Bocconi

unibocconi.it

WORKING PAPER
N. 238
FEBRUARY 2025

NON-RESIDENTIAL REAL ESTATE PRICES AND MACHINE LEARNING: THE HOW AND THE WHY

Raffaella Barone

Università Bocconi
BAFFI
Centre on Economics, Finance and Regulation

Via Röntgen 1 | 20136 Milano – Italia | Tel +39 02 5836.5908/5564
baffi@unibocconi.it | www.baffi.unibocconi.eu

NON-RESIDENTIAL REAL ESTATE PRICES AND MACHINE LEARNING: THE HOW AND THE WHY

Raffaella Barone[♣]

Abstract

This paper examines the relationship between non-residential property prices and various social, economic, and environmental indicators within the provinces where these properties are located. We focus on indicators from the Eni Enrico Mattei Foundation and SDSN Italia that track the 17 sustainable development goals, as well as additional factors like crime rates, per capita GDP, and sales frequency.

Using a machine learning algorithm, we predicted property sale prices and applied SHapley Additive exPlanations to assess the importance of each variable. Our findings highlight the strong influence of categorical variables and SDG indicators on prices. Finally, we used causal inference to explore how policy interventions might affect property prices.

Keywords: Machine Learning, Real estate market, Financial Stability, Sustainability, Crimes

JEL: B4: Economic Methodology, C1: Econometric and Statistical Methods and Methodology: General, G01: Financial crises, R33: Non-agricultural and Non-residential Real Estate Markets, K: Law and economics

---

[♣] Raffaella Barone, Department of Law, University of Salento; Baffi Centre, Bocconi University. E-mail: raffaella.barone@unisalento.it.

Acknowledgements: The author wishes to thank participants in: 1) The 5th International Conference on Corporate Governance and Risk Management – GRM 11-12 July 2024, which was organized by The University of Bari Aldo Moro and University of Foggia. 2) 1ST Modern Finance Conference, September 15-17, 2024, Kozminski University, Warsaw, Poland.

A special thanks goes to Professor Lorenz Kueng for his kind hospitality at USI Lugano, where I found the serenity and the right concentration to complete this work. I couldn't have done it without him.

Funding: The work was supported by the Apulia Region and ARTI REFIN project cod. 00CA919A through the action 10.4 POR Puglia FESR FSE 2014–2020. The final responsibility is up to the author, without any involvement of the founding source.

1. Introduction

The residential and non-residential real estate sector represents a significant portion of global GDP. Between 2007 and 2020, it accounted for a median value of 15.2% of GDP in the US, 17.5% in the UK, 14% in the EU, and 16.65% in Italy. Recent studies on residential real estate prices have generated substantial literature (Batalha et al. 2022; Beland et al. 2020; Gupta et al. 2022; Liu and Su 2021; Zhao 2020; Agarwal et al. 2020; Capponi and Rios 2020; D'Lima et al. 2021; Dave et al. 2020; Duca et al. 2021; Granja et al. 2020; Liu and Tang 2021). Some research has also explored the use of agent-based models to examine house price determinants and the impact of money laundering on the residential market (Barone, 2023; Cincotti, 2022; Carro et al., 2022; Ozel et al., 2019; Baptista et al., 2016; Geanakoplos et al., 2012).

However, studies on the determinants of non-residential property prices are limited, with a focus on sustainability, particularly green building construction (Mangialardo et al., 2019; Bowman and Wills, 2008; Fuerst and Mcallister, 2008a, b). A few studies have addressed the impact of environmental amenities on non-residential properties, such as office, commercial, industrial, and retail sales (Franco and Cutter, 2022). Good environmental performance is known to increase property value (Porter and van der Linde, 1995; Klassen and McLaughlin, 1996), and changes in environmental factors influence real estate prices and transactions.

This paper, using a supervised machine learning approach and in line with AGENDA 2030 and AGENDA 21, analyses the impact of the 17 sustainable development goals (SDGs) on non-residential property prices in Italy. We also incorporate additional variables, such as crime rate, GDP, and sales frequency (the ratio of sales volume to real estate stock) for each province. After developing and assessing the predictive model, we evaluated feature importance using the SHapley Additive exPlanations approach. Finally, using the DoWhy and EconML libraries for causal inference, we explored causal relationships, hypothesizing that only a limited number of features have a direct causal effect on property prices.

2. The Methodology

The implementation of the 2030 Agenda for sustainable development requires determining the efficient level of environmental exploitation. This assessment involves a cost-benefit analysis since environmental protection comes at a cost. Understanding the demand for environmental goods is essential, i.e., how much individuals are willing to pay for such protection. However, environmental goods are public goods, and the market does not provide information on the quantity of these goods at different prices. To estimate demand, scholars use direct and indirect evaluation methods, such as stated versus revealed preferences.

One common indirect method is the hedonic price model (Griliches, 1961; Lancaster, 1966; Rosen, 1974; Lucas, 1975), which relies on linear regression techniques. The model has been applied primarily to the residential real estate market (see Hill, 2013 for a review), while the non-residential sector remains underexplored. In this study, we apply a machine learning algorithm, gradient boosting trees, which efficiently captures non-linear dependencies (for an overview of machine learning, see De Liso, 2023).

Potrawa and Tetereva (2022) compared machine learning approaches with traditional hedonic regression models for the residential sector, arguing for the superiority of machine learning. However, machine learning models are often criticized as "black boxes" due to their lack of interpretability. Recently, this issue has been addressed through eXplainable Artificial Intelligence (XAI) methods. For example, Chen et al. (2020) used SHAP values (Lundberg & Lee, 2017) to analyze the residential sector in Shanghai. Despite its usefulness, SHAP has limitations, such as its inability to reveal correlations between variables or infer causal relationships (De Villa, 2023).

To overcome these limitations, we perform causal inference using the DoWhy and EconML libraries from the PyWhy project (Sharma & Kicirman, 2020). Our analysis combines two methods: 1) Directed Acyclic Graphs (DAGs), which visually represent causal assumptions (Pearl, 1995; Pearl et al., 2016), and 2) the Potential Outcomes framework (Splawa-Neyman, Dabrowska, & Speed, 1990; Rubin, 2005). The resulting causal estimate, known as the average treatment effect (ATE), helps clarify the impact of environmental factors on property prices.

3. Data collection and processing

The first step in the analysis involved collecting, processing, and analysing the data.

- ***Non-residential Real Estate Prices***

We gathered price data for various non-residential properties, including offices, shops, warehouses, laboratories, shopping centres, industrial warehouses, pensions, and similar types. The data, provided by the Real Estate Market Observatory (OMI) of the Revenue Agency, covered the years 2016-2020. Prices were categorized by provincial capital, urban area (B=centre, C=semi-centre, D=outskirts, E=suburban, R=extra-urban/agricultural), year, and conservation status (excellent, normal, poor). For each property type, province, conservation status, and micro zone, sales prices included a minimum and maximum, from which we calculated the average value, termed the "average purchase price."

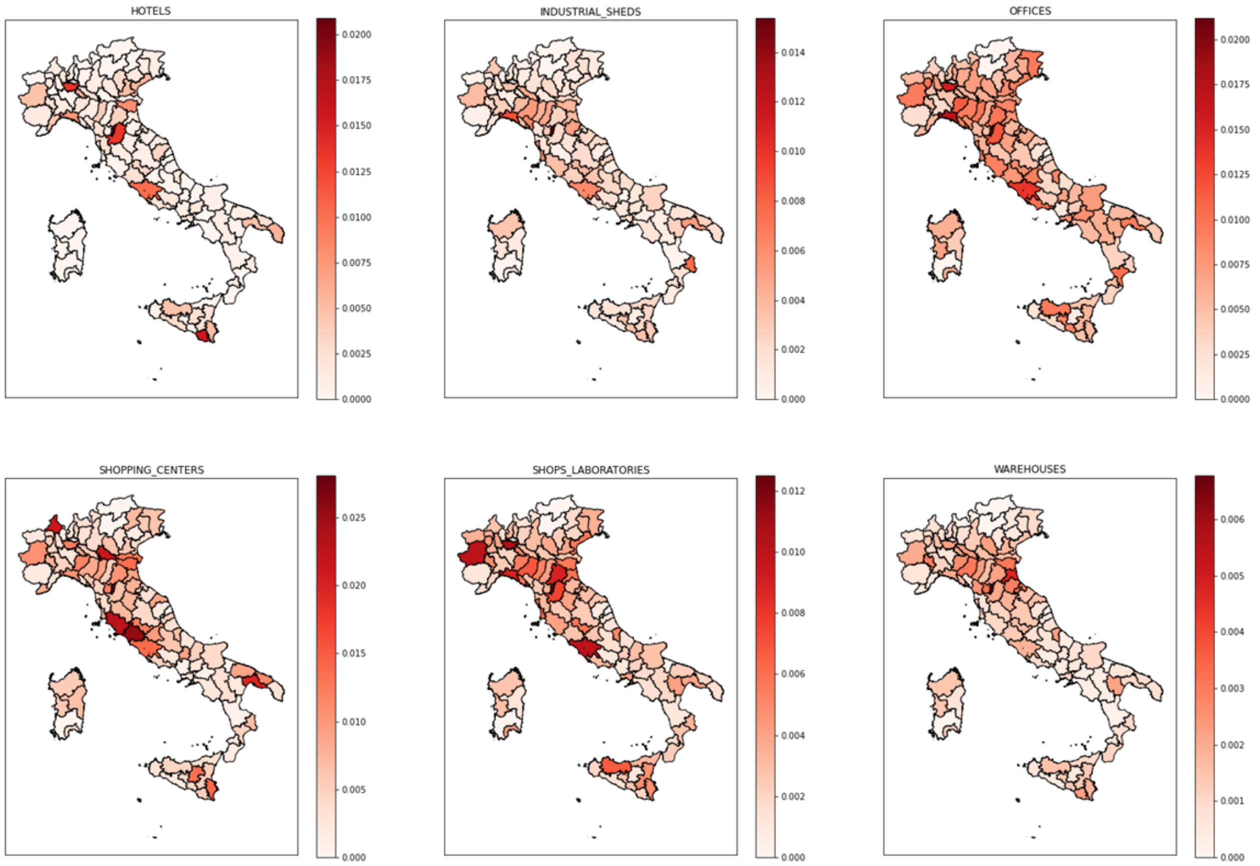
For instance, in the case of shops in Alessandria, Piedmont (normal condition), we found that in the semi-centre zone (C), there were three micro zones (C1, C2, C3). The minimum price for shops in zone C ranged from 1000 to 1300 €/sqm, while the maximum ranged from 1300 to 1950 €/sqm. Therefore, the minimum price for the zone was taken as 1000-1300 €/sqm, and the maximum as 1300-1950 €/sqm. We then computed the average purchase price for each zone (B, C, D, E, R).

- ***Real Estate Stock and Volume of Sales***

Next, we collected data on real estate stock from the Revenue Agency, available on their website: https://www.agenziaentrate.gov.it/portale/web/guest/schede/fabbricatiterreni/omi/pubblicazioni/statistiche-catastali. Additionally, we obtained sales volume data for each non-residential property in every provincial capital. From this, we calculated the sales-to-stock ratio for each province to determine the frequency of sales. The distribution of sales frequencies was then plotted using the geopandas library in Python, as shown in Figure 1.

Our analysis revealed that the most active cities were Prato (across all non-residential property types), Genoa (offices, shops, laboratories, industrial sheds), Milan (offices, shops, laboratories, hotels), Florence (offices, hotels), and Rome (offices, shops, laboratories, industrial sheds, hotels).

