HEDONIC PRICE MODEL OF COMMERCIAL REAL ESTATE LEASE IN UKRAINE

by

Yevhen Husakov

A thesis submitted in partial fulfillment of the requirements for the degree of

MA in Business and Financial Economics

Kyiv School of Economics

2024

Thesis Supervisor: ______________________________ Professor Elena Besedina

Approved by ____
Head of the KSE Defense Committee, Professor [Type surname, name]

____

____

____

Date ______________________________________

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

OLS Ordinary Least Squares

WLS Weighted Least Squares

GBR Gradient Boosting Regressor

XGBoost Extreme Gradient Boosting

Lasso Least Absolute Shrinkage and Selection Operator

MAE Mean Absolute Error

MSE Mean Squared Error

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

GLOSSARY

Hedonic pricing model – method used to estimate economic values for goods or services by decomposing the studied object into its characteristics and estimating the value and impact of each characteristic.

Regularization – a statistical technique used to prevent overfitting in regression models by introducing additional constraints or penalties to the model parameters.

Lasso regression – a type of regularized regression that applies L1 regularization, shrinking some coefficients to zero.

Ridge regression – a type of regularized regression that applies L2 regularization, which lower the coefficients of the predictors but does not set any to zero.

GBR – a machine learning technique that builds models sequentially by combining multiple weak learners (more often decision trees). Each subsequent model is trained to eliminate errors of the previous ones, resulting in a more accurate overall prediction.

XGBoost regressor – advanced implementation of gradient boosting technique with upgraded regularization to prevent overfitting and support parallel processing.

Hyperparameter optimization – a process of tuning the parameters of the learning process in a machine learning model (e.g., learning rate, number of trees) to find the best configuration that improves the model's performance.

Optuna – an open-source hyperparameter optimization framework that uses different strategies to automate the search for the best hyperparameter settings in machine learning models.

City's Economic Share – a percentage share of a city in the national economic activity, calculated as a ratio of the income tax (from wages) paid in the respective city to the total income tax (from wages) paid in Ukraine.

CHAPTER 1. INTRODUCTION

The commercial real estate market in Ukraine has faced significant changes due to the socio-economic shocks from the COVID-19 pandemic and the ongoing Russian full-scale invasion. These events have notably influenced market dynamics, changing demand and supply mechanics and subsequently affecting lease pricing across various regions of Ukraine. Given the above, this research aims to study the hedonic pricing model of commercial real estate leases, focusing on understanding how various factors contribute to lease pricing in Ukraine.

Thus, the research question is about the main factors influencing the lease prices for commercial real estate in Ukraine and how these factors vary in terms of regional dynamics in the context of the mentioned socio-economic shocks.

Speaking of the applicability of this study, in the first place, the current study may be useful for key players in the Ukrainian lease market. By identifying the variables impacting the lease prices, the research may provide investors and/or real estate managers with a better understanding of the current commercial lease market dynamics in Ukraine. Such investors or managers can use findings of this study to adjust their investment strategies.

The other purpose for which this research may be also useful is a state policy development. The governmental authorities (involved in the lease market regulation) can get from this study valuable information on the commercial real estate market behavior under the named conditions, helping them to make more informed regulations and/or incentives.

Finally, this study can be useful for academic development by including some of the current socio-economic events in the hedonic pricing model, offering a fresh view of the hedonic price model under exceptional war conditions. Furthermore, this research aimed to fill a gap in the existing literature on the Ukrainian commercial real estate market. Over the past few years, a very limited number of studies have been published regarding how the Ukrainian real estate lease market has changed in response to recent shocks like war.

The basis for this study is the work of Rosen (1974), developing the hedonic price theory. The core of this theory is that the value of a property can be viewed as the sum of the values of its individual attributes. The theory has since been applied widely across various markets, demonstrating its applicability in different contexts. In this regard, Adair, Berry, and McGreal (1996) adjusted the hedonic price theory, indicating the significance of submarkets, while Benson et al. (1998) introduced environmental factors in the hedonic price model, such as views from a building, etc.

To cover the respective topic, a dataset of about 10,000 commercial real estate objects from dim.ria.com (real estate marketplace) was received, including different variables such as size of the property, number of floors, location of the property, number of photos, etc. In addition, the distances from the frontline, based on the data from deepstatemap.live, were integrated into the dataset, capturing the effects of war on pricing. Furthermore, the dataset also included the City's Economic Share, i.e., percentage share of a city (where the property is located) in the national economic activity.

After cleaning and filtering the data, the dataset for price modeling contained 9,450 objects. The final dataset includes the property attributes mentioned above. To analyze the respective data this study relies on the approach of combining econometric methods and advanced machine learning techniques. The analysis was based on Ordinary Least Squares (OLS) regression to examine the linear relationships between the dependent variable (price per sqm) and independent variables like property characteristics and location. The White standard errors and Weighted Least Squares (WLS) were additionally applied to address heteroscedasticity by assigning weights to observations, improving the model's performance.

To extend the analysis, machine learning techniques such as Gradient Boosting Regression (GBR) and XGBoost were used to capture the non-linear relationships that OLS and WLS models might not cover. Optuna, an advanced optimization framework, was used to fine-tune the hyperparameters for these ML models. This mix of traditional econometrics and modern machine learning provided a complex assessment of both linear and non-linear factors affecting commercial real estate lease pricing.

The main goals of this research were: (i) to establish a comprehensive list of factors that influence the hedonic price of commercial real estate leases in Ukraine. This includes traditional factors, such as property size, number of floors, location, as well as non-traditional factors like distance to war zone, the impact of different real estate agencies, etc.; (ii) to assess the impact and relative importance of each factor on lease prices through the use of a hedonic pricing model; and (iii) to explore regional effect on the lease prices, particularly focusing on how regions closer to war zones are affected compared to safer regions like Kyiv or Lviv.

The results of this study provide various insights, e.g., that newer and taller buildings are typically associated with higher lease prices. In contrast, leased objects, e.g., priced in the UAH tend to be priced lower than objects priced in USD. The study also reveals regional disparities, with Kyiv and western cities showing relatively higher lease prices per sqm, while regions closer to the east and south show lower lease prices. Furthermore, the study demonstrates that real estate agencies also have a moderate impact on lease prices.

This thesis is structured as follows: Chapter 1 introduces the research questions and background; Chapter 2 reviews industry trends on the commercial lease market in Ukraine and relevant literature on the hedonic pricing model; Chapter 3 describes methodology used for this study; Chapter 4 details the data and provides descriptive statistics thereof; Chapter 5 presents the results; and Chapter 6 provides conclusions, business implications, and recommendations for future research.

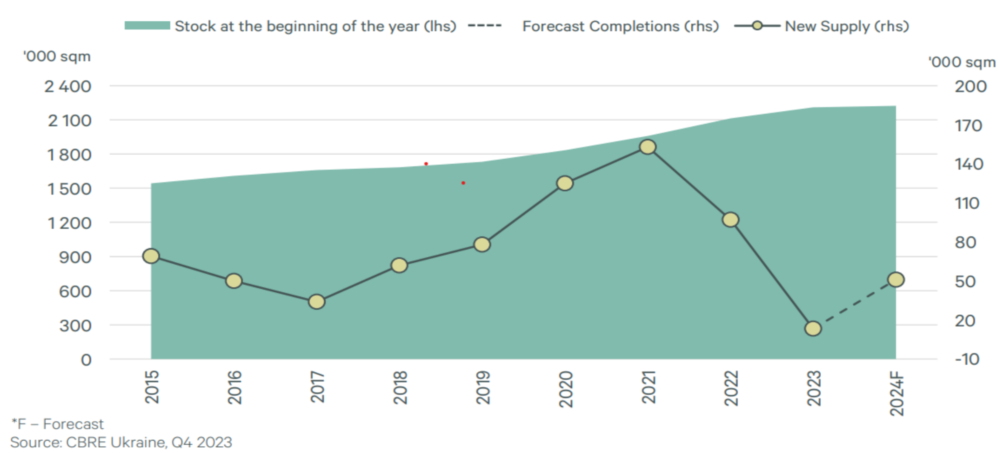
CHAPTER 2. INDUSTRY OVERVIEW AND RELATED STUDIES

2.1. Industry overview

In general, the commercial real estate market in Ukraine, especially the office spaces sector, continues developing despite the ongoing war and economic challenges affecting Kyiv and other regions. This is demonstrated by the continued development activities and the introduction of new office spaces in the market.

The construction of new commercial spaces has not been suspended. For instance, in Kyiv, the newly completed projects added 52,000 sq.m. of office space in 2022. This includes major developments like the Gradient business center and the Unit.City complex. By the end of 2024, an additional 50,000 sq.m. of office space is expected to be completed. These developments may be a strong indicator of market confidence and a driver for future growth, provided that the geopolitical situation will stabilize[1].

Figure 1. Kyiv annual and new supply of lease spaces, as of Q4 2023



[1] CBRE. 2023. "Kyiv Office Market: Resurgence of Occupier Demand Shape an Optimistic Outlook.". https://cbre-expandia.com/wp-content/uploads/2024/02/CBRE_Kyiv-Office-MarketView-H2_2023-ENG-2.pdf

Furthermore, Western regions of Ukraine like Lviv have benefited from its geographical advantage of being far from the frontline, which made it a preferred location for businesses, including major IT firms. On the other hand, cities closer to war zones faced a decrease in market activity due to ongoing security concerns and economic instability. Despite this, the overall market in Ukraine shows signs of partial recovery, although it has not yet returned to pre-war demand levels[2].

The IT sector continues to dominate the demand for office space, shaping the development and leasing strategies in urban centers. Despite IT companies being less interested in office spaces due to the trends for remote or flexible working regime, this sector still significantly influences the market. The changing work formats, including the increase in remote working arrangements, have prompted a shift in the office market dynamics, with companies now prioritizing flexibility and cost-efficiency in their real estate decisions[3].

