Impact of Built Environment Factors on House Prices in London Boroughs

Module: CASA0007

Burhan Ahmad Wani

Centre for Advanced Spatial Analysis (CASA)

University College London (UCL)

1. Introduction
2. Literature Review
3. Data
4. Methodology
5. Results & Analysis
6. Conclusion

1 Introduction

2 Literature Review

3 Data

4 Methodology

5 Results & Analysis

6 Conclusion

Background

- The built environment, encompassing various physical features of urban spaces such as residential buildings, commercial areas, transportation infrastructure, and green spaces, plays a crucial role in shaping the dynamics of urban housing markets Glaeser & Kahn (2004). In metropolitan regions like London, where real estate values are a significant concern for both residents and policymakers, understanding the impact of these environmental factors on house prices is essential.

- These built environment factors influence a city's aesthetic and functional aspects, thus its economic value.

Aim of Study

- This study aims to explore the relationship between different elements of the built environment and house prices in the boroughs of London. By analyzing the housing & built environment data, we seek to identify which factors most significantly influence housing values and provide insights that could inform urban planning and policy-making. The findings of this research have the potential to guide decisions that enhance the livability and economic value of urban spaces, ensuring that growth and development align with the needs of the population.

Research Question

'Which built environment factors most significantly influence house prices in London boroughs?'

This is the central question guiding our research and it will drive our data analysis and interpretation as we seek to uncover the most influential built environment factors contributing to house prices. Also, with our null hypothesis being:

"None of the built environment factors have a statistically significant impact on house prices in London boroughs."

1 Introduction

2 Literature Review

3 Data

4 Methodology

5 Results & Analysis

6 Conclusion

- The connection between the built environment and house prices has been thoroughly explored in urban economics and real estate studies.
- Cordera et al. (2019) demonstrated that roads and railways improve accessibility, which significantly increases demand for nearby properties, as observed in a comparison of two Spanish cities.
- Conway et al. (2010) found that proximity to commercial areas is positively associated with higher property values due to the convenience and economic activity they offer.
- The effect of green spaces on house prices is mixed: Lee et al. (2023) showed a positive impact, while Chen et al. (2023) highlighted a more complex relationship that depends on the urban context and distance from green spaces.
- Our study builds on this body of work by focusing on London boroughs. We employ advanced statistical methods to provide a clearer understanding of how different aspects of the built environment influence house prices in London Boroughs.

1 Introduction

2 Literature Review

3 Data

4 Methodology

5 Results & Analysis

6 Conclusion

Data Description

Our analysis is based on two key datasets:[1]

Built Environment Data	**House Price Data**
This includes the distribution of land use types across London wards, covering a percentage of domestic buildings, a percentage of gardens, a percentage of roads, and a percentage of green spaces in the respective wards.	We used house price data from the Land Registry, focusing on 2006, which provides a comprehensive view of the property market in that period, as the built environment data available is from 2006.
Source: Land Use by Borough - London Datastore *(n.d.)*	*Source:* Average House Prices by Borough - London Datastore *(n.d.)*

Table 1: Main Datasets used in Study

[1] Data & Analysis of Code available [here]

Data Processing

- To prepare the data for analysis, the built environment data was grouped into boroughs, and mean values of all the independent variables were obtained at borough level. This was done as our house price data is available at the borough level. Also, the latest built environment data available is from 2006, so we will use the house price data of 2006. Further, we cleaned the data which included converting string data type to float for house price. Also, for the current dataset, with independent variables as percentages and within the same scale, normalization is not required. Further, the variables were renamed for ease of interpretation. Also, missing values were checked & the current study was not having missing values in our interested columns. Further, the merging of datasets was done based on borough names, for carrying out our analysis.

Key Variables in the Final Dataset

Variable	Description	Units
house_price	Average house prices in London boroughs	Pounds(£)
domestic	Percentage of domestic buildings in the borough	Percentage (%)
gardens	Percentage of garden space in the borough	Percentage (%)
commercial	Percentage of commercial land use in the borough	Percentage (%)
road	Percentage of road space in the borough	Percentage (%)
rail	Percentage of rail infrastructure in the borough	Percentage (%)
path	Percentage of pedestrian paths in the borough	Percentage (%)
greenspace	Percentage of green spaces in the borough	Percentage (%)
water	Percentage of water bodies in the borough	Percentage (%)
other_areas	Percentage of other land use areas in the borough	Percentage (%)

Table 2: Key Variables in the Final Dataset

1 Introduction

2 Literature Review

3 Data

4 Methodology

5 Results & Analysis

6 Conclusion

Methods

Descriptive Statistics

- To summarize the data.