Figure 1: The average frequencies for each non-residential property



Source: Own elaboration based on data analysis

- ***Crime rate and GDP per-capita***

Additionally, we considered the impact of crime on property values, particularly for non-residential properties, which are often targeted by organized crime for money laundering. We obtained crime rate data for the provincial capitals, along with GDP per capita figures for the period 2016-2020 from the Istat website. The crime data included both total crime rates and data excluding financial crimes like money laundering and usury, which were analyzed separately. GDP per capita was categorized into four income bands: 0-15,000, 15,001-28,000, 28,001-50,000, and above 50,000.

- ***17 sustainable development goals***

In evaluating the environmental context of properties, we accounted for factors influencing territory desirability, buildability, and construction safety, including geology, seismicity, climate, environmental degradation, flood risks, and solar radiation. Economic and social conditions were also considered. To encompass these factors, we collected data from the Eni Enrico Mattei Foundation and the Sustainable Development Solutions Network (SDSN Italia), which monitors the 17 SDGs across Italian cities. These data were available on the website site https://sdg-portal.it/it. Descriptive statistics were generated for each indicator, and the results were visually represented with histograms (Figures 2, 3, and 4).

We grouped the SDGs into three pillars: environmental, economic, and social. The categorization did not impact the results of the analysis.

1) **The environmental pillar:**

Goal 2: Zero hunger

This goal focuses on ensuring food security, improving nutrition, and promoting sustainable agriculture. Specifically, the indicators considered include:

- *Urban gardens*: how many square meters of decommissioned or abandoned surface area per 100 inhabitants have been converted to the organic cultivation of fruit and vegetables.
- *Obesity and severe obesity*: the percentage of individuals aged 18+ who are obese or overweight.

Goal 6: Clean water and sanitation

This goal aims to ensure access to water and sanitation for all:

- *Network losses*: the percentage of water losses on the total volume released into the network.
- *Resident population connected to urban wastewater purification plants* (%).
- *Resident population served by the urban wastewater sewerage network* (%).

Goal 7: Affordable and clean energy

This goal seeks to ensure access to reliable, sustainable, and modern energy for all:

- *Solar thermal and photovoltaic every 1000 inhabitants*: the amount of solar thermal and photovoltaic power installed on public buildings per 1,000 inhabitants.
- *Photovoltaic solar panels per* km^2: the amount of photovoltaic solar power installed per km^2.

Goal 12: Responsible consumption and production

Promoting sustainable consumption and production patterns, focusing on waste management and resource use:

- *Separate waste collection*: the percentage of separate waste out of the total waste produced.
- *Production of urban waste*: kg of urban waste produced per inhabitant each year.
- *Residential waste collection*: the percentage of the population served by residential waste collection out of the total population.
- *Recycling points*: square meters per 100,000 inhabitants designed for ecological islands.

Goal 13: Climate action

This goal addresses climate change and its impacts:

- *CO2*: the total CO2 equivalent emitted by each inhabitant.
- *Population exposed to flood risk*: the percentage of the population exposed to flood risk out of the total population**.**

Goal 15: Life on land

This goal focuses on protecting, restoring, and promoting sustainable use of terrestrial ecosystems:

- *Ecolabel licenses*: the percentage of companies with the Ecolabel mark out of the total number of companies.
- *Usable urban greenery*: square meters of urban green space per inhabitant).

2) **The economic pillar**

These goals address economic development, growth, and social inclusion, focusing on reducing poverty, inequality, and promoting sustainable industrialization.

Goal 1: No poverty

This goal seeks to eradicate extreme poverty for all people everywhere:

- *Economic suffering index*: the percentage of families with an annual income between 0 and €10,000 at the time of the declaration.
- *Individuals in families with low work intensity*: the percentage of individuals in families with work intensity levels less than 0.20**.**

Goal 8: Decent work and economic growth

Promotes sustained, inclusive, and sustainable economic growth, employment, and decent work for all:

- *The average taxable income per capita*; *NEET between 15 and 29 years*: the percentage of young people aged 15-29 who neither study nor work out of the total number of young people aged 15-29.
- *Early exit from the education and training system*: the percentage of 18–24-year-olds who have completed secondary school and are not currently enrolled in any education or training program.

Goal 9: Industry, Innovation, and Infrastructure

Promotes sustainable industrialization and fosters innovation:

- *Mobility offered by transport public*: number of kilometers per inhabitant covered each year by public transport.

Goal 10: Reduced inequalities

This goal aims to reduce inequality within and among countries:

- *GINI index:* the GINI coefficient, which represents the distribution of wealth among residents - most used to measure inequality.
- *Digital divide*: the percentage of the population excluded from fixed and mobile broadband.

Goal 17: Partnership for the goals

Focuses on strengthening the means of implementation and revitalizing global partnerships for sustainable development:

- *Broadband access*: the percentage of households served by a potential internet connection of at least 30 Mbps.
- *Social cooperatives*: number of social cooperatives per 10,000 inhabitants.

3) The social pillar

These goals focus on improving human well-being, education, gender equality, and ensuring justice and strong institutions.

Goal 3: Good wealth and well-being

This goal aims to ensure healthy lives and promote well-being for all at all ages:

- *Life expectancy at birth*: how many years can a newborn baby expect to live assuming that in the future the mortality is constant at each age.
- *Life expectancy at 65 years*: how many years can a 65-year-old person expect to live assuming that in the future the mortality is constant at each age
- *Deaths and injuries in road accidents*: number of deaths or injuries in road accidents per 1,000 inhabitants.
- *Suicide mortality and intentional self-harm*: the total number of deaths due to suicide or intentional self-harm based on the territory of residence.
- *Infant mortality ratio*: the total number of children under one year of age who die for every 1,000 live births)**.**

Goal 4: **Quality education**

Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all:

- *Index of care of users of childcare services*: the percentage of children under 3 years old who attend childcare services.
- *Level of alphabetic competence*: the score that students in the second grade of secondary school obtained in the alphabetic proficiency tests.
- *Level of numerical competence*: the score that students in the second grade of secondary school obtained in numerical competence tests.
- *Population with a middle school diploma*: the percentage of students who have obtained a secondary school diploma out of the total population.
- *Population 0-16 enrolled in the public education cycle*: the percentage of individuals aged 0-16 who attend childcare services, first year of primary school or who are enrolled in a regular course of studies out of the total population aged 0-16.
- *Schools with a ramp*: the percentage of schools equipped with a ramp out of the total number of schools)**.**

Goal 5: Gender equality

Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all:

- *Absolute difference between male employment rate and female employment rate*: the difference between the percentage of men who work and the percentage of women who work.
- *Education level of women*: the percentage of women graduates out of the total number of graduates.
- *Women enrolled in university courses*: the percentage of women enrolled at university out of the total number of students enrolled.

Goal 11: Sustainable cities and communities

Make cities and human settlements inclusive, safe, resilient, and sustainable:

- *Cycle paths*: meters of surface used for cycle paths for every 100 inhabitants.
- *PM2.5 µg/m3*: µg/m3 of PM2.5 recorded on average every year.
- *Housing quality*: number of individuals living in homes without a toilet per 100,000 inhabitants.
- *PM10 µg/m3*: amount of µg/m3 of PM10 recorded on average each year.
- *Noise pollution*: number of noise pollution complaints per 100,000 inhabitants.
- *Nitrogen dioxide NO2 µg/m3*: amount of µg/m3 of NO2 recorded on average each year.

- *Pedestrianized road surfaces*: number of square meters of surface per inhabitant used for pedestrian traffic.
- *Dead, missing and people directly affected by disasters*: number of dead, missing and people directly affected by disasters per 100,000 inhabitants.

Goal 16: Peace, Justice and strong institutions

Promote peaceful and inclusive societies for sustainable development, provide access to justice for all, and build effective, accountable institutions:

- *Electoral participation in politics*: the percentage of the population that took part in political elections.
- *Court efficiency*: average number of days of pending civil proceedings in the year.

Due to the focus on urban and territorial data, **Goal 14** (Life Below Water) was excluded from the analysis for comparability reasons. This goal focuses on conserving and sustainably using the oceans, seas, and marine resources for sustainable development.

Finally, for each pillar, we measured the dispersion of the indicators, which highlights the variation or spread of values for each indicator across cities. As plotted in figures 2,3 and 4 it results that in the **environmental pillar**, urban gardens, solar energy installations, and flood exposure showed the greatest variation; in the **economic pillar**, the digital divide had the highest dispersion, reflecting differences in access to technology across cities; in the **social pillar**, indicators like cycle paths, noise pollution, and pedestrianized surfaces had the highest variation, pointing to urban planning and infrastructure disparities.

We merged all the previous data to obtain a database of 5500 rows by 61 columns. Then we added two columns to this database relating to the coordinate (longitude and latitude) of the various provinces considered.