Rental rates have been adjusted in response to the changing market conditions. In Kyiv, the prime effective rents have stabilized at around USD 20 per sqm per month. This stabilization is likely part of a broader market adaptation to the new economic realities, where companies are negotiating more favorable rental terms and seeking better-quality spaces at competitive prices[4].

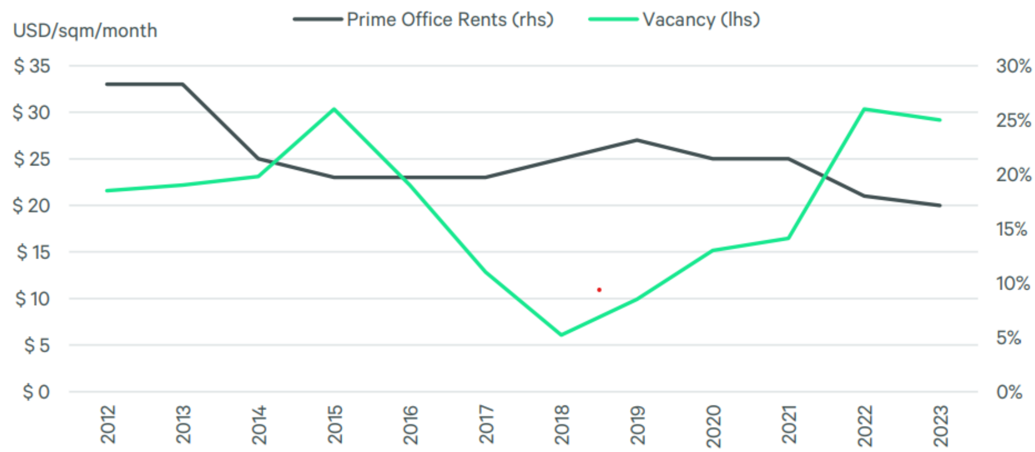
---

[2] UTG. 2023. "Challenges and Prospects of Ukrainian Commercial Real Estate During the War." https://en.interfax.com.ua/news/blog/987908.html

[3] Commercial Real Estate - Review of Office Markets." 2024. Accessed from Delo.ua. https://delo.ua/realty/zalizni-ofisi-comu-rinok-komerciinoyi-neruxomosti-prodovzuje-zrostati-popri-padinnya-stavok-429106/

[4] CBRE. 2023. "Kyiv Office Market: Resurgence of Occupier Demand Shape an Optimistic Outlook.". https://cbre-expandia.com/wp-content/uploads/2024/02/CBRE_Kyiv-Office-MarketView-H2_2023-ENG-2.pdf

Figure 2. Kyiv office prime rents and vacancy, as of Q4 2023



Source: CBRE Ukraine, Q4 2023

In summary, Ukraine's commercial real estate lease market goes through a complex period with significant challenges and changes. The strategic shifts in demand and supply dynamics evidence that the market is adaptive and potentially able for a gradual recovery when conditions stabilize.

2.2. Literature overview

The fundamental work of Rosen (1974) is a basis of the hedonic price theory, detailing how properties are valued based on their distinct attributes. This concept has been widely applied and tested in various markets globally, showing the model's robustness and adaptability to different economic contexts.

Further studies have developed Rosen's work. For instance, Adair, Berry, and McGreal (1996) explored hedonic modeling in the context of housing submarkets, emphasizing the importance of local market conditions and segmentation. This approach has been expanded in recent studies that use geospatial data to more accurately define and analyze submarkets, increasing the precision of hedonic models in real estate valuation.

Benson et al. (1998) examined the valuation of residential characteristics, particularly the value of a view from the building. This work has inspired further research into how natural factors, environmental quality, and urban planning decisions impact property values. Recent studies have used more comprehensive environmental quality indices to explore these effects in greater detail.

Bourassa and Peng (1999) study on the influence of Feng Shui on property prices in Asian markets highlighted cultural factors in hedonic valuations. This has opened up new research possibilities that consider a broader range of cultural and psychological factors affecting property prices, such as historical significance and neighborhood identity.

Clapp and Giaccotto (1998) used a rational expectations approach to model age effects in residential property values. Recent advancements in this area have integrated machine learning techniques to predict changes in property values based on anticipated future developments in neighborhoods, thereby providing more dynamic valuation models.

Goodman (1978) analysis of price indices and housing markets has been crucial for understanding market trends over time. Recent studies have built on this by using real-time data feeds from online real estate platforms to develop more timely and responsive housing market indices that can reflect rapid changes in market conditions.

Halvorsen and Pollakowski (1981) work on the choice of functional form for hedonic price equations has influenced recent studies to employ non-linear models and machine learning algorithms to find nonlinear relationships between property characteristics and prices.

Palmquist (1992) focused on valuing localized external factors using hedonic prices. This has been particularly relevant for recent studies on the impacts of climate change on property values, where researchers have used hedonic pricing to quantify the costs of increased flood risk and other climate-related risks.

Tyrvainen (1997) study on the amenity value of urban forests has inspired recent research into the "green premium" associated with urban green spaces, using more detailed satellite imagery and environmental data to assess their impact on urban property values more precisely.

Recent studies from 2019 to 2024 have broadened the hedonic models' application, integrating modern econometric techniques and innovative data sources to adapt to dynamic market environments. However, it is important to note that the majority of these studies focused on residential real estate, with relatively fewer studies of commercial properties, especially commercial leases.

This is crucial as the factors influencing commercial lease prices can differ significantly from those affecting residential property values. In particular, the features of commercial leases, involving longer lease terms or variable rates based on business performance introduce additional layers of complexity not commonly found in residential market analyses.

CHAPTER 3. METHODOLOGY

In this study a hedonic pricing model was used to analyze how various factors affect commercial real estate lease prices.

The analysis is based on a semi-logarithmic form of the hedonic pricing model to estimate the impact of various characteristics on lease prices for commercial real estate. This model form was chosen because it may effectively capture the percentage changes in the lease prices resulting from unit changes in the specified attributes of properties.

The regression model of hedonic price is typically structured as follows:

$$\ln(Price_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon_i \tag{1}$$

where $X_1, X_2, \ldots X_n$ represent the independent variables influencing the lease prices, like size, location, distance from the war zone, etc.; and

ε_i is the error term.

It is worth mentioning that the relations between dependent and independent variables may not be linear as the real estate objects and their environment might include much more characteristics influencing the real estate lease prices. In this case, machine learning techniques could be applied to enhance the model predictability and determine the most important predictors for the real estate lease price. The particular ML techniques will be described further in this section.

Independent variables used for this study include both internal and external characteristics such as the size (area) of the property, number of floors, number of rooms, quality ratings, whether the building is newly constructed, and the presence of additional features like photos and videos of the property. Location variables like city and district were also included to capture regional impact on pricing. Furthermore, untraditional variables were also introduced in the model, in particular, (1) distance to the frontline, (2) the impact of the real estate agency managing the leased objects, and (3) the City's Economic Share, as defined in the Glossary above.

Given the above, it is anticipated that the size of the leased object will likely have a negative relationship with price per square meter due to the economies of scale effect. New builds are expected to show a positive impact on prices as they often attract higher demand by offering modern amenities. Prime locations and taller buildings should also positively affect lease prices.

The number of photos may be another variable positively affecting prices, as more visual information gives transparency to potential tenants. The distance to the frontline is expected to have a negative impact on prices, i.e., properties farther from war zones may be offered at higher prices. The real estate agency effect may also play a role in influencing lease prices, depending on their reputation and status. In addition, the City's Economic Share in the national economic activity will likely positively impact the lease prices per sqm.

To improve the model's predictive power, handle characteristics selection, and, consequently, address the issues with multicollinearity Lasso and Ridge regressions were applied. Lasso (using L1 regularization) penalizes less important features, shrinking some coefficients to zero, while Ridge regression (using L2 regularization) penalizes the same but does not reduce them to zero. These techniques allowed for simplifying the model by keeping only the most significant predictors, thereby reducing overfitting and enhancing generalization.

To study heteroskedasticity, the residual plot was drawn to assess whether the variance of the residuals is constant across all levels of the fitted values. A pattern in the residual plot revealed the presence of heteroskedasticity suggesting the OLS model's assumptions were violated, potentially leading to inefficient estimates. The issue with residuals will be described in Chapter 5 of this report.

After identifying heteroskedasticity in the OLS model through residual plots, robust standard errors were applied to improve the reliability of the results. In particular, White standard errors were used to eliminate heteroskedasticity without changing the overall model. These adjustments made p-values and confidence intervals more accurate, allowing to improve the statistical significance of the model, which will be described in more detail in Chapter 5 of this report.

As a second approach, Weighted Least Squares (WLS) also were tested to see if it could help to further address the issue. The WLS approach slightly improved the model's fit and made the predictions more stable; however, it did not significantly reduce the heteroskedasticity. The residuals still showed non-random patterns, indicating that simply adjusting weights in a linear model might not be enough to capture the underlying pattern.

The persistence of heteroskedasticity in both approaches suggests that the relationships within the data may be more complex than a linear model can capture. The observed patterns in the residuals hint that a more advanced model might be needed. Thus, it was reasonable to explore non-linear modeling/machine learning techniques, like decision trees, gradient boosting, or neural networks, which are better suited for capturing the non-linear relations in the data.

The first ML technique used was Gradient Boosting Regressor (GBR), a machine learning technique that builds models sequentially (more often in the form of decision trees). Each tree aims to correct errors of the previous to gradually improve the model's performance. The final model constitutes a weighted sum of models, where more weight is given to better models. This method effectively captures complex patterns in the data, provided that the parameters such as learning rate (determining the share of each tree in the final prediction), number of trees (determining the number of boosting rounds), tree depth (determining the model's ability to capture patterns in the data), and subsample ratio (determining the diversity of trees built) are appropriately tuned to avoid overfitting or underfitting.

