Exploring the Effects of Host Attributes, Property Type, and Location on Airbnb Pricing in Boston

Hexiang Li

1 Introduction

As urban tourism becomes increasingly accessible, platforms like Airbnb play a significant role in shaping short-term rental markets. In a city like Boston—known for its high living costs and strong tourist appeal—understanding the determinants of Airbnb pricing is both economically and socially important. This paper addresses the question: **In what ways do host attributes, property type, and location contribute to the pricing structure of Airbnb listings in Boston?** This is an economic question because it explores how different supply-side and environmental variables affect prices in a quasi-competitive online marketplace.

Previous research has explored similar topics in various markets. For instance, Gibbs et al. (2018) proposed a hedonic pricing model that shows how amenities and host reputation affect Airbnb prices. Teubner et al. (2017) demonstrated that Superhosts tend to charge more due to higher perceived quality. Neighborhood conditions like poverty rate and household income were also found to strongly influence prices in multiple studies, including those by Chica-Olmo et al. (2020) and Zhang and Liu (2019). Zhang et al. (2021) further observed that properties in wealthier or more central areas generally command higher prices. While many of these studies focus on single-variable effects or specific case studies, our contribution is to combine host, property, and neighborhood variables into a unified model using both linear regressions and machine learning. We also integrate external socioeconomic indicators like per capita income and poverty rate with scraped Airbnb listings to capture location-based effects more effectively. Additionally, we use decision trees and random forest models to identify non-linearities and interaction effects missed in traditional OLS. By building on past findings and introducing richer modeling techniques, this paper aims to uncover nuanced relationships between Airbnb prices and a combination of host, property, and neighborhood characteristics. We now turn to the data and methods used to answer this question.

2 Data

Our primary dataset is scraped from Airbnb listings in Boston using the platform's web interface, with the scraping date being September 7, 2016 (Kaggle). This includes variables such as price, number of bedrooms and bathrooms, property type, host Superhost status, number of listings per host, and how long a host has been active. We merge this with socioeconomic data from the 2015–2019 "Neighborhood Tables" dataset provided by Analyze Boston. This includes per capita income, median household income, and poverty rate, aggregated at the neighborhood level (Analyze Boston, 2021). These external indicators allow us to explore the influence of neighborhood wealth and inequality on Airbnb pricing. Each observation is a unique Airbnb listing in Boston. After cleaning and merging, we obtained a sample of 3,611 observations with complete data across all relevant variables.

3 Summary Statistics and Visualizations

To better understand the economic landscape of Boston's Airbnb market, Figure 1 displays several housing investment indicators. Notably, the Mortgage as a Percentage of Income stands out at 64.51 percent, suggesting that housing costs constitute a large portion of household budgets. The Price-to-Rent Ratios, especially in city centers, are also elevated, pointing to a premium placed on central locations. Meanwhile, gross rental yields appear higher outside the city center, which may reflect either lower purchase prices or alternative host pricing strategies. These findings suggest spatial variation in investment returns, which might shape how hosts strategize across different neighborhoods.