Figure 2: Environmental pillar. A histogram for each indicator

Figure 3: Economic pillar. A histogram for each indicator

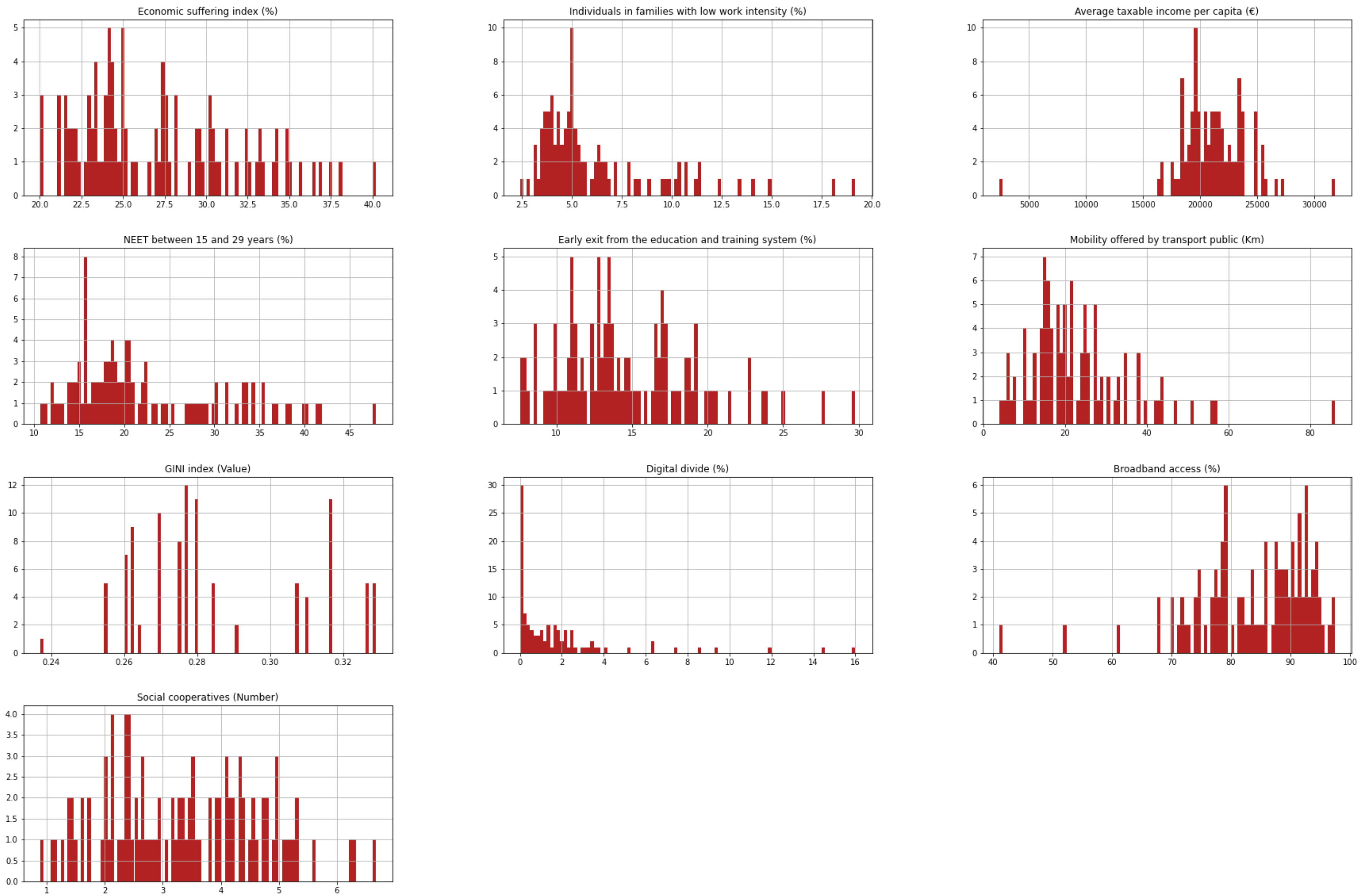
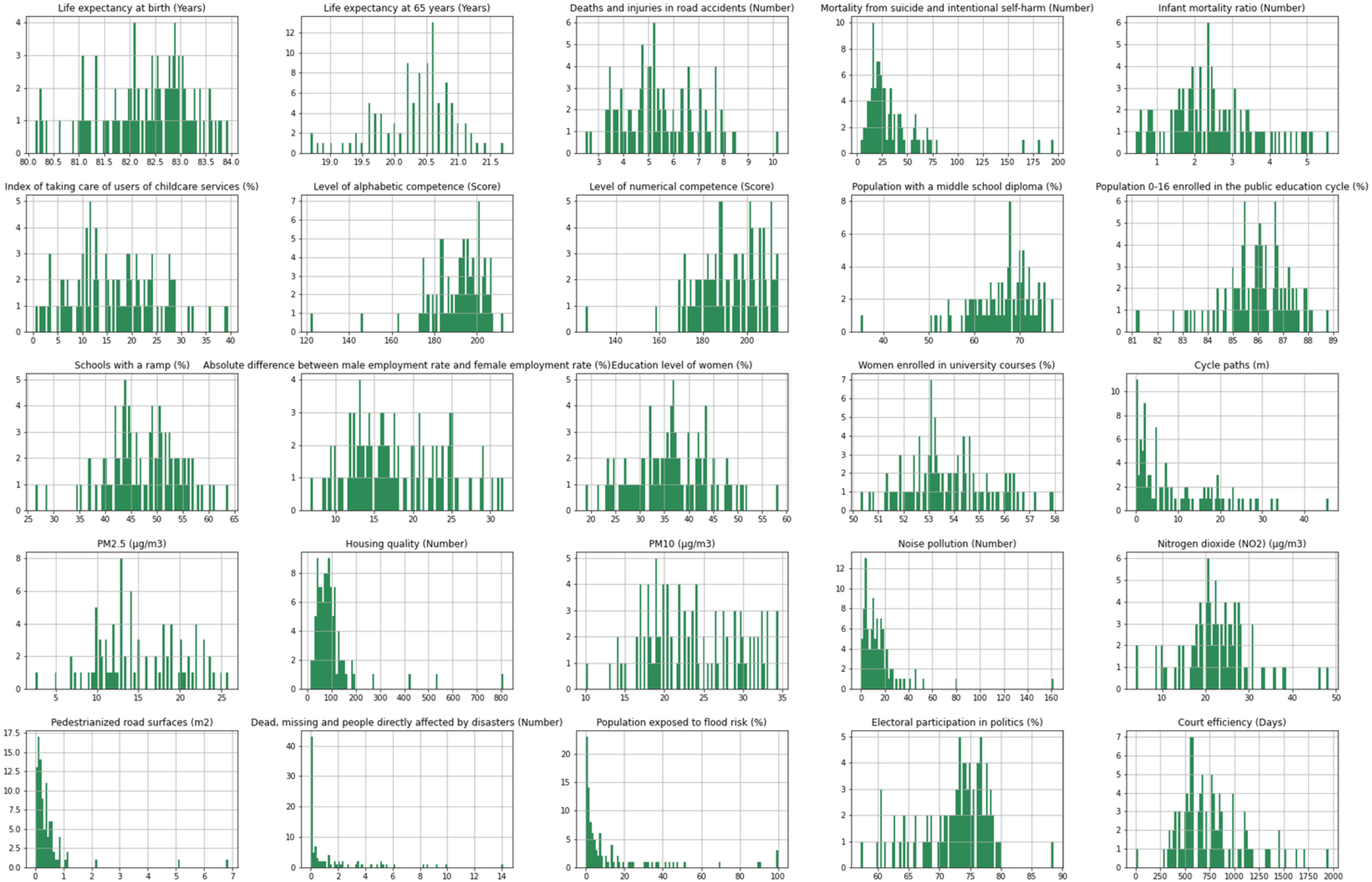


Figure 4: Social pillar. A histogram for each indicator



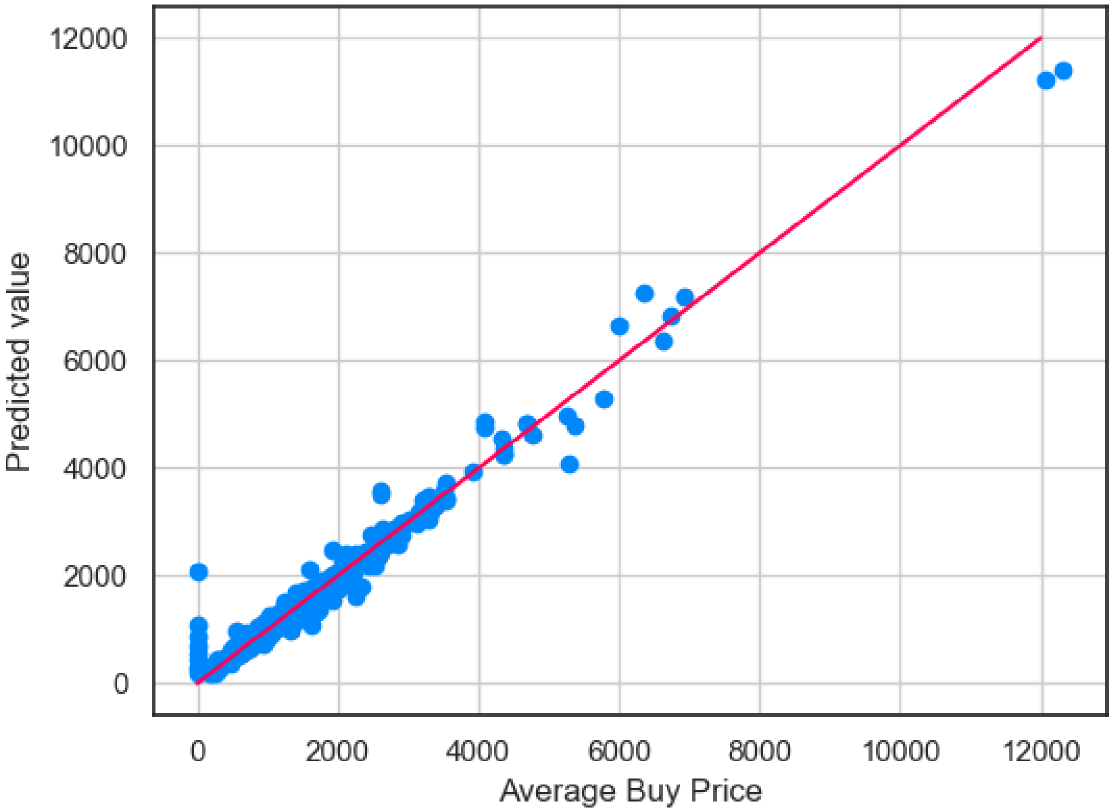
4. Machine learning model for price prediction and results discussion

We trained a machine learning model to predict property sales prices using the dataset from the previous section. The model employs Gradient Boosting Trees, an Ensemble Learning algorithm implemented in the Scikit-learn library (Hastie et al., 2009, 2013).

The dataset was split into training and test subsets, with 15% of the rows randomly selected for testing using a stratification criterion based on per-capita GDP income brackets. The rest of the rows constituted the training data set.
This split was automated through a routine available in the Scikit-learn library. The GDP per capita was categorized into four brackets: €0-€15,000 (category 1), €15,000-€28,000 (category 2), €28,000-€50,000 (category 3), and over €50,000 (category 4).

Since each algorithm has free-to-vary parameters that affect the performance of the model, we used a Scikit learn technique for parameter optimization, to make sure that the model has optimal performance. This technique is called hyper parameter tuning (an explanation is provided here: https://scikit-learn.org/stable/modules/grid_search.html#searching-for-optimal-parameters-with-successive-halving).
Using the parameters identified with this technique, the price values predicted by the model agree with the observed price values (see Figure 5). The goodness of the model measured through the R^2 on the test set is equal to 0.97. The graphical representation of the relationship between these two variables is available in figure 5.

Figure 5: Comparison between predicted and observed prices' value.



4.1.The Shapley

After obtaining the price prediction model, we sought to identify which variables in our dataset had the greatest impact on the prediction. To answer this, we utilized Python's SHAP library, using the test dataset as our database.

The SHAP explanation algorithm assigns a value to each input variable based on its contribution to predicting the output variable, i.e., the predicted price (Štrumbelj and Kononenko, 2014). This assignment is based on cooperative game theory. Specifically, Shapley values consider different subsets of input variables, adjusting the explanation model to compare the price prediction when including or excluding the input variable in question. Since the effect on the prediction varies depending on the subset considered, the algorithm calculates the difference between the predicted price with and without the variable for each subset and then computes a weighted average. For a theoretical analysis of SHAP values, see Lundberg and Lee, 2017.

To represent the SHAP value results, we used two types of graphs: the bar plot (Figure 6) and the beeswarm plot (Figure 7).

Figure 6 The summary plot to show the feature importance of each feature in the model.