The study experimented with various configurations of these parameters to find the optimal balance between model complexity and predictive performance. Although GBR showed significant improvements over linear models, capturing more complex interactions among variables, there still was room for improvement by using more advanced techniques.

To increase the model's performance even further, Extreme Gradient Boosting (XGBoost) Regressor was applied as an advanced version of GBR. It optimizes the gradient boosting process by (a) using regularization techniques to prevent overfitting, (b) addressing missing data more effectively, and (c) supporting parallel processing to speed up computation.

Similar to GBR, the key parameters in XGBoost were tuned, i.e., the number of trees, learning rate, maximum depth, subsample ratio, and column sample by tree (which controls the number of features used in each tree). Additionally, XGBoost allows for further customization through L1 and L2 regularization. These parameters were optimized to improve the model's robustness.

XGBoost provided better performance than GBR likely due to its advanced regularization techniques. However, to achieve the best possible model configuration, Optuna, an open-source hyperparameter optimization framework, was employed. Optuna is typically used to automate the search for optimal hyperparameters through Bayesian optimization and other strategies. This approach allowed to tune the parameters for both GBR and XGBoost, ensuring that the model configurations were optimized for maximum performance. Optuna conducted a series of trials, each testing a different combination of hyperparameters, and selected the configuration that showed the best performance based on the R^2 score.

Through this optimization process, hundreds of hyperparameter configurations were explored, ultimately identifying the best-performing set of parameters for both GBR and XGBoost. This tuning process additionally improved the model's predictive power.

After extensive experimentation and hyperparameter tuning, the best-performing model was achieved with the GBR, which captured a substantial amount of variance in the price per sqm. The model reached R^2 score of approximately 0.70, indicating that it explained 70% of the variance in real estate lease prices. This performance proves the effectiveness of ML techniques in capturing the complex relationships between variables of real estate objects.

CHAPTER 4. DATA

4.1. Data sources

The primary data for this analysis was used from the following sources:

Dim.ria.com API[5]. This database contains data regarding commercial lease objects, including:

(a) Property type, i.e., types of commercial real estate.

(b) Geographic details, which include state, city, and district of the property, as well as coordinates of the property.

(c) Property characteristics like the number of rooms, office type, total and useful area, floor number, etc.

(d) Additional attributes, including contractual price flexibility, price currency, etc.

More information on variables can be found in Appendix A to this report.

Deepstatemap.live[6]. The data from this website was used for establishing the geographic coordinates of the frontline, which helped in determining the distance of real estate objects from war zones. I selected an array of the respective coordinates to be the basis for the respective analyses. To calculate the distance from the frontline I compared/subtracted the geographic coordinates of the property from the closest coordinates of the war zone (using my array of coordinates from deepstatemap.live).

[5] Dim Ria API Documentation. https://docs-developers.ria.com/en/dim_ria

[6] Deep State Live Map. https://deepstatemap.live

Top Lead infographic published by Forbes[7]. The data from this infographic was used to determine the City's Economic Share, i.e., a percentage share of a city in the national economic activity, calculated as a ratio of the income tax (from wages) paid in the respective city to the total income tax (from wages) paid in Ukraine. The City's Economic Share was further added to the dataset as an independent variable for each real estate object to establish possible dependencies between the lease prices per sqm and the economic activity of the city, where the corresponding object is located.

4.2. Preliminary data preparation

As mentioned above, for this research, I got data from dim.ria.com API. I downloaded full information regarding the 10,000 most recent commercial real estate objects offered for the lease.

The way I got and cleaned the data was the following:

(a) First, I got 10,000 IDs of real estate commercials from dim.ria.com API by using appropriate GET requests.

(b) After obtaining the respective IDs, I proceeded with downloading the detailed data for each object associated with the respective IDs. The received dataset includes different variables like city, price per square meter, number of floors, number of rooms, quality, photo count, geographic coordinates (latitude and longitude), and other variables listed in Appendix A to this report.

(c) The data cleaning process was carefully designed to ensure that the dataset was accurate, focused, and free from inconsistencies. The first step involved removing

[7] Top Lead. 2024. "Regional Economic Structure and Impact Post-Invasion." https://toplead.com.ua/ru/get_fileinfographic/id/forbes-top-lead-2023-2024.pdf

all non-meaningful columns, such as various IDs, in order to eliminate any 'junk' values that could detract from the analysis.

(d) Additionally, all price-related columns were removed except for the "Price per sqm in USD," which was used as the primary (dependent) metric. This was done to prevent potential issues with multiple pricing formats which might have led to a high correlation between values.

(e) After that, the dataset was refined by removing rows with missing values in the "Price per sqm" column, ensuring that only complete and valid data were included. To further improve the quality of the dataset, outliers in the "Price per sqm" column were filtered out by removing the top 1% and bottom 1% of values. This step was crucial to avoid skewing the analysis with extreme values.

(f) The dataset was also filtered to exclude rows where 'city_id' was greater than 30, ensuring that only relevant geographical locations (big regional cities) were included.

(g) Finally, specific columns such as 'secondaryUtp', 'youtube_link', 'metro_station_name_uk', and 'user_newbuild_name_uk' were transformed into boolean values (0 or 1). This transformation was based on whether the content in these columns contained more than three characters, simplifying these fields for further analysis.

4.3. Descriptive statistics

After cleaning the data, I also removed the 1st and 99th percentiles for all variables to mitigate the influence of extreme outliers. This was done only for the descriptive statistics but not for the analysis (where the 'Price per sqm' was the only column with removed 1st and 99th percentiles).

After completing the data manipulations, I obtained data from about 9,450 real estate objects for modeling and 7,226 real estate objects for descriptive statistic purposes. Among these 7,226 real estate objects, 4,763 objects fell into the office premises category, 1,782 objects were categorized as commercial premises, and the remaining 681 were special premises.

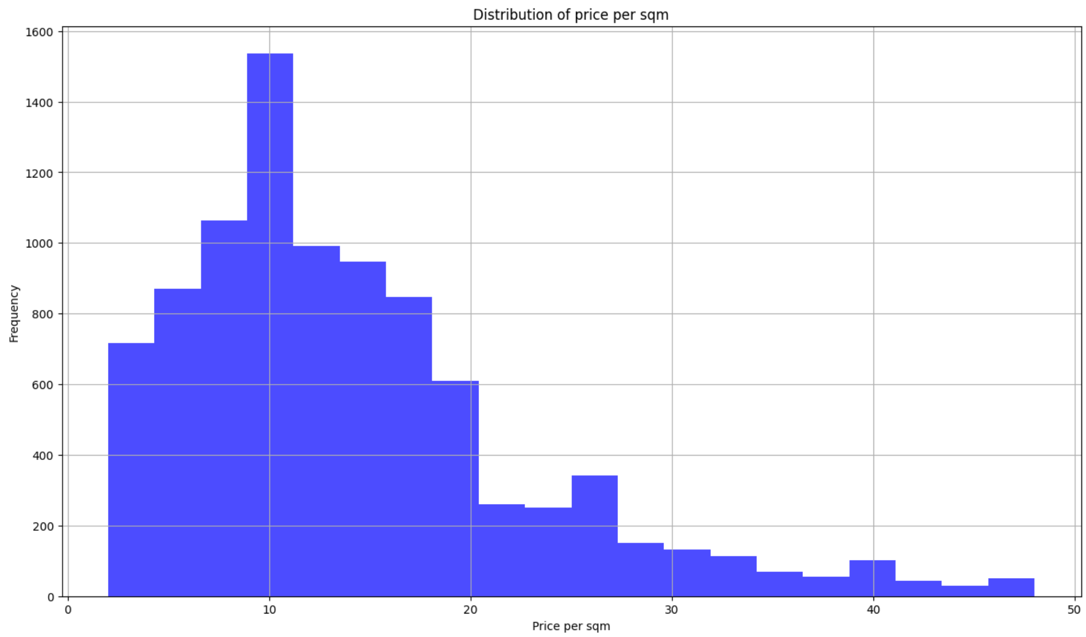
After that, using Python tools, I computed descriptive statistics for the key variables, as presented in Table 1 below.

Table 1. Descriptive statistics

	Count	**Mean**	**Std**	**Min**	**25%**	**Median**	**75%**	**Max**
Distance to frontline	7,226	320.78	136.46	35.07	284.99	286.58	289.91	915.87
Price per sqm	7,226	14.15	8.71	2.00	8.00	12.00	18.00	48.00
Total area (sqm)	7,226	328.31	358.51	15.00	100.00	201.00	415.00	2,492.00
Quality	7,226	80.69	11.74	42.00	75.00	84.00	90.00	97.00
Number of rooms	7,226	3.92	3.57	1.00	1.00	3.00	5.00	21.00
Number of floors	7,226	3.27	3.99	0.00	1.00	2.00	4.00	26.00
Number of photos	7,226	15.88	10.33	1.00	10.00	13.00	20.00	127.00

As provided in the descriptive statistics, the average lease price per square meter is about USD 14.15. At the same time, more than 75% of the objects are actually priced below USD 18.00. Table 1 shows a wide spread of prices with a standard deviation of USD 8.71, indicating that prices are widely distributed. Figure 3 below suggests a better view of the prices distribution evidencing that most of the prices are skewed to the left and are generally less than USD 20 per square meter.