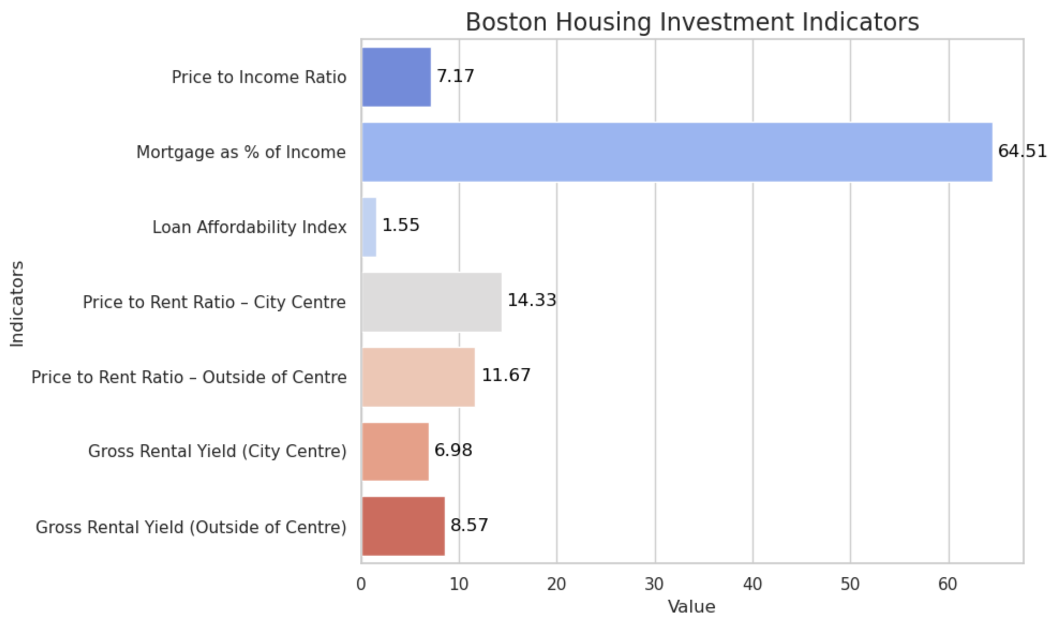

Figure 1: Boston Housing Investment Indicators

Next, Figure 2 visualizes the relationship between neighborhood income, poverty, and average Airbnb pricing. A clear positive correlation emerges between Per Capita Income and Airbnb prices: neighborhoods with wealthier populations, such as Beacon Hill and Back Bay, command significantly higher rental prices. However, this trend also reveals diminishing marginal returns—beyond a certain income threshold, Airbnb prices plateau. Furthermore, the color gradient highlights how poverty rate inversely correlates with price: neighborhoods with higher poverty such as Roxbury and Mattapan consistently fall in the lower-left quadrant of the chart. This pattern supports the hypothesis that neighborhood affluence level and poverty rates significantly influence Airbnb pricing structures in Boston. Overall, this plot provides early evidence for the economic complexity embedded in Airbnb pricing. It suggests that location interacts with socioeconomic variables in non-linear ways, reinforcing the need for multivariable regression analysis and machine learning tools in the following sections to capture these subtleties. This motivates our deeper investigation into the roles of host attributes, property types, and neighborhood characteristics in shaping the Airbnb pricing structure across Boston.