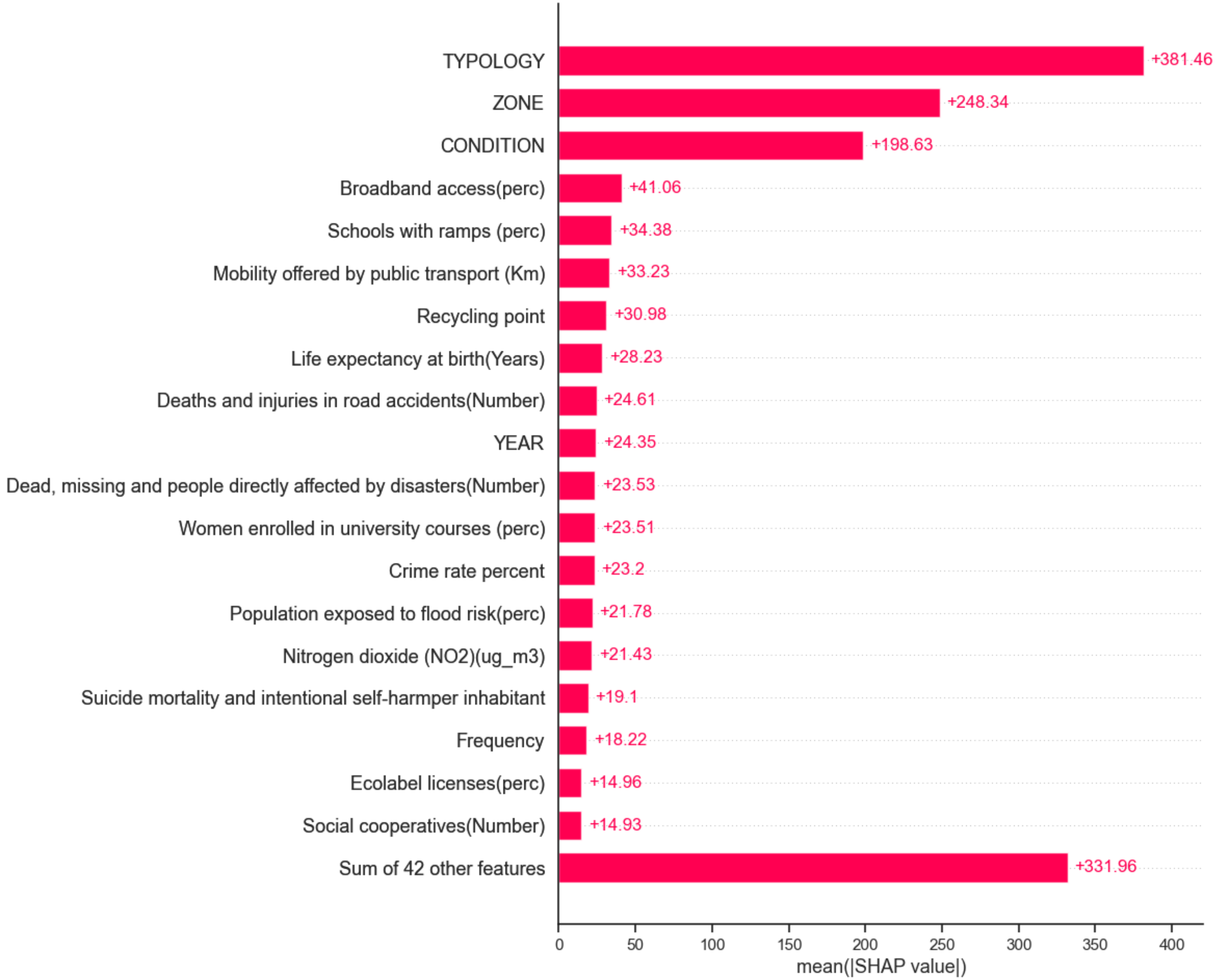
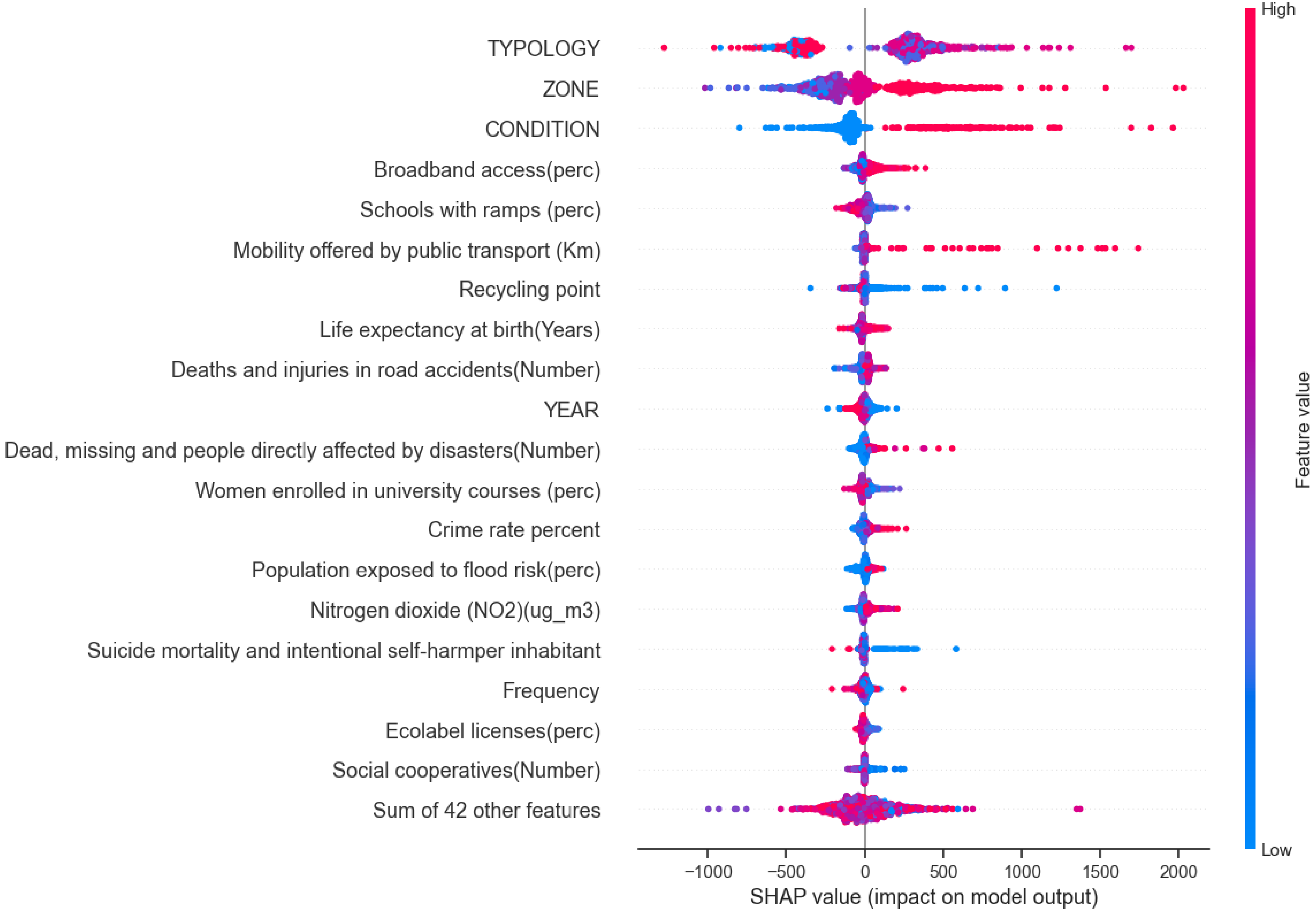


Figure 7: Beeswarm plot to show the most impactful features



The first graph shows the range of total variation of the average SHAP value for all the input variables on the x-axis. The input variables are sorted in decreasing order by their average SHAP value along the ordinate axis. Among the top 20 most important variables for the prediction of the model, we observe the following: categorical features (zone, typology, and condition), 'Broadband access (%) (goal 17)', 'Schools with ramps (%) (goal 4)', 'Mobility offered by public transport (Km) (goal 9)', 'Recycling point (m2) (goal 12)', 'Life expectancy at birth (Years) (goal 3)', 'Deaths and injuries in road accidents (Number) (goal 3)', 'Women enrolled in university courses (%) (goal 5)', 'Dead, missing, and people directly affected by disasters (Number) (goal 11)', 'Crime rate percent', 'Population exposed to flood risk (%) (goal 11)', 'Nitrogen dioxide (NO2) (μg/m3) (goal 11)', 'Suicide mortality and intentional self-harm per inhabitant (goal 3)', 'Frequency', 'Ecolabel licenses (%) (goal 15)', and 'Social cooperatives (Number) (goal 17)'. These variables span all three pillars considered: environmental, economic, and social. This indicates that all these aspects influence the value of non-residential properties.

Turning to figure 7, the beeswarm plot represents the distribution of SHAP values for the most relevant input variables. The color of each point indicates the value of the input variable, with blue indicating low values and red indicating high values. The range of SHAP values is shown on the x-axis, and the input variables are displayed in order of importance on the vertical axis. Upon inspecting the beeswarm plot, we observe that the color distribution for some variables is not clearly separated into blue or red but appears more heterogeneous. To investigate this, we examined potential interactions between these variables and others, which are discussed in figures 8, 9, 10, and 11.

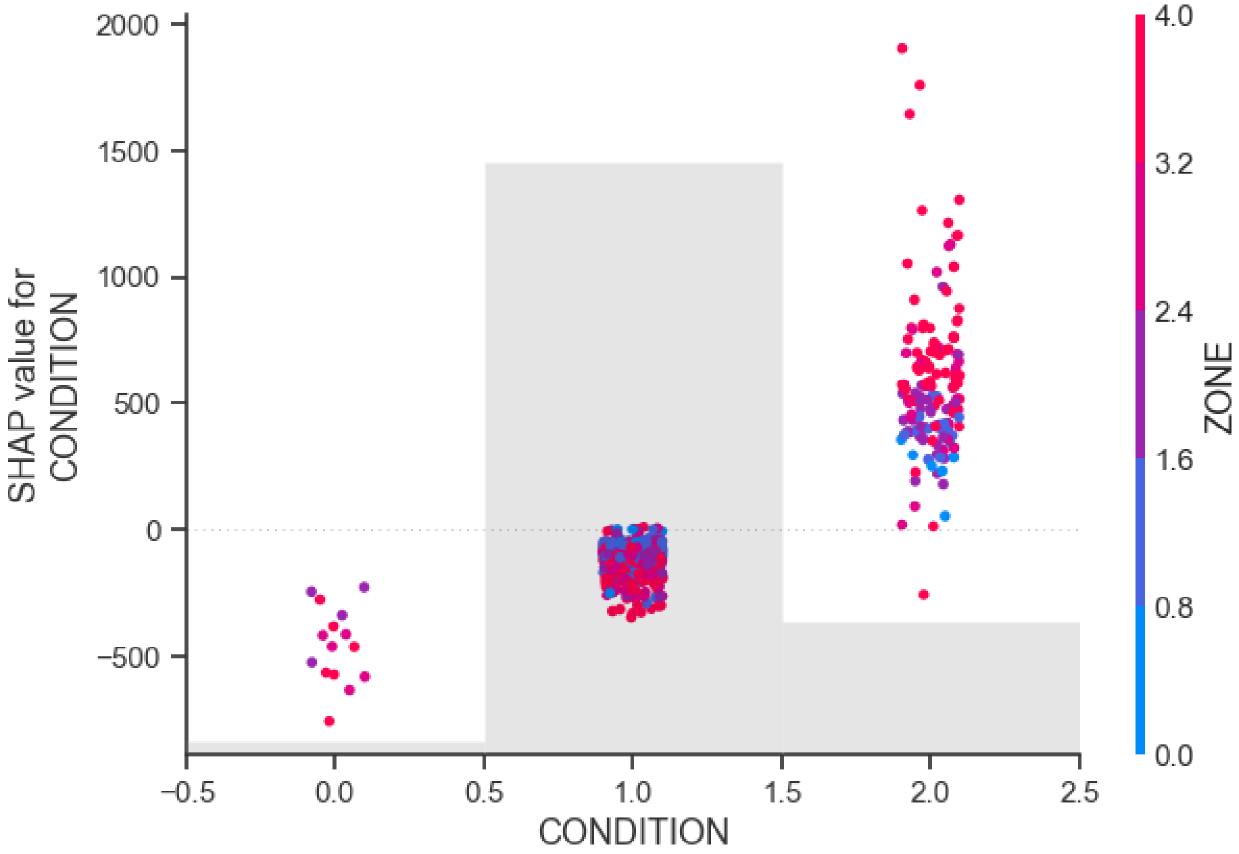
We now focus on the most relevant features as identified in figure 7, comparing our findings with results from the literature. For instance, studies such as Molnar et al. (2019) and Wolf and Irwin (2024) explored the relationship between broadband access and house values, showing that homes with access to faster broadband have higher sale prices. While there are no studies examining broadband access and non-residential property values, our analysis indicates that higher broadband access tends to positively affect the model's output, as reflected by the mostly positive SHAP values.