Figure 3. Distribution of price per sqm

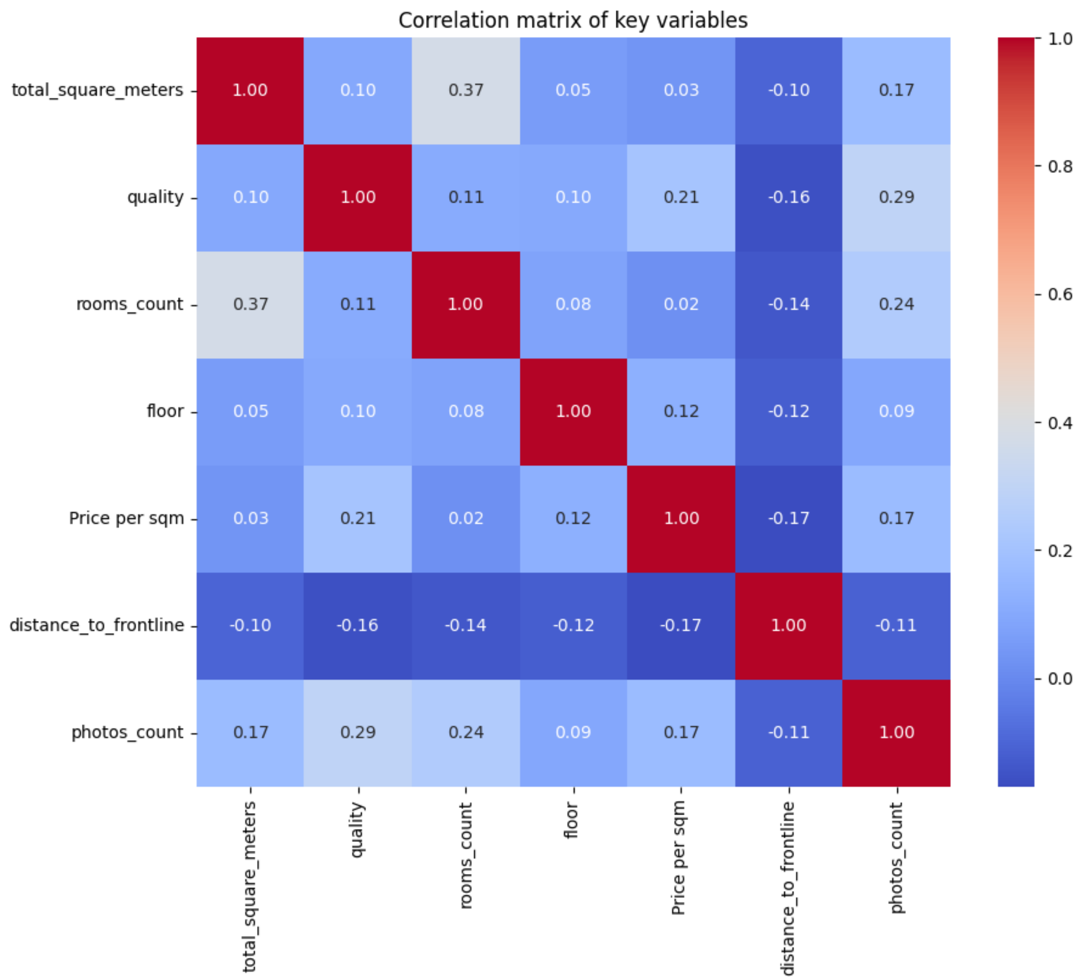


The received dataset also provides other insights. In particular, it shows a wide range in property sizes, from small units of 15 sqm to large properties of nearly 2,500 sqm, with an average of around 328 sqm, indicating a quite diverse market. In addition, quality ratings for the leased objects, computed by dim.ria based on their own methodology, are generally high, showing (on average) over 80 points out of 100, with a large part rated even higher, evidencing that leased objects are relatively well-maintained.

Most leased objects have around three to four rooms, though some have up to 21 rooms, probably in large business centers. Buildings, in which the leased objects are placed, typically have from one to four floors, with a few exceptions with up to 26 floors. The average number of photos per property is 16 photos.

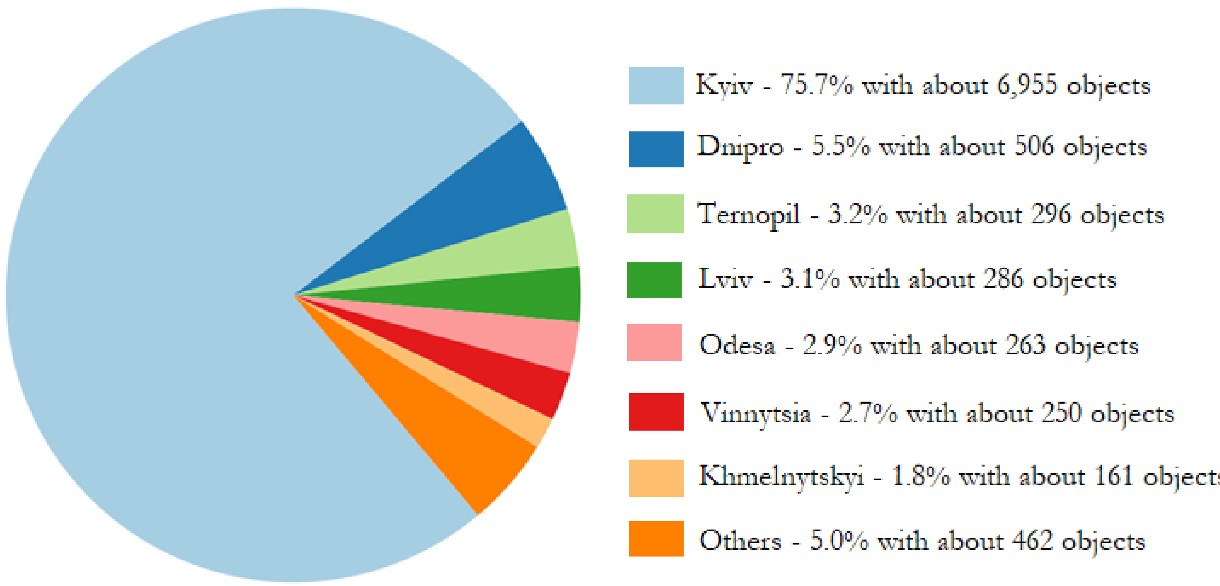
In addition to the above statistics, Figure 4 below shows the correlation matrix of the key variables to better understand pair relations between the respective variables.

Figure 4. Correlation matrix of key variables



Furthermore, I also checked the geographical distribution of the real estate offered for lease. The distribution of real estate by region shows that Kyivska oblast leads with 6,955 objects (which is more than 75% of the whole set of objects). Following Kyivska, the Dnipropetrovsk oblast has 506 objects, the Ternopilska oblast has 296 objects, and the Lviv oblast has 286 objects. This distribution highlights the dominance of Kyivska Oblast in the commercial real estate lease market, with other regions showing significantly fewer commercial offers. More insights are in Figure 5 below.

Figure 5. Distribution of real estate objects by city



CHAPTER 5. RESULTS

5.1. Linear model results

Looking at the coefficients after application of the linear regression model (OLS and WLS), several variables demonstrate statistically significant impacts on the log lease price per sqm.

The variable indicating if the property is newly built (coefficient = 0.0549 with $p < 0.001$) shows a positive relationship with the dependent variable (price per sqm). From economic perspective we have strong evidence that newly constructed properties tend to have 5.5% higher prices per square meter than old ones, probably due to the premium placed on newer constructions. Similarly, the number of floors in the building (coefficient = 0.0752 with $p < 0.001$) is also positively associated with higher prices, indicating that each floor can add an additional 7.5% to the lease price per sqm while holding other factors constant.

The analysis also reveals that the location of a property also matters in determining its price per sqm. For instance, properties located in Kyiv are associated with almost 15% higher prices per sqm (coefficient = 0.1474, $p < 0.001$), likely due to the status of capital city, being economic and administrative centre of Ukraine. Western cities like Lviv also show a positive effect on prices (coefficient = 0.1024, $p < 0.001$), adding 10% to the price per sqm, while holding other factors constant.

On the other hand, southern cities like Mykolaiv (coefficient = -0.1792, $p < 0.001$) and Odesa (coefficient = -0.1542, $p < 0.001$) show negative effects on prices per sqm decreasing them by 17.9% and 15.4% respectively. The same is for eastern cities like Dnipro (coefficient = -0.0525, $p = 0.033$) and Kharkiv (coefficient = -0.0475, $p = 0.041$), showing the negative effect of -5.3% and -4.75% on the price per sqm, again while holding other factors constant.

Nevertheless, the distance to the frontline does not show a significant impact on the property prices (coefficient = -0.0120, $p = 0.218$), suggesting that other factors might be more critical in determining property values in these regions. The same is true for the City's Economic Share – the regression analysis shows a p-value of 0.068, meaning that the impact of the City's Economic Share is not statistically significant at the level of 0.05. In other words, the impact of the City's Economic Share on the lease prices per sqm is likely negligible for this regression analysis.

At the same time, realtor agency effects appeared to have a statistically significant influence on the lease prices per sqm, though this influence is moderate. This is evidenced by the respective coefficients for most real estate agencies. For instance, properties listed by a particular agency, such as 'City Realty' (coefficient = 0,0412 with $p < 0.001$), tend to be priced higher. In this case, the price per sqm for objects offered by the respective agency tends to be 4% higher possibly due to the agency's reputation or operational practices, like personal fees of realtors. This finding aligns with the observation that more institutional and established agencies generally have a positive impact on pricing.

The other interesting finding is an impact of currency. Thus, properties priced in UAH tend to decrease price by 7.5% (coefficient = -0.0753, $p < 0.001$), while those priced in USD tend to increase price by 7% (coefficient = 0.0709, $p < 0.001$). This may be due to the lower stability of the local currency compared to the USD.

The other OLS and WLS results summarized in tables, containing the top significant variables by T, are attached in Appendix B and Appendix C to this report (respectively).

5.2. ML model results

The feature importance analysis in ML models of GBR and XGBoost was crucial for identifying the most influential variables and improving the model's predictions. Feature importance assigns a score to each input (independent) variable based on its contribution to the prediction accuracy of the model. In tree-based models, it is achieved by measuring how much each feature reduces the error in the trees of the ensemble. Features with higher importance scores are those that the model relies on more heavily to make accurate predictions.

The analysis of feature importance in the GBR model shows several significant factors influencing property prices. The most important feature was the currency type (namely UAH), with an importance score of 0.3274. This result may confirm the significant currency impact on lease prices, indicating that properties priced in UAH are particularly sensitive to the factors driving the model.