Correlation Analysis

1. To identify significant relationships between variables.

Regression Analysis

1. To determine the impact of each built environment factor on house prices.

Variance Inflation Factor (VIF)

1. To check for multicollinearity among variables.

1 Introduction

2 Literature Review

3 Data

4 Methodology

5 Results & Analysis

6 Conclusion

Descriptive Statistics - Built Environment

- The following table summarizes the statistical information of independent variables for various built environment factors. All variables are in percentage.

Variable	Mean (%)	Min (%)	Max (%)	Std. Dev. (%)
domestic	11.87	6.99	20.17	3.19
gardens	25.33	0.10	38.86	8.28
commercial	7.58	2.33	37.50	6.51
road	16.46	10.21	26.05	4.28
rail	1.36	0.20	3.03	0.80
path	0.94	0.40	1.45	0.28
greenspace	25.34	4.80	44.24	9.40
water	2.29	0.14	13.30	2.66
other areas	8.83	4.66	18.29	2.99

Table 3: Statistics - Independent Variables

Interpretation - Built Environment Data

- Mean Values:
  - On average in London boroughs in 2006, the percentage of domestic buildings was around 11.87%, and for gardens it was 25.33%, and for commercial building the percentage was around 7.57%.
- Standard Deviation:
  - Built Environment Variables while showed a high standard deviation in London Boroughs included garden space around (8.28%) and commercial space (6.51%) also show high variability.
- Minimum and Maximum Values:
  - The maximum values for the domestic, gardens, and commercial spaces was around 20.17%, 38.86%, and 37.5%, respectively while minimum was around 6.995%, 0.100%, 2.332% respectively.

Descriptive Statistics - House Prices

- The following table summarizes the statistical information of the dependent variable, i.e., house price:

Variable	**Mean (£)**	**Min (£)**	**Max (£)**	**Std. Dev. (£)**
house price	261658	167998	525000	71424

Table 4: Statistics - Dependent Variable

Interpretation - House Prices Data

- Mean Values:
  - On average, house prices in London boroughs were around £261,658 in 2006.
- Standard Deviation:
  - House prices have a high standard deviation of £71,424, indicating considerable variation across boroughs.
- Minimum and Maximum Values:
  - The minimum house price is £167,998, and the maximum is £525,000.

Distribution of Dependent Variables

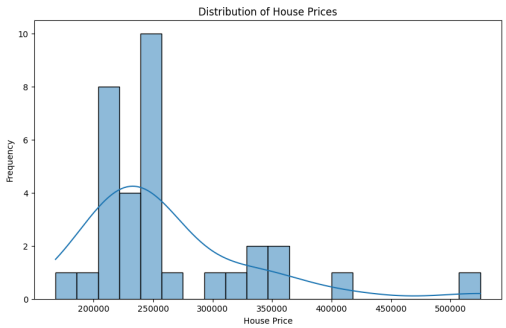

Figure 1: House Prices

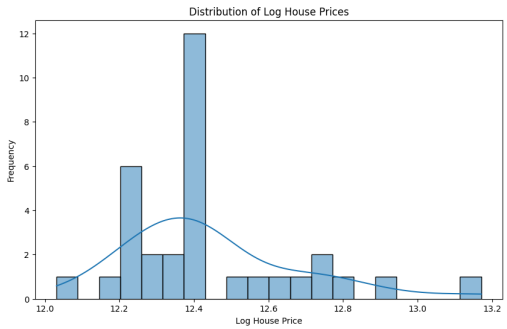

Figure 2: Log House Prices

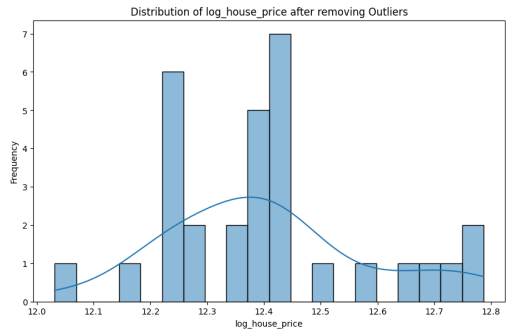

Figure 3: Log House Prices After Removing Outliers

Explanation: The distribution of the dependent variable, 'house_price', was initially skewed. To enhance the reliability of our regression model, the variable was logarithmically transformed to achieve a more normal distribution. Furthermore, outliers were removed from the logged data, further refining the dataset to prevent potential distortions in the analysis and ensure more robust, interpretable results.

