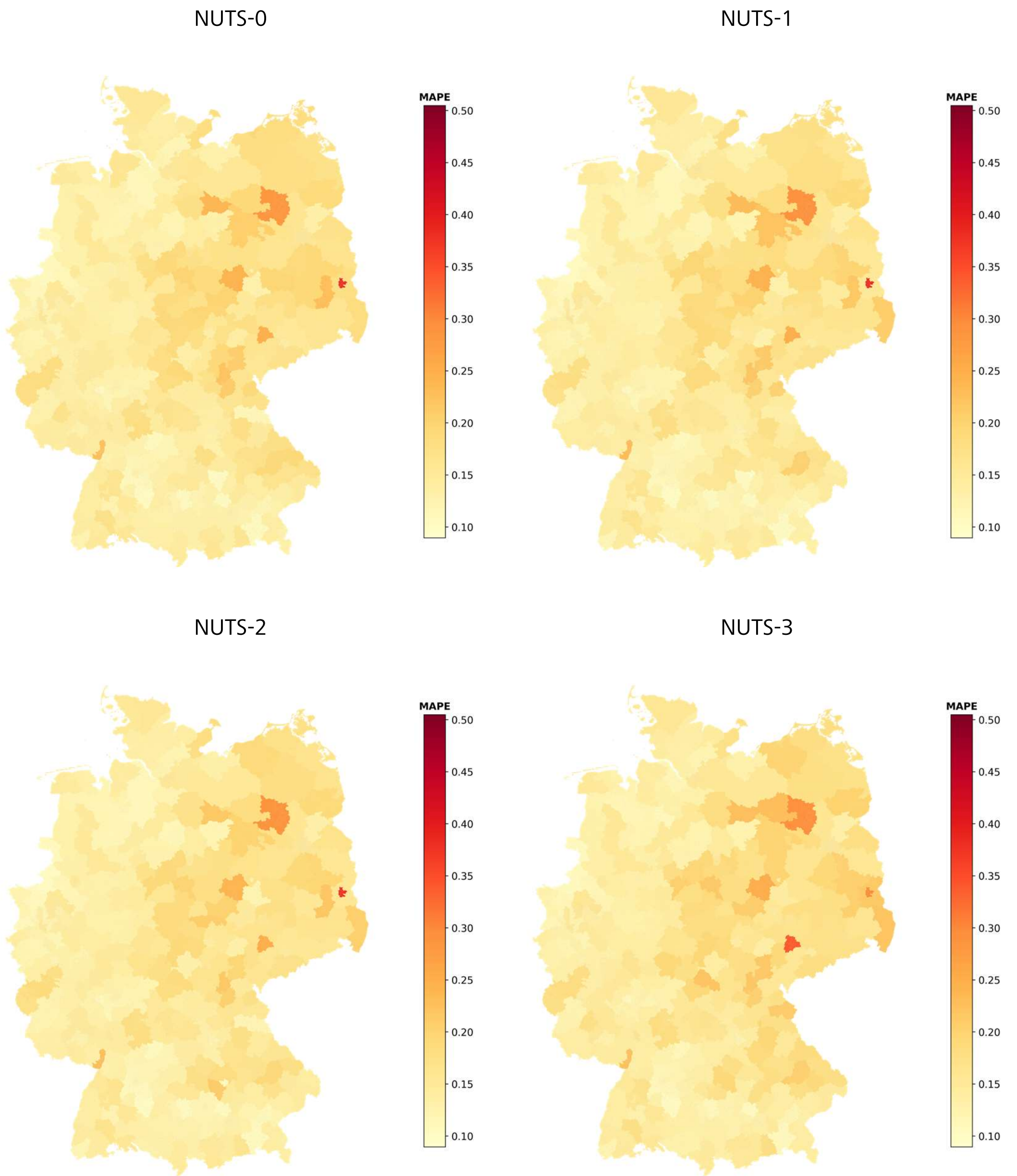
Models	MAPE	MdAPE	PE(10)	PE(20)
$XGB_{NUTS\text{-}0}$	0.1426	0.1077	0.4693	0.7780
$XGB_{NUTS\text{-}1}$	0.1402	0.1064	0.4739	0.7869
$XGB_{NUTS\text{-}2}$	0.1407	0.1071	0.4719	0.7850
$XGB_{NUTS\text{-}3}$	0.1442	0.1107	0.4578	0.7733

Notes: This table reports the model prediction errors for the XGBoost. Here, too, the results are the same across all evaluation metrics. Unlike the first two methods, however, the model performance of the XGBoost does not improve with a decreasing spatial training level but is relatively constant across all levels. The best performance is achieved at the NUTS-1 level, indicating that the XGBoost can gain more explanatory power by adding more data.

The analysis of the maps from Figure 3.6 shows, in particular, that in the parts of Germany where few observations are available (see Figure 3.2), the choice of a higher spatial level for training the models leads to a performance improvement. It is an important implication that in regions where little data are available, it can be useful in the case of the XGBoost to include data from other surrounding districts. This represents an essential difference between the results of the GAM and the OLS. For them, especially in parts of Germany with low data availability, the results deteriorate with a higher spatial level for training the models.

In summary, the heterogeneity of local real estate markets can still be detected by the XGBoost when trained at a higher spatial level, as the XGBoost can combine several local models into one large global model. In some cases, the use of additional data even leads to a further improvement of the estimation accuracy as more nuanced relationships can be learned, and the risk of overfitting decreases. Therefore, unlike for the OLS and the GAM, the NUTS-3 level is not the optimal spatial level for training the XGBoost, but the NUTS-1 level. However, the results of Table 3.7 also show that there seems to be a limit regarding the optimal size of the spatial level. The results at NUTS-0 level are still better than those at NUTS-3 level but not as good as on NUTS-1 and NUTS-2 level.

Figure 3.6: MAPE of the different XGB models



Notes: This figure visualises the MAPE of the four different XGBoost models. The maps show the average absolute percentage error obtained when applying the individual models within a given region. For a granular representation, the 327 NUTS-3 regions were selected as the corresponding levels of representation. The representation of the scale is chosen so that the minimum and maximum are the largest and smallest errors, respectively, of all four methods.

3.6.4 Results of the neural network

Finally, in analysing the results of the DNN, we again see a different picture. The evaluation metrics presented in Table 3.8 show that the DNN can improve its valuation accuracy as the spatial training level increases. This is the exact opposite of the OLS and GAM results and a different result than the XGBoost. Although the results of the MAPE indicate that the NUTS-1 level performs best here as well, the three other metrics yield a slightly different picture for this specific algorithm. They evaluate the NUTS-0 level as the best suited. In principle, therefore, the situation between the NUTS-0 and NUTS-1 levels is quite similar, influenced only by marginal changes. Compared to the other modern ML algorithm, the XGBoost, the number of observations used to optimise the algorithm seems more important. This is also logical from the point of view of the complexity of the method. The DNN can only show its strength in recognising non-linear relationships and multi-layer interactions if a sufficiently large number of observations is available. This finding is also in line with those of Nghiep & Cripps (2001), which show, based on a dataset for Rutherford County, Tennessee, that neural networks perform better than multiple regression analysis only with increasing dataset size.[25]

Table 3.8: DNN – model prediction errors 2020 throughout Germany

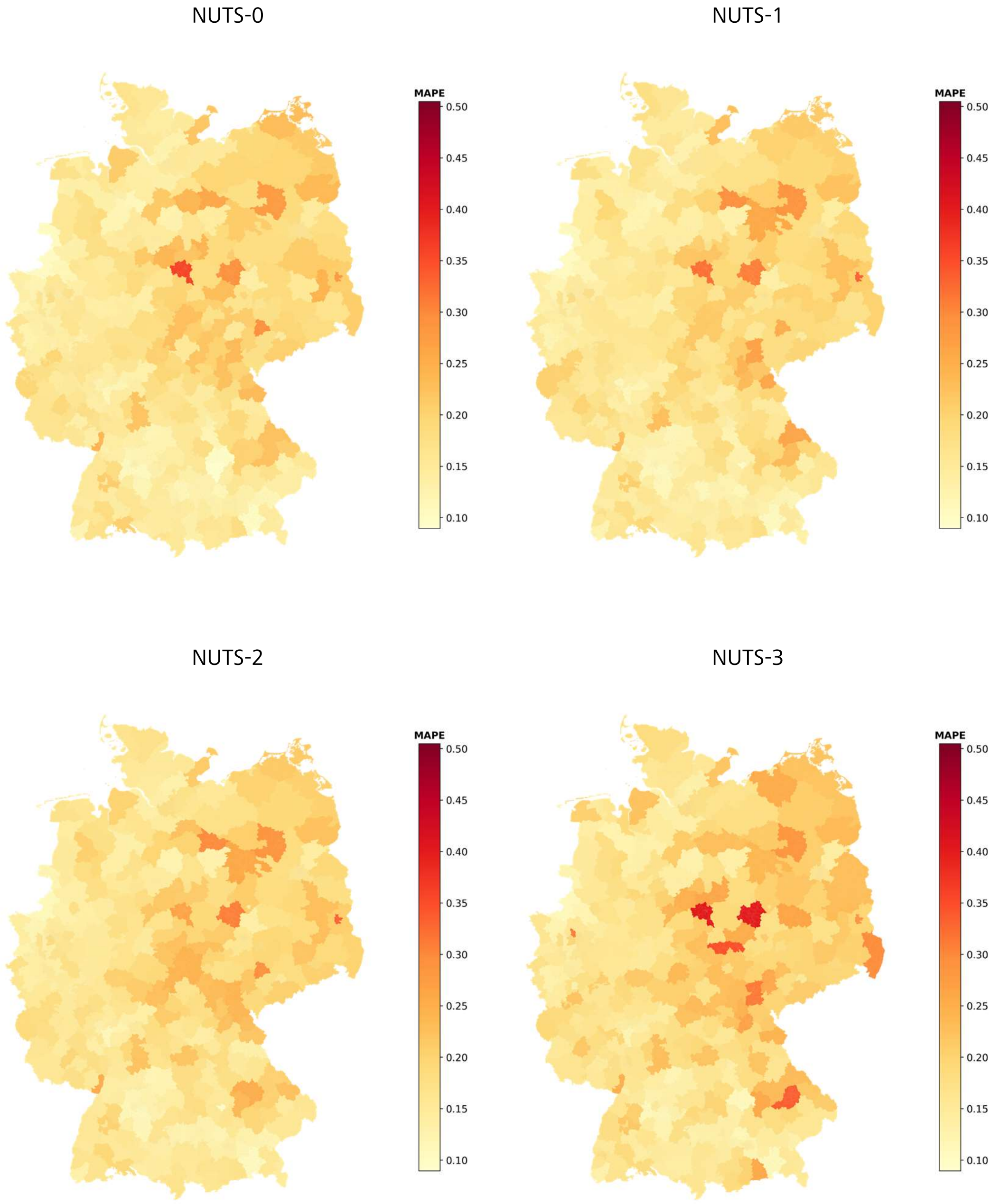
Models	MAPE	MdAPE	PE(10)	PE(20)
DNN_{NUTS-0}	0.1551	0.1080	0.4700	0.7620
DNN_{NUTS-1}	0.1542	0.1090	0.4648	0.7595
DNN_{NUTS-2}	0.1595	0.1142	0.4471	0.7448
DNN_{NUTS-3}	0.1656	0.1176	0.4356	0.7281

Notes: This figure visualizes the MAPE of the four different DNN models. The maps show the average absolute percentage error obtained when applying the individual models within a given region. For a granular representation, the 327 NUTS-3 regions were selected as the corresponding levels of representation. The representation of the scale is chosen so that the minimum and maximum are the largest and smallest errors, respectively, of all four methods.

The visual representation of the results in Figure 3.7 yields a similar picture to the XGBoost results. By choosing a higher training level for the DNN, the valuation performance can be increased, especially in areas with few observations. Again, the same implication emerges as with the XGBoost: in regions where little data is available, including data from other surrounding districts can be helpful. Concerning the four algorithms used for the analysis, this can only be empirically proven for the modern ML algorithms, which represents a significant contribution to the literature of this study.

[25] However, unlike in our study, these authors only work at a county level and only vary the data available within the county. In our case, the amount of data is varied by adding observations from other spatial levels.

Figure 3.7: MAPE of the different DNN models



Notes: This figure visualizes the MAPE of the four different DNN models. The maps show the average absolute percentage error obtained when applying the individual models within a given region. For a granular representation, the 327 NUTS-3 regions were selected as the corresponding levels of representation. The representation of the scale is chosen so that the minimum and maximum are the largest and smallest errors, respectively, of all four methods.

In summary, the DNN can only estimate property values as accurately as possible once a certain number of observations has been used. Adding more observations, therefore, outweighs the effect of local heterogeneity. Thus, the DNN can independently generate additional explanatory power for a specific real estate market, even from data outside the specific market. Regional effects can therefore be more effectively detected, extracted and extrapolated by modern ML algorithms. The reason could be that neural networks are designed to handle large datasets and can benefit from more data by learning more nuanced relationships. Besides that, larger datasets help to reduce the variance of the neural networks' predictions. This variance reduction helps prevent overfitting, making the model more robust and accurate. Concerning the DNN, it is advisable to choose as high as possible a training level or to maximize the available observations for training the algorithm.

3.7 Conclusion

This study is intended to answer whether the right choice of the appropriate spatial level for training AVM algorithms also plays an important and underestimated role in improving the valuation accuracy of AVMs. We use a dataset of 1.2 million residential properties across Germany to test our hypotheses for four different typical AVM algorithms (OLS, GAM, XGBoost, DNN). All four are each trained on four different spatial levels, after which the results are evaluated. The four spatial levels are based on the NUTS nomenclature of the European Union. We use the NUTS-0, NUTS-1, NUTS-2 and NUTS-3 levels to train our models on a country, state, cross-regional, and county level, respectively.

Our results indicate that the correct choice of spatial training level can significantly influence the model performance, and that this can vary considerably, depending on the type of method. Concerning the OLS results, selecting a training level that is as granular as possible is the only way to ensure that the most accurate valuations are attained. There are regional differences and, thus, certain heterogeneities, which the OLS can only recognise as accurately as possible if they are locally limited.

The results for the GAM yield a similar picture to the OLS. The model performance correlates positively with a smaller spatial training level. Accordingly, the same findings can be generated for the parametric and the semi-parametric approaches. These confirm the correctness of the trend in academic publications and in practice of choosing the most granular analysis level possible for traditional econometric methods. These two methods cannot draw additional explanatory power from observations that lie outside a region. On the contrary, they even suffer from it.

The results of the two applied modern ML algorithms are quite different. Concerning the XGBoost, the evaluation metrics show that the choice of the most suitable spatial level can be made with relative indifference. Although there are marginal differences concerning the evaluation accuracy, these are only minor compared to OLS and GAM. In contrast to the parametric and semi-parametric approaches, the non-parametric XGBoost shows that the performance increases slightly with increasing spatial training levels. The NUTS-1 level seems the most appropriate level. This trend can be observed even more clearly in the results of the DNN. Here, it can be seen that the performance does not decrease with an increasing training level, as is the case with the OLS and the GAM, but it improves.