Regarding recycling points, we observe that lower values of this feature positively impact the model's prediction. This result is consistent with existing literature, which suggests that the presence of recycling points generally lowers real estate prices, particularly when the sites are classified as hazardous. Non-residential properties, often located near waste sites, may face more significant price impacts (Ihlanfeldt and Taylor, 2004; Braden et al., 2011).

The 'Mobility offered by public transport' feature also shows a positive correlation with the predicted price. This outcome aligns with empirical evidence, which supports the idea that better public transport mobility tends to enhance property values, both in medium-sized cities without mobility issues and in larger cities with congestion problems (Cordera et al., 2019).

When examining the categorical variables, we notice a clear positive relationship with SHAP values. Better conservation states of properties are associated with higher prices. Furthermore, when analyzing the interaction between the "condition" and "zone" features (see figure 8), we observe that properties in excellent condition located in central areas produce a positive impact on the model's output. Previous studies, such as Nanda and Ross (2012), have explored how property conditions affect home prices. Miller et al. (2018) found that the price differences based on property conditions are more significant in weaker market conditions, whereas in favorable market conditions, these differences are less pronounced.

Figure 8: SHAP interaction plot between ‘condition’ and ‘zone’



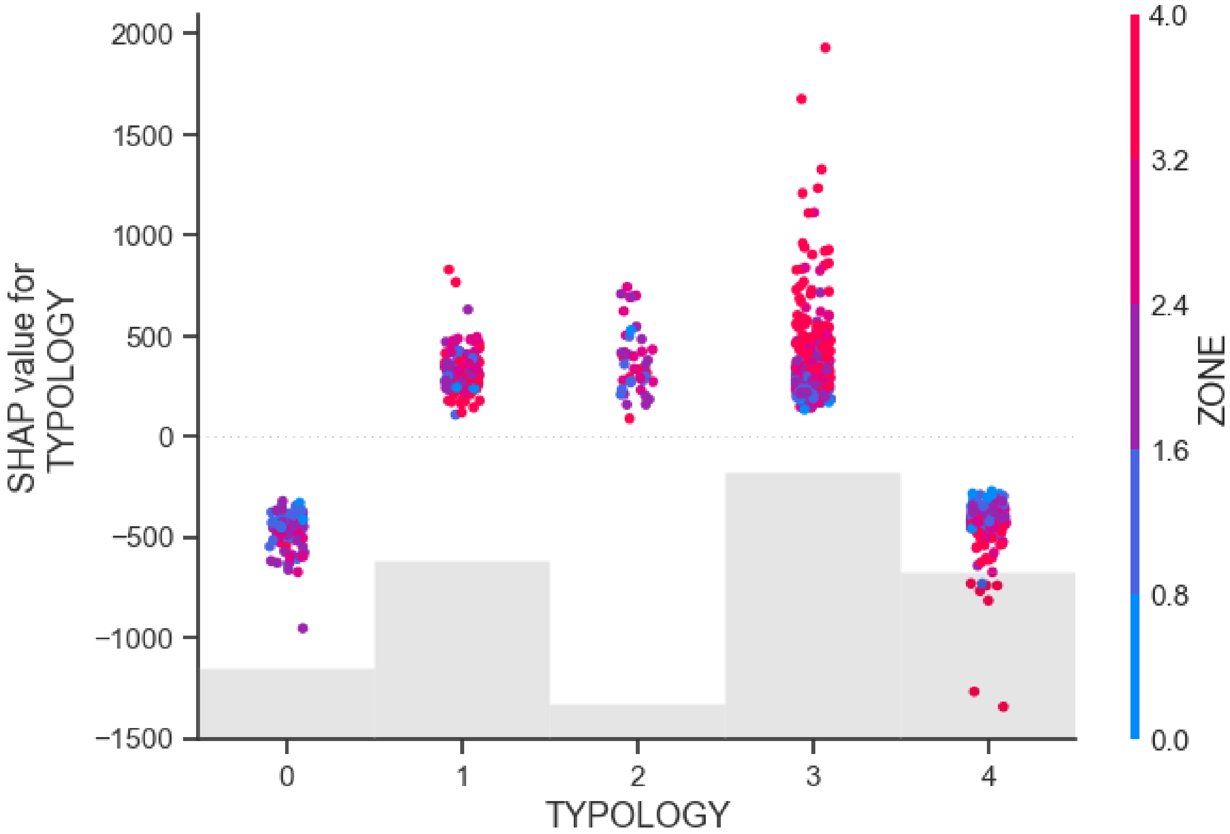
n

Note.

The zone considered are 0=R; 1=E; 2=D; 3=C; 4=B, where B=centre, C=semi-centre, D=outskirts, E=suburban area, R=extra-urban area/agricultural area. The conditions are 0=POOR; 1=NORMAL; 2=EXCELLENT.

Another important result, related to the categorical variables, concerns the typology of properties considered. Among the various non-residential property types analyzed, offices, shopping centers, shops, and laboratories—primarily located in central areas—appear to have the most significant influence on SHAP values (see Figure 9). The fact that the SHAP values for properties in this sector are consistently higher than those for warehouses and industrial sheds suggests a direct association between these property types and higher prices.

Figure 9: SHAP interaction plot between 'typology' and 'zone'



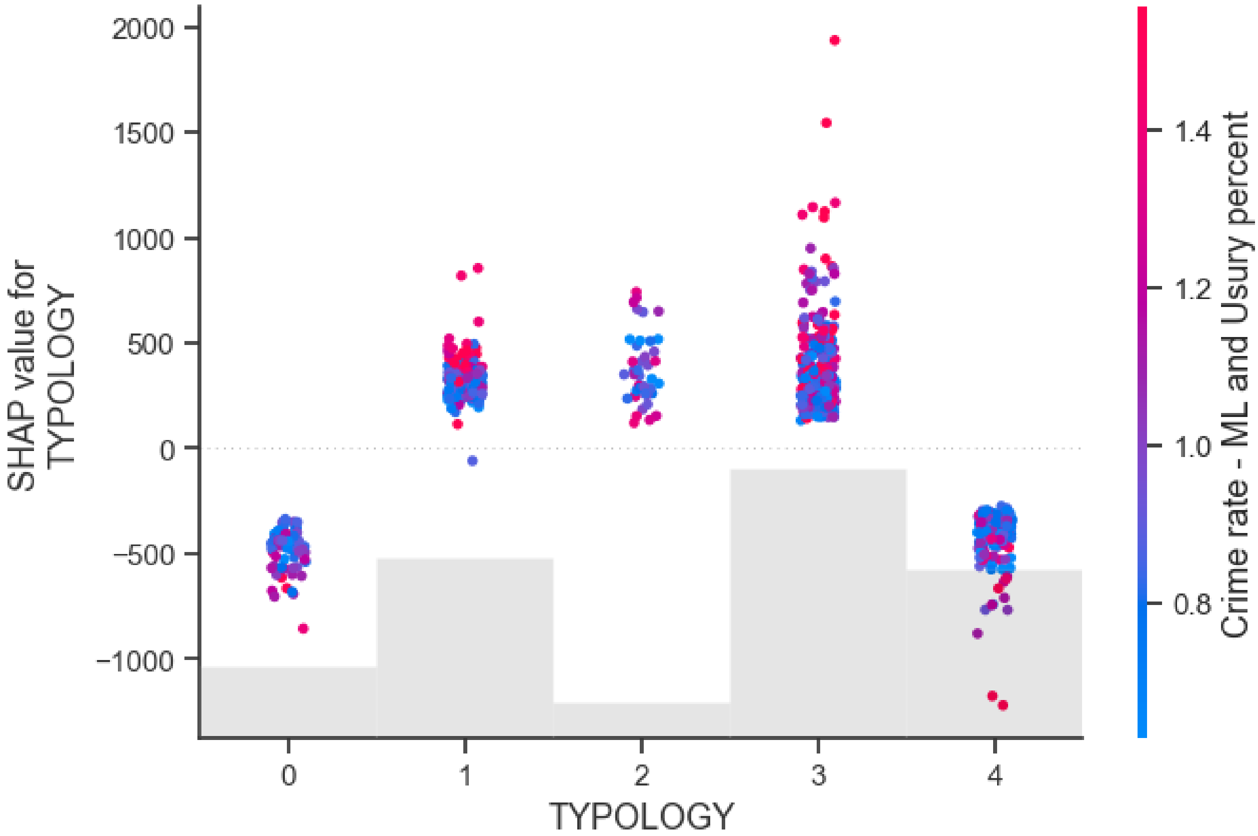
Note.

The typologies considered are 0=INDUSTRIAL SHEDS; 1=OFFICES; 2=SHOP

PING CENTERS; 3= SHOPS AND LABORATORIES; 4= Warehouses

The interaction with the crime rate, particularly for offices, shops, and laboratories, is illustrated in Figure 10. This can be seen through the increasing vertical distinction between blue (low crime rate) and red (high crime rate) colors. Notably, we observe that the SHAP value for these properties is higher when associated with high crime rates.

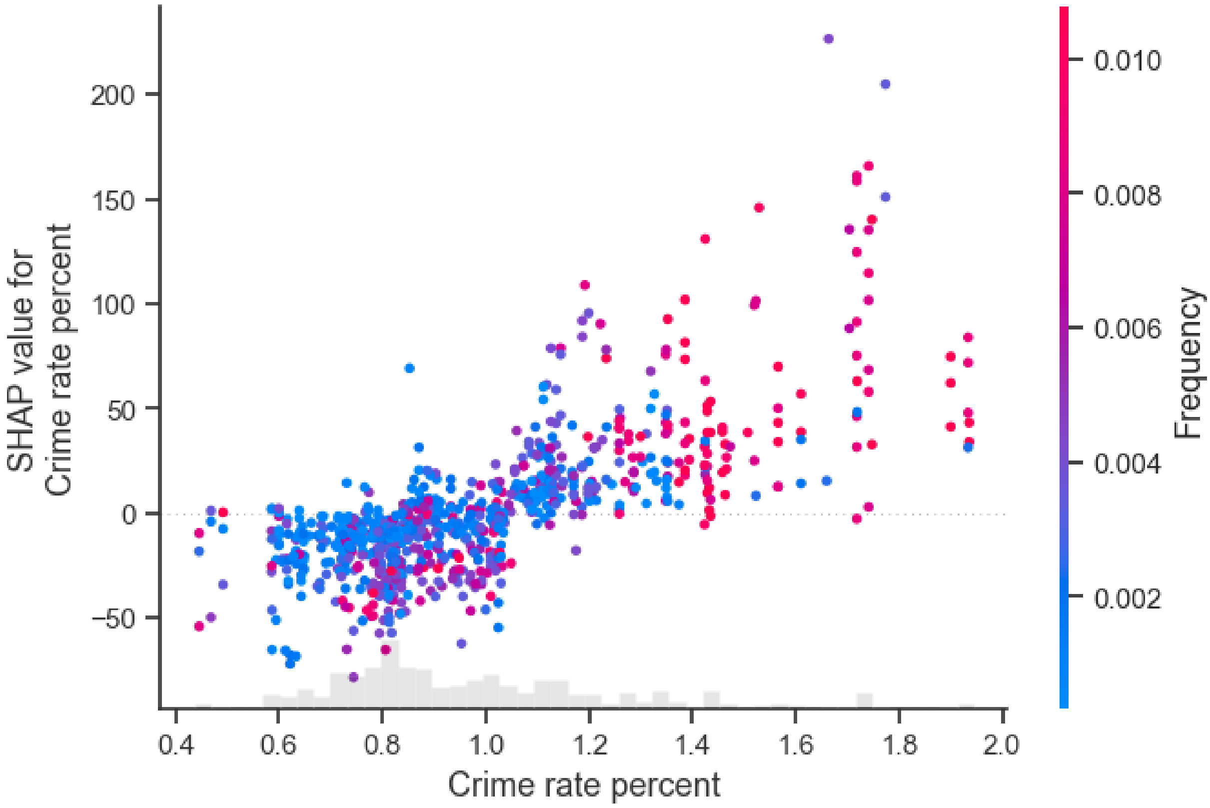
Figure 10: SHAP interaction plot between ‘typology’ and ‘crime rate minus ML and usury’



In Figure 11, we observe a growing relationship between crime rate and transaction frequency. Higher crime rates correspond to higher SHAP values, while lower crime rates are associated with lower SHAP values. At higher crime levels, there is a greater frequency of real estate transactions.