Correlation Analysis

- **Positive Correlations:**
  - **Road and Commercial (0.76):** A strong positive correlation indicates that areas with more commercial development tend to have more road infrastructure.
  - **Road and Other Areas (0.78):** A strong positive correlation indicates that areas with more road infrastructure also have more land categorized under "other areas."
- **Negative Correlations:**
  - **Greenspace and Roads (-0.93):** There is a powerful negative correlation between greenspace and road infrastructure, suggesting that areas with more greenspace have significantly less road coverage.
  - **Gardens and Commercial (-0.84):** Areas with more gardens tend to have less commercial space, showing a trade-off between these land uses.
  - **Gardens and Road (-0.71):** A strong negative correlation shows that areas with more garden space generally have less road infrastructure.

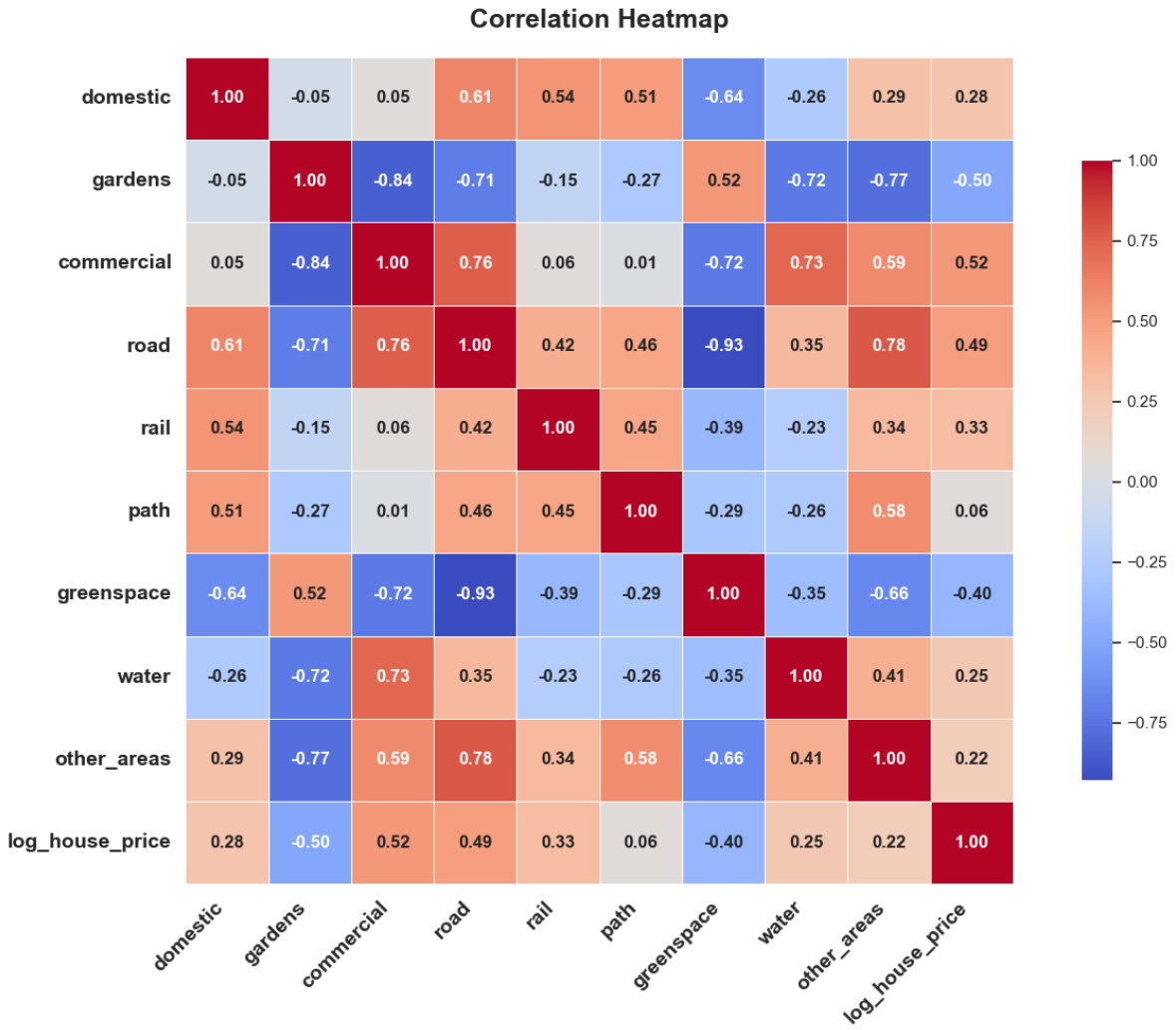

Figure 4: Correlation Matrix of Built Environment Factors and House Prices

Visualizing Relationships- Dependent Vs Independent

- These plots illustrate the relationships between each built environment factor and log house prices in London boroughs.
- The blue line represents the regression line, and the shaded area indicates the confidence interval.

