Airbnb in London: A Spatial Analysis of Distribution, Typologies, and Pricing Mechanisms

1. Introduction

With the rise of shared accommodation platforms, Airbnb has become an important component of the urban short-term rental market, especially in international metropolises such as London. Its impact is not only reflected in the improvement of tourism reception capacity, but also brings changes to the urban spatial structure and housing price mechanism.

Against the backdrop of rapid expansion in short-term leasing, urban managers and researchers are increasingly concerned about the following key issues:

1. Are Airbnb listings showing a spatial clustering trend? Is there a hotspot area?
2. Are there significant differences in landlord operation strategies among different regions?
3. What factors affect the price of Airbnb and are their effects consistent in terms of space?

This study takes London as a case study and comprehensively applies spatial analysis methods, including point pattern analysis, cluster analysis, and geographically weighted regression (GWR), to explore the spatial distribution characteristics and price mechanism of Airbnb, aiming to provide data support and spatial perspective for urban tourism management and rental policy formulation.

2. Data&Methods

2.1 Data sources

This study used Airbnb data from London published on the Kaggle website (version: 2022).

https://www.kaggle.com/datasets/whenamancodes/london-uk-airbnb-open-data

The data contains nearly 70000 property records, covering the following core fields:

- House price

- Monthly number of comments (reviews_per_ month)
- Number of days available for booking (availability 365)
- The number of properties owned by the landlord (calculatedhosted listing count)
- Property type, geographical location, minimum stay nights, etc

After cleaning, the data is converted to GeoDataFrame format and uniformly projected onto the UK National Grid Coordinate System (EPSG: 27700) to support subsequent distance and density analysis.

2.2 Analysis Methods

This study mainly adopts the following three spatial analysis methods to gradually construct an explanatory path for Airbnb spatial behavior:

① Point pattern analysis (KDE+CSR verification)

Using Kernel Density Estimation to draw Airbnb hotspot maps;

Conduct a complete spatial randomness (CSR) test on the distribution of housing resources through Quadrant Analysis to determine their clustering significance.

② Cluster analysis (KMeans)

Select variables such as price, activity level, available booking days, and number of landlord listings;

Using KMeans clustering to identify operational types and explore the spatial "ecological differentiation" of landlords;

③ Spatial Regression Model (GWR)

Using Geographically Weighted Regression method;

Using housing prices as the dependent variable, activity level, available booking days, and number of properties as explanatory variables;

Analyze whether the mechanisms that affect prices in different regions are consistent.

All analyses were completed using the Python toolchain, including GeoPandas, Scikit learn, pointpatches, mgwr and other modules, supplemented by the visualization library matplotlib, Folium and other graphics support expression. All code is displayed in the Jupyter notebook.

3. Spatial Distribution Analysis: Agglomeration and Hotspot Recognition (KDE+CSR)

In order to understand the spatial distribution pattern of Airbnb listings in London, this paper first uses kernel density estimation (KDE) to visualize the spatial clustering of listings, and further tests whether it significantly deviates from the assumption of complete spatial randomness (CSR) through Quadrant Analysis.

3.1 Kernel Density Estimation (KDE) Analysis

By performing KDE analysis on the location points of approximately 70000 Airbnb listings, the following spatial heatmap was obtained:

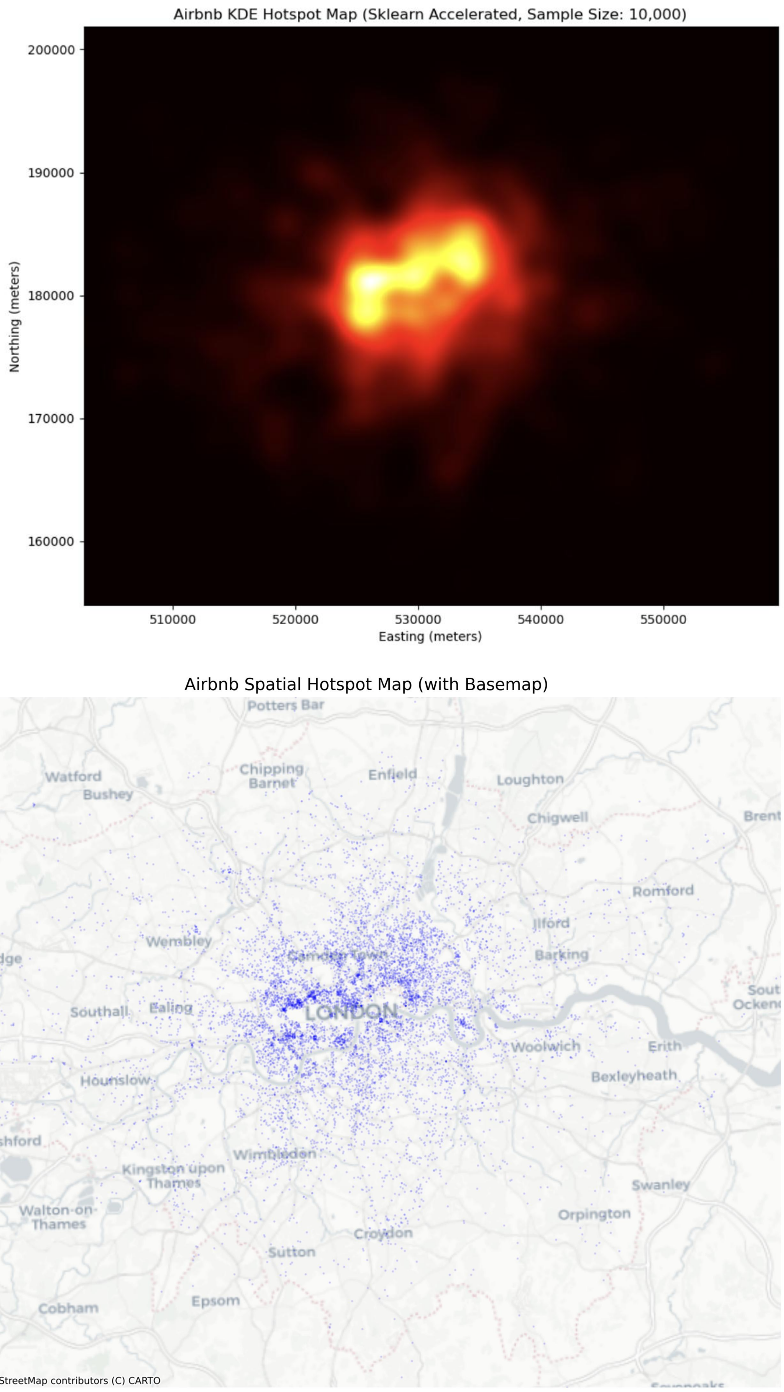

Figure 3.1 Hotspot Map of Airbnb Property Space Density (KDE)

The graph shows that Airbnb listings in London are clearly clustered, with hotspots concentrated in:

- ◆ Central London (Soho, Covent Garden, Westminster)
- ◆ Eastern District Cultural and Creative Cluster (Shoreditch, Hackney)

◆ South Bank Landscape Area (London Bridge, Borough)

◆ Camden and King's Cross surroundings

These areas are mostly urban areas with concentrated tourist attractions, convenient transportation, or a strong cultural atmosphere, demonstrating the strategic choice of Airbnb hosts' layout.

3.2 CSR test (spatial complete randomness)

To verify whether the above visual aggregation has statistical significance, the Quadrant Analysis method was used to test the spatial point patterns.

```
Quadrat Chi-squared: 392621.84238158434
Degrees of Freedom: 99
P-value: 0.0
```

Figure 3.2 Hotspot Map of Airbnb Property Space Density (KDE)

◆ Quadrat Chi-squared: 392621.84238158434

◆ Degrees of Freedom: 99

◆ p-value < 0.001

The results indicate that we can reject the hypothesis of complete spatial randomness at a 99.9% confidence level, which suggests that Airbnb's spatial distribution exhibits statistically significant clustering structures.

3.3 Summary

Airbnb listings are not evenly distributed throughout the city, but significantly clustered in specific core functional areas;

The gathering areas are mostly commercial, cultural, and tourism hotspots, which have a dependency on the urban spatial structure;

This conclusion provides a spatial structure foundation for subsequent analysis, such as type identification and price modeling

4. Identification of landlord operation types: KMeans clustering analysis

After clarifying that Airbnb listings have significant spatial clustering characteristics, this article further explores the landlord operation model behind them. We use KMeans clustering method to identify multiple Airbnb host "ecological types" that exist in the city of London based on host behavior characteristics.

4.1 Selection of Cluster Variables and Method Explanation

Cluster analysis selects the following four variables as feature dimensions:

Explanation of variable	name	meaning indicators
Price	house price	profit strategy
Reviews_per_month	Monthly Comment	Count Activity and Demand
availability_365	Number of days available for booking each year	Is it open for operation throughout the year
calculated_host_listings_count	Number of landlord properties	Commercialization level (individual vs. multi property enterprise type)

After standardization, the data is clustered into 5 categories using the KMeans algorithm. The number of categories is determined by a combination of silhouette coefficient and interpretability, ultimately reflecting strong differences in housing behavior.