Another critical variable was the distance to the frontline, with an importance score of 0.1571. This finding highlights the role of distance to war zones in determining property prices. Additionally, the size of the property was another significant factor with an importance score of 0.1036, emphasizing that this was also an essential factor for the model.

Other notable features included the number of floors in the building with an importance score of 0.0520, the location in Kyiv with an importance score of 0.0409, and the specific floor on which the property is located with an importance score of 0.0277. Further features such as the number of photos (importance score of 0.0257) also showed relevance, though to a lesser extent. The quality of the property, whether it is a new build, and the number of rooms were also influential, though with lower importance scores of 0.0187, 0.0170, and 0.0167 respectively.

In the end of the list, the influence of certain agencies and the regional significance of cities like Lviv also have an impact on the price with an importance score of approx. 0.004. You can find importance scores of all significant variables for GBR and XGBoost models in Appendix D to this report.

It worth mentioning that these machine learning results are typically challenging to interpret from an economic perspective due to the complexity and non-linear nature of the model.

However, they provide valuable insights into the relative importance of specific features, which can be used to validate and reinforce the findings from the linear regression model. By identifying which variables most strongly influence lease prices per sqm, such as currency type, location, etc., the ML model mostly confirms the key factors observed in the linear analysis, offering an additional layer of support for the linear model's results.

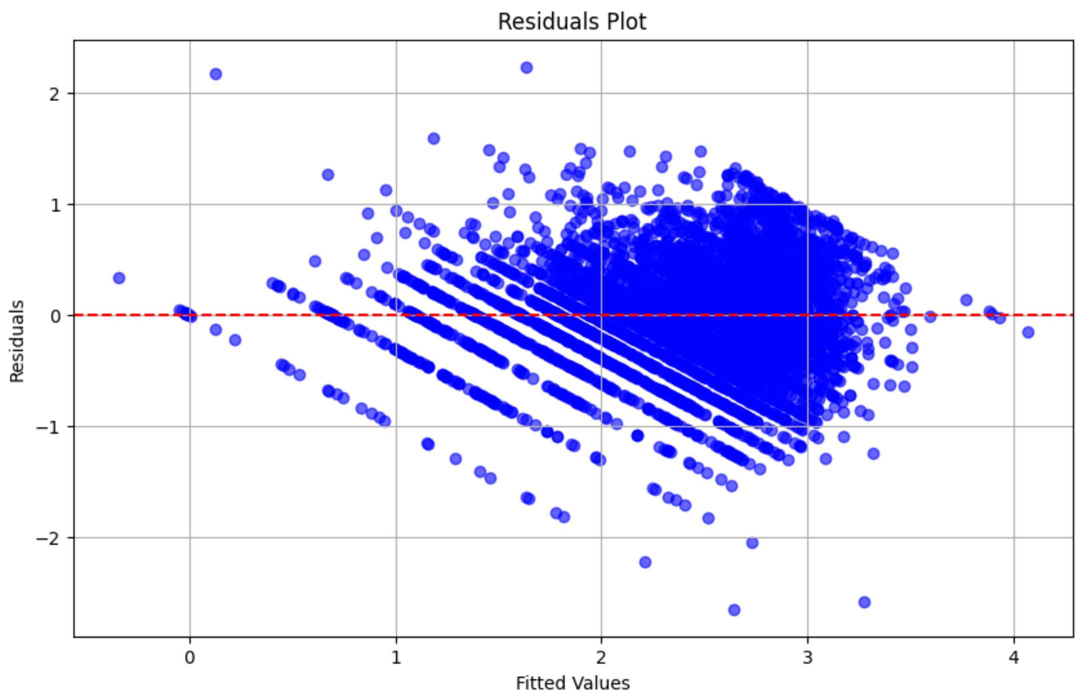
5.3. Justification of the results based on the econometric indicators

The OLS model achieved an R-squared value of 0.556, meaning that 55.6% of the variability [in log price per sqm] can be explained by the independent variables included in the respective OLS model. The model's F-statistic of 11.76, with a corresponding p-value of less than 0.001, confirms that the regression model (in general) is statistically significant. This means that at least one of the predictors is significantly related to the dependent variable (log price per sqm), rejecting the null hypothesis that all regression coefficients are equal to zero. The other econometric indicators describing this linear model are listed in Appendix E to this report.

Lasso regression technique, used to address potential multicollinearity and overfitting, produced MAE of 0.3863, MSE of 0.2445, and R-squared value of 0.538, while Ridge regression had MAE of 0.3854, MSE of 0.2498, and R-squared of 0.529. As we can see, the results suggest that the Lasso and Ridge models explain even less variance in the dependent variable compared to the OLS model, indicating that while Lasso or Ridge may reduce overfitting, it might also not capture some important variables. At the same time, both Lasso and Ridge significantly improved F-statistic (27.57) and decreased AIC and BIC of the model evidencing reduce in the overfitting and model complexity. The other econometric indicators describing this linear model are listed in Appendix E to this report.

Despite relatively solid econometric results for OLS the residual plot for OLS, shown in Figure 6 below, reveals several issues with the OLS regression model, indicating that it may not capture the full pattern and complexity of the data.

Figure 6. Residual plot for OLS regression

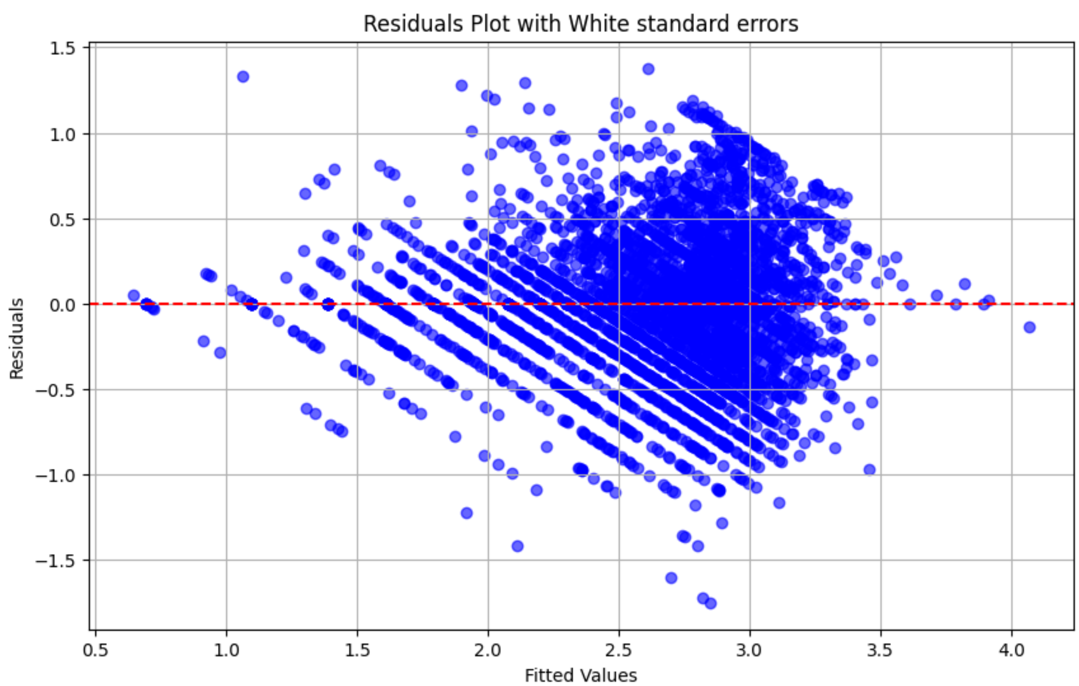


The first problem with this plot is a non-random pattern of the residuals showing a clear picture where the residuals form several downward-sloping lines. This pattern likely suggests that the relationships between variables may be non-linear.

Another issue is the spread of residuals. In a well-fitting model, the spread of residuals should be constant at all levels of fitted values. At the same time, the more focused spread in the residuals can indicate the non-constant variance of the errors, violating one of the fundamental assumptions of OLS regression. In other words, this can be evidence that the model's predictions may be less accurate for higher values of the dependent variable, reducing the overall reliability of the model.

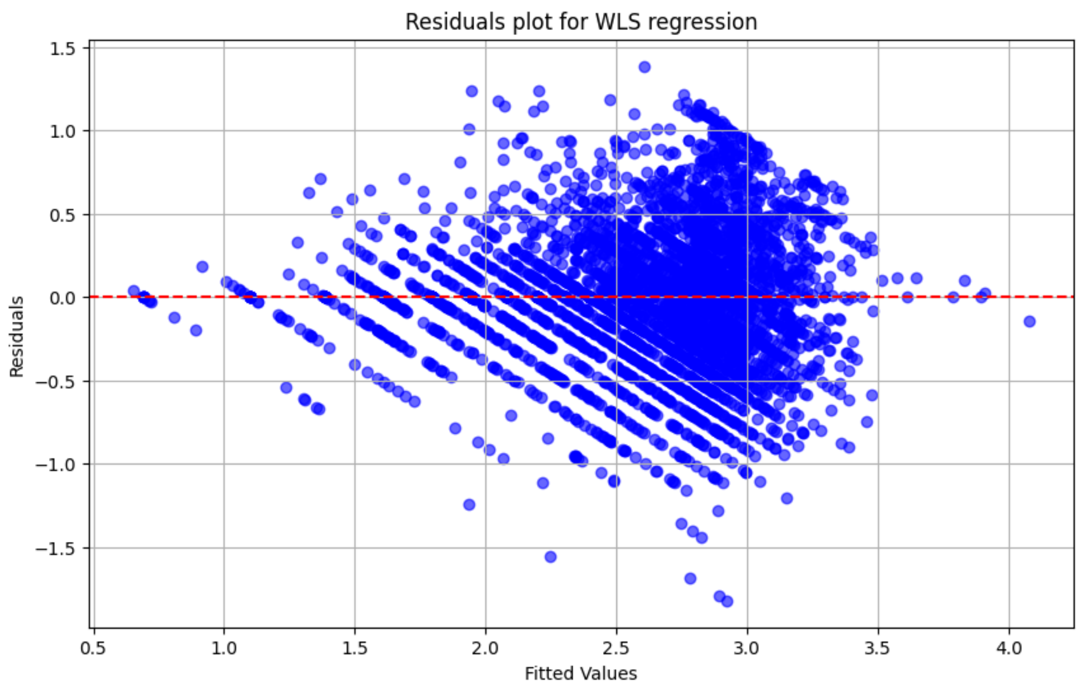
Given the issues with heteroskedasticity identified in the residuals of the OLS model, White standard errors were used to address them. The idea of this method is to adjust the standard errors and confidence intervals to be more reliable without changing the model's structure. As a result of implementation of White standard errors, the model showed even less AIC and BIC values of 5271 and 8103 respectively and the highest F-statistic value of 2089 among other models, indicating improvements in fit and simplification of the model. The other econometric indicators describing this linear model are listed in Appendix E to this report.