The relationship between crime and property value has been analyzed in the literature, particularly in the residential market. The findings show a mixed effect, with the relationship being either positive or negative depending on the type of crime considered. For example, acts of vandalism and other violent crimes tend to lower housing prices by creating a climate of fear and insecurity (Dugato et al., 2015; Kruisbergen et al., 2015). On the other hand, the impact of money laundering in the real estate sector has the opposite effect, often driving up property values (Barone, 2023). Money laundering typically enters the real estate market through mortgages or cash, with private digital currencies acting as close substitutes (Schneider, 2004; Borgonovo et al., 2021).

Figure 11: SHAP interaction plot between 'crime rate and 'frequency'



4.2.The causal inference

Once we have identified, through Shapley, the variables that most significantly impact the determination of the output, we must consider whether some of these variables might not directly affect the output variable. It is possible that these variables impact other input variables, which in turn have a direct link to the output.

To answer this question by estimating the causal effect between the treatment and the outcome, we need to proceed with causal inference. For this analysis, we used the DoWhy and EconML libraries from the PyWhy project (Sharma, 2020).

Focusing on the variables identified as relevant based on the SHAP values, we estimate the causal effect for the following features: Nitrogen dioxide concentration (NO2, in $\mu g/m^3$), the crime rate (%), and the mobility offered by public transport (in km). The idea is that these variables may be more directly controlled by the government, which could, in turn, influence prices.

Starting with the first input variable, Nitrogen dioxide concentration, we evaluate the possible causal effect between this variable and the price level of the properties considered. We also account for other variables (from the 20 most significant ones identified by Shapley) that could confound the causal relationship. These variables include: zone, typology, life expectancy at birth (in years), women enrolled in university courses (as a percentage), and the number of deaths, missing persons, and individuals directly affected by disasters. For the selection of variables, we consulted relevant literature.

- **Nitrogen Dioxide Concentration (NO2) – Non-residential Property Price**

The levels of NO2 concentration, both outdoors and indoors, appear to be linked to different types of properties. Residential properties, offices, and schools produce different levels of NO2, depending on the relationship between indoor and outdoor NO2 concentrations. Residential properties are more likely to have higher indoor concentrations of NO2, which leads to a noticeable difference between residential properties and schools or offices. Additionally, factors such as ventilation strength in different buildings contribute to varying ratios of indoor to outdoor NO2 concentrations (Hu and Zhao, 2020; Linaker et al., 1996).

Another factor affecting NO2 levels is the area where the buildings are located. Areas with lower socio-economic status (often suburban) tend to experience higher exposure to NO2 (Lai et al., 2006; Temam et al., 2017). Exposure to nitrogen dioxide is linked to a variety of health issues, including heart disease, respiratory problems, diabetes, and cognitive decline (Kaplan & Keil, 1993; Roffia et al., 2023), and an increased risk of natural disasters, thereby reducing life expectancy. Research indicates that higher education levels are often associated with healthier behaviors, lower exposure to life-threatening factors (such as smoking, alcohol consumption, and poor diet). To assess health status, we use life expectancy at birth as a proxy for health (Cochrane et al., 1978; Jaba et al., 2014). We also use the percentage of women enrolled in university courses as a proxy for education level.

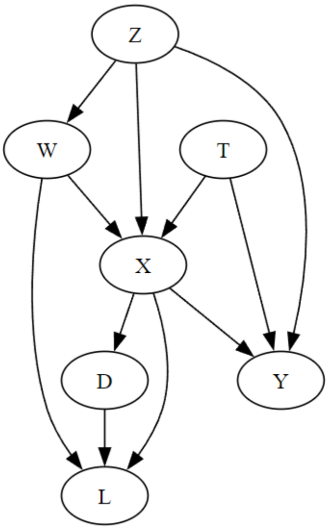
Here are the hypotheses:

1. The typology of the building ('T') affects NO2 concentration (X) and the average price of commercial properties ('Y') (based on the Shapley analysis).
2. The zone ('Z') affects NO2 concentration ('X'), education level (proxied by the percentage of women enrolled in university courses, 'W'), and the average price of commercial properties ('Y') (based on the Shapley analysis).

3. Exposure to NO2 (‘X’) affects life expectancy at birth (‘L’) and the number of deaths, missing persons, and individuals directly affected by disasters (‘D’).
4. The number of deaths, missing persons, and individuals directly affected by disasters (‘D’) affects life expectancy at birth (‘L’).
5. The level of education (‘W’) affects both NO2 concentration (‘X’) and life expectancy at birth (‘L’).

We present a graphical representation of these relationships using a directed acyclic graph (DAG) generated with Graphviz (see Figure 12).

Figure 12:DAG that show the causal effect between NO2 and the commercial properties price



To identify causal estimands, we use the back-door criterion, implemented in the double machine learning (DML) method from EconML. The goal of DML is to estimate (heterogeneous) treatment effects in the presence of many observed confounders (Chernozhukov, 2016).

Our estimate finds a causal effect of 29.878 €/sqm. To assess the robustness of the estimate, we use the “add a random common cause” refute test. In this test, we add an independent random variable that is correlated with both the treatment and the outcome. If the model is properly accounting for confounders, the causal estimates should remain stable. The original effect estimate of 29.878 €/sqm is very close to the new estimate of 30.0505 €/sqm, with a p-value of 0.41. Since this p-value exceeds the threshold of 0.05, we cannot reject the null hypothesis.

- **Mobility Offered by Public Transport – Non-residential Property Price**

For the second analysis, existing literature on transport and mobility suggests that the growing number of road traffic accidents in cities has drawn increasing attention to urban mobility issues among policymakers, government bodies, transport operators, researchers, and user groups. The choice to use cars over public transport often stems from a lack of knowledge about available alternatives. Therefore, providing information about the benefits of public transport or walking is crucial (Horeni et al., 2007; Gardner and Abraham, 2007; Tyrinopoulos and Antoniou, 2013). Certain groups, such as women, travelers with children, young people, rural dwellers, and people with mobility impairments, are more likely to rely on public transport (Punzo et al., 2022).

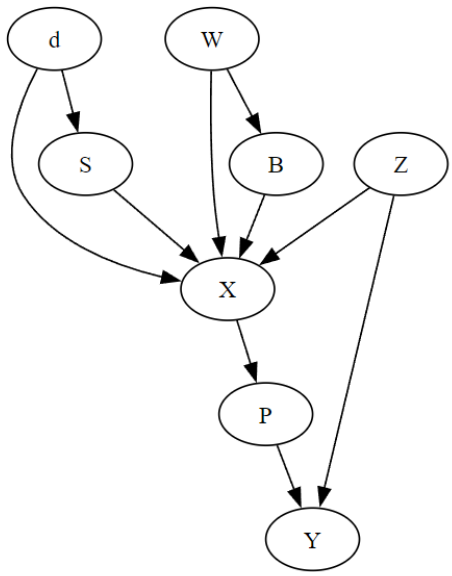
Despite the benefits of public transport, the mobility and transport sector were responsible for the second-highest level of CO2 emissions in Europe in 2010, contributing to global warming, an increase in natural disasters, and heightened risks for the population (Schoenau and Müller, 2017; Prillwitz and Barr, 2011). Some studies have explored the impact of flood risk on property values, with many respondents perceiving that properties at risk of flooding lose utility, thus affecting their value (Bhattacharya-Mis and Lamond, 2016; Porter et al., 2022).

Based on these findings, our hypotheses are:

1. Mobility offered by public transport (‘X’) is influenced by ‘deaths and injuries in road accidents’ (‘d’), ‘schools with ramps’ (‘S’), ‘broadband access’ (‘B’), ‘women enrolled in university courses’ (‘W’), and ‘zone’ (‘Z’).
2. Broadband access is influenced by women enrolled in university courses.
3. The percentage of schools with ramps (‘S’) is influenced by deaths and injuries in road accidents (‘d’).
4. Mobility offered by public transport (‘X’) affects the percentage of the population exposed to flood risk (‘P’).

We present the DAG for this analysis in Figure 13.

Figure 13:DAG that show the causal effect between public transport mobility and the commercial properties price



Using the back-door criterion, we find a causal effect of 17.05 €/sqm. After running the robustness test, we obtain a new effect of 17.21 €/sqm, with a p-value of 0.3. Since the difference is not significant, we do not reject the null hypothesis.

- **Crime – Non-residential Property Price**

Finally, we analyse the relationship between crime and commercial property prices. Several determinants of crime are identified in the literature, including deterrence variables (e.g., law enforcement presence, risk of discovery), pecuniary variables (e.g., unemployment rate, income inequality), and socio-demographic variables (e.g., urbanization, education). It has been shown that even small differences in the probability of apprehension and income inequality can lead to significant differences in crime rates across similar environments (İmrohoroğlu et al., 2006).

Higher crime rates increase both direct and indirect costs at the economic and social levels. Studies also find that crime causes mental distress, including depression and anxiety, among individuals living or working in affected areas (Dustmann and Fasani, 2016).

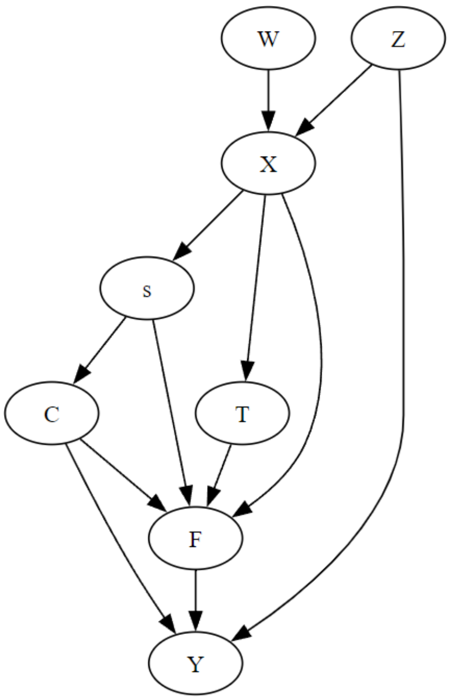
Organized crime often uses the real estate sector for money laundering, exploiting factors such as the difficulty of valuing properties, especially high-value ones (e.g., hotels, shopping centers), and the ability to buy or sell properties above or below market value (FATF/GAFI, 2008).