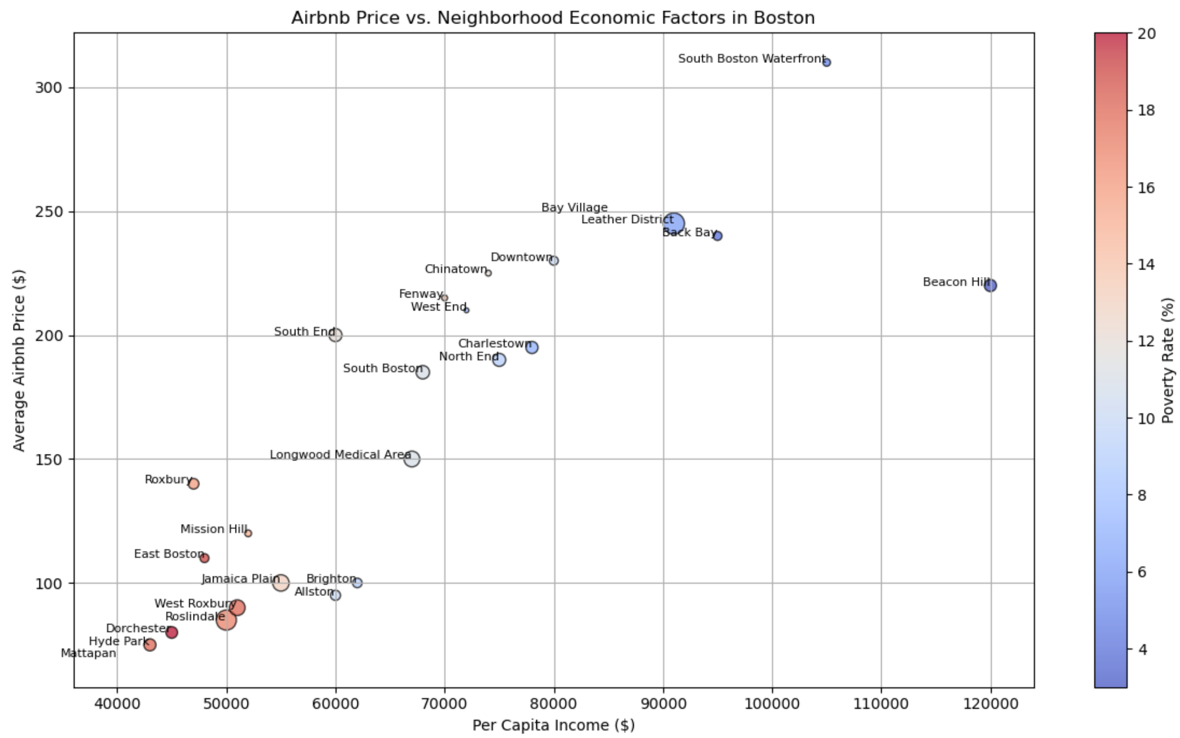

Figure 2: Airbnb Price vs. Neighborhood Economic Factors in Boston. The x-axis represents per capita income, the y-axis shows average Airbnb price, and the color gradient indicates poverty rate by neighborhood.

4 OLS Regressions

Here, let's go deeper into the variables. Table 1 reports the results from four sequential models. In column (1), we find that Per Capita Income is positively and significantly associated with Airbnb listing prices. However, the R^2 value is relatively low (0.085), indicating that other important variables are missing. Also, column (2) adds property characteristics, the number of bedrooms and bathrooms, which significantly improve the fit of the model, and R2 increases to 0.244. The coefficients for both variables are positive and highly significant, suggesting that larger properties indeed command higher prices, consistent with standard housing theory.

Meanwhile, column (3) introduces market level factors, particularly the number of listings managed by a host. This variable is also positively associated with price, suggesting that more experienced or commercial hosts tend to charge more. Also, as we might notice that interestingly "host is superhost"(variable) remains insignificant, implying that simply being labeled a Superhost may not directly increase price when other attributes are controlled for.

Also, column (4) introduces "host is superhost" (variable), which turns out to be statistically insignificant after controlling for other variables. With a coefficient of 1.19 and a high standard error, it appears that the Superhost label itself does not command a meaningful price premium once income, property size, and host activity are taken into account. This supports the idea that Superhost status might play a larger role in influencing guest trust or review scores than in directly affecting price.

Together, these regressions suggest that property features are the dominant drivers of Airbnb pricing, while income and host competition have modest but consistent effects. Superhost status, although important for platform reputation, does not translate into higher prices once other host and neighborhood characteristics are accounted for.

Table 1: OLS Regression Results - Table 1

Dependent variable:	price			
	(1)	(2)	(3)	(4)
	Basic Regression	*Adding Property Characteristics*	*Adding Market Factors*	*Considering Superhost Effect*
Intercept	87.342***	-36.196***	-34.106***	-34.184***
	(5.353)	(7.437)	(7.413)	(7.428)
Per_Capita_Income	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
bathrooms		27.835***	23.960***	23.918***
		(4.875)	(4.902)	(4.909)
bedrooms		69.074***	68.419***	68.407***
		(3.238)	(3.226)	(3.227)
calculated_host_listings_count			0.460***	0.461***
			(0.081)	(0.082)
host_is_superhost				1.185
				(6.935)
Observations	3476	3452	3452	3452
R^2	0.085	0.244	0.250	0.250
Adjusted R^2	0.084	0.243	0.250	0.249
Residual Std. Error	142.619	129.980	129.401	129.419
F Statistic	160.784***	370.045***	287.994***	230.336***

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

To deepen the analysis, Table 2 presents an expanded set of regressions that incorporate neighborhood fixed effects, poverty rate, and interaction terms to uncover more nuanced price determinants. Compared to the earlier models, these regressions help distinguish between the direct impact of income and the contextual effects driven by neighborhood characteristics. The Fixed Effects Model shows that even after controlling for regional heterogeneity, Per Capita Income remains a significant and positive predictor of Airbnb prices. This reinforces the idea that higher-income areas generally command higher prices, aligning with previous studies (Zhang Liu, 2019). However, when Poverty Rate is introduced, we see that income's explanatory power weakens—indicating that localized deprivation more directly drives price variation. In fact, Poverty Rate itself becomes significant, with a positive effect, likely capturing scarcity of listings or higher willingness to pay in lower-supply areas.

Moreover, the Non-Linear Regression confirms a concave relationship between income and price, meaning that price increases level off in wealthier neighborhoods. This diminishing marginal return is important from an economic perspective because it shows that while income matters, its influence is not unlimited. Finally, the Interaction Regression investigates whether Superhost status moderates the income-price relationship. The insignificant interaction term implies that Superhost identity does not significantly alter the income effect—suggesting that income and host status may act independently in shaping pricing dynamics. These layered regressions provide a richer picture of Airbnb pricing, where neighborhood conditions like poverty play a more direct role than initially expected, and income effects are nuanced by local saturation and affluence thresholds.

