Spatial Dependencies and Pricing Dynamics in Palermo's Airbnb Market

Samuele Carrubba

February 08 2026

Abstract

This study investigates the price determinants of Airbnbs in Palermo, Italy, with a specific focus on the role of spatial dependence. While traditional Ordinary Least Square (OLS) models assume independence between observations, this research acknowledges that the price of a listing is inherently influenced by the pricing strategies of its geographic neighbors. Utilizing a dataset of over 5,500 listings, the study employs a multistage spatial econometric approach, moving from exploratory spatial data analysis to the estimation of a Spatial Autoregressive (SAR) model. the results indicated the existence of a moderate spillover effect in Palermo. Results indicate a significant spatial autoregressive coefficient (p = 0.31), suggesting that a 10% increase in the average price of neighboring listings is associated with a 3.1% increase in a listing's own price. These findings provide critical insights for hosts and policymakers regarding the localized nature of the digital tourism economy in Mediterranean urban centers.

keywords: airbnb, bnb price, spatial regression, spillover effect, spatial lag model

1 1. Specification of a research question

1.1 1.1 Introduction

The rise of the "sharing economy" has transformed the housing market in historic European cities. Palermo, a city characterized by high urban density and a rapidly growing tourism sector, presents a unique case for studying how digital platforms like Airbnb interact with physical geography. Airbnb is a P2P platform connecting directly travelers and hosts, its listings include the most diverse units, spread across the territory. With a number of active listing of around 8 million across 220 Countries, it's clear the relevance Airbnb assumes in the context of hosting services.

Palermo is the main province of the region of Sicily, Italy. The city accounts for over 5800 active listings, 10% of the total listings in the region, a popular and vibrant touristic destination of the island.

1.2 1.2 Research questions

This research aims to bridge the gap between traditional hedonic analysis and spatial econometrics in the new context of Palermo, by answering three interconnected questions:

- RQ1: Does the Airbnb market in Palermo exhibit significant spatial autocorrelation, and to what extent does the Ordinary Least Squares (OLS) model fail to satisfy the assumption of independent observations?

- RQ2: Through a top down specification strategy, which model can better explain the spatial price dynamics?
- RQ3: To what extent do internal structural attributes and external neighboring attributes determine the listing prices of Airbnb accommodations in Palermo?

2 2. Description of data

2.1 2.1 Exploratory preliminary analysis

Data have been taken by the official website of Airbnb and are updated at 29 September 2025. The dataset accounts for 57531 entries and 83 different features. As shown in the literature, only a relevant subset of these features has been taken into account, furthermore the original dataset has been reduced to account for the only location of Palermo, the main geographical context of the analysis.

	price	accommodates	bedrooms	number_of_reviews	review_scores_rating	log_price
count	5571.00	5571.00	5571.00	5571.00	5571.00	5571.00
mean	136.68	3.94	1.51	46.57	4.77	4.40
std	609.78	1.98	0.86	76.06	0.35	0.62
min	14.00	1.00	0.00	1.00	1.00	2.64
25%	58.00	2.00	1.00	4.00	4.70	4.06
50%	75.00	4.00	1.00	16.00	4.87	4.32
75%	103.00	5.00	2.00	56.00	5.00	4.63
max	10000.00	16.00	12.00	845.00	5.00	9.21

Figure 1: Numerical variables stats

Further preparation of data has been performed:

- Data has been cleaned of missing values.
- The two main datasets have been joined in order to have a single one with both the features and the geometry. This has been done through a left join on the "city" column, after checking both datasets contained the same cities in the same format, with no mismatch.
- Strings formatting has been applied to all the columns for which it was necessary (e.g: city).
- The response variable (price), being strongly skewed, has been transformed into its logarithmic form.