Figure 7. Residual plot with White standard errors



At the same time, even after applying White standard errors the non-random patterns remain visible, as it is shown in Figure 7 above. As an alternative, WLS regression was tested to address heteroskedasticity. WLS regression demonstrated a significant improvement in the model's explanatory power, with an R-squared value of 0.937 and an adjusted R-squared of 0.930. The F-statistic value of 135.6 further confirms that the model (in general) is statistically significant. The other econometric indicators describing this linear model are listed in Appendix E to this report. However, the residual plot after applying WLS, as shown in Figure 8 below, almost did not change the residual pattern compared to the previous residual plots.

Figure 8. Residual plot for WLS regression



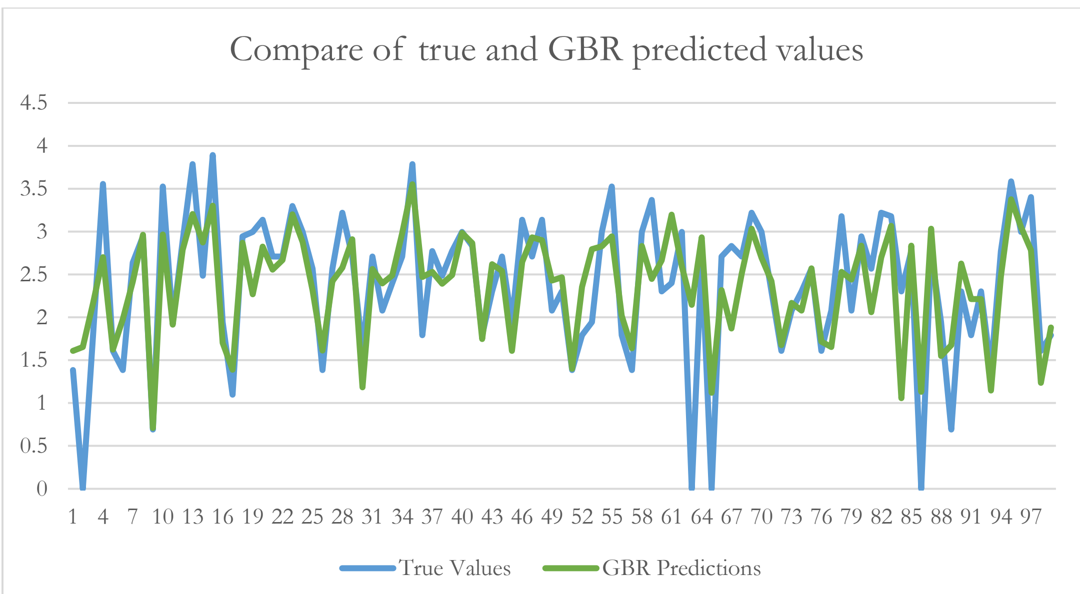
Given the above, the econometric results of linear models were not fully satisfying. Therefore, it was decided to improve the model by using the advanced ML techniques allowing it to capture non-linear dependencies between the price per sqm and other variables.

The machine learning models applied in this analysis, specifically the GBR regressor and XGBoost regressor, provided valuable insights into predicting the log price per sqm. These models were optimized using advanced techniques such as Optuna to enhance their predictive performance.

The initial application of the GBR to the dataset resulted in MAE of 0.3652, MSE of 0.2164, and R-squared score of 0.5261. These results indicate that the model explained approximately 52.6% of the variance in the dependent variable, reflecting moderate predictive power. Recognizing the room for improvement, the model was further improved through hyperparameter tuning.

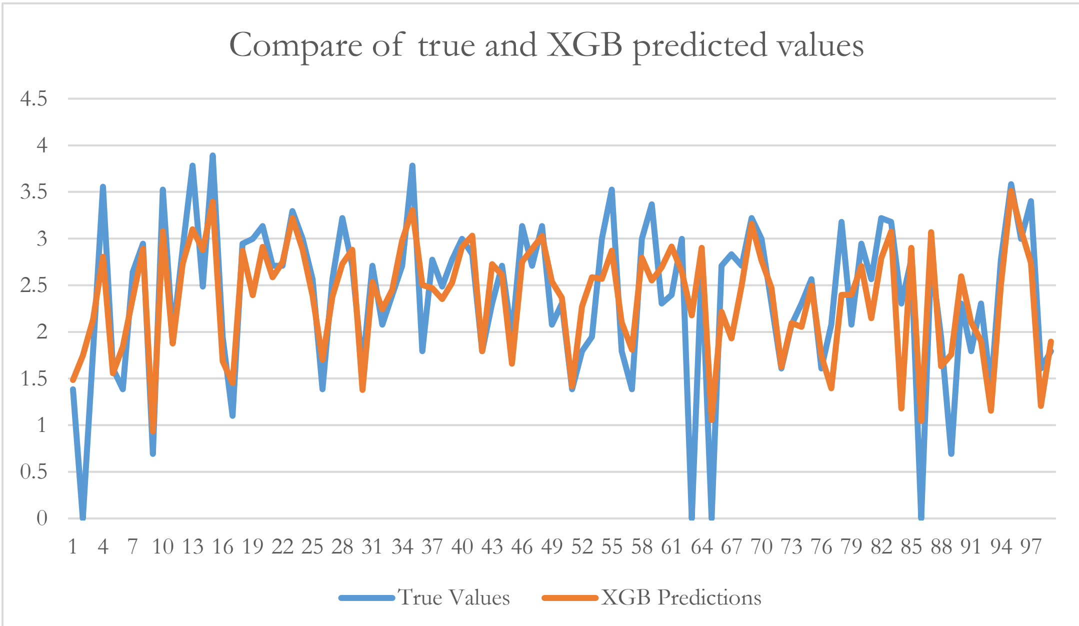
Initial optimization was completed through Bayesian optimization, which resulted in improved model performance. The GBR model, after Bayesian optimization, achieved MAE of 0.2832, MSE of 0.1477, and an R-squared of 0.6765. But most importantly the model better catches the pattern of non-linear relations between variables, as you can see in Figure 9 below, where the true and GBR predicted values are compared.

Figure 9. Comparative graph of true and GBR predicted values



Following the GBR, the XGBoost regressor was applied to the dataset, being initially optimized with GridSearchCV. The resulting model shows MAE of 0.3213, MSE of 0.1728, and R-squared of 0.6217. This model explained 62.2% of the variance in the price per sqm, slightly outperforming the initial GBR model but underperforming the GBR after Bayesian optimization. Again, the results of the comparison between true and predicted values are reflected in Figure 10 below.

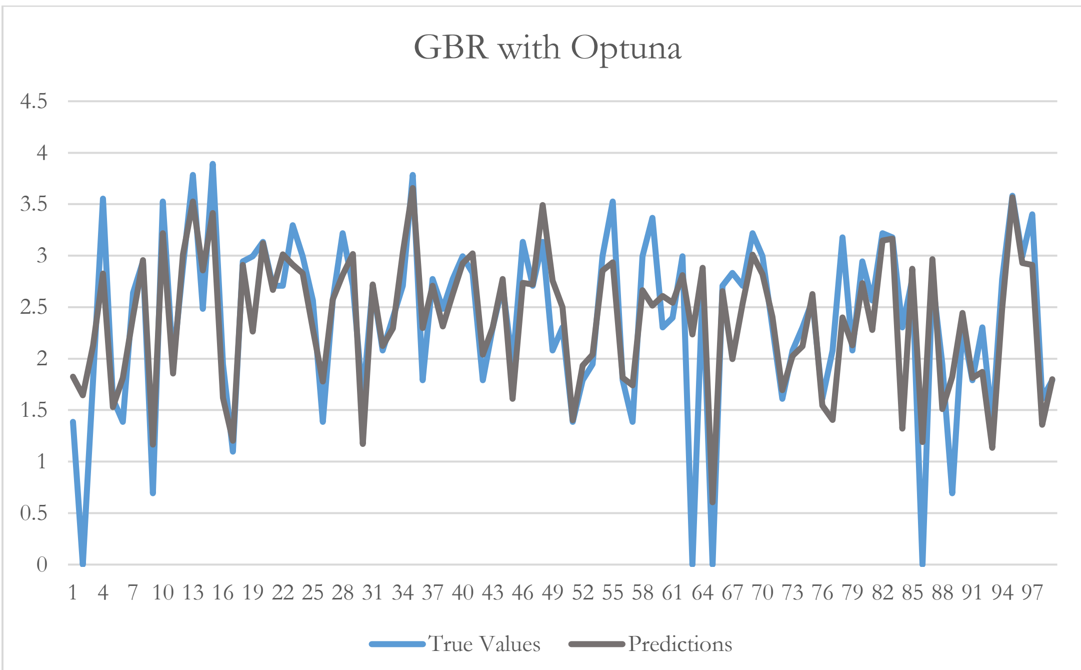
Figure 10. Comparative graph of true and XGB predicted values



However, the most significant enhancement in model performance was achieved by applying Optuna, a more advanced optimization framework, to GBR. Optuna allowed for a more comprehensive search of the hyperparameter space, leading to the best results, with MAE of 0.2638, MSE 0.1374, and an R-squared of 0.6992, i.e., a model with almost 70% explanatory power and the lowest deviations in the errors. These improvements demonstrate Optuna's effectiveness in identifying the optimal hyperparameters, enabling the model to suggest more accurate predictions.