Table 2: OLS Regression Results - Table 2

Dependent variable:	price			
	(1)	(2)	(3)	(4)
	Fixed Effects	*Adding Poverty Rate*	*Non-Linear Regression*	*Interaction Regression*
Intercept	-61.984***	-67.726***	-20.714**	-65.565***
	(12.695)	(10.118)	(8.794)	(10.202)
Per_Capita_Income	0.002***	0.002***	-0.000	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
bedrooms	74.552***	74.552***	74.552***	74.141***
	(3.154)	(3.154)	(3.154)	(3.163)
bathrooms	24.151***	24.151***	24.151***	24.397***
	(4.770)	(4.770)	(4.770)	(4.772)
calculated_host_listings_count	0.030	0.030	0.030	0.036
	(0.087)	(0.087)	(0.087)	(0.087)
host_is_superhost	11.578*	11.578*	11.578*	-13.068
	(6.783)	(6.783)	(6.783)	(16.547)
Poverty_Rate		15.903*	24.516***	16.104*
		(9.324)	(9.147)	(9.323)
`I(Per_Capita_Income ^ 2)`			0.000***	
			(0.000)	
Per_Capita_Income:host_is_superhost				0.000
				(0.000)
Observations	3452	3452	3452	3452
R^2	0.310	0.310	0.310	0.311
Adjusted R^2	0.305	0.305	0.305	0.306
Residual Std. Error	124.490	124.490	124.490	124.460
F Statistic	64.250***	64.250***	64.250***	61.816***

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

5 Decision Tree

To explore non-linearities in Airbnb pricing, I train a regression tree using all ten variables. Here, this tree splits first on Median Household Income, followed by Per Capita Income and Region. This suggests that economic conditions and location matter more than host level features. Listings in high-income neighborhoods tend to be more expensive, but even within rich areas, location still drives differences. For example, some high-income areas average 300 dollars, while others average 250 dollars. In lower-income branches, Region and income still differentiate pricing. Overall, the tree reveals that broader neighborhood characteristics play the strongest role in shaping Airbnb prices in Boston rather than host reputation.

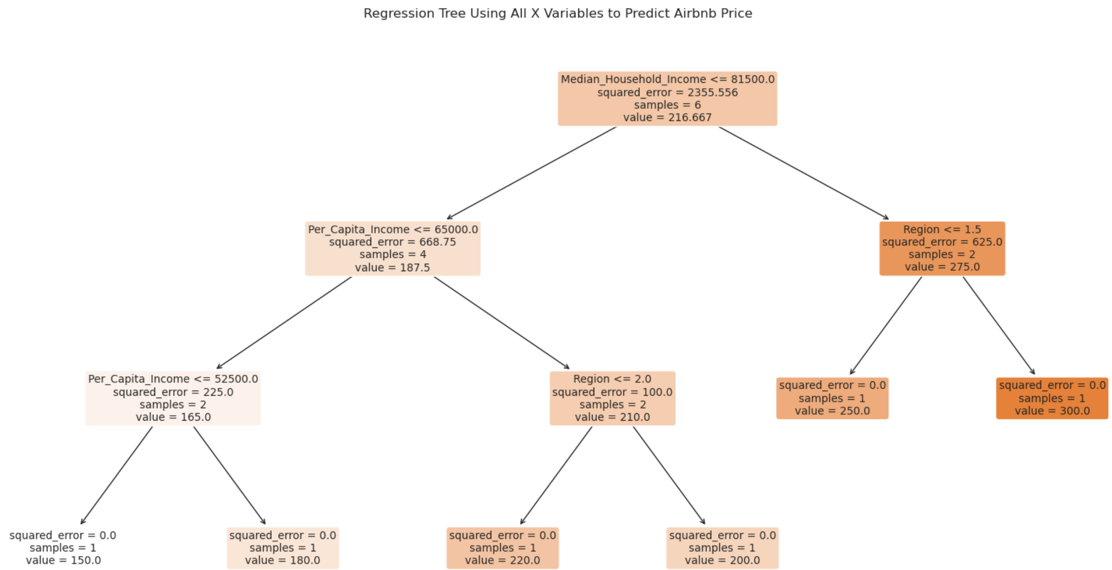

Figure 3: Regression Tree Using All X Variables to Predict Airbnb Price. This tree splits listings based on median household income, per capita income, and region, illustrating how different income brackets and locations influence price.