The analysis employed two different coordinate reference systems (CRS):

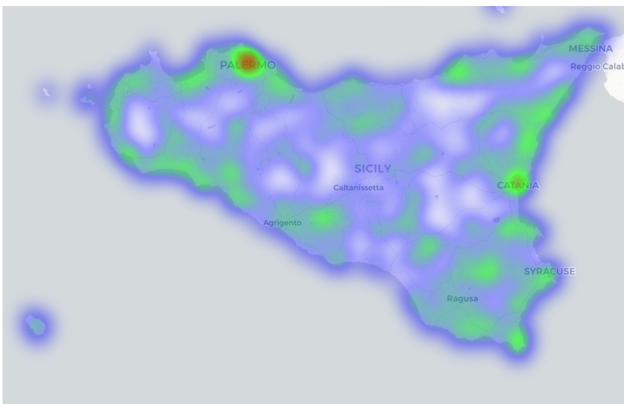
- epsg: 4326
- epsg: 32633

The approach used is: "store in geography, analyze in projected", meaning that the first one has been used for global positioning and visualization, while the second for spatial analysis. This distinction has revealed necessary because of how spatial analysis is performed: this relies on distance calculation between neighbors, so a more appropriate system based precise calculations was needed.

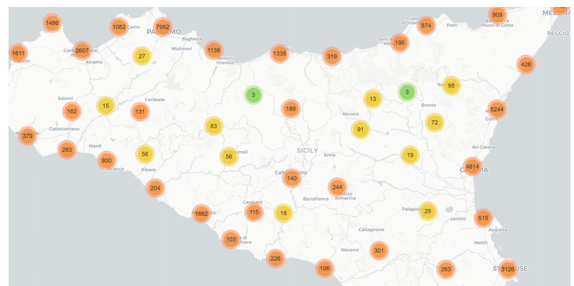
Two final datasets were created: the primary one for the city of Palermo and a secondary one for the whole Sicily.

3 3. Data analysis oriented by the research questions

The main analysis started with a visual understanding of Airbnbs distribution across Sicily and in the city of Palermo. The island showed an uneven distribution of listings, with a concentration in the major provinces on the coasts, and a considerably minor presence in the hinterland.



(a) Figure 2: Airbnbs density in Sicily



(b) Figure 3: Airbnbs in Sicily, overview

To validate the research questions, a preliminary analysis on the target variable "price" was necessary and performed as shown in the previous chapter. In its raw form the listings prices assumed a strong right skeweness, a common symptomph in the real estate market where the majority of prices options are overshowded by a few outliers. To satisfy the Gauss-Markov assumptions for linear regression and to stabilize the variance, a logarithmic transformation was applied, assuming an approximately normal distribution (see figure below).

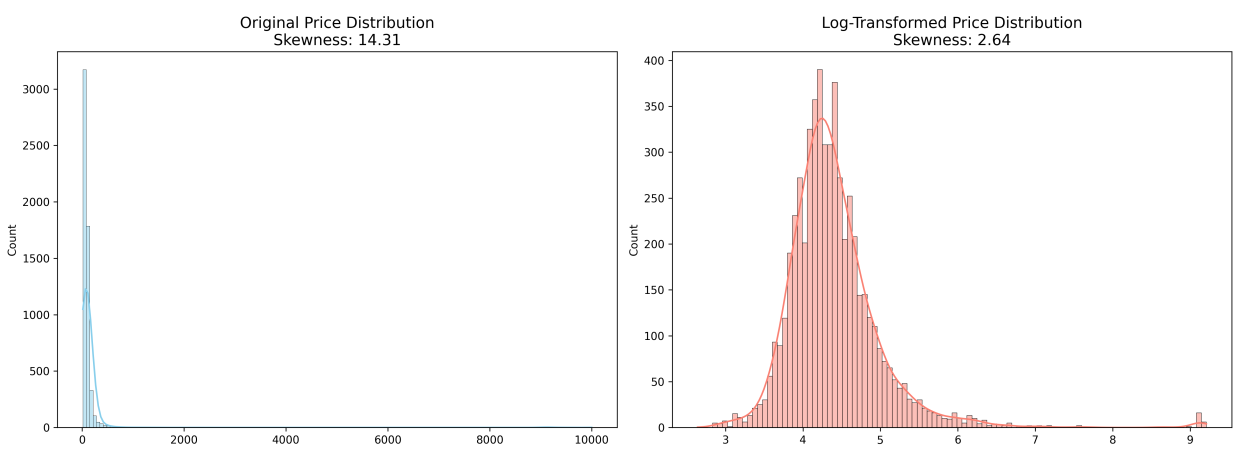


Figure 3: Price logarithm transformation

The variables have been tested against multicollinearity visually, through a correlation matrix, and more rigorously through the variance inflation factor (vif).