Based on the literature, we hypothesize the following relationships:

1. Women enrolled in university courses (as proxies for education, 'W') and the zone ('Z') affect the crime rate ('X').
2. The crime rate ('X') affects property typology ('T') and the frequency of transactions ('F').
3. The crime rate ('X') impacts mental health (proxied by suicide mortality and intentional self-harm rates, 's').
4. Suicide mortality and self-harm rates ('s') and property typology ('T') affect transaction frequency ('F').
5. Transaction frequency ('F') affects non-residential property prices ('Y').

We present the DAG for this analysis in Figure 14.

Figure 14: DAG that show the causal effect between crime rate and the commercial properties price



Using the back-door criterion, we estimate a causal effect of 1029.35 €/sqm. After the robustness test, the new estimate is 1027.24 €/sqm, with a p-value of 0.44. Since the difference is not significant, we do not reject the null hypothesis.

In conclusion, despite certain variables intrinsic to non-residential buildings—such as property type and condition—the presence of amenities in the surrounding area (e.g., broadband access, public transport mobility, recycling points) and some social and environmental variables (e.g., life expectancy, traffic accidents, educational enrolment, natural disasters, flood risk, suicide rates) significantly affect commercial property prices. However, the most important variable influencing non-residential property prices remains the provincial crime rate.

5. Conclusions

In this paper, we propose a machine learning approach to analyze the determinants of non-residential real estate market prices. To predict prices, we used a gradient boosting regressor trained on a custom dataset composed of data from the Real Estate Market Observatory (OMI) of the Italian Revenue Agency, as well as data encompassing the three main pillars of sustainability—environmental, social, and economic—for each Italian province. The best predictive model achieved an R^2 of 0.972. We then evaluated feature importance and, to explain the model's results, we used the Shapley Additive exPlanations (SHAP) approach and causal inference methods.

Our analysis reveals that all three pillars (social, environmental, and economic) impact non-residential property prices. The most significant goals are Goal 3 (Good Health and Well-being), Goal 11 (Sustainable Cities and Communities), Goal 12 (Responsible Consumption and Production), and Goal 17 (Partnership for the Goals). Among the most important variables, beyond the sustainability pillars, are the categorical variables: location of the properties, typology, and conservation state. For certain property types (offices, shops, and laboratories) that have the greatest positive impact on prices, the crime rate appears to play a key role. Specifically, we observe that property prices are higher in areas with higher crime rates. However, the SHAP values allowed us to assess the correlation between variables. It is important to note that correlation and causation are distinct concepts. To determine which variables have an actual causal relationship with property prices, we performed a causal inference analysis. In this analysis, we hypothesized that the variables most likely to have a causal relationship with prices are those that can be more easily controlled by policymakers—specifically, the level of NO2, the mobility offered by public transport, and the crime rate. Our hypothesis was confirmed by the data.

References

- Agarwal S, Ambrose BW, Lopez LA, Xiao X (2020) Did the paycheck protection program help small businesses? Evidence from commercial mortgage-backed securities. Health & the Economy eJournal. doi:10.2139/ssrn.3674960 Corpus ID: 234650284
- Barone R (2023) Home sweet home, how money laundering pollutes the real estate market: an agent-based model. Journal of Economic Interaction and Coordination 18, 779–806. https://doi.org/10.1007/s11403-023-00391-y
- Baptista, R., Farmer, J. D., Hinterschweiger, M., Low, K., Tang, D., & Uluc, A. (2016). Macroprudential policy in an agent-based model of the UK housing market. Staff Working Paper 619, Bank of England.
- Batalha M, Goncalves D, Peralta S, dos Santos JP (2022) The virus that devastated tourism: The impact of covid-19 on the housing market. Regional Science and Urban Economics 103774. ISSN 0166-0462, in press, available online. https://doi.org/10.1016/j.regsciurbeco.2022.103774
- Beland LP, Brodeur A, Wright T (2020) COVID-19, stay-at-home orders and employment: Evidence fromCPS data. IZA DP No. 13282.
- Bhattacharya-Mis N and Lamond J (2016). Risk perception and vulnerability of value: a study in the context of commercial property sector. International Journal of Strategic Property Management, 20(3), pp: 252–264. https://doi.org/10.3846/1648715X.2016.1188174
- Borgonovo E, Caselli S, Cillo A, Masciandaro D, Rabitti G (2021) Money, privacy, anonymity: What do experiments tell us?. Journal of financial stability, 56. https://doi.org/10.1016/j.jfs.2021.100934
- Bowman R and Wills J (2008) Valuing Green: How Green Buildings Affect Property Values and Getting the Valuation Method Right. Australian Green Building Council, Melbourne.
- Braden J B, Feng X and Won D (2011) Waste Sites and Property Values: A Meta-Analysis. Environ Resource Econ 50, 175–201. https://doi.org/10.1007/s10640-011-9467-9
- Breiman L (2001) Random Forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
- Buonanno P and Montolio D (2008) Identifying the socio-economic and demographic determinants of crime across Spanish provinces. International Review of Law and Economics, 28(2), pp: 89-97. ISSN 0144-8188, https://doi.org/10.1016/j.irle.2008.02.005.
- Capponi A, Rios DA (2020) COVID-19 mortgage forbearance: Implications on the housing market. SSRN Electronic Journal 3618776.

- Carro A, Hinterschweiger M, Uluc A, Farmer J D (2022) Heterogeneous effects and spillovers of macroprudential policy in an agent-based model of the UK housing market. Industrial and Corporate Change. dtac030, https://doi.org/10.1093/icc/dtac030
- Chapman B (2018) Criminals targeting property in university towns to launder dirty cash. The Independent
- Chen L, Yao X, Liu Y, Zhu Y, Chen W, Zhao X, and Chi T (2020) Measuring impacts of urban environmental elements on housing prices based on multisource data—A case study of Shanghai, China. International Journal of Geo-Information, 9(2), p. 106. https://doi.org/10.3390/ijgi9020106
- Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, and. Newey W (2016) Double Machine Learning for Treatment and Causal Parameters. ArXiv e-prints, July 2016.
- Cincotti S, Raberto M and Teglio A (2022) Why do we need agent-based macroeconomics?. Review of Evolutionary Political Economy 3, 5-29. https://doi.org/10.1007/s43253-022-00071-w
- Cochrane A L, St Leger A S, Moore F (1978) Health service 'input' and mortality 'output' in developed countries. Journal of Epidemiology & Community Health, Vol. 32(3) pp:200-205. https://doi.org/10.1136/jech.32.3.200
- Cordera R, Coppola P, Dell' Olio L, Ibeas A (2019) The impact of accessibility by public transport on real estate values: A comparison between the cities of Rome and Santander. Transportation Research Part A: Policy and Practice, 125, pp: 308-319. ISSN 0965-8564, https://doi.org/10.1016/j.tra.2018.07.015.
- De Villa A R (2023) Causal Inference for Data Science. MEAP. ISBN 9781633439658
- D'Lima W, Lopez LA, Pradhan A (2021) COVID-19 and housing market effects: Evidence from U.S. shutdown orders. Real Estate Economics 50(2), pp:303–339. https://doi.org/10.1111/1540-6229.12368
- Dave D, Friedson A, Matsuzawa K, Sabia J (2020) When do shelter-in-place orders fight COVID-19 best? Policy heterogeneity across states and adoption time (Technical Report w27091). National Bureau of Economic Research. Cambridge, MA. Retrieved fromhttp://www.nber.org/papers/w27091.pdf
- De Liso N (2023) Artificial Intelligence and Inequality. In Global Handbook of Inequality Jodhka, AG, Rehbein (eds.), Springer Nature Switzerland. 2023 S. S. B, https://.org/10.1007/978-3-030-97417-6_49-1.

- Duca JV, Hoesli M, Montezuma J (2021) The resilience and realignment of house prices in the era of COVID-19. Journal of European Real Estate Research 14(3), pp:421–431. https://doi.org/10.1108/JERER-11-2020-0055
- Dugato M, Savarin S, Giommoni L (2015) The Risks and Rewards of Organized Crime Investments in Real Estate. The British Journal of Criminology, 55(5), pp:944–965. https://doi.org/10.1093/bjc/azv002
- Dustmann C and Fasani F (2016) The Effect of Local Area Crime on Mental Health. The Economic Journal, 126(593), pp: 978–1017. https://doi.org/10.1111/ecoj.12205
- Franco and Cutter (2022) The determinants of non-residential real estate values with special reference to environmental local amenities. Ecological Economics, 201. https://doi.org/10.1016/j.ecolecon.2022.107485
- Fuerst F and Mcallister P (2008a) Green noise or green value? Measuring the price effects of environmental certification in commercial buildings. School of Real Estate and Planning, Henley Business School, University of Reading, Reading, MA.
- Fuerst F and Mcallister P (2008b) Pricing sustainability: an empirical investigation of the value impacts of green building certification. Working paper from the Proceedings of the American Real Estate Society Conference, April, Florida, ARES.
- Gardner B and Abraham C (2007) What drives car use? A grounded theory analysis of commuters' reasons for driving. Transportation Research Part F-traffic Psychology and Behaviour, 10(3), pp: 187-200. DOI:10.1016/J.TRF.2006.09.004
- Geanakoplos J, Axtell R, Farmer JD, Howitt P, Conlee B, Goldstein J, Hendrey M, Palmer NM, and Yang CY (2012). Getting at Systemic Risk via an Agent-Based Model of the Housing Market. American Economic Review, 102 (3), pp: 53-58. https://doi.org/10.1257/aer.102.3.53
- Granja J, Makridis C, Yannelis C, Zwick E (2020) Did the pay check protection program hit the target?. National Bureau of Economic Research. Series: Working Paper Series n. 27095. Retrieved from http://www.nber.org/papers/w27095
- Griliches Z (1961) Hedonic price indexes for automobiles: An econometric of quality change, Price Statistics of the Federal Government, 173-196, NBER.
- Guenster and Koegst (2018) Environmental responsibility and firm value, in Toine Spapens, Rob White, Daan van Uhm and Wim Huisman (Eds), Green Crimes and Dirty Money, Routledge, London and New York, ISBN: 978-0-8153-7221-9 (hbk); ISBN: 978-1-351-24574-6 (ebk)