Concerning the two modern ML algorithms, they can gain a higher degree of explanatory power by adding further observations, and this effect outweighs that of local heterogeneity. In particular, their ability to recognise and map non-linear relationships and multi-layered interactions allows them to exploit overlapping effects of different regions to achieve more accurate real estate valuations. This is particularly evident in regions where there are few observations. In these cases, training a modern ML algorithm with additional regions is advisable to benefit from their basic commonalities.

In summary, the right training level should always depend on the method. For parametric and semi-parametric methods, we recommend using a spatial level that is as granular as possible for training the models, since these can only separate local heterogeneities from each other to a limited extent. For non-parametric modern ML methods, however, we generally recommend a higher training level. These complex methods can detect regional differences independently and separate them. Furthermore, they benefit from the fact that there are basic commonalities in the functioning of local real estate markets, which can be used to increase their explanatory power. Concerning the practical application and implementation of AVM algorithms, this offers the additional advantage that the higher training level means fewer models must be trained and calibrated overall. For example, less effort is required for data preparation and processing. Thus, efficiencies can be increased for AVM providers operating nationwide, and significant economic advantages can be achieved.

Our findings empower real estate researchers to make more informed decisions about the appropriate spatial level when using and analysing different machine learning algorithms. As such, the main contribution of this paper is to update the standard guidelines for applying both traditional econometric and modern ML algorithms and setting new guidelines. On top of that, the contributions of our paper are not limited to scientific purposes but also provide practitioners in the field of AVM application with a new set of guidelines that can help them to improve the accuracy of their AVMs and reduce their implementation efforts at the same time.

3.8 Appendix

3.8.1 Appendix I – Macro Score

Table 3.9: Feature description

Variable	Description
Market value	Market value of the property determined by appraiser
Modernization year	Year of the last major refurbishment
Year of construction	Year in which the property was built
Year of valuation	Year in which the property was assessed
Quarter of valuation	Quarter in which the property was assessed
Quality grade	Grade concerning the quality of the property ranging from 1 (very poor) to 5 (very good)
Living area	Size of the property in square meters
Lot size	Size of the property plot in square meters
Longitude	Longitude of the property
Latitude	Latitude of the property
Micro score	Rates the quality of the micro location
Unemployment ratio	Variable that describes the unemployment ratio on a zip code level
Time on market	Measurement of the length of real estate listings in weeks at the zip code level
Basement	Variable that describes whether the property has a basement or not
Owner-occupied	Variable that describes whether the property is rented or owner-occupied
Object subtype	Variable that describes whether the property is a condominium, a detached single-family house, or a townhouse
Condition	Variable that describes the general condition of the property
Regiotype	Variable that describes the type of area in which the property is located
NUTS	Variable that specifies the NUTS 1/2/3 region associated with the property

Note: The table provides an overview of the variables used.

3.8.2 Appendix II – Micro Score

The micro score of a location is calculated via a gravity model and reflects the accessibility in the sense of proximity to selected everyday destinations. A gravity model is a standard method for approximating the accessibility of a location and is based on the assumption that nearby destinations play a more significant role in everyday life than more distant destinations (Handy & Clifton (2001). The relevant points-of-interest (POIs) are selected from the findings of Powe et al. (1995), Metzner & Kindt (2018), Yang et al. (2018), Nobis & Kuhnimhof (2018) and Huang & Dall'erba (2021) and are provided in Table 3.10.

Our gravity model can be described using an activity function $f(A_p)$ and a distance function $f(D_{i,p})$:

$$A_{i,p} = \sum f(A_p) f(D_{i,p}). \qquad (14)$$

$A_{i,p} \in [0{,}100]$ denotes the accessibility of point i for the POI p, whereby the activity function $f(A_p)$ specifies the relative importance of POI p, with $f(A_p) \in [0{,}1]$. $f(D_{i,p})$ measuring the travel time from point i to the POI p by using a non-symmetric sigmoidal distance function. The travel time was obtained for the selected POIs via Open Street Map and normalized using the following function:

$$L(x) = \frac{K}{(1 + Qe^{0.5x})^{\frac{1}{v}}}, \tag{15}$$

where $K, Q \in \mathbb{R}$ and $v \in \mathbb{R}^{+}$ are defined for all possible distances $x \in \mathbb{R}$. Furthermore, we have:

$$\begin{aligned} K &= (1 + Q)^{1+v}, \\ Q &= v \cdot \exp(B \cdot x^{*}), \\ v &= \frac{\exp(B \cdot x^{*}) - 1}{\ln(y_i) - 1}, \end{aligned} \tag{16}$$

where x^{*} denotes a feature specific point of inflection and y^{*} is 0.5.

Table 3.10: Features of the micro score of a location

Points-of-Interests	Category	Description
University	Education & Work	University campus: institute of higher education
School	Education & Work	Facility for education
Kindergarten	Education & Work	Facility for early childhood care
CBD	Education & Work	Centre of the next city
Supermarket	Local Supply	Supermarket – a large shop with groceries
Marketplace	Local Supply	A marketplace where goods are traded daily or weekly
Chemist	Local Supply	Shop focused on selling articles for personal hygiene, cosmetics, and household cleaning products
Bakery	Local Supply	Place for fresh bakery items
ATM	Local Supply	ATM or cash point
Hospital	Local Supply	Facility providing in-patient medical treatment
Doctors	Local Supply	Doctor's practice / surgery
Pharmacy	Local Supply	Shop where a pharmacist sells medications
Restaurant	Leisure & Food	Facility to go out to eat
Café	Leisure & Food	Place that offers casual meals and beverages
Park	Leisure & Food	A park, usually urban (municipal)
Fitness Centre	Leisure & Food	Fitness Centre, health club or gym
Movie Theatre	Leisure & Food	Place where films are shown
Theatre	Leisure & Food	Theatre where live performances take place
Shopping Mall	Leisure & Food	Shopping Centre – multiple shops under one roof
Department Store	Leisure & Food	Single large shop selling a large variety of goods
Subway Station	Transportation	City passenger rail service
Tram Station	Transportation	City passenger rail service
Railway Station	Transportation	Railway passenger only station
Bus Stop	Transportation	Bus stops of local bus lines
E-Charging Station	Transportation	Charging facility for electric vehicles

Note: The descriptions of the selected Points-of-Interest is based on the explanations of Open Street Map.[26]

[26] See https://wiki.openstreetmap.org/wiki/Map_features.

3.9 References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna: A next-generation hyperparameter optimization framework.* In proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2623-2631.

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, *39*(2), 1772–1778.

Baldominos, A., Blanco, I., Moreno, A., Iturrarte, R., Bernárdez, Ó. & Afonso, C. (2018). Identifying Real Estate Opportunities Using Machine Learning. *Applied Sciences, 8*(11), 2321.

Bourassa, S.C., Cantoni, E. & Hoesli, M. (2010). Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions. *Journal of Real Estate Research, 32*(2), 139–159.

Bunke, O., Droge, B., & Polzehl, J. (1999). Model selection, Transformations and Variance Estimation in Nonlinear Regression. *Statistics, 33*(3), 197–240.

Cajias, M., Willwersch, J. & Lorenz, F. (2019). *I know where you will invest in the next year – Forecasting real estate investments with machine learning methods* [Conference Paper]. European Real Estate Society Meeting, Cergy-Pontoise Cedex, France.

Chau, K. W., & Chin, T. L. (2003). A critical review of literature on the hedonic price model. International Journal for Housing Science and its applications, 27(2), 145-165.

Chun Lin, C., & Mohan, S. B. (2011). Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis*, *4*(3), 224–243.

Dąbrowski, J., & Adamczyk, T. (2010). Application of GAM additive non-linear models to estimate real estate market value. *Geomatics and Environmental Engineering*, *4*(2), 55-62.

Fahrmeir, L., Kneib, T., Lang, S. & Marx, B. (2013). *Regression: Models, Methods and Applications*. Berlin: Springer.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Handy, S. L., & Clifton, K. J. (2001). Evaluating Neighborhood Accessibility: Possibilities and Practicalities. *Journal of Transportation and Statistics*, *4*(2), 67-78.

Hastie, T., Friedman, J. & Tibshirani, R. (2001). *The Elements of Statistical Learning.* New York: Springer.

Hastie, T., & Tibshirani, R (1987). Generalized additive models: some applications. *Journal of the American Statistical Association, 82*(398), 371-386.

Hong, J., Choi, H., & Kim, W. (2020). A House price valuation based on random forest approach: The mass appraisal of residential property in south korea. *International Journal of Strategic Property Management*, *24*(3), 140–152.

Huang, Y. & Dall'erba, S. (2021). Does proximity to school still matter once access to your preferred school zone has already been secured? *The Journal of Real Estate Finance and Economics, 62*(4), 548–577.

Just, T. & Maennig, W. (Eds.) (2012). *Understanding German Real Estate Markets*. Berlin/Heidelberg: Springer.

Just, T. & Schaefer, P. (2017). *Germany's Regional Structure*. In Understanding German Real Estate Markets, Berlin/Heidelberg: Springer, 41–57.

Kok, N., Koponen, E.-L. & Martínez-Barbosa, C.A. (2017). Big Data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management, 43*(6), 202–211.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*(2), 132–157.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. *Housing Economics and Public Policy, 1*, 67–89.

Mayer, M., Bourassa, S.C., Hoesli, M. & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research, 12*(1), 134–150.

McCluskey, W.J., McCord, M., Davis, P.T., Haran, M. & McIlhatton, D. (2013). Prediction accuracy in mass appraisal: a comparison of modern approaches. *Journal of Property Research, 30*(4), 239–265.

McCluskey, W., Davis, P., Haran, M., McCord, M., & McIlhatton, D. (2012). The potential of artificial neural networks in mass appraisal: The case revisited. *Journal of Financial Management of Property and Construction, 17*(3), 274–292.

Metzner, S., & Kindt, A. (2018). Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties. *International Journal of Housing Markets and Analysis, 11*(1), 73–100.

Mortgage Bankers Association. (2019). *The State of Automated Valuation Models in the Age of Big Data.* Mortgage Bank Association. Washington, D.C.

Nghiep, N., & Cripps, A. (2001). Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research, 22*(3), 313–336.

Nobis, C. & Kuhnimhof, T. (2018). *Mobilität in Deutschland – MiD: Ergebnisbericht*. Bundesministerium für Verkehr und digitale Infrastruktur, Bonn.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, *32*(3), 289–320.

Pace, R. K., & Hayunga, D. (2020). Examining the information content of residuals from hedonic and spatial models using trees and forests. *The Journal of Real Estate Finance and Economics*, *60*(1-2), 170–180.

Pace, R. K. (1998). Appraisal using generalized additive models. *Journal of Real Estate Research*, *15*(1), 77–99.

Pace, R. K., & LeSage, J. (2004). Spatial Statistics and Real Estate. *The Journal of Real Estate Finance and Economics, 29(2)*, 147–148.

Páez, A., Long, F., & Farber, S. (2008). Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques. *Urban Studies, 45*(8), 1565–1581.

Pham, D. T. (1970). Neural Networks In Engineering. *WIT Transactions on Information and Communication Technologies, 6*, 5–36.

Powe, N.A., Garrod, G.D. & Willis, K.G. (1995). Valuation of urban amenities using an hedonic price model. *Journal of Property Research, 12*(2), 137–147.

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

Schulz, R., Wersing, M., & Werwatz, A. (2014). Automated valuation modelling: a specification exercise. *Journal of Property Research*, *31*(2), 131–153.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, *13*(1), 1–44.

Stang, M., Krämer, B., Nagl, C., & Schäfers, W. (2022). From human business to machine learning—Methods for automating real estate appraisals and their practical implications. *Zeitschrift Für Immobilienökonomie*, Advanced online publication.

Tse, R. Y. C. (2002). Estimating Neighbourhood Effects in House Prices: Towards a New Hedonic Model Approach. *Urban* Studies*, 39*(7)*,* 1165–1180.

Wood, S.N. (2017). *Generalized Additive Models: An Introduction with R*. Boca Raton: CRC Press.

Yang, J., Bao, Y., Zhang, Y., Li, X., & Ge, Q. (2018). Impact of Accessibility on Housing Prices in Dalian City of China Based on a Geographically Weighted Regression Model. *Chinese Geographical Science*, *28*(3), 505–515.

Yao, Y., Zhang, J., Hong, Y., Liang, H., & He, J. (2018). Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. *Transactions in GIS, 22*(2), 561–581.

Yilmazer, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy*, 99, 104889

Yoo, S., Im., J., & Wagner, J. E. (2012). Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, NY. *Landscape and Urban Planning, 107*(3), 293–306.

Zurada, J., Levitan, A. & Guan, J. (2011). A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context. *Journal of Real Estate Research*, *33*(3), 349–388.

4 Changing the Location Game – Improving Location Analytics with the Help of Explainable AI

4.1 Abstract

Besides its structural and economic characteristics, the location of a property is probably one of the most important determinants of its underlying value. In contrast to property valuations, there are hardly any approaches to date that evaluate the quality of a real estate location in an automated manner. The reasons are the complexity, the number of interactions and the non-linearities underlying the quality specifications of a certain location. By combining a state-of-the-art machine learning algorithm and the local post-hoc model agnostic method of Shapley Additive Explanations, this paper introduces a newly developed approach – called SHAP location score – that is able to detect these complexities and enables assessing real estate locations in a data-based manner. The SHAP location score represents an intuitive and flexible approach based on econometric modeling techniques and the basic assumptions of hedonic pricing theory. The approach can be applied post-hoc to any common machine learning method and can be flexibly adapted to the respective needs. This constitutes a significant extension of traditional urban models and offers many advantages for a wide range of real estate players.

Keywords: Location Analytics, Explainable AI, Machine Learning, Shapley Values, Automated Location Valuation Model

4.2 Introduction

Besides its structural and economic characteristics, the location of a property is probably one of the most important determinants of its underlying value. While there have been standardized and globally accepted methods for assessing property values (see, e.g., Parker, 2016), the assessment of the location quality at a given site remains a kind of arbitrary estimation based on more or less subjective individual opinions. Interestingly, these opinions are mostly formed on the basis of factors that can in principle be objectified. One example of this is the accessibility of certain amenities, such as the nearest supermarket, the nearest public transport stops and the nearest park. This accessibility or centrality of a location is analyzed by real estate experts such as brokers, real estate developers and investment managers and combined with other location-related objectifiable information – e.g., population structure, employment structure – in order to make a final judgment about the quality of a particular location.

The use of these location-specific factors that can in principle be objectified, raises the question of whether this process can be standardized and automated by means of computer-based models. Such models have been used in real estate valuation, for example, for several decades. They are usually referred to as Automated Valuation Models (AVM) and the subject of numerous scientific studies (see, e.g., Glumac & Des Rosiers, 2021). However, in contrast to property valuations, the process of evaluating the quality of a location is much more unstructured, requiring the linking of complex interactions and multilayered non-linear relationships. Modern machine learning models have the advantage that they can automatically capture such complex structures and thus enable their measurement, whereas this requires much more manual effort in traditional econometric models and is simply not feasible in practice. Deep neural networks, for example, are modeled on the functioning of the human brain and can independently generate insights that were previously difficult to capture with parametric and semi-parametric econometric models. In addition to neural networks, complex tree-based models, such as the XGBoost algorithm, have repeatedly shown that they are capable of delivering more accurate results than more restrictive models (see, e.g., Sangani et al., 2017; Mayer et al., 2019; Pace & Hayunga, 2020)

To the best of our knowledge, there is no approach to date that leverages the capabilities of modern machine learning algorithms to capture the quality of real estate locations in an automated manner. The objective of this paper is therefore to present a new methodology based on a state-of-the-art machine learning model and a post-hoc model agnostic[27] explainable artificial intelligence (XAI) approach, namely Shapley Additive Explanations, to evaluate real estate locations. Based on the assumption that the quality of a property's location is reflected in the individual willingness to pay for the property, and that the quality can be measured by means of objectifiable factors, this paper introduces a new approach that enables the rating of individual property locations. We call this approach "SHAP location score" (SHAP-LS). The approach is characterized by its high degree of flexibility and can be implemented in a model-agnostic manner for any machine learning algorithm and for any feature set.