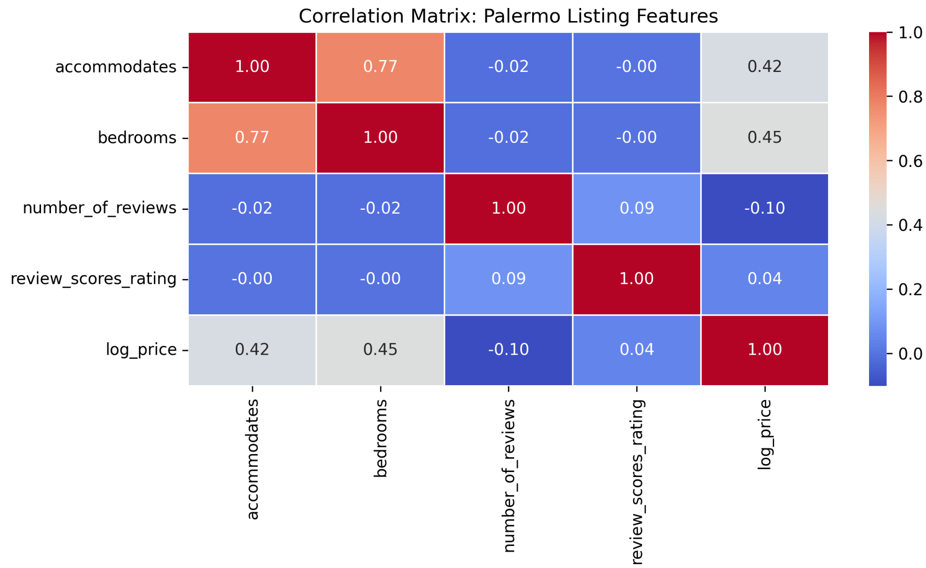


Figure 4: Variables correlation matrix

The number of accomodates and bedrooms showed the highest signs of positive correlation with price, while surprisingly enough the number of reviews were slightly negatively associated with prices. Neither the matrix nor the VIF showed signs of multicollinearity, so being everything acceptable, the dataset has been considered as analysis ready.

Before addressing the spatial component, it was necessary to identify which structural characteristics drive value in the Palermo market. As shown in the boxplots below, there is a clear hierarchical structure to pricing in Palermo: entire homes and apartments listings command a significant price premium and exhibit the highest internal variance, suggesting a diverse market ranging from small apartments to luxury villas.

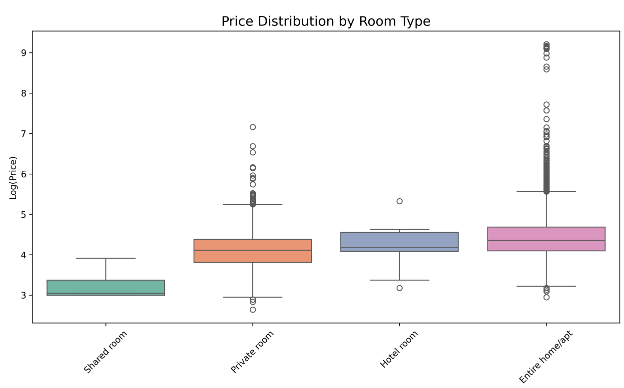
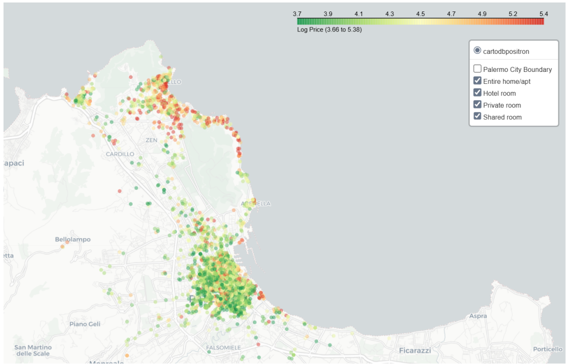