- Gupta A, Mittal V, Peeters J, Van Nieuwerburgh S (2022) Flattening the curve: Pandemic Induced revaluation of urban real estate. Journal of Financial Economics, 146(2), pp: 594-636, ISSN 0304-405X, https://doi.org/10.1016/j.jfineco.2021.10.008.
- Hastie T, Tibshirani R and Friedman J (2013) The elements of statistical learning, Springer, Stanford, California (chapter 16) https://hastie.su.domains/Papers/ESLII.pdf
- Hastie T, Tibshirani R, Friefidman J (2009). The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Springer Series in Statistics, Second Edition, New York, USA. ISBN: 978-0-387-84857-0.
- Hill R J (2013) Hedonic Price Indexes For Residential Housing: A Survey, Evaluation And Taxonomy. Journal of Economic Surveys, 27(5), pp: 879-914. doi: 10.1111/j.1467-6419.2012.00731.x
- Horeni O, Gärling T, Loukopoulos P, Fujii S (2007) An experimental simulation of adaptations to increased car-use costs. Transportation Research Part F: Traffic Psychology and Behaviour, 10(4), pp: 300-320. ISSN 1369-8478, https://doi.org/10.1016/j.trf.2006.12.001
- Hu Y and Zhao B (2020) Relationship between indoor and outdoor NO2: A review. Building and Environment, 180, 106909. ISSN 0360-1323. https://doi.org/10.1016/j.buildenv.2020.106909)
- Kaplan G A and Keil J E (1993). Socioeconomics factors and cardiovascular disease: A review of the literature. Circulation, 88(4), pp: 1973–1998. https://doi.org/10.1161/01.CIR.88.4.1973
- Klassen R and McLaughlin C (1996) The Impact of Environmental Management on Firm Performance. Management Science, 42, pp: 1199-1214. https://doi.org/10.1287/mnsc.42.8.1199
- Klügl F, Bazzan ALC (2012) Agent-Based Modeling and Simulation. Association for the Advancement of Artificial Intelligence, AI Magazine: 29-40. ISSN: 0738-4602.
- Kruisbergen EW, Kleemans ER, Kouwenberg RF (2015) Profitability, power, or proximity? Organized crime offenders investing their money in legal economy. European Journal on Criminal Policy and Research, 21(2), pp: 237–256.
- Ihlanfeldt KR, Taylor LO (2004) Externality effects of small-scale hazardous waste sites: evidence from urban commercial property markets. J Environ Econ Manage, 47, pp: 117–139 https://doi.org/10.1016/S0095-0696(03)00070-6

- İmrohoroĝlu A, Merlo A and Rupert P (2006) Understanding the determinants of crime. Journal of Economics and Finance, 30(2), pp: 270–284. https://doi.org/10.1007/BF02761491
- Jaba E, Balan C and Robu I (2014). The relationship between life expectancy at birth and health expenditures estimated by a cross-country and time-series analysis. Procedia Economics and Finance, 15, pp:108–114. https://doi.org/10.1016/S2212-5671(14)00454-7
- La Gala CG (2000) Il riciclaggio di denaro. Strumenti di contrasto e misure patrimoniali. Rassegna dell'Arma dei Carabinieri. https:// www. carab inieri. it/ docs/ defau lt- source/ defau lt- docum ent- libra ry/suppl emento- al-n- 4. pdf? sfvrsn= e31e6 d23_2
- Lai H K, Bayer-Oglesby L, Colvile R, Götschi T, Jantunen M J, Künzli N, Kulinskaya E, Schweizer C, Nieuwenhuijsen M J (2006) Determinants of indoor air concentrations of PM2.5, black smoke and NO2 in six European cities (EXPOLIS study). Atmospheric Environment, 40(7), pp: 1299-1313. ISSN 1352-2310, https://doi.org/10.1016/j.atmosenv.2005.10.030.
- Lancaster K. (1966) A new approach to consumer theory, Journal of Political Economy, 74 (2), pp. 132-157. https://doi.org/10.1086/259131
- Linaker C H, Chauhan A J, Inskip H, Frew A J, Sillence A, Coggon D, Holgate S T (1996) Distribution and determinants of personal exposure to nitrogen dioxide in school children. Occupational and Environmental Medicine, 53(3), pp: 200-203. https://doi.org/10.1136/oem.53.3.200
- Liu S, Su Y (2021) The impact of the COVID-19 pandemic on the demand for density: Evidence from the US housing market. Economics Letters, 207, 110010.
- Liu Y, Tang Y (2021) Epidemic shocks and housing price responses: Evidence from China's urban residential communities. Regional Science and Urban Economics 89, 103695. ISSN 0166-0462. https://doi.org/10.1016/j.regsciurbeco.2021.103695.
- Lucas R E B (1975) Hedonic Price Functions, Economic Inquiry, 13(2), pp.157-178. https://doi.org/10.1111/j.1465-7295.1975.tb00985.x
- Lundberg S.M. and Lee S.I., 2017, A Unified Approach to Interpreting Model Predictions, 31st Conference on Neural Information Processing Systems, available online at https://arxiv.org/pdf/1705.07874
- Mangialardo A, Micelli E, Saccani F (2019). Does Sustainability Affect Real Estate Market Values? Empirical Evidence from the Office Buildings Market in Milan (Italy). Sustainability, 11(12), pp: 1-14. doi:10.3390/su11010012

- Meera A K and Jayakumar M D (1995) Determinants of crime in a developing country: a regression model. Applied Economics, 27(5), pp: 455–460. https://doi.org/10.1080/00036849500000151
- Miller N, Vivek S and Sklarz M (2018) Estimating Property Condition Effect on Residential Property Value: Evidence from U.S. Home Sales Data, Journal of Real Estate Research, 40(2), pp: 179-198. DOI: 10.1080/10835547.2018.12091497
- Molnar G, Savage S J and Sicker D C (2019) High-speed Internet access and housing values, Applied Economics, 51(55), pp: 5923-5936, DOI: 10.1080/00036846.2019.1631443
- Mooya M M (2016) Real Estate Valuation Theory: A Critical Appraisal, Springer-Verlag, Berlin Heidelberg 2016, ISBN 978-3-662-49163-8. DOI 10.1007/978-3-662-49164-5
- Mooya, M M (2022) Property Valuation, Transaction Prices and Market Activity. In: D'Amato, M., Coskun, Y. (eds) Property Valuation and Market Cycle. Springer, Cham. https://doi.org/10.1007/978-3-031-09450-7_17
- Mullainathan S and Spiess J (2017) Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives, 31 (2), pp: 87-106. DOI: 10.1257/jep.31.2.87.
- Nanda A and Ross S L (2012) The Impact of Property Condition Disclosure Laws on Housing Prices: Evidence from an Event Study Using Propensity Scores. J Real Estate Finan Econ 45, 88–109. https://doi.org/10.1007/s11146-009-9206-y
- Norton J (2019) UK property register ’needed urgently’ to stop money laundering. The Guardian, 2019 May 20
- Ozel B, Nathanael RC, Raberto M, Teglio A, Cincotti S (2019) Macroeconomic implications of mortgage loan requirements: an agent-based approach. Journal of Economic Interaction and Coordination, 14, pp: 7-46. https://doi.org/10.1007/s11403-019-00238-5
- Pearl J (1995) Causal diagrams for empirical research. Biometrika, 82(4), pp: 669–688. https://doi.org/10.2307/2337329
- Pearl J, Glymour M and Jewell N P (2016) Causal inference in statistics: A primer. Chichester, England: John Wiley & Sons.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. (2011) Scikit-learn: Machine learning in Python. Journal of machine learning research:12(Oct), pp: 2825–30.
- Porter JR, Shu EG, Amodeo MF, Freeman N, Bauer M, Almufti I, Ackerson M and Mehta J (2022) Commercial Real-Estate at Risk: An Examination of Commercial Building and Economic Impacts in the United States Using a High-Precision Flood Risk Assessment Tool. Front. Water 4:875995. doi: 10.3389/frwa.2022.875995

- Prillwitz J and Barr S (2011) Moving towards sustainability? Mobility styles, attitudes and individual travel behaviour. Journal of Transport Geography, 19(6), pp: 1590-1600. ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2011.06.011.
- Porter M E and van der Linde C (1995) Toward a New Conception of the Environment-Competitiveness Relationship. Journal of Economic Perspectives, 9 (4), pp. 97–118. DOI: 10.1257/jep.9.4.97
- Potrawa T and Tetereva A (2022) How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market, Journal of Business Research, 144, pp.50-65 https://doi.org/10.1016/j.jbusres.2022.01.027
- Punzo G, Panarello D and Castellano R (2022) Sustainable urban mobility: evidence from three developed European countries. Quality & Quantity, 56, 3135–3157. https://doi.org/10.1007/s11135-021-01253-0
- Roffia P, Bucciol A and Hashlamoun S (2023) Determinants of life expectancy at birth: a longitudinal study on OECD countries. International Journal of Health Economics and Management, 23, pp: 189–212. https://doi.org/10.1007/s10754-022-09338-5
- Rosen S (1974) Hedonic prices and implicit markets: Product differentiation in pure competition, Journal of Political Economy, 82(1), pp: 35–55. https://doi.org/10.1086/260169
- Rubin D B (2005) “Causal Inference Using Potential Outcomes: Design, Modeling, Decisions”. In: Journal of the American Statistical Association 100(469), pp: 322–331. https://www.jstor.org/stable/27590541
- Savona EU, Riccardi M (2017) Assessing the risk of money laundering in Europe. Final Report of Project
- IARM. Transcrime–Università Cattolica del Sacro Cuore, Milano. ISBN 978-88-99719-07-4
- Schneider S (2004) Organized crime, money laundering, and the real estate market in Canada. Journal of Property Research, 21(2), pp: 99-118. DOI: 10.1080/0959991042000328801
- Schoenau M and Müller M (2017) What affects our urban travel behavior? A GPS-based evaluation of internal and external determinants of sustainable mobility in Stuttgart (Germany). Transportation Research Part F: Traffic Psychology and Behaviour, 48, pp: 61-73. ISSN 1369-8478, https://doi.org/10.1016/j.trf.2017.05.004.
- Sharma A and Kiciman E (2020) DoWhy: An End-to-End Library for Causal Inference. 2020. https://arxiv.org/abs/2011.04216
- Spapens T, White R, van Uhm D and Huisman W (Eds), Green Crimes and Dirty Money, Routledge, London and New York, ISBN: 978-0-8153-7221-9 (hbk); ISBN: 978-1-351-24574-6 (ebk)

- Splawa-Neyman J, Dabrowska D M, and Speed T P (1990) “On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.” In: Statistical Science 5(4), pp. 465–472. https://www.jstor.org/stable/2245382
- Štrumbelj E, Kononenko I (2014) Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst 41, 647–665 https://doi.org/10.1007/s10115-013-0679-x.
- Temam S, Burte E, Adam M, Antó J M, Basagaña X, Bousquet J, Carsin A-E, Galobardes B, Keidel D , Künzli N, Le Moual N, Sanchez M, Sunyer J, Bono R, Brunekreef B, Heinrich J, de Hoogh K , Jarvis D, Marcon A, Modig L, Nadif R, Nieuwenhuijsen M, Pin I, Siroux V, Stempfelet M, Tsai M-Y, Probst-Hensch N, Jacquemin B (2017) Socioeconomic position and outdoor nitrogen dioxide (NO2) exposure in Western Europe: A multi-city analysis. Environment International, 101, pp: 117-124. ISSN 0160-4120. https://doi.org/10.1016/j.envint.2016.12.026.
- Tyrinopoulos Y and Antoniou C (2013) Factors affecting modal choice in urban mobility. European Transport Research Review 5, pp: 27–39. https://doi.org/10.1007/s12544-012-0088-3
- Wolf D and Irwin N (2024) Is it really bridging the gap? Fiber internet's impact on housing values and homebuyer demographics. Journal of Regional Science, 64, pp: 238–271. https://doi.org/10.1111/jors.12670
- Zhao Y (2020) US housing market during COVID-19: Aggregate and distributional evidence. IMF Working Paper No: WPIEA2020212. ISBN/ISSN: 9781513557816/1018-5941. https://doi.org/10.1016/j.econlet.2021.110010