Throughout the paper, we first present the theoretical foundation of the SHAP-LS and then introduce the approach in more detail. Finally, using a dataset of 26,860 residential rental listings for the city of London from 2020, and a comprehensive set of location-specific points-of-interests (POIs), we show how an empirical implementation of the SHAP-LS is possible and what results can be obtained. The results of our empirical example demonstrate that the SHAP-LS method can identify the key factors influencing residential real estate location quality at a detailed (street) level, revealing areas of high and low location quality within the city. The findings indicate significant variations in individuals' willingness to pay for different levels of location quality, with the highest willingness observed in central areas. By categorizing the different POIs, the impact of different points of interest on location quality can be examined. The analysis reveals that proximity to public transport stops, educational institutions, shopping facilities, and negative POIs strongly influence the willingness to pay for location quality, while other categories like food shopping POIs have a lesser impact.

The SHAP-LS approach offers an initial automated valuation of real estate locations suitable for various real estate stakeholders. In academia, this approach extends traditional urban models by incorporating modern non-parametric valuation algorithms, allowing for a comprehensive examination of economic and econometric phenomena. For practitioners, the approach provides a quick empirical assessment of location quality within a city, enhancing market transparency and facilitating more informed decision-making. Accordingly, the SHAP-LS can be used for various purposes in real estate practice, such as the evaluation of different locations within a city for investment decisions.

[27] This term describes the fact that this technique is applied after the actual training of an algorithm (= post-hoc) and can be applied for different algorithms (= model-agnostic).

4.3 Theoretical Foundation

The SHAP-LS is based on two different fundamental assumptions. Assumption I: '*The quality of a certain location is reflected in the price of a property through the individual willingness to pay of individual market participants and their competition with each other.'* Assumption II: '*The quality of a particular location can be measured by means of individual features describing the location.'*

Assumption one represents an area that has already been much researched. Various studies show that market participants have different levels of willingness to pay with regard to different locations. A recent example can be found in Gabe et al. (2021), who examine residential consumer willingness to pay for location efficiency, which represents a normative component of new urbanism and describes a location based on the following five dimensions: urban design, density, land-use diversity, access to transit, and destination accessibility. Their results show that residential renters are willing to pay for multiple attributes of more efficient locations and require discounts for less desirable attributes. Other examples within the residential space can be found in Bartholomew & Ewing (2011), Seo et al. (2014), Freybote et al. (2015) and Jauregui et al. (2019). It is therefore evident that, in addition to the property itself, location-related attributes also play an important role and thus decisively determine the quality of a location.

In order to deal with the second assumption, a theory or methodology is required that allows measuring the marginal price effects of individual location-describing features. Such a theoretical foundation can be found in hedonic pricing theory. Similar to the basic assumptions mentioned above, hedonic price theory assumes that the value or price of an economic heterogenous good can be decomposed and determined on the basis of the sum of the marginally observable prices of the individual components of the good. One of the first applications of the hedonic pricing theory can be found in Court (1939) who used the theory to determine automobile prices. The first adaptations and implementations within a real estate context can be found in Lancaster (1966) and Rosen (1974), who assume that consumers derive value from different housing characteristics and that this value can be priced. Regarding the consumption of housing, consumers maximize their utility within their budget constraint. As Sirmans et al. (2005) show, the hedonic model generally takes the form:

$$Price = f(Physical\ Characteristics, Other\ Factors) \tag{17}$$

It is therefore generally assumed that, in addition to the physical characteristics of a property, such as the number of rooms or of bathrooms, other factors also play a price-determining role. These factors are generally referred to as location and neighborhood variables (Can, 1992 and Stamou et al., 2017). Location variables define the geographic location of a property, while neighborhood variables describe its social and economic environment. With regard to all three groups of variables, there is vast scientific evidence on the effect of individual variables on the price of a property. From a residential perspective for physical aspects, for example, the studies by Sirmans et al. (2005), Kestens et al. (2006), Randeniya et al. (2017) and Metzner & Kindt (2018) provide a good overview. Location-specific aspects, on the other hand, are the focus of Hoen & Atkinson-Palombo (2016), Dumm et al. (2016), W. Seo (2018) and Turner & Seo (2021). Accordingly, the applicability of the hedonic pricing theory can be seen as well proven and suitable for locational aspects in the real estate context. Thus, the theoretical framework of our approach is based on the principles of hedonic price theory. We make use of this theory and extract the location-specific effects of real estate prices and relate them to the quality of their respective location.

Whilst the majority of empirical papers use parametric or semi-parametric models to determine marginal prices in hedonic price studies, non-parametric models are not used that often, mainly due to their lack of intrinsic explanatory power. One category of models that falls into this group are the so-called modern machine learning approaches, such as neural networks or decision-tree based models. As Chun Lin & Mohan (2011), Kok et al. (2017), Mayer et al. (2019) and Stang et al. (2022) show, these models provide a high degree of accuracy, but the results obtained cannot be explained by the model itself. However, by using so-called model-agnostic approaches, post-hoc explainability can be created (Lorenz et al., 2022; Krämer et al., 2023a). Model-agnostic approaches describe a set of different new methods, which can in principle be applied to any machine learning algorithm and allow an investigation of the relationships learned by the model (Molnar, 2020).

We make use of these methods and, based on them, present a new methodology called SHAP-LS, in order to determine the quality of a location for real estate purposes. Thus, the novelty and contribution to the literature of the SHAP-LS lies essentially in the extension of traditional urban models by means of modern machine learning and model-agnostic methods. Compared to the previous parametric and semi-parametric approaches, the SHAP-LS enables taking complex and non-linear relationships into account. Especially with regard to the assessment of the quality of a location, this creates advantages, since location structures are often characterized by a multitude of different interactions (Gabe et al., 2021). With regard to the SHAP-LS, we decided to use the local model-agnostic method of Shapley Additive Explanations. Shapley values allow us to examine how much individual features contribute to the difference between an individual and the average prediction. This in turn allows us to capture the marginal effects of the location-determining features at the level of individual observations and thus to measure the influence of the location quality for any given location within a dataset.

4.4 Methodology

Following our theoretical foundation, the aim of this paper is to combine hedonic pricing theory and modern machine learning algorithms to develop a purely data-driven approach for assessing the quality of a property location. The approach is designed with special attention to the necessary flexibility, adaptability, simplicity and easy implementation. The mentioned criteria are therefore important to fulfill, because only this can guarantee that the approach can be adapted for a multitude of different purposes and thus be integrated into general scientific and practical use.

As the SHAP-LS is based on a post-hoc XAI methodology, initially the training of an appropriate valuation algorithm is required. To ensure the applicability of the SHAP-LS methodology, several points regarding the training need to be considered. The first step is the selection of a dataset suitable for analysis. In addition to the price of the property[28] and its physical characteristics, especially a comprehensive set of location-related features is required. Furthermore, the dataset has to be as spatially and temporally dense as possible. Since the quality of individual locations can change over time, the temporal density prevents possible bias over the years. The spatial density allows the presentation of granular results, which are visually easier to show and additionally allow extrapolation and interpolation of the results. Due to the model agnostic character of the SHAP-LS methodology, in principle any machine learning methodology can be applied to train the underlying valuation algorithm. Given their performance characteristics, generally, the use of modern tree-based models, such as the extreme Gradient Boosting algorithm, or the use of neural networks (Antipov & Pokryshevskaya, 2012; McCluskey et al., 2012; Yoo et al., 2012; Yilmazer & Kocaman, 2020 and Ho et al., 2021) is recommended.

Post-hoc to the model training, the Shapley Additive Explanations (SHAP) have to be calculated on the basis of the trained machine learning algorithm. Developed by Lundberg & Lee (2017), SHAP is a local post-hoc model-agnostic technique based on coalitional game theory named Shapley values in order to detect the contribution of a feature on a single prediction compared to the average prediction. One main advantage of this method

[28] In the context of the SHAP-LS methodology, it is in principle possible to use both purchase or rental prices. Both reflect the observable willingness to pay for a property with certain characteristics and a certain location and can thus be used in this logic in an arbitrary manner.

is that it can be applied to all machine learning models. The SHAP value of a feature value is its contribution to the payout, which is weighted and summed over all possible feature value combinations. In its original game theoretical context, the SHAP value of an observation i and feature j can be computed as follows:

$$\phi_{i,j} = \sum_{S \subseteq \{x_{i,1},\dots,x_{i,p}\}\{x_{i,j}\}} \frac{|S|!\,(p-|S|-1)!}{p!}(f\left(S \cup \{x_{i,j}\} - f(S)\right), \tag{18}$$

where $\phi_{i,j}$ denotes the features contribution, S is a coalition, $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{p} \times \boldsymbol{n}}$ is the full feature space containing p variables and n observations. Furthermore, $f(S)$ represents the prediction of a model f on S. As a result, the SHAP methodology shows a vector of the individual SHAP values for all features. It is important to note that the values are not valid on an aggregated level, but are determined separately for each individual observation in the dataset.[29] SHAP values show how the individual features influence the prediction made by the machine learning model in comparison to the average prediction. A comprehensive discussion of SHAP values and an example of how to interpret their results can be found in Appendix I.

To finally determine the SHAP-LSs, the identified SHAP values are used. In non-technical terms, the SHAP-LS of a single observation i is computed by extracting the SHAP values of all location-specific features j and adding them up. In order to increase the significance of the results, the top and bottom percentile of all SHAP-LSs is removed and in a last step, the SHAP-LS is scaled on the basis of all other SHAP-LSs. Technically, this can be expressed as follows:

Assume that there is a machine learning model f, a full feature space $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{p} \times \boldsymbol{n}}$ containing p variables and n observations. The SHAP values of each data point in $\boldsymbol{x}$ can be computed by using model f and are stored in the SHAP matrix $\boldsymbol{\phi} \in \mathbb{R}^{\boldsymbol{p} \times \boldsymbol{n}}$. Let $\boldsymbol{\phi_M} \in \mathbb{R}^{\boldsymbol{q} \times \boldsymbol{n}}$, $q \leq p$, be a subset of $\boldsymbol{\phi}$ containing the SHAP values of all location-specific features. For all observations, $i \in \{1, \dots, n\}$, all SHAP values $\phi_{i,j} \in \boldsymbol{\phi_M}$ are summed:

$$\phi_i = \sum_{j=1}^{q} \phi_{i,j}. \tag{19}$$

To obtain the final overall SHAP-LS, the top and bottom percentile of all ϕ_i are dropped and the SHAP-LSs are scaled between -1 and 1 in order to provide an easily interpretable scoring:

[29] An example of the identification of aggregated results would be the Permutation Feature Importance (see, e.g., Krämer et al., 2023a).

$$\phi_{scaled,i} = \frac{\phi_i - \min_{i\,\in\{1,\dots,n\}}(\phi_i)}{\max_{i\,\in\{1,\dots,n\}}(\phi_i) - \min_{i\,\in\{1,\dots,n\}}(\phi_i)} \cdot 2 - 1. \tag{20}$$

$\phi_{scaled,i}$ is set to be the final SHAP-LS of observation i and can be interpreted as the relative importance of the location-specific features of observation i. The higher the value, the higher the marginal willingness to pay for the features used as location quality descriptors. The relative comparison across all available observations makes it possible to determine the quality of different locations on the basis of the prevailing marginal willingness to pay for all characteristics describing the location.

Besides analyzing the overall location quality, the SHAP-LS methodology can also be used to assess the individual features that drive the quality of a particular location. This can be done by creating categorical SHAP-LSs for individual related feature groups.[30] In order to compute these categorical SHAP-LSs, a subset of all location-specific SHAP values containing the SHAP values of the features in this category has to be selected, $\boldsymbol{\phi_{CAT}} \subseteq \boldsymbol{\phi_M}, \boldsymbol{\phi_{CAT}} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}}$, $m \leq q$. The following steps are similar to those above. First, for every observation $i \in \{1, \dots, n\}$ the SHAP values $\phi_{i,j} \in \boldsymbol{\phi_{CAT}}$ are summed:

$$\phi_{CAT,i} = \sum_{j=1}^{m} \phi_{i,j}. \tag{21}$$

Again, the top and bottom percentile of all $\phi_{CAT,i}$ is excluded and all $\phi_{CAT,i}$ are scaled between -1 and 1 to ensure easy interpretation. Therefore, $\phi_{CAT,scaled,i}$ defines the final categorical SHAP-LS and can be interpreted as the relative importance of a feature category for an observation i regarding the marginal willingness to pay. [31]

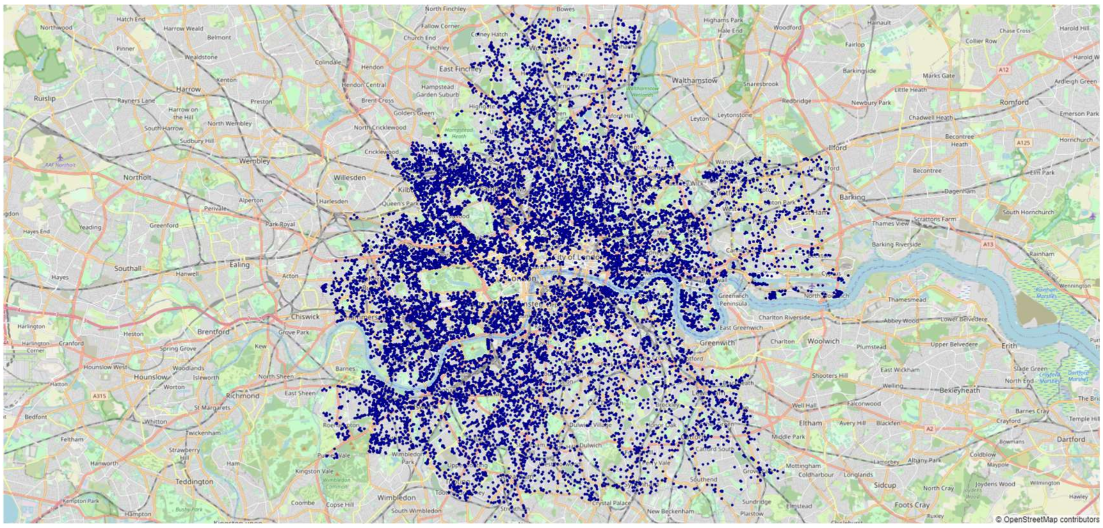
[30] Theoretically, the SHAP-LS of single features could be used for this kind of analysis. However, this is not recommended, as one is exposed to the capriciousness of the algorithms and data providers. Often, locational features such as the distance to the next bus stop and to the next subway station correlate highly. Consequently, the algorithm cannot distinguish perfectly between these correlated features, which can lead to a blurring of the individual SHAP values. Another reason that should not be neglected is the dependence on the categorization of the location characteristics of the data providers. In some cases, individual amenities overlap considerably, e.g., the classification of restaurants, pubs or bars. Combining several individual characteristics into categories can counteract this blurring. As a rule of thumb, it can be stated that the more data available, the smaller the categories that can be used.