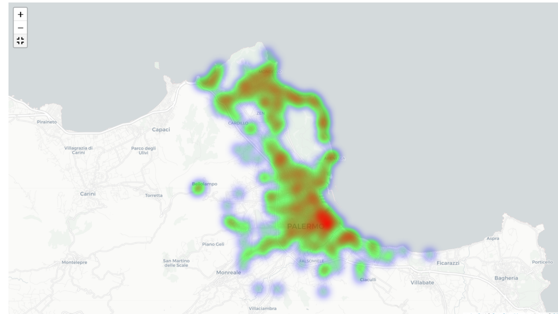
Figure 5: Price distribution by room type

It's immediately noticeable how the distribution of prices can be geographically split into two parts: the coastal part of Palermo and the urban one, with this last having generally lower prices.

Putting the attention on the urban part, it's noticeable how prices increase near the historical center, while decreasing in the suburban zones.



(a) Figure 7: Airbnbs distribution in Palermo filtered by log. price



(b) Figure 8: Airbnbs price density distribution in Palermo

3.1 Global Moran's I

To move from visual observation to statistical relevance, Global Moran's I test has been employed. This is the standard econometric tool here used to determine if the spatial pattern of Airbnb prices in Palermo is a result of structural clustering or mere random chance.

Global Moran's I is a correlation coefficient that measures spatial dependency. While a standard correlation coefficient (r) measures the relationship between two different variables, Moran's I measures the relationship between a single variable (price) and its spatial lag (the average price of its neighbors). It ranges from -1 (perfect dispersion) to $+1$ (perfect clustering), with 0 indicating the absence of spatial autocorrelation and so a completely random spatial distribution.

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

The test brought to the following values:

- Moran's I statistic: 0.154
- p-value = 0.001

With p-value < 0.05 the null hypothesis have been rejected at 99% confidence level in favor of the alternative hypothesis stating that spatial autocorrelation exists, meaning that the prices of listings are partially dependent on the prices of neighboring listings. While autocorrelation exists, the low Moran's I statistic value implies that the bnbs internal features still account for the majority of price determination.

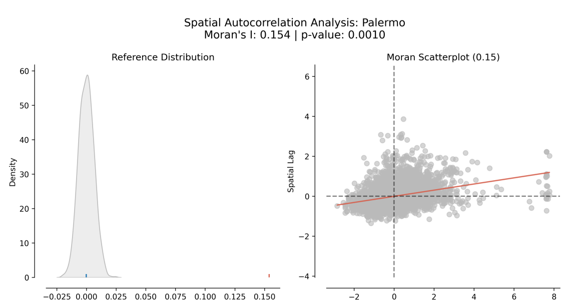


Figure 7: Moran's I results

To locate the specific geographic drivers of the global autocorrelation, a Local Indicator of Spatial Association (LISA) was calculated. Unlike the global test, the local statistic identifies significant clusters where prices are significantly higher (High-High) or lower (Low-Low) than the spatial mean. This allows for a granular mapping of Palermo's hotspots.

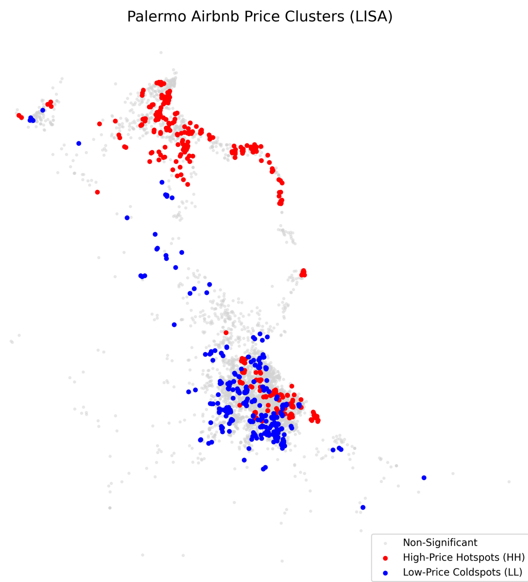


Figure 8: Price clusters

3.2 Spatial regression analysis

Spatial dependence in price determination violates the independence assumption of OLS, so a spatial regression model was needed.