The parameters for the GBR model optimized by Optuna reflect a solid approach to capture non-linear relationships in real estate data. Specifically, the model uses 840 trees, which indicates a high number of boosting rounds, allowing the model to learn specific patterns in the data over multiple iterations. The learning rate was 0.0249 (relatively low), ensuring that the model makes gradual adjustments, which helps to avoid overfitting. The max depth of 15 provides that each decision tree is allowed to grow deep, capturing detailed relations between the variables. Finally, a subsample rate was determined as 0.6487, meaning that only about 65% of the data is used in each boosting round, which introduces randomness and helps in improving the model's generalization. The results of the comparison between true and predicted values for GBR with Optuna are reflected in Figure 11 below.

Figure 11. Comparative graph of true and GBR predicted values with Optuna



CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

This study aimed to analyse the factors influencing real estate lease prices per square meter in Ukraine, with a particular focus on understanding the impact of various property characteristics both traditional and non-traditional (like distance to the frontline or City's Economic Share). The analysis employed a range of statistical and machine learning models, including OLS regression (with Lasso and Ridge regularization), WLS regression, GBR, and XGBoost. The findings offer valuable insights for stakeholders in the real estate market, including investors, developers, and real estate agencies.

6.1. Summary of results

The linear regression analysis revealed that several key factors significantly influence the price per square meter of real estate. Newer buildings were associated with higher prices, likely reflecting the premium buyers are willing to pay for modern infrastructure. The analysis also showed that properties in taller buildings generally have higher prices for lease, likely due to better views and 'prestigious' status.

Realtor agencies were found to play a moderate role in pricing, i.e., well-established agencies may positively influence property lease prices likely due to their reputations. Furthermore, lease objects priced in UAH were generally less expensive than those priced in USD, likely reflecting currency-related risks.

Regional differences have also a significant impact on the lease prices, i.e., the capital and western cities show a positive impact on the pricing while south and eastern cities show a negative impact. These findings highlight the importance of regional economic conditions and demand dynamics in commercial real estate lease prices.

The machine learning models, particularly those optimized using techniques like Optuna, reinforced these findings and provided more nuanced insights. For instance, feature importance analysis in the GBR and XGBoost models confirmed the critical role of currency type, property size, and regional location in determining commercial real estate lease prices.

6.2. Implications for business

The results of this study have several important implications for stakeholders in the real estate lease market, in particular:

Realtor agencies. Institutional and well-established realtor agencies should improve their reputations to keep or increase their pricing strategies. Agencies with less influence may need to focus on building trust and demonstrating value to offer higher lease prices. Additionally, agencies should consider the impact of their fees on overall pricing strategies.

Property developers. Developers should take into account the finding that newer buildings and taller structures are generally associated with higher prices. This suggests that investments in modern construction and high-rise buildings may result in better returns, particularly in regions where demand for such properties is high.

Regional impact. Investors and developers should pay close attention to regional dynamics. Understanding regional economic conditions, infrastructure development, and local demand trends will be crucial in making informed decisions.

Currency considerations. The study highlights the importance of currency in real estate lease pricing. For prices in local currency UAH, stakeholders should be aware of the potential for lower pricing.

6.3. Recommendations

Based on the findings, the following recommendations are proposed for stakeholders in the Ukrainian real estate lease market.

Enhance realtor agency credibility: Realtor agencies should invest in building their brand and credibility, as these factors may influence property pricing. This could include marketing strategies that emphasize the agency's reputation, successful transactions, and customer satisfaction.

Focus on newer and taller developments. Developers should prioritize projects involving new constructions and higher buildings (instead of buying old ones), as these are associated with higher prices. This strategy is particularly relevant in urban areas with a demand for modern living spaces.

Regional impact. The regional impact of particular cities should be carefully considered when developing real estate strategies, as they prices may vary from city to city as shown in the current research.

6.4. Suggestions for future work

This study provides a foundational analysis of factors driving real estate prices in Ukraine, but there are several areas where further research could enhance understanding and provide more details:

Broader geographic scope. Including other cities and regions (i.e., not only regional centres) in the analyses could provide a more comprehensive picture of the Ukrainian real estate lease market. In addition, one can consider excluding Kyiv from the analysis or focusing on Kyiv market only as Kyiv apparently has more developed and complex economic than other cities in Ukraine.

Incorporating macroeconomic variables. Future research could incorporate macroeconomic variables such as inflation and interest rates, GDP, and similar factors to better understand their influence on real estate lease prices and to refine the predictive models.

In conclusion, this study has provided valuable insights into the factors influencing real estate lease prices in Ukraine and offers practical recommendations for stakeholders in the market. By applying these findings, businesses can make more informed decisions, optimize their strategies, and better navigate the complexities of the Ukrainian real estate market. Future research in this area may continue to focus on these findings, providing deeper insights and more robust predictive models.

REFERENCES

Adair, A., Berry, J., and McGreal, W. 1996. "Hedonic Modeling, Housing Submarkets, and Residential Valuation." *Journal of Property Research* 13: 67-83.

Benson, E. D., Hansen, J. L., Schwartz, A. L., and Smersh, G. T. 1998. "Pricing Residential Amenities: The Value of a View." *Journal of Real Estate Finance and Economics* 16 (1): 55-73.

Bourassa, S. C., and Peng, V. S. 1999. "Hedonic Prices and House Numbers: The Influence of Feng Shui." *International Real Estate Review* 2 (1): 79-93.

CBRE. 2023. "Kyiv Office Market: Resurgence of Occupier Demand Shape an Optimistic Outlook." Accessed February 2024. https://cbre-expandia.com/wp-content/uploads/2024/02/CBRE_Kyiv-Office-MarketView-H2_2023-ENG-2.pdf

Clapp, J. M., and Giaccotto, C. 1998. "Residential Hedonic Models: A Rational Expectations Approach to Age Effects." *Journal of Urban Economics* 44: 415-437.

Goodman, A. C. 1978. "Hedonic Prices, Price Indices, and Housing Markets." *Journal of Urban Economics* 5: 471-482.

Halvorsen, R., and Pollakowski, H. O. 1981. "Choice of Functional Form for Hedonic Price Equations." *Journal of Urban Economics* 10: 37-49.

"Iron offices: why the commercial real estate market continues to grow despite falling rates" 2024. Accessed from Delo.ua. https://delo.ua/realty/zalizni-ofisi-comu-rinok-komerciinoyi-neruxomosti-prodovzuje-zrostati-popri-padinnya-stavok-429106/

Palmquist, R. B. 1992. "Valuing Localized Externalities." *Journal of Urban Economics* 31: 59-68.

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

Tyrvainen, L. 1997. "The Amenity Value of the Urban Forest: An Application of the Hedonic Pricing Method." *Landscape and Urban Planning* 37: 211-222.

UTG. 2023. "Challenges and Prospects of Ukrainian Commercial Real Estate During the War." https://en.interfax.com.ua/news/blog/987908.html

APPENDIX A.

VARIABLES FROM DOM.RIA.COM API

Table 2. Variables from Dom.Ria.com API

Field Identifier	**English Description**	**Type**
category	Type of Object	Number
realty_type	Type of Realty	Number
operation_type	Type of Operation	Number
state_id	Region	Number
city_id	City	Number[]
district_id	District	Number[]
characteristic[210][from] - characteristic[210][to]	Number of Rooms	Number[]
characteristic[214][from] - characteristic[214][to]	Total Area	Number[]
characteristic[217][from] - characteristic[217][to]	Usable Area	Number[]
characteristic[227][from] - characteristic[227][to]	Floor	Number[]
characteristic[228][from] - characteristic[228][to]	Number of Floors	Number[]
characteristic[235][from] - characteristic[235][to]	Price per Month	Number[]
characteristic[1011]	Price Negotiable	Number
characteristic[254]	Additional Payments	Number
characteristic[273]	Possible Negotiation	Number
characteristic[242]	Type of Currency	Number
characteristic[247]	Price per (object/sqm.)	Number
characteristic[1437]	Type of Offer	Number
characteristic[274]	Possible Installment	Number
with_map	Only with Map	Number
with_video	Only with Video	Number

Table 2 continues on the next page

Field Identifier	English Description	Type
with_photo	Only with Photo	Number
photos_count_from	Only with Photos	Number
urgent_only	Only Top Listings	Number
banks_only	Pledged Property	Number
secondary	Secondary Housing	Number
newbuildings	New Buildings	Number
date_from - date_to	Date of Submission	Number
page	Page	Number

APPENDIX B.