[31] It should be noted that the overall and categorical SHAP-LSs can also be utilized without scaling, providing absolute values instead of relative comparisons. This alternative application of the SHAP-LS methodology, the absolute SHAP-LS, leads to the derivation of absolute contributions of location-specific features, expressed in the unit of the property's price. This interpretation presents a quantifiable measure of the value placed on the location characteristics compared to the average prediction by market participants, thus providing an additional perspective for analysis. It also broadens the scope for data interpretation, thereby offering new ways for investigation within the realms of real estate valuation and location quality assessment. An exemplary implementation of this can be found in Appendix VI.

4.5 Empirical Application

The following section deals with an empirical implementation of the presented SHAP-LS. The example pertains to the application in the residential real estate sector. [32] We use an exemplary dataset from the city of London for this purpose. The dataset consists of 26,860 residential rental condominiums that were listed on different Multiple Listing Systems (MLS) in 2020. Figure 4.1 shows how the observations are distributed across the city.

Figure 4.1: Distribution of individual observations across the city of London



Notes: This figure shows the spatial distribution of all available observations within the city of London. The 26,860 observations are distributed over the entire city, whereby the spatial density is particularly high in the center and decreases towards the outskirts. The data was generated by webscrapping from different MLSs, and cleaned using standardized procedures.

In addition to the rental price, a set of hedonic features defining the physical characteristics of the individual condominiums and a time variable to capture temporal and seasonal effects are available. To compute the SHAP-LS, a comprehensive set of location-specific features is enriched, using the geocoordinates of the condominiums. We capture the location quality by means of the distance to the nearest major amenities. The distance is calculated using Euclidean proximity. It is important to understand that the SHAP values alone do not provide information about whether the distance is positively or negatively correlated with the predicted rent. Therefore, it cannot be determined whether a change in distance would have a positive or negative impact on the price. As a result, there is no need for data engineering on the different distance features to reflect the fundamental differences in the assumed economic desirability of these features. A total of 41 relevant POIs are selected and used for model estimation. These individual POIs can be grouped into a total of 9 categories (transportation, healthcare, education, food shopping, local

[32] A further application in other asset classes, such as retail, industrial, hotel, or office, is also theoretically possible but beyond the scope of this study.

supply, shopping, eateries, leisure and negative POIs[33]) and, as a whole, provide a comprehensive and conclusive assessment of the accessibility of a location. The integration of further location-related features is in principle arbitrarily representable, but beyond the scope of this study. The descriptive statistics of the dataset can be seen in Table 4.1. The mean asking rent is £2,111 per month, with rental values ranging from £346 per month to £4,546 per month. The average apartment has roughly two bedrooms and has a combined area of 64.68 sqm. Regarding the nearest major amenities, it is observed that, on average, they are all located in the immediate vicinity of the observations in the dataset. The farthest amenity, on average, is the nearest wastewater plant, located away 5.48 km, while the nearest bus stop is on average only 0.12 km away, indicating the high urbanity of the observations and the overall excellent availability of public transportation in London.

Table 4.1: Descriptive statistics

Variable	Category	Unit	Minimum	Median	Mean	Maximum	Standard Deviation
Rent GPB/p.m.	Price	Metric	346.00	1948.00	2111.42	4546.00	769.85
Bedrooms	Physical	Integer	1.00	2.00	1.97	5.00	0.80
New build	Physical	Binary	0.00	0.00	0.07	1.00	0.25
Size sqm	Physical	Metric	20.07	61.04	64.68	248.7	23.44
Month of listing	Time	Integer	1.00	6.00	5.63	12.00	3.32
Bus stop_dist	Transportation	Km	0.00	0.11	0.12	0.79	0.08
Bus station_dist	Transportation	Km	0.03	2.03	2.33	8.09	1.44
Railway station_dist	Transportation	Km	0.01	0.43	0.49	3.18	0.30
Car sharing_dist	Transportation	Km	0.00	0.31	0.42	4.43	0.42
Bicycle rental_dist	Transportation	Km	0.00	0.32	0.94	7.12	1.21
Motorway junction_dist	Transportation	Km	0.04	4.49	4.50	10.87	2.23
Doctors_dist	Healthcare	Km	0.00	0.48	0.55	2.77	0.36
Hospital_dist	Healthcare	Km	0.01	0.89	1.01	4.28	0.61
Pharmacy_dist	Healthcare	Km	0.00	0.33	0.39	2.56	0.27
Dentist_dist	Healthcare	Km	0.00	0.44	0.55	3.96	0.43
School_dist	Education	Km	0.00	0.21	0.24	0.99	0.13
Kindergarten_dist	Education	Km	0.00	0.52	0.61	3.77	0.42
University_dist	Education	Km	0.01	1.19	1.44	6.73	1.08
College_dist	Education	Km	0.01	0.84	1.02	5.27	0.79
Greengrocer_dist	Food Shopping	Km	0.00	0.78	0.95	6.00	0.73
Bakery_dist	Food Shopping	Km	0.00	0.51	0.62	4.28	0.49
Supermarket_dist	Food Shopping	Km	0.00	0.32	0.36	1.93	0.24
Kiosk_dist	Local Supply	Km	0.01	1.20	1.48	6.97	1.05
Atm_dist	Local Supply	Km	0.00	0.31	0.39	3.07	0.31
Post office_dist	Local Supply	Km	0.00	0.44	0.48	2.35	0.27
Post box_dist	Local Supply	Km	0.00	0.13	0.14	0.87	0.09
Hairdresser_dist	Local Supply	Km	0.00	0.26	0.35	3.90	0.36
Laundry_dist	Local Supply	Km	0.00	0.28	0.36	3.41	0.32
Optician_dist	Local Supply	Km	0.00	0.60	0.72	4.66	0.56
Department store_dist	Shopping	Km	0.01	1.19	1.39	6.73	1.00
Mall_dist	Shopping	Km	0.02	1.31	1.62	6.69	1.13
Clothes_dist	Shopping	Km	0.00	0.46	0.55	4.23	0.43
Restaurant_dist	Eateries	Km	0.00	0.18	0.24	2.18	0.22
Fast food_dist	Eateries	Km	0.00	0.24	0.30	1.79	0.22

[33] Since the SHAP-LS methodology is used to analyze the extent of influence of individual location-descriptive features, there is no need to transform the negative POIs. No statement is made or needed beforehand regarding whether a greater distance to these POIs generally has a positive or negative impact on the quality of a location. The term "negative POIs" is simply chosen because they are generally perceived as negative by humans. The logic and use of these POIs are therefore arbitrary in relation to the other POIs.

Cafe_dist	Eateries	Km	0.00	0.19	0.24	2.23	0.21
Beer garden_dist	Eateries	Km	0.01	2.24	2.51	9.05	1.42
Sports centre_dist	Leisure	Km	0.00	0.47	0.52	2.49	0.31
Park_dist	Leisure	Km	0.00	0.23	0.27	1.26	0.16
Playground_dist	Leisure	Km	0.01	0.27	0.32	1.44	0.21
Swimming pool_dist	Leisure	Km	0.01	1.02	1.14	5.54	0.71
Bar_dist	Leisure	Km	0.00	0.41	0.53	4.10	0.45
Nightclub_dist	Leisure	Km	0.01	1.20	1.53	7.15	1.18
Pub_dist	Leisure	Km	0.00	0.18	0.22	1.71	0.17
Prison_dist	Negative POIs	Km	0.09	3.81	4.15	14.08	2.36
Wastewater plant_dist	Negative POIs	Km	0.03	5.46	5.48	14.38	2.77
Graveyard_dist	Negative POIs	Km	0.04	1.37	1.43	3.99	0.72
Windmill_dist	Negative POIs	Km	0.02	3.95	3.91	10.25	1.78

Notes: This table reports the descriptive statistics of the dataset. In addition to the physical features describing the condominiums itself and a temporal variable, distances to the most important POIs were added for all observations, so as to describe the spatial location of the individual condominiums. The distances were determined by means of the Euclidean distance and are given in kilometers (Km). The selection of the parameters was in accordance with other publications in the real estate literature (see, e.g., Metzner & Kindt, 2018). The parameter "Rent GPB/p.m." is the dependent variable in our model and describes the asking rent for the individual condominiums.

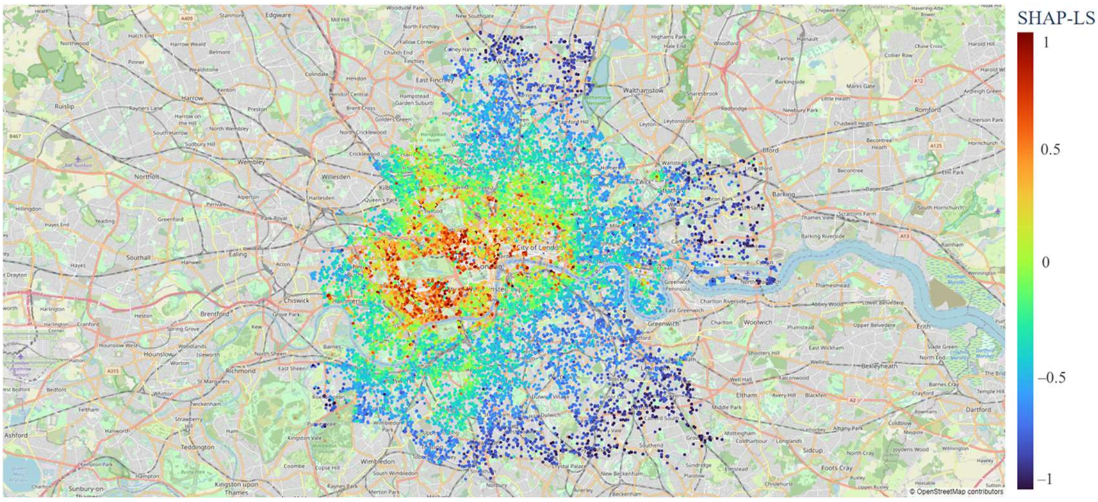
As an underlying machine learning model, the XGBoost algorithm is chosen. This is a computationally effective and highly efficient sequential tree-based ensemble learning method. For more information about tree-based methods, ensemble learning and (extreme) gradient boosting, the reader is advised to read Hastie et al. (2001) and Chen & Guestrin (2016). In order to ensure that the machine learning model provides reliable results, the model was tested by using five-fold cross validation on four different evaluation metrics, namely the Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE) and the 10/20 Percent Error Buckets (PE(10) and PE(20)). A more detailed description of the error metrics can be found in Appendix II. The results indicate a strong and robust performance of the XGBoost. The MAPE of 15.36% falls within a very good range, indicating that the XGBoost model is able to estimate rents with only a 15.36% deviation on average. Additionally, the MdAPE of 9.68% and the PE(10) of 51.24% and the PE(20) of 78.51% also suggest a significantly robust and accurate estimation. Overall, the performance level of the XGBoost model is consistent with the results reported in other studies (see e.g., Mayer et al., 2019; Stang et al., 2022; Krämer et al., 2023a). Therefore, it is well suited for a post-hoc analysis and the SHAP-LS can be calculated.[34]

Figure 4.2 visualizes the results of the calculated SHAP-LS.[35] With regard to the visual representation, relatively attractive locations are shown in red, and locations seen as relatively unattractive in blue. The selected color gradient makes it easy to recognize differences between the individual subareas.

[34] To obtain shap values of the features of interest the shap package (https://shap.readthedocs.io/en/latest/index.html) is used.

[35] To ensure the reliability of results obtained from the newly introduced SHAP-LS across various model specifications, hyperparameter choices, and other potential variations, two distinct robustness checks are performed. The results can be seen in Appendix III.

Figure 4.2: Overall SHAP location score



Notes: This figure visualizes the final results of the calculated SHAP-LSs. For each individual observation within the dataset, the corresponding SHAP-LS is determined using the methodology described above. The visual representation of the scores clearly shows that there are key differences with regard to the quality of different real estate locations. It is particularly noticeable that the central locations near the CBD seem to be characterized by above-average quality.

As the picture shows, the SHAP-LS methodology recognizes different levels of location attractiveness or location quality within the city of London. It can be seen that particularly attractive locations are distributed around the center of London and are predominantly found in the districts of Mayfair, Kensington, South Kensington, Knightsbridge, Brompton, Chelsea and Convent Garden. In terms of the SHAP-LS methodology, this shows that in these parts of the city, the marginal willingness to pay for location-defining features is particularly high and takes up a large share of the rent paid in each case.[36] In contrast, relatively unattractive locations are found in particular in the outer boroughs of London. This empirical result is consistent with the theoretical assumption that attractive locations are found in the central areas of a city. With reference to the city of London, it can be seen that the SHAP-LS is capable of identifying the theoretical and presumed attractiveness levels of different locations on the basis of empirical data. Since the score is always calculated at the level of individual observations, the attractiveness of a location can be determined at the block or street level, in contrast to a conventional manual analysis, which usually refers to individual neighborhoods.[37] For practitioners, the approach provides a

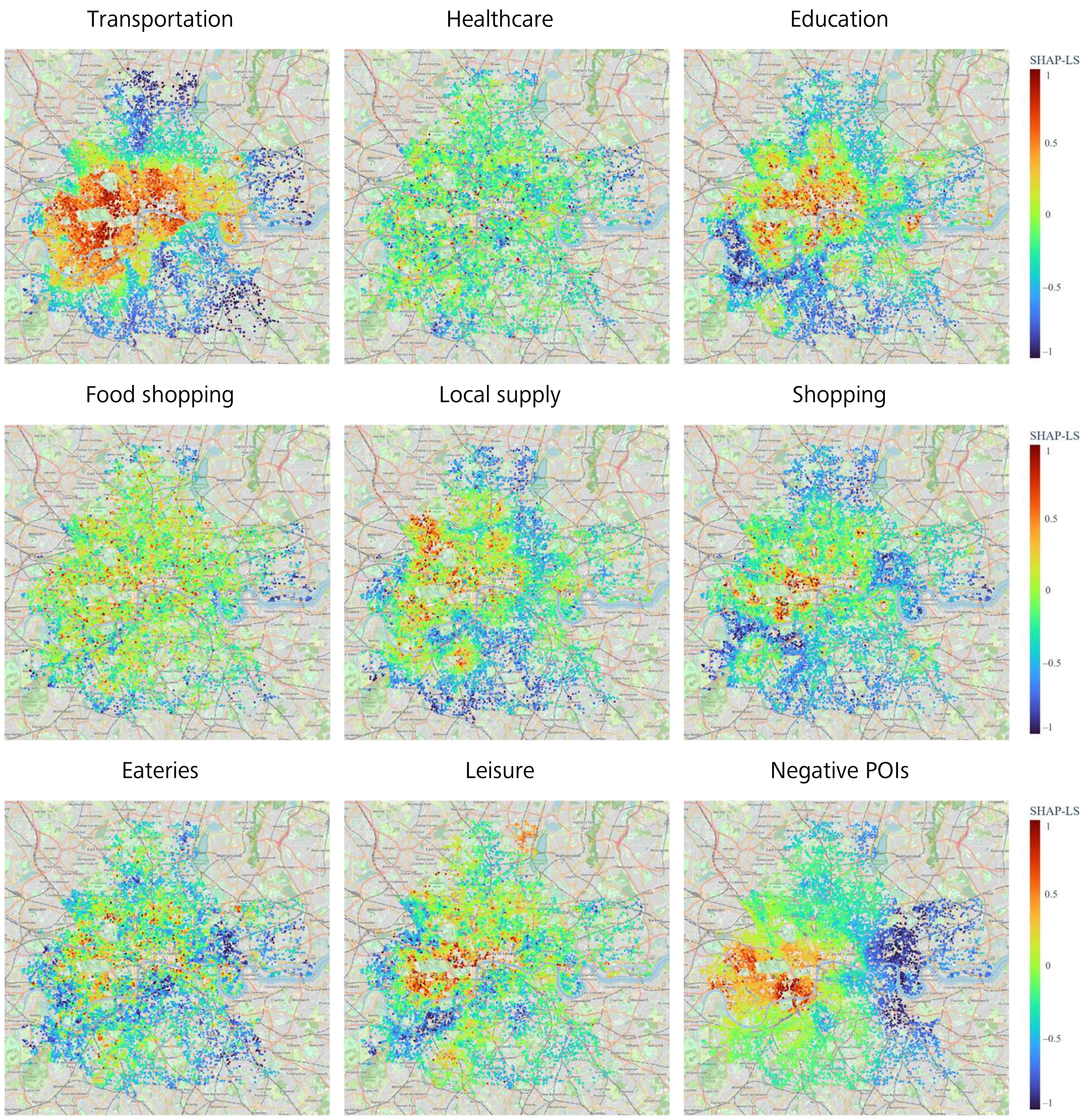
[36] To verify that the SHAP-LS does not merely reflect the rental price per square meter, the correlation between the SHAP-LS and the rent per square meter is calculated. A coefficient of 0.58 signifies that the SHAP-LS effectively captures the relative attractiveness of a location, rather than solely relying only on the prices. The presence of a discernible correlation between the appeal of a location and the rental rates should not be regarded as unexpected.