4.2 Clustering Results and Feature Interpretation

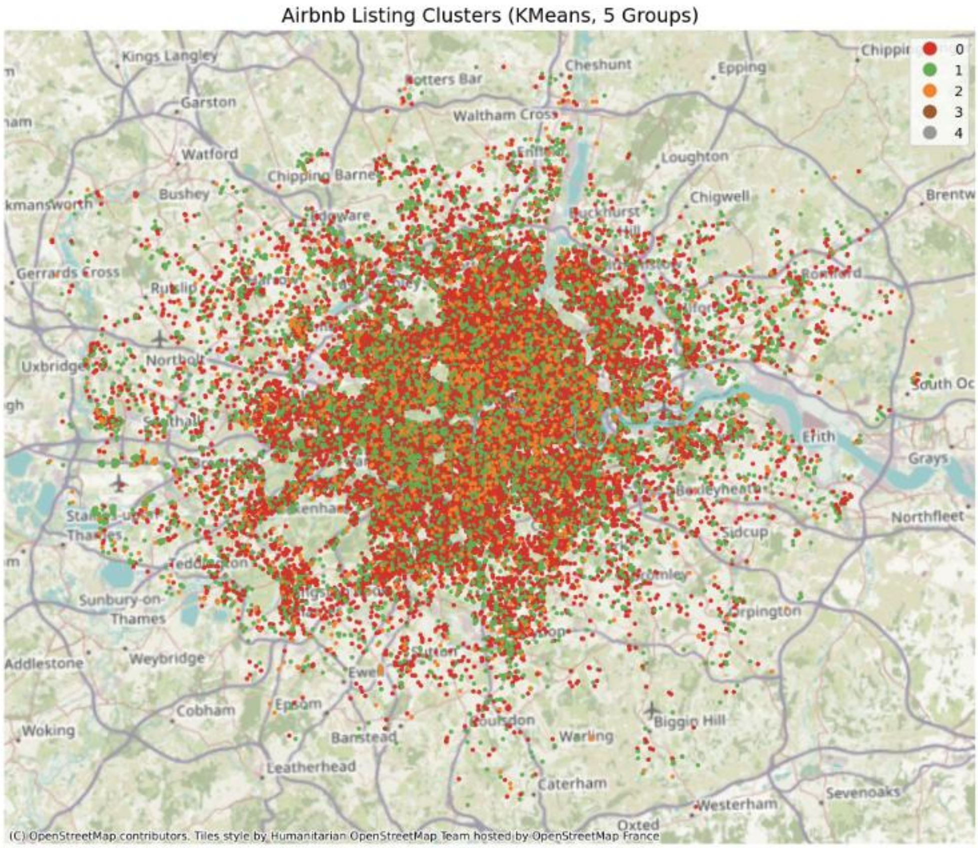

Figure 4.1 Airbnb Property Clustering Map and Attribute Mean Table (KMeans, 5 Categories)

The figure shows the spatial distribution of clustering Airbnb listings into 5 categories, with different colors representing different types of operations. The map clearly presents:

Cluster 3 (orange): concentrated in the central area of London, with the highest housing prices (average £ 550) and the largest number of landlord listings (183.90), speculated to be large-scale corporate landlords;

Cluster 0 (red): widely distributed around the outskirts of London, with the lowest price (£ 115) and extremely low booking days (21.98), more likely to be occasional renters or inactive listings;

Cluster 1 (green): With a wide coverage area, medium to high housing prices, and year-round availability (availability=365 ≈ 282), it is speculated to be a stable operating small and medium-sized landlord;

Cluster 2 and Cluster 4 exhibit more marginalization or special operational strategies:

Cluster 2： Low to medium price active landlords (reviews ≈ 4.03) may be short-term flexible;Cluster 4： The price is extremely high (average £ 5024.9), but the activity is low and the number of landlord properties is relatively high, which is likely to represent luxury/investment asset properties.

The following table shows the average levels of each cluster in terms of price, activity, available booking time, and landlord size dimensions:

cluster	price	reviews_per_month	availability_365	calculated_host_listings_count
0	115.02	0.47	21.98	3.90
1	191.36	0.85	281.96	8.90
2	129.85	4.03	153.07	6.41
3	550.02	0.49	101.44	183.90
4	5024.90	0.59	182.43	75.09

Figure 4.2 Airbnb Property Clustering Result

Summary of Chart Analysis:

Significant spatial heterogeneity: Different types of landlords exhibit a clear "zoning" phenomenon in space;

**The enterprise type operating landlord (Cluster 3) * * shows a strong concentration trend in both space and variables;

**Luxury high priced landlords (Cluster 4) * * Although the quantity is very small, the price is extremely high and belongs to the abnormal category;

**The widest group (Cluster 0) * * presents a typical "individual amateur operator" pattern of low price and low activity.