The spatial weights matrix formalizes the neighborhood structure needed for a spatial regression model, while initially two methods were compared (KNN and distance based), KNN was finally adopted for two reasons: the approach is easily interpretable and it guarantees that each listing has the same amount of neighbors independently from variation in density which characterizes the territory.

There exists a variety of different spatial regression models, and in order to use the most suited one the following procedure has been employed:

1. Estimate a OLS and use LM-test.
2. If OLS is rejected, estimate SDM.
3. Conduct a likelihood ratio test (LRT) to check whether SDM can be reduced to SAR.
4. if it cannot be reduced, go for SDM.

The likelihood ratio test assumed a value of 4.97 with a p-value of 0.29 (p-value > 0.05), so the null hypothesis has not been rejected and SDM has been reduced to SAR.

SAR showed to be a better fit, with an AIC drop of over 300 compared to OLS, and a rho value of 0.31, meaning that a 1% percent increase of a neighboring listing price would increase the current listing price of about 0.31%, this is the final proof and quantification of the spatial impact on the determination of Airbnbs prices. the neighborhood effect exerts more influence on price than any individual physical attribute of the listing.

Each of the predictors directly influence its listing's price and indirectly influence neighboring listings prices through a spillover effect. The following tables summarize the findings (github for more details):

Measure	OLS	SAR
AIC	9037.53	8718.88
log likelihood	-4513.76	-4353.44

Table 1: OLS VS SAR

Predictor	Direct	Indirect
Accomodates	0.06	0.03
Number of reviews	-0.001	-0.001
Rating	0.09	0.04
Number of bedrooms	0.20	0.09

Table 2: Spatial lag model impacts

4 Conclusions

This study set out to investigate the spatial dynamics of Airbnb pricing in Palermo, with the aim of understanding whether listing prices are spatially interdependent and to what extent do internal structural attributes and neighboring listing attributes drive price formation.

By integrating spatial econometric techniques, this analysis provides empirical evidence that pricing in the Palermo's Airbnb market is shaped by both intrinsic listing characteristics and neighborhood spatial influences.

All three research questions found an answer:

- RQ1: Palermo Airbnb market showed significant moderate levels of spatial autocorrelation, with high priced listings tending to cluster geographically. This evidence failed the OLS independence assumption.
- RQ2: Through a methodologically robust procedure, OLS was rejected in favor of SDM, which was then proved to be reducible to SAR.
- The neighborhood effect has been confirmed and quantified, with the market showing signs of spatial spillovers, furthermore the direct and indirect effects of a variation in predictors values were quantified, showing their individual impact on neighbors.

The findings provide clear answers to the research questions posed in this study, but despite its contributions, this study is subject to some limitations: the analysis relies on cross sectional data, which restricts the ability to capture dynamic pricing behavior over time; additionally, the choice of spatial weights matrix and neighborhood definition, while grounded in geographic reasoning, may have influenced the magnitude of estimated spillovers. Future research may extend this framework by incorporating panel data and go for a more fine grained analysis on a district level.

5 Codes, scripts, softwares

All the analysis and maps is available at the following github page:
https://github.com/SamC-dev/Airbnb-spatial-regression-analysis/tree/main

6 References

1. https://www.statista.com/statistics/1193532/airbnb-nights-experiences-booked-worldwide/
2. https://www.demandsage.com/airbnb-statistics/
3. Raymond Lagonigro, Joan Carles Martori, Philippe Apparicio, Understanding Airbnb spatial distribution in a southern European city: The case of Barcelona, https://www.sciencedirect.com/science/artic
4. https://insideairbnb.com/get-the-data/
5. Teruel-Gutierrez, R., Maté-Sánchez-Val, M. The impact of Instagram on Airbnb's listing prices in the city of Barcelona. Ann Reg Sci 67, 737–763 (2021). https://doi.org/10.1007/s00168-021-01064-z