[37] A corresponding example can be seen in Appendix IV.

quick empirical assessment of location quality within a city on a more granular level, enhancing market transparency and facilitating more informed decision-making.[38]

In a next step, we divide the location-specific features into different categories to provide more insight into how the different amenities and disamenities affect the individual SHAP-LSs. Overall, we divide the amenities into nine categories: Transportation, healthcare, education, food shopping, local supply, shopping, eateries, leisure and negative POIs, like the distance to the next prison or graveyard. Figure 4.3 visualizes the final results.

Figure 4.3: Categorical SHAP location score



Notes: This figure maps the results of the calculated categorical SHAP-LSs. The results show how the individual amenities and disamenities of the individual categories influence the asking rent. The maps indicate that the effect varies by category and location.

[38] Another potential application for practitioners would be the further implementation of the SHAP-LS within an automated real estate valuation process to further improve their valuation accuracy. An empirical example of this can be found in Appendix V.

Focusing on the highlights of the maps, the visualization of the category "Education", for example, shows some interesting patterns. It is clearly evident that, apart from the city center, there seem to be individual clusters within London that are characterized by a high level of payment dispersion with regard to the spatial distance to educational institutions. A more detailed analysis shows that these clusters are located close to certain educational centers (e.g., universities). The spatial proximity to higher educational institutions thus initially has a positive influence on the quality of a property location in London. With regard to the category "Transportation", a clear pattern also emerges. The marginal price influence of POIs in this category is almost ring-shaped around the city center and is arranged according to the local public transport axes. Furthermore, it can be seen that there are also individual clusters here, which in turn are determined by local traffic centers in their midst. This picture clearly shows that a good public transport connection is of importance and represents an essential aspect for the determination of location quality in London.

In contrast, the effect for POIs in the categories "Food Shopping" and "Healthcare", for example, is very small and the marginal willingness to pay is almost the same across the city. Only at the outermost edges does this effect reverse. For both categories, this indicates either that these types of POIs are available equally throughout the city or that they do not play an important role in the individual determination of willingness to pay. Due to the frequency of necessary food shopping and the generally good supply within the city, the assumption for the category "Food Shopping" lies with the former argumentation. For the category "Healthcare", on the other hand, the presumption lies with the second argument, since these typically do not determine a person's everyday life and thus initially appear less important. Last but not least, it is also worth taking a look at the category "Negative POIs". Here, an east-west gradient is particularly noticeable, which is not the case in any of the other categories. It can be seen that there is a greater willingness to pay for a particular location, especially in the city center, as this is influenced only minimally by negative POIs. This shows that it is not only important to analyze positive aspects to determine the quality of a location, but also to include points that are perceived as negative. The selected examples illustrate well how the categorical SHAP-LSs can be used to explain the results of the overall SHAP-LS and to perform more detailed analyses.[39]

[39] As outlined in the methodology section, the SHAP-LS technique can be employed for location analysis without the necessity for scaling to gain further insights. An illustrative example of this approach which we call absolute SHAP-LS can be found in Appendix VI.

4.6 Conclusion

While Automated valuation models (AVM) have become popular in the field of real estate valuation, there are hardly any applications for the valuation of the quality of a real estate location. Accordingly, the objective of our work is to apply the principles of AVMs to the field of location analytics for real estate purposes. For this purpose, we have developed a new approach, named SHAP location score (SHAP-LS), which is able to assess different real estate locations on the basis of their quality. The approach results in an overall score that indicates whether the quality of a particular location is strong or weak. In addition to the overall score, the SHAP-LS method also allows us to disaggregate the score into different categories and thus analyze which factors have a greater influence on the quality of the location and which play only a subordinate role. The SHAP-LS approach is based on two basic assumptions. The first implies that the quality of a certain location is reflected in the price of a property by means of the individual willingness to pay of individual market participants and their competition with each other. The second assumption proposes that the individual differences in quality can be measured by objectifiable factors. These two basic assumptions essentially allow the theoretical framework of hedonic price theory to be applied. The combination of a non-parametric machine learning algorithm and the local post-hoc model-agnostic explainable artificial intelligence (XAI) method of Shapley Additive Explanations allows us for the first time to exploit the properties of modern machine learning methods in this particular domain, and thus to capture non-linearities and higher-order interactions, which appropriately play an important role in location analytics. Previously, this was only possible to a limited extent using parametric or semi-parametric methods.

An application of our methodology to a residential dataset of the city of London shows that the SHAP-LS method is able to extract the value - and thus also location quality - determining factors of residential real estate prices on a granular level and thus to show where within the city the quality of a location is particularly good and where it is not. The results show that there are significant differences with regard to the individual willingness to pay for different levels of location quality and that this willingness seems to be highest in the central locations of the city. While this result seems to be nothing new in principle, this time an analysis can be made not only at the level of districts, as is usually the case, but also at the street level. A presentation of the results according to different categories also allows us to further analyze which group of points of interest contributes how strongly to the location quality. The results obtained show that the degree of willingness to pay for location quality essentially depends on the spatial proximity to public transport stops, educational institutions, shopping facilities and negative POIs. Other frequently used categories, such as proximity to food shopping POIs, play a lesser role.

The results of our robustness test demonstrate that the outcomes presented in the given empirical example are generally stable concerning the chosen machine learning model, thereby enabling an initial assessment of location quality across different points within the city of London. However, it is important to note at this juncture that the overall findings of SHAP values depend on the previously selected machine learning model. Variations in SHAP values can emerge depending on the specific model employed. Hence, the robustness of the results primarily applies intramodel-specifically and not intermodel-specifically for various machine learning models. This, in a way, presents a limitation, which can, nevertheless, be mitigated to a certain extent through an appropriate model selection process, including assessing model performance.

The SHAP-LS represents an initial approach to an automated valuation of real estate locations, which is suitable for a variety of different real estate players. The approach can be applied post-hoc to any common machine learning method and can be flexibly adapted to the respective needs. The approach is not limited to certain features, and can be adapted to the existing data structure by the modeler. Regarding academia, the presented approach extends traditional urban models by the use of modern non-parametric valuation algorithms and thus allows an extended consideration of economic and econometric phenomena. For practitioners, the approach enables a way to quickly obtain an empirical assessment of the quality of different locations within a city. Thus, the transparency of local markets can be increased, and more well-founded or more objectified decisions can be made.

While the initial intention of this paper is the introduction of the newly developed SHAP-LS, there are already other use cases that are worth investigating in terms of future work. For example, the application of the approach to different asset classes needs to be examined and investigated. While the results of the empirical example have demonstrated the feasibility of applying the SHAP-LS methodology in the residential sector, it is worth noting that the application in the commercial properties sector may pose challenges due to the necessary data density and the current limited availability of data. However, if there is a sufficient density of data available, there should be no impediment to applying the methodology in the commercial sector. Furthermore, the use of the methodology to investigate the quality of a location over time seems very promising. Also, an in-depth examination of the respective effect of potential negative externalities, such as road noise or air pollution, could provide valuable insights. Additional, analyzing, if the effect of the different location describing features varies across different price levels can be done using the SHAP-LS and may provide interesting results. Lastly, regarding the categorical SHAP-LSs, investigating the spatial variability of the scores in comparison to the spatial distribution of the utilized spatial features presents an intriguing research objective. This provides information on whether there is a positive or negative correlation concerning certain spatial feature groups. Furthermore, considering the results of the categorical SHAP-LS from the empirical example in this paper, it appears intriguing to delve deeper into the patterns identified therein. At first glance, it seems to suggest a sort of local vs. global story, wherein the POIs from the categories "Transportation", "Education", "Local Supply", "Shopping", "Leisure" and "Negative POIs" exhibit kind of local patterns, while the remaining three categories show more global patterns. This aspect warrants further investigation, including a potential exploration of underlying causes. Accordingly, we encourage other authors to explore further and yet unknown areas on the basis of the SHAP-LS methodology.

4.7 Appendix

4.7.1 Appendix I – Shapley Additive Explanations

The Shapley Additive Explanations (SHAP) is a relatively new local post-hoc model-agnostic technique, that has been developed by Lundberg & Lee (2017). Post-hoc model-agnostic techniques aim to provide insights into why a machine learning model made a particular prediction for a given input. In particular, the SHAP values provide information about the influence that a specific feature has on a particular observation compared to the average prediction. Technically, the SHAP value of a feature value is its contribution to the overall payout, which is weighted and summed over all possible feature value combinations. In contrast to other post-hoc model-agnostic methods, such as Accumulated Local Effects (ALE) plots developed by Apley & Zhu (2020), SHAP values represent a local rather than a global methodology. This means that SHAP values are capable of determining the results or individual feature influences at the level of individual observations. When comparing ALE plots with SHAP values, for example, ALE plots provide insights into how the effect of a feature changes across its distribution across all observations in a dataset (see, e.g., Krämer et al., 2023a). On the other hand, SHAP values provide information about the effect of an individual feature for each individual observation. Therefore, the results are always dependent on the observed value that the individual feature takes on for each respective observation.

The SHAP values have various advantages that make their application highly beneficial. Following Molnar (2020), first and foremost, SHAP values provide results at the level of individual observations. This means that for each prediction in a dataset, the effects of individual features on the final prediction can be analyzed. In the context of real estate valuation, for example, it allows for a precise analysis of the drivers of property values and how they individually contribute. Next to the beneficial local interpretability, the SHAP values can also be used to obtain global interpretability. To do so, the individual SHAP values need to be summed across all observations of a dataset. By analyzing the distribution of SHAP values for each feature, we can determine which features have the most consistent impact on the model's predictions, enabling model evaluation and feature selection. Besides that, the SHAP values provide consistent and fair results. Consistency means, that the sum of the SHAP values of all features equals the difference between the model prediction and the expected average prediction. Fairness means, that each features fair contribution to the overall effect is extracted using the SHAP values. Finally, the SHAP values are able to account for non-linearities and higher order interactions. By using complex machine learning models, the non-linearities and interactions can be detected and made visible by the SHAP values. This ensures, that a more nuanced understanding of how different features influence the model's prediction can be achieved.

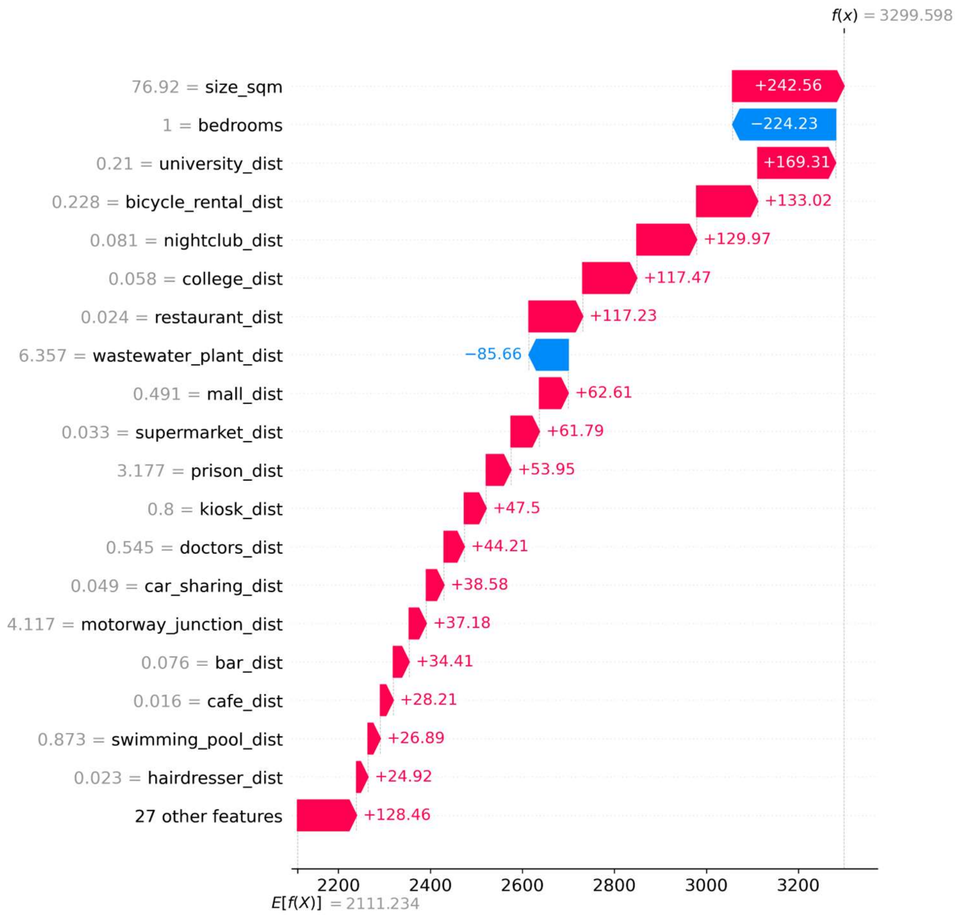
Alongside the advantages, there are also certain limitations to consider when it comes to SHAP values. In general, SHAP values can be computationally intensive, especially for complex models or large datasets. The process involves evaluating the model multiple times for different combinations of features, which can be time-consuming. Next, the results of the SHAP values depend on the performance of the previously selected machine learning model. Depending on the model chosen, different results can appear in the SHAP values. Therefore, the initial model selection is an important aspect and should always be carefully justified. Another significant critical aspect is that SHAP values have difficulties handling strongly correlated features.[40] SHAP values treat features as independent entities, assuming no correlation between them. Similar to other feature importance techniques, when calculating SHAP values, unrealistic data pairs can arise in the presence of highly correlated features, leading to erroneous overall statements. This is particularly problematic in the real estate sector, where features often exhibit strong correlations (e.g., size of the apartment and number of rooms) (Molnar, 2020). However, this problem can be addressed up to a certain extent by applying dimensionality reduction techniques, feature selection methods or as it is done within this study, by grouping the SHAP values of different features. These techniques can help in reducing the number of correlated features or capturing the most informative features, which can improve the interpretability and stability of the SHAP values.