OLS MOST SIGNIFICANT VARIABLES BY T

Table 3. OLS most significant variables by t

Variable	Coef.	Std.Err.	t	P>|t|
const	2.449471	0.005879	416.657500	0.000000
floors_count	0.061444	0.008205	7.489051	0.000000
district_name_uk_Дніпровський	-0.052610	0.007647	-6.879110	0.000000
total_square_meters	-0.041770	0.006403	-6.522630	0.000000
district_name_uk_Святошинський	-0.043680	0.006727	-6.493180	0.000000
is_new_build	0.045445	0.007169	6.339476	0.000000
agency.name_City Realty	0.039985	0.007130	5.608073	0.000000
district_name_uk_Солом'янський	-0.045970	0.008290	-5.545740	0.000000
agency.name_Space Realty	0.036833	0.006900	5.338194	0.000000
district_name_uk_Печерськ	0.042486	0.008048	5.279032	0.000000
district_name_uk_Бортничі	-0.030320	0.005988	-5.063910	0.000000
district_name_uk_Печерський	0.052556	0.010477	5.016499	0.000001
agency.name_Primes Real Estate	0.069073	0.013917	4.963234	0.000001
district_name_uk_Солом'янка	-0.032160	0.006536	-4.920900	0.000001
district_name_uk_ДВРЗ	-0.024430	0.005907	-4.135870	0.000036
agency.name_ Dominanta	0.027143	0.006643	4.086045	0.000045
agency.name_T.H.E. CAPITAL	0.038930	0.009951	3.912057	0.000093
district_name_uk_Дарницький	-0.030110	0.007731	-3.895250	0.000099
agency.name_Місто Мрій	-0.023780	0.006152	-3.864770	0.000112
agency.name_Маяк	0.043140	0.011479	3.758175	0.000173
rooms_count	-0.025020	0.006668	-3.751600	0.000177
district_name_uk_Оболонський	-0.028170	0.007656	-3.679720	0.000236
agency.name_Аверс	-0.045180	0.013250	-3.410120	0.000654
photos_count	0.023199	0.006832	3.395539	0.000689
district_name_uk_Голосіївський	0.032318	0.009793	3.300101	0.000972

APPENDIX C.
WLS MOST SIGNIFICANT VARIABLES BY T

Table 4. WLS most significant variables by t

Variable	Coef.	Std.Err.	t	P>|t|
const	2.32960	0.00786	296.49890	0.00000
district_name_uk_Голосіїв	-0.32146	0.00510	-62.99060	0.00000
city_name_uk_Миколаїв	-0.17923	0.00632	-28.34060	0.00000
agency.name_ОВИЛ	-0.04525	0.00186	-24.29310	0.00000
district_name_uk_Ковпаковський	-0.00983	0.00042	-23.37890	0.00000
agency.name_АН Realterra	-0.04419	0.00204	-21.66620	0.00000
district_name_uk_Личаків	0.00000	0.00000	19.74794	0.00000
agency.name_ INVEST DREAM	0.00000	0.00000	18.50058	0.00000
district_name_uk_Шулявка	-0.11326	0.00642	-17.65150	0.00000
district_name_uk_Вишгородський Масив	0.00000	0.00000	-15.88380	0.00000
district_name_uk_Західний	0.00000	0.00000	-14.91380	0.00000
district_name_uk_Лиманський	0.00000	0.00000	-13.40890	0.00000
district_name_uk_Чечелівський	-0.04883	0.00385	-12.69270	0.00000
agency.name_ АН "М2"	0.00000	0.00000	-12.47930	0.00000
district_name_uk_Самарський	-0.02683	0.00233	-11.50110	0.00000
agency.name_Malinka Real Estate Lviv	-0.07837	0.00686	-11.42230	0.00000
agency.name_Центр нерухомості "М2"	0.00000	0.00000	-11.03630	0.00000
district_name_uk_Голоско	0.00000	0.00000	10.38152	0.00000
agency.name_Місто Мрій	-0.02220	0.00223	-9.93595	0.00000
district_name_uk_Багринова гора	0.00000	0.00000	-9.67378	0.00000
currency_type_id_UAH	-0.07528	0.00814	-9.24974	0.00000
currency_type_id_USD	0.07094	0.00803	8.83572	0.00000
agency.name_Василь	0.00000	0.00000	-8.77545	0.00000
agency.name_Маркіян	0.00000	0.00000	-8.22701	0.00000

Table 4 continues on the next page

Variable	Coef.	Std.Err.	t	P>|t|
floors_count	0.07518	0.00967	7.77428	0.00000
agency.name_Nataly	-0.01722	0.00234	-7.37520	0.00000
district_name_uk_Гніванське шосе	0.00000	0.00000	7.08506	0.00000
district_name_uk_Інгульський	0.01696	0.00244	6.95034	0.00000
district_name_uk_Академмістечко	-0.03914	0.00570	-6.86887	0.00000
agency.name_АН "Квартал"	-0.02295	0.00361	-6.35308	0.00000
district_name_uk_Майзлі	0.00000	0.00000	-6.28271	0.00000
agency.name_Маклер	0.00000	0.00000	-6.23578	0.00000
district_name_uk_Богунський	0.00000	0.00000	-6.22168	0.00000
district_name_uk_Центральний	0.07618	0.01239	6.14880	0.00000
city_name_uk_Київ	0.14737	0.02426	6.07433	0.00000
agency.name_Твій Дім	-0.02190	0.00361	-6.06473	0.00000
district_name_uk_Центр	0.08286	0.01373	6.03737	0.00000
district_name_uk_Суворовський	-0.03285	0.00545	-6.03339	0.00000
district_name_uk_Корабельний	0.04366	0.00753	5.79619	0.00000
agency.name_БЦ "Святошин"	0.00000	0.00000	-5.74584	0.00000
is_new_build	0.05491	0.00989	5.55534	0.00000

APPENDIX D.
GBR AND XGB MOST IMPORTANT VARIABLES

Table 5. GBR and XGB most important variables

GBR most important variables		XGB most important variables	
Variable	**Importance**	**Variable**	**Importance**
currency_type_id_UAH	0.32737	currency_type_id_UAH	0.22518
distance_to_frontline	0.15707	currency_type_id_USD	0.11426
total_square_meters	0.10365	city_name_uk_Київ	0.05113
floors_count	0.05203	is_new_build	0.01315
city_name_uk_Київ	0.04086	city_name_uk_Харків	0.01033
floor	0.02771	district_name_uk_Вишенька	0.00928
photos_count	0.02569	city_name_uk_Львів	0.00926
complete_time	0.01926	distance_to_frontline	0.00817
quality	0.01870	district_name_uk_Оболонський	0.00805
is_new_build	0.01699	agency.name_Трофей	0.00718
rooms_count	0.01674	city_name_uk_Дніпро	0.00693
district_name_uk_Центр	0.00716	district_name_uk_Голосіївський	0.00637
district_name_uk_Оболонський	0.00578	district_name_uk_Промисловий	0.00634
district_name_uk_Голосіївський	0.00456	agency.name_Ольга	0.00631
city_name_uk_Львів	0.00427	city_name_uk_Тернопіль	0.00630
realty_type_name_uk_Спеціальне приміщення	0.00415	district_name_uk_Центр	0.00619
district_name_uk_Шевченківський	0.00368	city_name_uk_Вінниця	0.00611
isBinotel	0.00351	district_name_uk_Дніпровський	0.00583
district_name_uk_Печерський	0.00330	city_name_uk_Миколаїв	0.00534
agency.name_ОВИЛ	0.00323	agency.name_Оселя	0.00529
is_show_building_no	0.00301	district_name_uk_Тяжилів	0.00509
agency.name_Rork Realty Invest	0.00298	agency.name_Аверс	0.00499
agency.name_Аверс	0.00272	agency.name_Rork Realty Invest	0.00488

Table 5 continues on the next page

GBR most important variables		XGB most important variables	
Variable	**Importance**	**Variable**	**Importance**
district_name_uk_Дніпровський	0.00266	floors_count	0.00483
district_name_uk_Печерськ	0.00254	agency.name_Primes Real Estate	0.00454
city_name_uk_Тернопіль	0.00213	agency.name_АН Европейское	0.00447
district_name_uk_Вишенька	0.00202	agency.name_Топаз	0.00436
district_name_uk_Промисловий	0.00194	total_square_meters	0.00427
have_additional_features	0.00192	district_name_uk_Чечелівський	0.00425
agency.name_AH Realterra	0.00181	agency.name_Space Realty	0.00423
agency.name_Primes Real Estate	0.00173	agency.name_BIG	0.00413
city_name_uk_Дніпро	0.00169	district_name_uk_Голосіїв	0.00411
district_name_uk_Академмістечко	0.00167	agency.name_ОВИЛ	0.00401
city_name_uk_Вінниця	0.00167	district_name_uk_Печерськ	0.00396
district_name_uk_Суворовський	0.00158	agency.name_AH Concept Estate	0.00394
have_youtube_link	0.00150	realty_type_name_uk_Спеціальне приміщення	0.00393
district_name_uk_Чечелівський	0.00149	city_name_uk_Черкаси	0.00384
agency.name_Space Realty	0.00148	floor	0.00383
district_name_uk_Тяжилів	0.00148	district_name_uk_Корабельний	0.00362
agency.name_CITY SALE	0.00148	district_name_uk_Дарницький	0.00362
agency.name_INVEST	0.00144	district_name_uk_Галицький	0.00360
have_metro	0.00143	agency.name_Diamond Estate	0.00349

APPENDIX E.

ECONOMETRIC RESULT OF THE LINEAR MODELS

Table 6. OLS regression key metrics

Metric	Value
Dep. Variable:	log_price_per_sqm
Model:	OLS
R-squared:	0.556
Adj. R-squared:	0.508
F-statistic:	11.46
AIC:	9,674
Df Residuals:	5,981
BIC:	14,130
Df Model:	654

Table 7. OLS regression key metrics after Lasso

Metric	Value
Dep. Variable:	log_price_per_sqm
Model:	OLS
R-squared:	0.538
Adj. R-squared:	0.519
F-statistic:	27.57
AIC:	9,169
Df Residuals:	6,366
BIC:	11,000
Df Model:	269

Table 8. OLS regression key metrics with White standard errors

Metric	Value
Dep. Variable	log_price_per_sqm
Model	OLS
R-squared	0.535
Adj. R-squared	0.480
F-statistic	2089
AIC	5271
Df Residuals	3718
BIC	8103
Df Model	446

Table 9. WLS regression key metrics

Metric	Value
Dep. Variable:	log_price_per_sqm
Model:	WLS
R-squared:	0.937
Adj. R-squared:	0.930
F-statistic:	135.6
AIC:	12,840
Df Residuals:	5,981
BIC:	17,300
Df Model:	654