Economically, the tree reveals that in high-income areas, neighborhood (Region) further differentiates prices. Meanwhile, in low-income areas, price variation is better explained by per capita income ranges. Host attributes like Superhost status do not appear among the top splits, suggesting their limited marginal impact.

6 Random Forest

To further explore non-linear relationships, we train a Random Forest with 100 trees. The feature importance plot in Figure 4 shows that Per Capita Income, host listings count, and Median Household Income are the top predictors of Airbnb price, confirming our earlier findings that neighborhood economic factors are more influential than host identity. Interestingly, Superhost status and property type rank among the least important, suggesting they play only a minor role once broader economic and regional differences are controlled for.

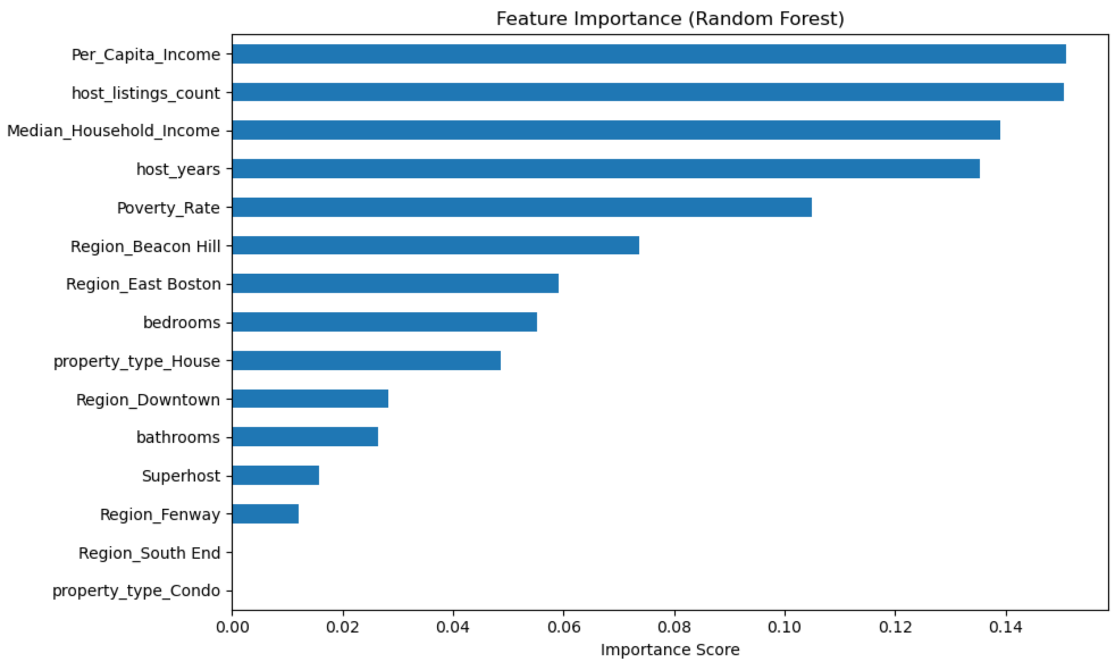

Figure 4: Feature Importance from Random Forest. Per capita income and host listing count are the most influential predictors of Airbnb price, while property type and Superhost status contribute less.

These results suggest that professional hosts and hosts in wealthier neighborhoods tend to charge more, but Superhost status is not a major pricing driver. The Random Forest model achieves an R^2 of 0.73 and a Mean Squared Error (MSE) of 330, outperforming the OLS model and confirming the value of allowing flexible interactions across predictors.

7 Conclusion

Airbnb pricing in Boston is influenced by a combination of property types, host attributes, and neighborhood economic conditions. Larger properties, such as those with more bedrooms and bathrooms, tend to have higher prices, while apartments, the most common type of listing, show a more standardized pricing pattern. Regression trees and random forest models confirm that property characteristics matter, but their influence is conditional: bedrooms and bathrooms only become important within certain income brackets or regions.

The analysis suggests that host factors, such as Superhost status and the number of listings managed, have a limited impact on pricing compared to neighborhood-level economic factors. However, random forest results reveal that host-related variables, such as the number of listings and the experience of hosting, are among the top predictors of price, more important than some property features. So this suggests that while host attributes may not show strong marginal effects in linear models, their contribution becomes clearer in non-linear, interaction-driven models.