To gain a better understanding of the functionality and interpretation of SHAP values, Figure 4.4 illustrates an exemplary calculation of the SHAP values for the model and dataset utilized in this paper i.e., a rent prediction for the city of London. The example shows the results of one single observation that was randomly selected from the available dataset. On the y-axis, the features are ranked starting with the feature that contributes least (bottom left) to the individual estimate compared to the average rent price prediction up to the feature that represents the largest differentiating factor (up right). The observed characteristics of the single observation can be seen on the left-hand side in light grey font. On the x-axis, the predicted rental price is shown. E[f(x)] denotes the price of the average prediction, £2,111, and f(x) the prediction of the selected observation, £3,300. Red arrows indicate a positive, and blue arrows a negative contribution of the individual feature. In this example, the apartment size of 77 square meters has a positive impact of £243,

[40] In the context of the SHAP-LS, the correlation between different features can be seen as only a minor concern because the results are presented at an aggregated overall or group level. The individual interpretation challenges of specific features are thus avoided.

compared to the average prediction, and the fact that the apartment has only one bedroom has a negative impact of about -£224. Regarding the individual distance features, we observe that all the shown features, except the distance to the next wastewater plant, have a positive impact on the predicted rent. For instance, considering the feature that represents the distance to the nearest university, it can be observed that the observed distance (0.21 km) positively impacts the predicted rent by £169.31. Again, it is important to understand that the SHAP values alone do not provide information about whether the distance is positively or negatively correlated with the predicted rent. Therefore, it cannot be determined whether a change in distance would have a positive or negative impact on the price. As a result, there is no need for data engineering on the different distance features to reflect the fundamental differences in the assumed economic desirability of these features. All outputs can be interpreted in the same manner.

Figure 4.4: SHAP value example



Notes: This figure shows an example of how SHAP values work. The example is an actual observation from the London dataset used for this paper. The plot shows that the average prediction of the XGBoost across all observations is £2,111, and for the observation shown here, £3,300. How the difference between the average prediction and this observation is composed can be easily seen thanks to the waterfall-like structure of the SHAP values.

4.7.2 Appendix II – Evaluation Metrics

Table 4.2: Evaluation metrics

Error	Formula	Description
Mean Absolute Percentage Error (MAPE)	$MAPE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|$	Mean of all absolute percentage errors. A lower MAPE signals higher overall prediction accuracy in percent.
Median Absolute Percentage Error (MdAPE)	$MdAPE(y,\hat{y}) = median(\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|)$	Median of all absolute percentage errors. A lower MdAPE denotes greater precision in percent without being sensitive to outliers.
Error buckets (PE(x))	$PE(x) = 100\left\|\frac{y_i - \hat{y}_i}{y_i}\right\| < x$	Percentage of predictions where the relative deviation is less than %, with x being 10 and 20. A larger PE(x) signals a lower variation in the predictions.

Notes: This table reports the evaluation metrics used to determine the valuation accuracy of the XGBoost and OLS algorithm. All three metrics are regularly used to assess the quality of different valuation algorithms. The choice of several metrics in total, allows a more differentiated statement to be made than would be the case with just one metric.

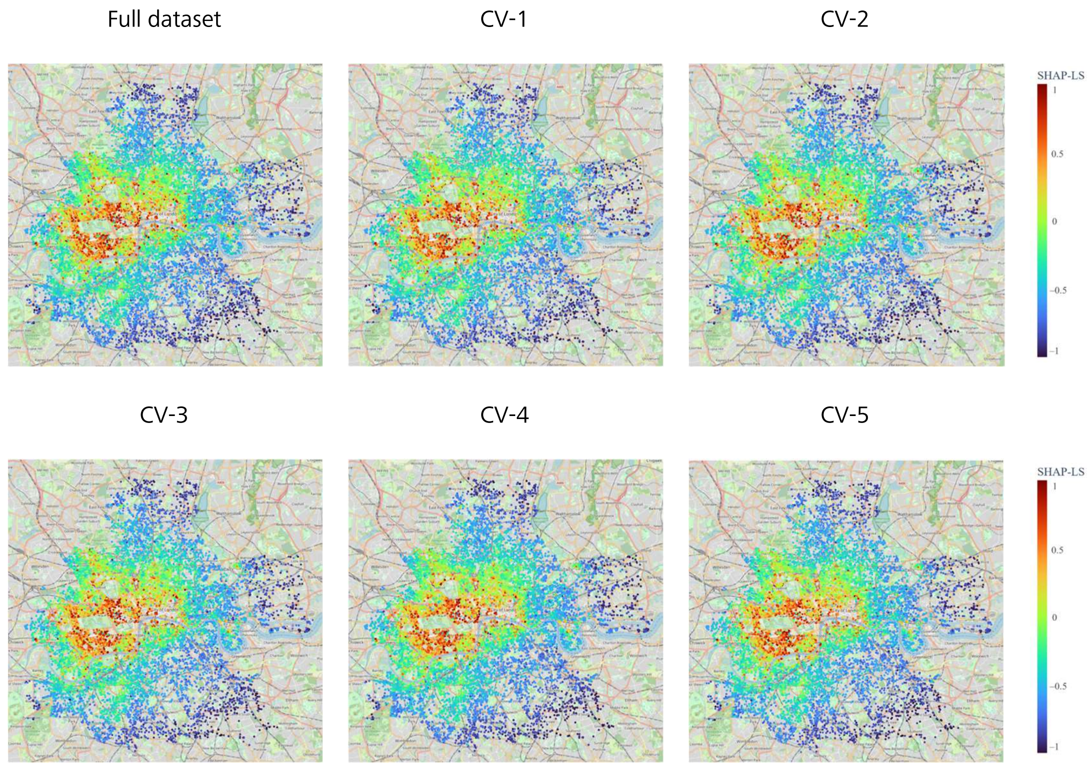
4.7.3 Appendix III – Robustness

To ensure the reliability of results obtained from the newly introduced SHAP-LS, two distinct robustness checks are performed.

Cross-validation testing

First, the results of the SHAP-LS are evaluated by implementing the score on five different datasets used in the five-fold cross-validation model testing in the chapter Empirical Application. Figure 4.5 shows these SHAP-LSs alongside those received when using the full datasets. The results exhibit minimal variation, thereby indicating a robust methodology.

Figure 4.5: Cross validation testing of the SHAP location scores



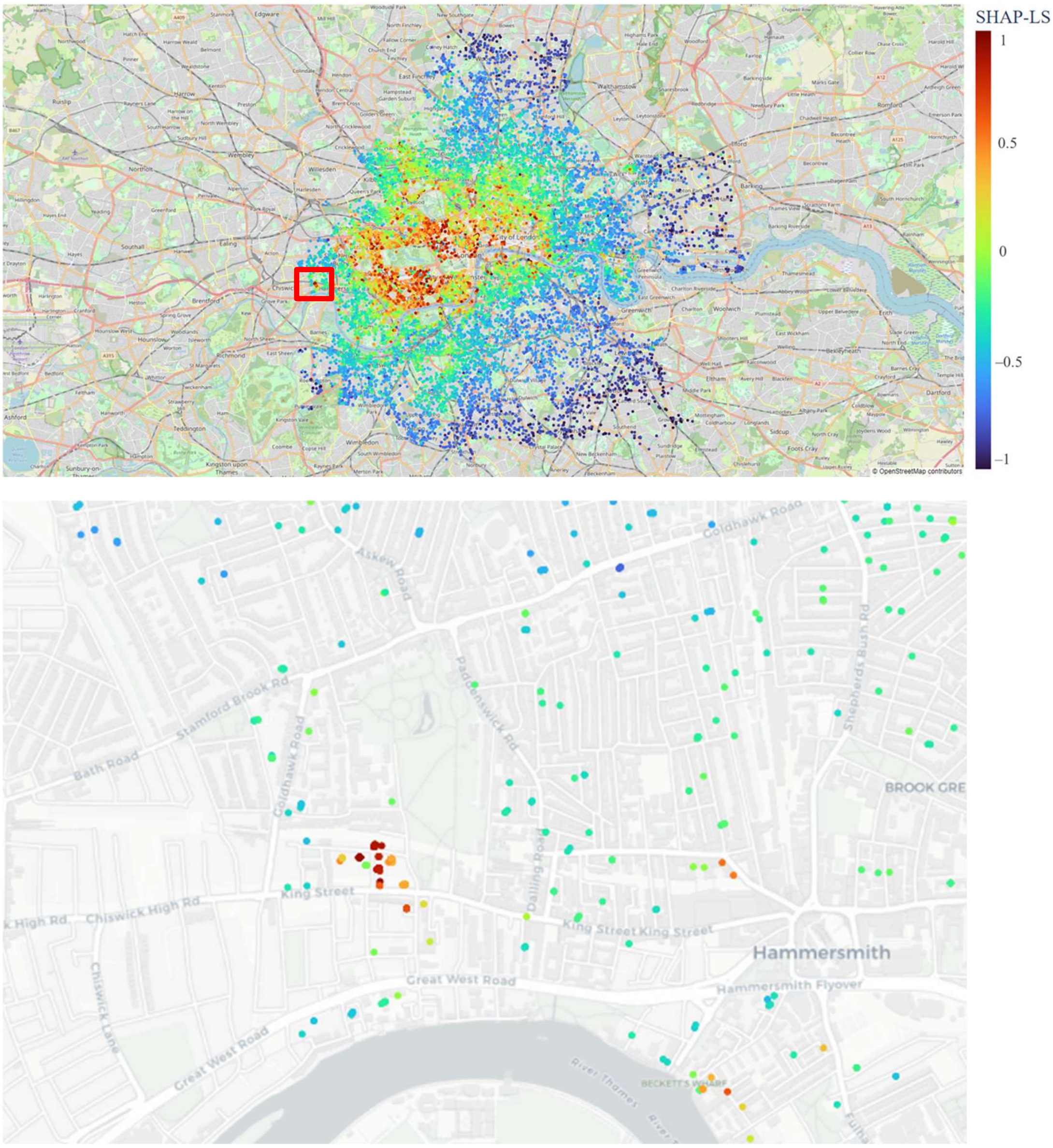
Notes: This figure shows the results of the SHAP-LS when being implemented on the full data set and five subsets. The results hardly vary and therefore indicate a robust methodology.

Spatial Autocorrelation

Second, Moran's I for the SHAP-LS is calculated. It measures the spatial autocorrelation based on feature positions and feature values simultaneously by evaluating whether the pattern formed by the SHAP-LS is clustered, dispersed, or random. Like the well-known Pearson correlation coefficient, Moran's I can take values between -1 and 1. A high Moran's I indicates that the willingness to pay for a location does not vary strongly over a short distance. The objective of the SHAP-LS is to determine the quality of a location, aiming for homogeneous evaluations for neighboring observations. As a result, it is expected that the calculated SHAP-LSs would exhibit a high degree of spatial autocorrelation. For the empirical example of the city of London, the calculated Moran's I is 0.88 with a p-value of 0.001. This demonstrates with high significance that the expected high spatial autocorrelation is indeed present. Overall, this indicates the robustness of the SHAP-LS methodology and further highlights its ability to classify individual property locations according to their quality. For more information about Moran’s I, Getis & Ord (1992) is recommended.

4.7.4 Appendix IV – Zoomed-in SHAP-LS analysis

Figure 4.6: Zoomed-in SHAP-LS analysis



Notes: The lower image offers an illustration of how the SHAP-LS can be employed to analyze location quality at the granularity of individual street segments. The corresponding position (marked in red in the upper part of the figure) was exemplarily selected from the map of the overall SHAP-LS already shown in Figure 4.2. This presentation allows for a detailed view of the location quality determined for individual streets. The results indicate that the quality of a location seems to vary, in part, at the level of individual streets and such differences in quality can be well identified using the SHAP-LS methodology.

4.7.5 Appendix V – Improving the accuracy of automated real estate valuations by applying SHAP-LSs

To investigate whether the precision of automated real estate valuations could be potentially enhanced by a more comprehensive incorporation of SHAP-LSs into the evaluation process, three distinct XGBoost models were developed, based on an extended version of the London dataset (see Table 4.1).[41] Each model utilizes the same structural and temporal characteristics as outlined in Table 4.1. The primary distinction lies in the use of a single locational variable: one model employs SHAP-LSs[42], another uses the ZIP codes, and the third incorporates the so-called NUTS-3 regions as a spatial separator.[43]

The models' efficacy is assessed via a moving window approach, commencing with a training set from 2020 and testing on the data from the first quarter of 2021. This assessment method is repeated quarterly until the fourth quarter of 2022, which served as the final testing set.[44] Finally, the accuracy of the three different models is calculated across all testing sets.

Table 4.3 provides an overview of the performance of these models. Across the four metrics - MAPE, MdAPE, PE(10), and PE(20) - the models using the SHAP-LS shows the lowest error rates, indicating superior accuracy. Conversely, it also exhibits higher PE(10) and PE(20) scores, suggesting that it delivers more reliable predictions within the given error margins compared to the other two models. These findings affirm the potential of the SHAP-LS in enhancing the precision and consistency of automated real estate valuation systems, a proposition that merits further academic exploration.

[41] Compared to the empirical example conducted in the upper part of the study, this illustrative demonstration utilized data not only from the year 2020 but also from the years 2021 and 2022. This allows for the implementation of the moving window approach and further enables a more robust testing of the results.

[42] To utilize the SHAP-LSs in the context of automated real estate valuation, they need to be computed beforehand. To enable a fair comparison, the calculation of SHAP-LSs for the test data is performed out-of-sample. In the first step, following the logic outlined in the methodology section, the SHAP-LSs are calculated for a training dataset. In the second step, the SHAP-LSs for the unseen test data are determined using a k-nearest neighbors approach (k=5) based on the previously computed SHAP-LSs of the training data. Therefore, our approach can be seen as a feature selection or feature aggregation method, as it systematically identifies and incorporates the most relevant locational variables into the model. By concentrating on key features, our methodology minimizes the risk of overfitting, leading to models that potentially generalize better to new data.

[43] The NUTS (Nomenclature of territorial units for statistics) classification is a hierarchical system for dividing up the economic territory of the EU and the UK. Overall, there are four different subdivision levels, called NUTS-0, NUTS-1, NUTS-2 and NUTS-3. The NUTS-3 regions in the UK constitute local administrative units including counties, unitary authorities and London boroughs. For a more detailed about the NUTS regions, we refer to Krämer et al. (2023b).

[44] The same logic is applied to calculate the SHAP-LSs.