5. Spatial modeling: Interpretation of spatial mechanisms of housing prices (GWR)

In order to further explain the spatial difference mechanism of Airbnb room prices, this paper uses Geographically Weighted Regression (GWR) model to model the relationship between room prices and operational variables. The GWR model allows regression coefficients to vary in space, capturing spatial heterogeneity that traditional OLS cannot reflect.

5.1 Model Construction and Variable Explanation

The model takes price as the dependent variable and selects the following three variables as explanatory variables:

independent variable	meaning	Explanation
Reviews_per_month	Monthly Comment	Count Activity and Demand
availability_365	Number of days available for booking each year	Is it open for operation throughout the year
calculated_host_listings_count	Number of landlord properties	Commercialization level (individual vs. multi property enterprise type)

All variables were standardized before modeling. Firstly, the empirical bandwidth value (150 meters) is used for fitting, and then the Sel_SW automatic bandwidth selection method is used for optimization, resulting in the optimal bandwidth of 162 meters. The final model uses automatic bandwidth results for interpretation and visualization, while manual models are used for comparative analysis.

5.2 Model Performance Comparison: OLS vs GWR (Manual vs Automatic Bandwidth)

Model Type	Bandwidth	R^2	Adjusted R^2	AIC
OLS		0.093	0.090	13652
GWR (Manual Bandwidth)	150	0.298	0.242	13536.47
GWR (Automatic Optimization)	162	0.291	0.239	13536.23

From the results, it can be seen that compared to traditional regression models, GWR significantly improved the explanatory power of the model (R^2 increased from 0.093 to 0.291), indicating that Airbnb room prices are affected by variables and have significant spatial heterogeneity.

Although the R^2 of the automatic bandwidth model is slightly lower than the manual version, its AIC is lower and the model is more stable, so it will be used as the main model for further explanation.

5.3 Summary

- The impact mechanism of Airbnb room prices varies by location, and the direction and intensity of variables have significant spatial variations;
- GWR better characterizes the spatial relationship between housing prices and operational behavior compared to OLS;
- Policy interventions in different regions should consider the effects of local variables (such as "activity=competitive pressure" in which regions, and "landlord=price dominance" in which regions).

6. Conclusion and policy recommendations

This study systematically analyzed the spatial structure of Airbnb in London from three aspects: housing distribution, landlord types, and pricing mechanisms. A spatial analysis framework consisting of kernel density estimation (KDE), cluster analysis, and geographically weighted regression (GWR) was constructed.

Firstly, at the spatial distribution level, both the KDE plot and CSR test indicate that Airbnb listings are significantly clustered in London, mainly concentrated in the city center and tourist hotspots such as Soho and Shoreditch. This distribution characteristic indicates that the short-term rental market has a clear spatial preference and is closely related to urban functional areas.

Secondly, through KMeans clustering, five types of landlords were identified, including occasional, stable, corporate, and luxury investment types. Different types not only have significant differences in operational characteristics, but also exhibit spatial clustering. For example, enterprise landlords are concentrated in the core area, while occasional landlords are distributed more discretely.

In terms of price mechanism modeling, the GWR model shows that housing prices are influenced by factors such as activity level, number of open days, and landlord size, but this influence varies by location and spatial heterogeneity is significant. For example, the activity of comments is negatively correlated in some areas, reflecting intense price competition in popular regions; Corporate landlords have stronger pricing power in core areas, while the effect may be opposite in peripheral areas.

Overall, the Airbnb market is not a spatially homogeneous system, but a composite structure composed of multiple actors and geographical locations. Urban management policies should be more geographically sensitive, adopting layered and differentiated strategies for different regions and types of landlords, rather than a one size fits all regulation. Future research can combine time series and urban accessibility data to further enrich our understanding of the evolution and governance strategies of short-term rental spaces.

Reference

Gutiérrez, A., & Domènech, A. (2020). Understanding the spatiality of short-term rentals in Spain: Airbnb and the intensification of the commodification of housing. Geografisk Tidsskrift-Danish Journal of Geography, 120(2), 98–113.

Roelofsen, M. (2018). Exploring the socio-spatial inequalities of Airbnb in Sofia, Bulgaria. Erdkunde, 72(4), 313–328.

Shabrina, Z., Zhang, Y., Saito, K., & Yoshikawa, K. (2021). A Geographically Weighted Regression (GWR) and a Multiscale GWR (MGWR) approaches. Geographical Analysis. https://doi.org/10.1111/gean.12302

Wachsmuth, D., & Weisler, A. (2018). Airbnb and the rent gap: Gentrification through the sharing economy. Environment and Planning A: Economy and Space, 50(6), 1147–1170.