Regression results show that Per Capita Income initially has a positive and significant effect on price, but this effect becomes smaller and even turns slightly negative after accounting for poverty rate and neighborhood fixed effects. This suggests that poverty rate plays a more direct role in price determination and that income alone does not fully explain price variation. The regression tree reinforces this, showing that neighborhood median and per capita income are the first decision splits, and that poverty rate and region further shape the pricing path.

Moreover, the non-linear regression reveals a diminishing return to income and Airbnb prices rise with income but level off at higher income levels. This non-linearity is captured only through machine learning models like regression trees and random forests, which uncover threshold effects missed by OLS. Interaction regression results also indicate that Superhost status does not significantly alter the relationship between income and price.

The pricing competition among Airbnb hosts reflects features of Cournot or Bertrand models. While high-income area concentration may worsen housing shortages, lower-priced listings in disadvantaged neighborhoods offer economic opportunities but risk displacement and affordability issues.

For the **FINAL ANSWER** to the research question. Host attributes such as number of listings and years of experience, contribute moderately to price differentiation, especially in non-linear models. Property type affects pricing primarily through size of properties which relate to number of bedrooms and bathrooms, but not all types show clear price differences such as condos and houses. Location and neighborhood-level economic indicators such as Per Capita Income, Poverty Rate and Region are the most consistent and powerful predictors of Airbnb prices, both in linear and machine learning models. **So these findings suggest that location and economic context dominate pricing structures, while host and property factors matter more in specific contexts or through interaction effects.**

References

Data Sources

1. Airbnb. (n.d.). *Boston Airbnb Open Data.* Kaggle. Retrieved from https://www.kaggle.com/datasets/airbnb/boston

2. City of Boston. (n.d.). *Neighborhood Demographics.* Boston.gov. Retrieved from https://data.boston.gov/dataset/neighborhood-demographics/resource/d8c23c6a-b868-4ba4-8a3b-b9615a21be07

Literature Sources

1. Chica-Olmo, J., Chica-Olmo, M., Cano-Guervos, R. (2020). Location determinants of Airbnb prices. Tourism Management, 79, 104098. https://doi.org/10.1016/j.tourman.2020.104098

2. Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., Goodwill, A. (2018). Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings. Journal of Travel Tourism Marketing, 35(1), 46–56. https://doi.org/10.1080/10548408.2017.1308292

3. Teubner, T., Hawlitschek, F., Dann, D. (2017). Price determinants on Airbnb: How reputation pays off in the sharing economy. Journal of Self-Governance and Management Economics, 5(4), 53–80. https://publikationen.bibliothek.kit.edu/1000068696

4. Wang, D., Nicolau, J. L. (2017). Price determinants of sharing economy-based accommodation rental: A study of listings from 33 cities on Airbnb.com. International Journal of Hospitality Management, 62, 120–131. https://doi.org/10.1016/j.ijhm.2016.12.007

5. Piao, R. (2023). The new age of collusion? An empirical study into Airbnb's pricing behavior. arXiv preprint arXiv:2312.05633. https://arxiv.org/abs/2312.05633

6. Pittala, T. S. S. R., Meleti, U. M. R., Vasireddy, H. (2024). Unveiling patterns in European Airbnb prices: A study using machine learning. arXiv preprint arXiv:2407.01555.

https://arxiv.org/abs/2407.01555

7. Zhang, Z., Zhang, Z., Lu, L. (2021). Study of price determinants of Airbnb services: A case study in Beijing. Journal of Theoretical and Applied Electronic Commerce Research, 16(4), 973–987. https://doi.org/10.3390/jtaer16040035

8. Toader, V., Negrușa, A. L. (2021). Analysis of price factors in Airbnb listings. Economic Research-Ekonomska Istraživanja, 34(1), 1238–1255. https://doi.org/10.1080/1331677X.2021.1962380

9. Zhang, Z., Liu, Y. (2019). Investigating the impact of neighborhood conditions on Airbnb listing prices: A case study of Sydney. Environment and Planning B: Urban Analytics and City Science, 46(6), 1174–1195. https://doi.org/10.1177/2399808319828730

10. Kakar, V., Franco, J., Voelz, J., Wu, J. (2016). Effects of host race information on Airbnb listing prices in San Francisco. MPRA Paper No. 69974. https://mpra.ub.uni-muenchen.de/69974/