Table 4.3: Overview model performance

Metrics	$Model_{SHAP-LS}$	$Model_{Postcode}$	$Model_{NUTS-3}$
MAPE	0.1549	0.1562	0.1568
MdAPE	0.1040	0.1219	0.1248
PE(10)	0.4847	0.4250	0.4142
PE(20)	0.7738	0.7084	0.7056

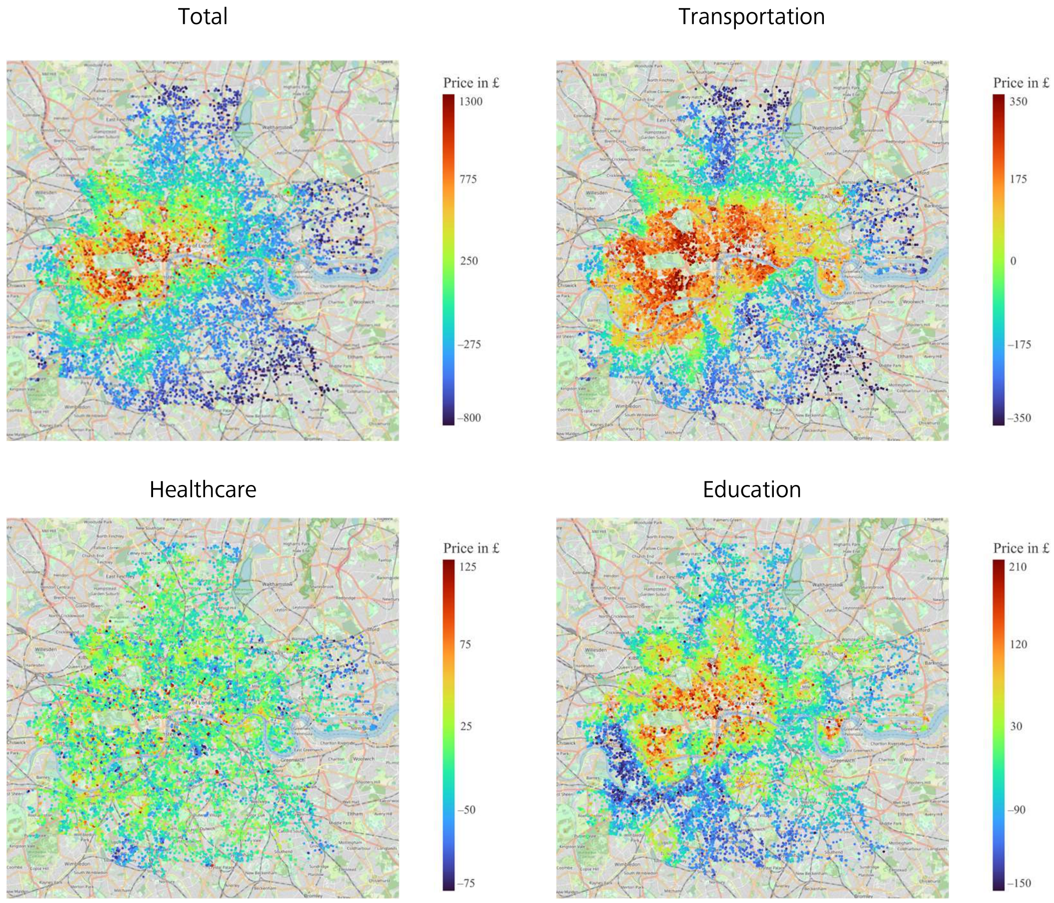
Notes: This table shows the results of the XGBoost models. The results indicate that the SHAP-LS can be used to improve the model performance of automated valuation models.

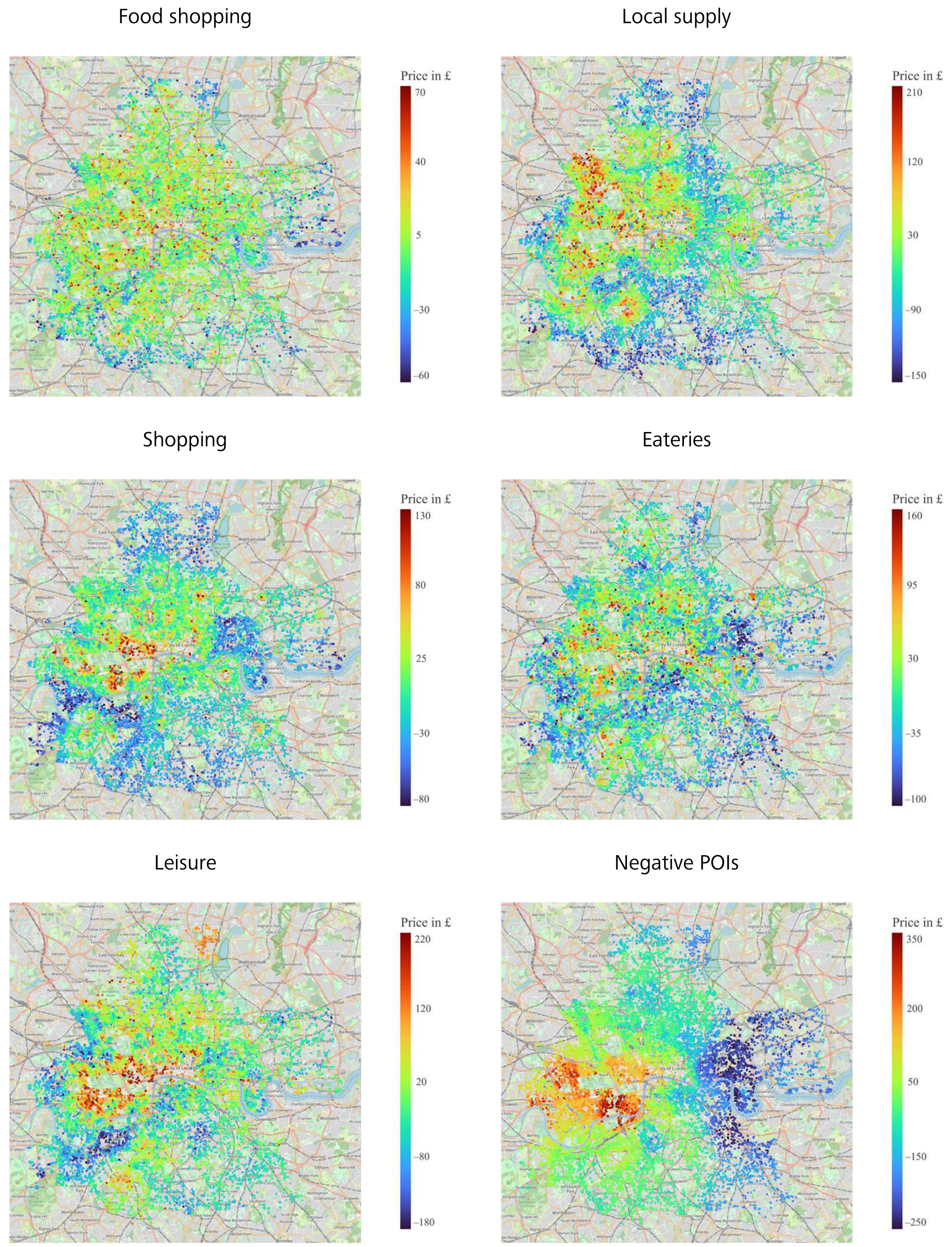
4.7.6 Appendix VI – Absolute SHAP-LS

The SHAP-LS technique can also be implemented without scaling the values between -1 and 1 to procure additional insights. By doing so, the absolute SHAP-LS is calculated and provides the absolute contribution of the locational features to the selected dependent variable (e.g., rental price, purchase price, market value) compared to the average prediction across all observations.

An illustration of this approach can be found in Figure 4.7, which displays the results of the absolute SHAP-LSs, utilizing the previously used London dataset. The initial image, located at the top left, presents the overall absolute SHAP-LSs. Evidence suggests that prime locations demand a rental premium, with costs escalating up to £1,300 above the average prediction. Conversely, locations deemed less desirable register a deficit of up to £900 compared to the average rental price prediction. The subsequent images display the categorical absolute SHAP-LSs. It becomes apparent that certain variables exert a greater influence than others. For example, transportation-related features have a higher impact than those related to food shopping. This can be observed by the fact that the former varies from -£250 to £350, while the latter fluctuates between -£60 and £70.

Figure 4.7: Absolute SHAP-LS





Notes: This figure maps the results of the overall and categorical absolute SHAP-LSs. The results show how the individual amenities and disamenities of the individual categories influence the asking rent. The maps indicate that the effect varies by category and location.

4.8 References

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications, 39*(2), 1772–1778.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(4), 1059–1086.

Bartholomew, K., & Ewing, R. (2011). Hedonic Price Effects of Pedestrian- and Transit-Oriented Development. *Journal of Planning Literature, 26*(1), 18–34.

Can, A. (1992). Specification and estimation of hedonic housing price models*. Regional Science and Urban Economics, 22*(3), 453–474.

Chen, T., & Guestrin, C. (2016). XGBoost.

Chun Lin, C., & Mohan, S. B. (2011). Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis, 4*(3), 224–243.

Court, A. T. (1939). *Hedonic price indexes with automotive examples*. In The Dynamics of Automobile Demand, General Motors Corporation, New York, 99–117.

Dumm, R. E., Sirmans, G. S., & Smersh, G. T. (2016). Price Variation in Waterfront Properties Over the Economic Cycle. *Journal of Real Estate Research, 38*(1), 1–26.

Freybote, J., Sun, H., & Yang, X. (2015). The Impact of LEED Neighborhood Certification on Condo Prices. *Real Estate Economics*, *43*(3), 586–608.

Gabe, J., Robinson, S., & Sanderford, A. (2021). Willingness to Pay for Attributes of Location Efficiency*. The Journal of Real Estate Finance and Economics*, 1–35.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical analysis, 24*(3), 189-206.

Glumac, B., & Des Rosiers, F. (2021). Practice briefing – Automated valuation models (AVMs): Their role, their advantages and their limitations. *Journal of Property Investment & Finance*, *39*(5), 481–491.

Hastie, T., Friedman, J. & Tibshirani, R. (2001). *The Elements of Statistical Learning.* New York: Springer.

Ho, W. K., Tang, B.-S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research, 38*(1), 48–70.

Hoen, B., & Atkinson-Palombo, C. (2016). Wind Turbines, Amenities and Disamenitites: Astudy of Home Value Impacts in Densely Populated Massachusetts. *Journal of Real Estate Research*, *38*(4), 473–504.

Jauregui, A., Allen, M. T., & Weeks, H. S. (2019). A Spatial Analysis of the Impact of Float Distance on the Values of Canal-Front Houses. *Journal of Real Estate Research*, *41*(2), 285–318.

Kestens, Y., Thériault, M., & Des Rosiers, F. (2006). Heterogeneity in hedonic modelling of house prices: Looking at buyers' household profiles. *Journal of Geographical Systems*, *8*(1), 61–96.

Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big Data in Real Estate? From Manual Appraisal to Automated Valuation. *The Journal of Portfolio Management, 43*(6), 202–211.

Krämer, B., Stang, M., Nagl, C., & Schäfers, W. (2023a). Explainable AI in a Real Estate Context - Exploring the Determinants of Residential Real Estate Values. *Journal of Housing Research*, Advanced online publication.

Krämer B., Stang M., Doskoč V., Schäfers W. & Friedrich F. (2023b). Automated valuation models: improving model performance by choosing the optimal spatial training level. *Journal of Property Research.*

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy, 74*(2), 132–157.

Lorenz, F., Willwersch, J., Cajias, M., & Fuerst, F. (2022). Interpretable machine learning for real estate market analysis. Journal of *Real Estate Economics*, Forthcoming.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Proc. *Advances in Neural Information Processing Systems*, *30*, 4768–4777.

Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*, *12*(1), 134–150.

McCluskey, W., Davis, P., Haran, M., McCord, M., & McIlhatton, D. (2012). The potential of artificial neural networks in mass appraisal: The case revisited. *Journal of Financial Management of Property and Construction, 17*(3), 274–292.

Metzner, S., & Kindt, A. (2018). Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties. *International Journal of Housing Markets and Analysis, 11*(1), 73–100.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

Pace, R. K., & Hayunga, D. (2020). Examining the Information Content of Residuals from Hedonic and Spatial Models Using Trees and Forests. *The Journal of Real Estate Finance and Economics, 60*(1-2), 170–180.

Parker, D. (2016). *International Valuation Standards: A Guide to the Valuation of Real Property Assets*. Chichester: Wiley Blackwell.

Randeniya, T. D., Ranasinghe, G., & Amarawickrama, S. (2017). A model to Estimate the Implicit Values of Housing Attributes by Applying the Hedonic Pricing Method. *International Journal of Built Environment and Sustainability, 4*(2), 113–120.

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

Sangani, D., Erickson, K. & Hasan, M.A. (2017). *Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting*. In IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Orlando.

Seo, K., Golub, A., & Kuby, M. (2014). Combined impacts of highways and light rail transit on residential property values: a spatial hedonic price model for Phoenix, Arizona. *Journal of Transport Geography, 41*, 53–62.

Seo, W. (2018). Does Neighborhood Condition Create a Discount Effect on House List Prices? Evidence from Physical Disorder. *Journal of Real Estate Research, 40*(1), 69–88.

Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature, 13*(1), 1–44.

Stamou, M., Mimis, A., & Rovolis, A. (2017). House price determinants in Athens: A spatial econometric approach. *Journal of Property Research, 34*(4), 269–284.

Stang, M., Krämer, B., Nagl, C., & Schäfers, W. (2022). From human business to machine learning—Methods for automating real estate appraisals and their practical implications. *Zeitschrift Für Immobilienökonomie*, Advanced online publication.

Turner, T. M., & Seo, Y. (2021). House Prices, Open Space, and Household Characteristics. *Journal of Real Estate Research, 43*(2), 204–225.

Yilmazer, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy, 99*, 104889.

Yoo, S., Im., J., & Wagner, J. E. (2012). Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, NY. *Landscape and Urban Planning, 107*(3), 293–306.

5 Conclusion

5.1 Executive Summary

This section provides a comprehensive summary of all three papers included in this thesis. It discusses the problems and objectives, the data and methodologies used, as well as the results and their contribution for both science and practice.

Paper 1: Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values

Problems and Objective

Understanding real estate markets and its drivers is arguably one of the most important areas of real estate research. Compared to other asset classes, real estate is a heterogeneous asset and properties differ from one another in terms of their features. There is a large body of literature dealing with the factors which have a significant influence on the value or price of a property, subsumed under the term Hedonic Price Models. They are usually based on parametric and semi-parametric methods like the Ordinary Least Squares approach (see, e.g., Malpezzi, 2003; Sirmans et al., 2005; Schulz et al., 2014) or the Generalized Additive Models (see, e.g., Bourassa et al., 2007; Bourassa et al., 2010; Brunauer et al., 2010). In recent years, more advanced statistical and modern machine learning methods have attracted interest in the real estate community, as they are often less restrictive in terms of their model structure and thus are more flexible. Especially, bagging techniques like random forest (RF) and boosting algorithms like the eXtreme Gradient Boosting (XGBoost) algorithm, seem better suited to real estate valuation (see, e.g., Cajias et al. 2019; Stang et. al, 2022). However, machine learning applications are usually criticized for their lack of transparency and are therefore often referred to as black boxes (see, e.g., Din et al., 2001; McCluskey et al., 2013). While parametric and semi-parametric applications are comprehensible to humans, the calculations of modern machine learning applications can only be understood with difficulty if at all. To overcome this problem, so called eXplainable Artificial Intelligence (XAI) approaches have been developed. These approaches use model-agnostic frameworks to reveal the modes of operations of machine learning algorithms and thus help to make their mode of action more transparent.