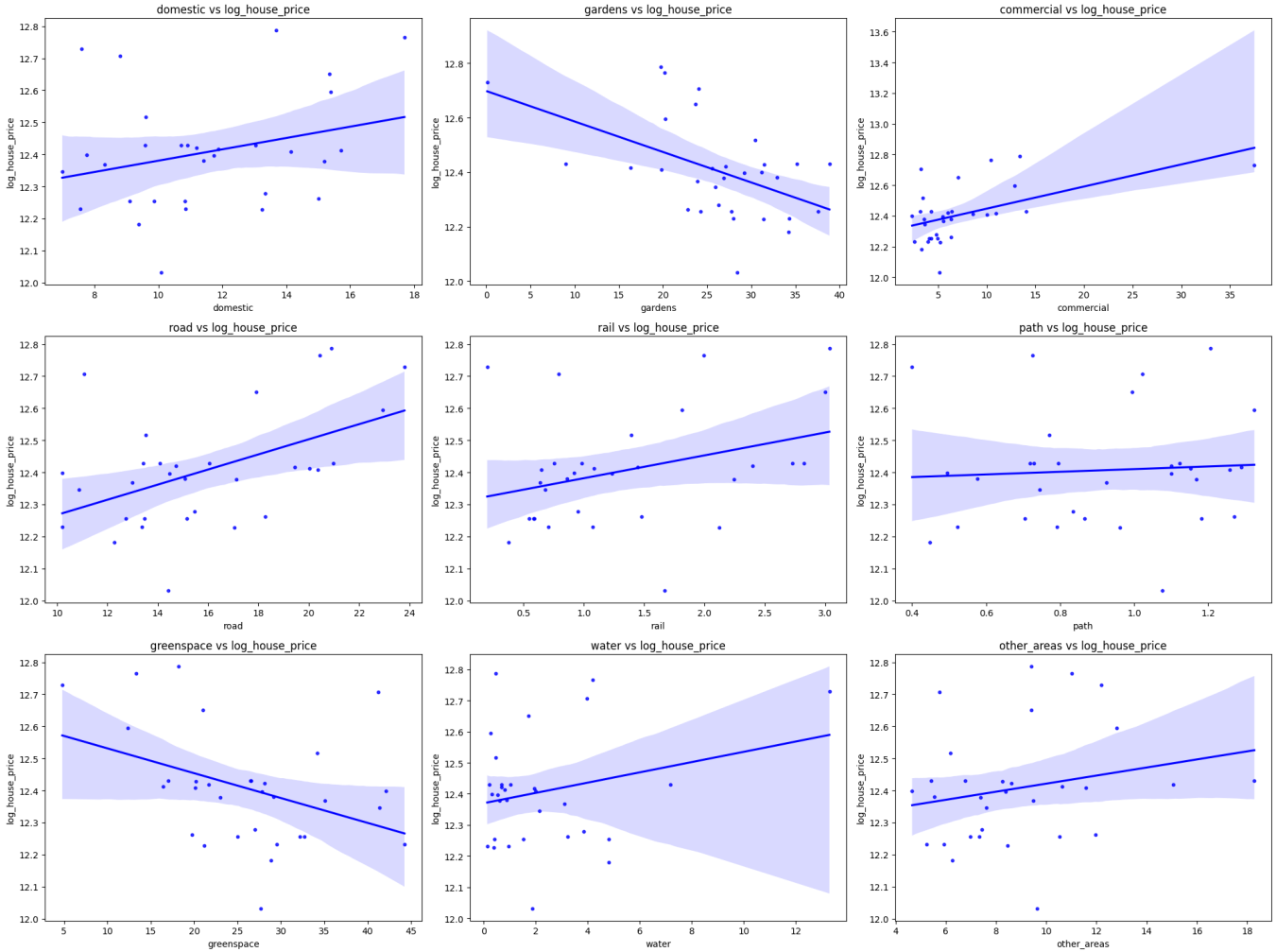

Figure 5: Built Environment Factors vs. Log House Prices

Regression Analysis

- The regression model was used to quantify the impact of built environment factors on house prices.
- Model summary: $R^2 = 0.60$, Adj. $R^2 = 0.43$, Prob. < 0.05
- Since the p-value is very small (0.00991), we can reject the null hypothesis. This suggests that, collectively, the built environment factors do significantly affect house prices in London boroughs.
- But, the Cond. No (1.31e+05) is very large, thus might indicate strong multicollinearity. So, we would proceed with VIF for removing the variables having strong multicollinearity .