In real estate, so far, XAI approaches have been explored only to a limited extent, but we believe they offer several benefits. First, they shed light on the mechanism underlying machine learning algorithms, thus overcoming their image of black boxes, and therefore increasing their acceptance in different regulated and unregulated areas within the real estate industry, for example in the mortgage lending industry. Second, XAI methods are able to support research in understanding the key drivers of real estate markets. Accordingly, this paper not only focuses on identifying reliable and unbiased relations between features and residential property prices, but also discusses their economic implications.

Methodology and Data

This study uses a unique dataset of 81,166 residential properties from the largest cities in Germany. The dataset can be divided into 61,763 condominiums and 19,403 single-family homes. It encompasses several structural, locational, temporal, and socio-economic features and was collected between 2014 and 2020 by professional appraisers from a large German banking group. This dataset is analyzed using two global post-hoc, model-agnostic XAI methods: Permutation Feature Importance (PFI), implemented by Breiman (2001), and Accumulated Local Effects (ALE) plots, introduced by Apley & Zhu (2020). These methods are applied to an eXtreme Gradient Boosting tree (XGBoost). First, the PFI measures the relevance of a feature to the predictions. Subsequently, the ALE plots are utilized to provide further insights into how the most important features affect the predictions.

Results and their Contribution to Science and Practice

The study reveals that the same value-determining features play an important role for both condominiums and single-family homes. However, there are fundamental differences within the two property types with regard to the shape of the individual ALE plots and thus the influence of the respective feature on the market value of a property. Furthermore, we identify non-linear relationships for the majority of features. Generalized rules of thumb such as "the larger the living area, the lower the market value per square meter" are refuted by our findings for condominiums, but can be confirmed for single-family homes.

In summary, our results show how important it is for both real estate research and practice to conduct data-driven analyses with the help of modern machine learning and XAI approaches, in order to gain important market insights and, if necessary, to update long-established assumptions regarding the determinants of real estate market values. This is especially important for mortgage underwriters, valuation firms and regulatory authorities and, thus, of considerable interest to most of the real estate community.

Paper 2: Automated Valuation Models: Improving Model Performance by Choosing the Optimal Spatial Training Level

Problems and Objective

The academic community has discussed using Automated Valuation Models (AVMs) in the context of traditional real estate valuations and their performance for several decades, and practitioners are also now increasingly scrutinizing it. Most studies focus on the comparison of different statistical methods. Accordingly, a large body of literature compares traditional hedonic models with more modern machine learning approaches or approaches from spatial econometrics (see, e.g., Pace & Hayunga, 2020). These studies aim to identify which method is best suited for estimating real estate values or prices.

Besides the method selection, AVMs can be optimized in many other areas. For example, the selection, cleaning and preparation of data play an important role for the overall performance of the AVM. Another aspect is the choice of spatial level to train the selected methods. This is decisive for determining which data are ultimately included in the estimation of the AVM and, thus, what information is used or ignored. Thanks to georeferencing, models can, in principle, be trained at any level. For example, a model can be trained at the city level, the associated commuter belt, or even nationwide. However, this aspect has received little to no attention from the academic community until now.

The published research on AVMs usually deals only with a manually defined region and fails to test the methods used on different spatial levels. Therefore, it raises the question of whether considering different spatial levels for training AVMs could be an important and undervalued factor in enhancing their valuation accuracy.

Methodology and Data

The methods selected for this study are commonly used in academic research related to Automated Valuation Models (AVMs). Alongside Ordinary Least Squares (OLS) regression, this analysis incorporates Generalized Additive Models (GAM), eXtreme Gradient Boosting (XGBoost) algorithms, and Deep Neural Networks (DNN). The data, gathered by professional real estate appraisers from a large German banking group, encompasses 1,212,546 residential properties across Germany, containing a range of structural, temporal, locational, and socio-economic features. These four methodologies are evaluated across different spatial layers, which are defined according to the Nomenclature of Territorial Units for Statistics (NUTS) of the European Union.

Results and their Contribution to Science and Practice

The results show that the right choice of spatial training level can significantly influence the model performance of different AVM algorithms, and that this influence varies considerably, depending on the type of method. The results indicate that for parametric and semi-parametric approaches, choosing a relatively small training level is advisable. This shows that the trained OLS and GAM cannot draw additional explanatory power from observations outside a particular region. The results for the two modern machine learning algorithms are quite different. We observe that they can gain more explanatory power by adding further observations, and that this effect outweighs local heterogeneity. Therefore, we recommend, choosing a generally higher training level for modern machine learning algorithms.

The contributions of our paper are manifold. First and foremost, our findings provide further evidence that when it comes to testing more traditional versus modern machine learning methods, fundamental differences in their application should be considered to achieve the best model performance. Our findings indicate that assumptions valid for applying traditional machine learning methods may not be suitable for modern methods.

This provides real estate researchers and practitioners with new guidelines for using different AVM algorithms, which can help improve the performance of their valuation results. Additionally, our findings also shed light on the question of whether real estate markets are characterized by high local heterogeneity. The results of our OLS and GAM models study suggest significant heterogeneity in local real estate markets. Still, the results of the XGBoost and DNN indicate that there are overall patterns that apply to all real estate markets. In summary, our paper provides a new set of guidelines that can be used to answer various real estate-related questions more accurately. These new guidelines are a starting point for further research into the analysis of real estate markets using modern machine learning algorithms.

Paper 3: Changing the Location Game – Improving Location Analytics with the Help of Explainable AI

Problems and Objective

Besides its structural and economic characteristics, the location of a property is probably one of the most important determinants of its underlying value. While there have been standardized and globally accepted methods for assessing property values (see e.g., Parker (2016)), the assessment of the location quality at a given site remains a kind of arbitrary estimation based on more or less subjective individual opinions. Interestingly, these opinions are mostly formed on the basis of factors that can in principle be objectified. One example of this is the accessibility of certain amenities, such as the nearest supermarket, the nearest public transport stops and the nearest park. Therefore, the question arises if this process can be standardized and automated by means of computer-based models. Machine learning based models have been used in real estate valuation, for example, for several decades.

As of yet, there is no approach to date that leverages the capabilities of modern machine learning algorithms to capture the quality of a real estate location in an automated manner. The objective of this paper is therefore to present a new purely data-driven approach methodology based on a state-of-the-art machine learning model and a post-hoc model agnostic explainable artificial intelligence (XAI) approach to introduce to evaluate real estate locations and its drivers.

Methodology and Data

Based on the assumption that the quality of a property's location is reflected in the individual willingness to pay for the property, and that the quality can be measured by means of objectifiable factors, the newly introduced approach named "SHAP location score" (SHAP-LS) enables the rating of individual property locations and its drivers. This approach combines a non-parametric machine learning algorithm and the local post-hoc model-agnostic explainable artificial intelligence (XAI) method of Shapley Additive Explanations and is characterized by its high degree of flexibility and can be implemented in a model-agnostic manner for any machine learning algorithm and for any feature set.

By using a dataset of 26,860 residential rental condominiums for the city of London from 2020, and a comprehensive set of location-specific points-of-interests (POIs), it is shown how an empirical implementation of the SHAP-LS is possible and what results can be obtained.

Results and their Contribution to Science and Practice

The results of our empirical example demonstrate that the SHAP-LS method can identify the key factors influencing residential real estate location quality at a detailed (street) level, revealing areas of high and low location quality within the city. The findings indicate significant variations in individuals' willingness to pay for different levels of location quality, with the highest willingness observed in central areas. By categorizing the different POIs, the impact of different points of interest on location quality can be examined. The analysis reveals that proximity to public transport stops, educational institutions, shopping facilities, and negative POIs strongly influence the willingness to pay for location quality, while other categories like food shopping POIs have a lesser impact.

The SHAP-LS approach offers an initial automated valuation of real estate locations suitable for various real estate stakeholders. In academia, this approach extends traditional urban models by incorporating modern non-parametric valuation algorithms, allowing for a comprehensive examination of economic and econometric phenomena. For practitioners, the approach provides a quick empirical assessment of location quality within a city, enhancing market transparency and facilitating more informed decision-making. Accordingly, the SHAP-LS can be used for various purposes in real estate practice, such as the evaluation of different locations within a city for investment decisions.

5.2 Final Remarks

As we move further into the 21st century, the real estate valuation process is facing a significant turning point. Traditionally, this sector has been labor-intensive, requiring professional appraisers to conduct extensive and often time-consuming analyses to determine property values. This approach is costly and problematic given the declining number of professional appraisers in the market (Coyle, 2015; Lausberg & Dust, 2017; Yeh & Hsu, 2018).

Given the inherently standardized nature of real estate and the increasing availability of data, the residential valuation process, in particular, can benefit significantly from the incorporation of data-driven methods (Royal Institution of Chartered Surveyors, 2017). The first steps toward automating these processes are already visible and promise to reduce the resource and time constraints associated with property valuation (see e.g., Stang et al., 2022). However, the shift to a more automated process is not without challenges. Current automated systems need further refinements to increase their acceptance among appraisers and regulatory authorities alike (Royal Institution of Chartered Surveyors, 2017; European Banking Authority, 2020). This thesis, which consists of three papers, contributes to these efforts by offering insights into the potential of machine learning and explainable artificial intelligence to enhance property valuation. The first paper takes an in-depth look at the key determinants of property values and opens the black box of machine learning. It provides a comprehensive analysis that goes beyond traditional linear models to captures the influences on property values. The second paper investigates how the accuracy of valuation models can be improved by choosing the appropriate spatial training level. The third paper demonstrates the application of explainable artificial intelligence in assessing location quality, showing how this advanced technology can provide granular, accurate, and transparent results.

From a broader perspective, this thesis pioneers the integration of machine learning and explainable artificial intelligence into residential property valuations. It successfully uncovers patterns and derives valuable knowledge from comprehensive and multi-layered datasets, highlighting the potential of these methods and their applications in enhancing the residential property valuation process. This work represents a significant interdisciplinary collaboration linking the fields of real estate, economics, computer science, and statistics. This critical connection allows to realize the full potential of new technologies in a traditionally rather manual sector. By integrating those technologies into the necessary economic and analytical concepts, the outcomes become more valuable and purposeful. This dissertation, thus, bridges the gap between theoretical understanding and practical application. It not only provides guidance but also sparks critical discourse and encourages further research in this field. This represents an important step towards a future in which the property valuation process is not merely more efficient and accurate, but also fairer and transparent. The aim is to utilize the full potential of machine learning and explainable artificial intelligence in property valuation process, setting up the stage for the next phase of innovation and transformation in the sector.

Yet, the work in this thesis is only the start. There are several crucial areas that need further research when it comes to implementing machine learning and explainable artificial intelligence in the property valuation process. One important issue is data availability, especially in rural regions. The lack of comprehensive datasets compromises the effectiveness of data-driven valuation methods. At the same time, the burgeoning field of artificial intelligence brings forth opportunities for leveraging advanced machine learning techniques to refine automated valuation models. For example, the advent of large language models such as GPT offers a promising avenue for property valuation, particularly for managing and interpreting large-scale textual data. Moreover, as we move towards a more automated property valuation process, it is also crucial to establish a strong legal framework. Such a system should advocate for fairness, transparency, and accountability, accommodating both regulatory and ethical considerations. An important issue in this context is the inclusion of sustainability considerations in property valuation. This thesis focuses on residential property valuation, but initial steps to extend these data-driven methodologies to other asset classes, such as commercial real estate, have already been taken (see, e.g., Deppner et al., 2023) and have to be further explored.

While this thesis focuses on machine learning and explainable artificial intelligence, the future requires the integration of other advancing technologies. Building Information Modeling (BIM), Blockchain, and others hold the potential to significantly disrupt and improve the property valuation process (Royal Institution of Chartered Surveyors, 2017). However, integrating these technologies will require substantial time and research, further expanding the scope of this exploration.

In conclusion, it can be said that the current period is an exciting phase of transition and innovation in the field of property valuation. This thesis is a contribution to the evolving discourse and provides a foundation upon which further research can be built. With the prospect of countless advancements, this field has a promising future. As the technology and methodologies continue to evolve, so will our understanding of property valuation and the potential it holds.

5.3 References

Apley, D.W. & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology*), *82*(4), 1059–1086.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial Dependence, Housing Submarkets, and House Price Prediction. *The Journal of Real Estate Finance and Economics*, *35*(2), 143–160.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2010). Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions. *Journal of Real Estate Research*, *32*(2), 139–159.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Brunauer, W. A., Lang, S., Wechselberger, P., & Bienert, S. (2010). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *The Journal of Real Estate Finance and Economics*, *41*(4), 390–411.

Cajias, M., Willwersch, J. & Lorenz, F. (2019). *I know where you will invest in the next year – Forecasting real estate investments with machine learning methods* [Conference Paper]. European Real Estate Society Meeting, Cergy-Pontoise Cedex, France.

Coyle, M.L. (2015). *Future of Valuation*. Working RE magazine, Editor's note, San Diego.

Deppner, J., von Ahlefeldt-Dehn, B., Beracha, E. & Schaefers, W. (2023). Boosting the Accuracy of Commercial Real Estate Appraisals: An Interpretable Machine Learning Approach. *The Journal of Real Estate Finance and Economics*, Advanced online publication.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental Variables and Real Estate Prices. *Urban Studies*, *38*(11), 1989–2000.

European Banking Authority (2020). *Final Report: Guidelines on loan origination and monitoring.* European Banking Authority, Paris.

Lausberg, C., & Dust, A. (2017). Reducing the appraisal bias in manual valuations with decision support systems. *Advances in Automated Valuation Modeling: AVM After the Non-Agency Mortgage Crisis*, 331-343.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. *Housing Economics and Public Policy*, *1*, 67–89.

McCluskey, W. J., McCord, M., Davis, P. T., Haran, M., & McIlhatton, D. (2013). Prediction accuracy in mass appraisal: A comparison of modern approaches. *Journal of Property Research*, *30*(4), 239–265.

Pace, R. K., & Hayunga, D. (2020). Examining the information content of residuals from hedonic and spatial models using trees and forests. *The Journal of Real Estate Finance and Economics*, *60*(1-2), 170–180.

Parker, D. (2016). *International Valuation Standards: A Guide to the Valuation of Real Property Assets*. Chichester: Wiley Blackwell.

Royal Institution of Chartered Surveyors (2017). *The Future of Valuations: The relevance of real estate valuations for institutional investors and banks – views from a European expert group*. RICS, London.

Schulz, R., Wersing, M., & Werwatz, A. (2014). Automated valuation modelling: a specification exercise. *Journal of Property Research*, *31*(2), 131–153.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, *13*(1), 1–44.

Stang, M., Krämer, B., Nagl, C., & Schäfers, W. (2022). From human business to machine learning—Methods for automating real estate appraisals and their practical implications. *Zeitschrift Für Immobilienökonomie*, Advanced online publication.

Yeh, I. C., & Hsu, T. K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing, 65*, 260-271.