Variance Inflation Factor (VIF)

- VIF analysis was conducted to check for multicollinearity among variables.
- Variables with high VIF values (e.g., Greenspace, Roads & Gardens) were removed to improve the model's robustness.

Run	VIF $>$ 10	5 $<$ VIF $<$ 10	VIF $<$ 5	Variable Dropped
run-1	*domestic, gardens, commercial, road, rail, path, greenspace, water & other_areas*	-	-	-
run-2	*commercial, road*	*domestic, gardens, water & other_areas*	*rail, path*	greenspace
run-3	-	*gardens*	*domestic,commercial, water,rail,path & other_areas*	road
run-4	-	-	*domestic, commercial, rail, path, water & other_areas*	gardens

Table 5: VIF Analysis Results

Refit Regression after VIF

- After removing variables with high VIF, the regression model was refit.
- Model summary: The model explains 44% of the variance in log house prices $R^2 = 0.44$, which suggests a moderate fit. The Adj. $R^2 = 0.30$ indicates that the model can explain a good portion of the variability in house prices even after adjusting for the number of predictors. Model Equation:

$$\begin{aligned}\text{log\_house\_price} = {} & 12.2149 + 0.0102 \cdot \text{domestic} \\ & + 0.0200 \cdot \text{commercial} + 0.0665 \cdot \text{rail} \\ & + 0.0617 \cdot \text{path} + 0.0020 \cdot \text{water} \\ & - 0.0249 \cdot \text{other\_areas} + \epsilon\end{aligned}$$

- The condition number has been significantly reduced from 1.31e+05 to 119, indicating that multicollinearity has been addressed effectively by removing highly collinear variables. This makes the model more robust and reliable.

Final Regression Table

Dep. Variable: log house price

Variable	**Coef**	**Std Err**	**t**	**P>|t|**	**[0.025, 0.975]**	
const	12.2149	0.143	85.289	0.000	[11.919, 12.510]	
domestic	0.0102	0.013	0.803	0.430	[-0.016, 0.036]	
commercial	0.0200	0.007	2.748	0.011	[0.005, 0.035]	
rail	0.0665	0.043	1.558	0.132	[-0.022, 0.155]	
path	0.0617	0.175	0.352	0.728	[-0.300, 0.423]	
water	0.0020	0.019	0.106	0.917	[-0.037, 0.041]	
other_areas	-0.0249	0.017	-1.439	0.163	[-0.061, 0.011]	

Table 6: Final OLS Regression after VIF

Interpretation - Final Regression Table

- **Significance of Variables:**
  - Commercial: The variable commercial has a statistically significant and positive impact on house prices, indicating that an increase in commercial areas is likely to lead to higher house prices, holding other factors constant which were also confirmed by Conway et al. (2010) in their study. Also, in our study - For each unit increase in commercial percentage, the log of house price increases by 0.0200, holding other factors constant.

- **Insignificant Variables:**
  - Domestic, Rail, Path, Water, Other Areas: These variables have p-values greater than 0.05, indicating that they are not statistically significant predictors of house prices in this model. This means changes in these factors do not have a statistically significant impact on house prices within the context of this model after removing multicollinearity.

QQ Plot

- The QQ Plot was used to check the normality of the residuals, supporting the assumptions of the linear regression model.

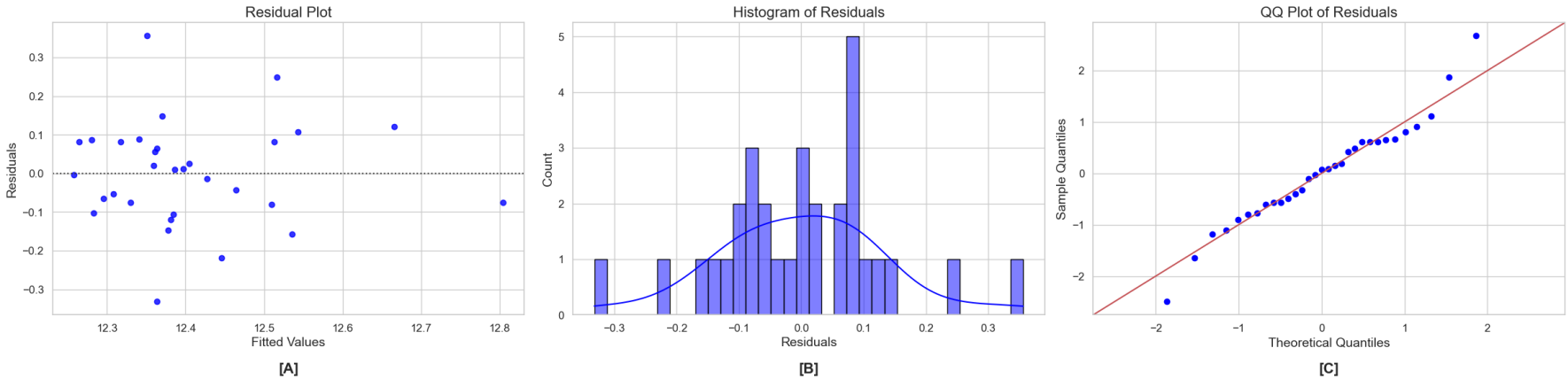

Figure 6: QQ Plot of Residuals

1 Introduction

2 Literature Review

3 Data

4 Methodology

5 Results & Analysis

6 Conclusion

Answering the Research Question

The analysis reveals that commercial areas have a statistically significant and positive impact on house prices in London boroughs, underscoring their role in urban economic vitality. However, other factors, such as domestic buildings, rail proximity, and paths, did not show significant effects once multicollinearity was accounted for. This suggests that while these elements are critical to urban design, their direct influence on house prices might be more complex or context-dependent than initially hypothesized. The negative impact of greenspaces (also reported by Chen et al. (2023) in their study) and water bodies on house prices was unexpected and warrants further investigation to understand the underlying trade-offs between urban density and built environment.

Policy Implications

Given the significant impact of commercial areas on house prices, urban planners should prioritize the strategic development of commercial zones to foster economic growth and enhance property values. However, a balanced approach is necessary, ensuring that commercial development does not come at the expense of residential and recreational spaces, which are also essential for long-term urban sustainability. Moreover, while factors like rail proximity did not show significant effects in this model, enhancing transport infrastructure remains crucial for overall urban livability and could have indirect benefits for property values.

Limitations and Future Research

This studys cross-sectional nature, focusing on 2006 data, limits its ability to capture temporal dynamics in the relationship between built environment factors and house prices. Future research should adopt a longitudinal approach to understand how these relationships evolve overtime. Additionally, integrating socio-economic variables into the analysis could provide a more comprehensive understanding of the factors influencing house prices. The geographic scope of the study is another limitation; expanding the analysis to other cities or regions would allow for the generalization of findings and the identification of broader urban trends.

References

Average House Prices by Borough - London Datastore (n.d.).
URL: *https://data.london.gov.uk/dataset/average-house-prices*

Chen, Y., Jones, C. A., Dunse, N. A., Li, E. & Liu, Y. (2023), 'Housing Prices and the Characteristics of Nearby Green Space: Does Landscape Pattern Index Matter? Evidence from Metropolitan Area', *Land* **12**(2), 496. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

Conway, D., Li, C., Wolch, J., Kahle, C. & Jerrett, M. (2010), 'A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values', *The Journal of Real Estate Finance and Economics* **41**(2), 150–169. Publisher: Springer.

Cordera, R., Coppola, P., dell'Olio, L. & Ibeas, . (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**, 308–319.

Glaeser, E. L. & Kahn, M. E. (2004), 'Sprawl and urban growth', *Handbook of Regional and Urban Economics* **4**, 2481–2527. Publisher: Elsevier.

Land Use by Borough - London Datastore (n.d.).
URL: *https://data.london.gov.uk/dataset/land-use-ward*

Lee, J.-W., Lee, S.-W., Kim, H. G., Jo, H.-K. & Park, S.-R. (2023), 'Green Space and Apartment Prices: Exploring the Effects of the Green Space Ratio and Visual Greenery', *Land* **12**(11), 2069. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
URL: *https://www.mdpi.com/2073-445X/12/11/2069*

Thanks!