European Management Studies
Vol. 22, No. 4, p. 33–58, e-ISSN: 2956-7602
https://doi.org/10.7172/2956-7602.106.2

Private Rooms in Peer-to-Peer Accommodation: Employing Machine Learning to Determine Revenue Drivers

Ewa E. Kiczmachowska

Kozminski University, Poland
https://orcid.org/0000-0002-7336-875X

Submitted: 22.07.2024 | Accepted: 23.01.2025

Abstract

Purpose: The goal of this study was to identify the performance determinants of private rooms offered via peer-to-peer accommodation (P2PA) platforms.

Design/methodology/approach: Based on property level data of Airbnb, the applications of machine learning clustering and artificial neural networks methods were proposed to identify the performance determinants of private room performance defined as yearly revenue.

Findings: The results indicate that the most important revenue determinants were maximum number of guests, number of photos, number of bathrooms, location, and cancellation policy, as well as high availability of the property throughout the year. The highest revenues were achieved by 2–3 person rooms with at least two bathrooms, 4-person rooms, or rooms located around the Central Railway Station. However, properties with at least 2 bathrooms tend to deliver high revenue, regardless of location. Furthermore, a sufficient number of photos and a flexible cancellation policy could offset location disadvantages and deliver higher revenues in distant vs. central locations.

Research limitations/implications: The limitations of this study are that it covered only one destination during a 12-month period, which encompasses yearly seasonality, but might include some exceptional events. Replication for other destinations (urban or non-urban) or timeframes would be valuable future research recommendations.

Originality/value: This study adds threefold to the P2PA performance determinants stream. First, it offers an application of machine learning methods to identify property and host features and behaviours that contribute to high performance. Second, as opposite to the majority of available research, it defines performance as revenue rather than price. Third, building on the hedonic price theory, it sheds light on how to manage a complex offer for private rooms in P2PA, in contrast to entire homes/apartments that researchers have focused on to date.

Keywords: peer-to-peer accommodation, machine learning, revenue, performance determinants, Airbnb.

JEL: D22, L25, L83, M10, Z33

Correspondence address: Kozminski University, 59 Jagiellońska St., 03-301 Warsaw, Poland; e-mail address: ekiczmachowska@kozminski.edu.pl.

Suggested Citation: Kiczmachowska, E.E. (2024). Private Rooms in Peer-to-Peer Accommodation: Employing Machine Learning to Determine Revenue Drivers. *European Management Studies*, *22*(4), 33–58.
https://doi.org/10.7172/2956-7602.106.2

Prywatne pokoje w usługach noclegowych typu peer-to-peer: zastosowanie uczenia maszynowego do określenia determinantów przychodu

Streszczenie

Cel: celem tego projektu było zidentyfikowanie determinantów wyników dla prywatnych pokoi oferowanych w ramach usług noclegowych typu peer-to-peer.

Projekt/metodologia/podejście: bazując na danych Airbnb na poziomie pojedynczych lokali, projekt proponuje zastosowanie uczenia maszynowego i sztucznych sieci neuronowych do określenia determinantów wyników, zdefiniowanych jako roczne przychody, dla prywatnych pokoi oferowanych na platformie usług noclegowych typu peer-to-peer.

Wyniki: wyniki wskazują, że najważniejszymi determinantami przychodów były: maksymalna liczba gości, zdjęć, łazienek, lokalizacja nieruchomości i polityka odwołania rezerwacji, a także wysoka dostępność lokalu do wynajmu w ciągu roku. Najwyższe przychody zostały osiągnięte przez 2–3 osobowe pokoje wyposażone w co najmniej dwie łazienki, pokoje 4-osobowe, lub pokoje zlokalizowane najbliżej stacji kolejowej Warszawa Centralna. Jednakże nieruchomości wyposażone w co najmniej dwie łazienki wykazały tendencję do osiągania wyższych przychodów niezależnie od lokalizacji. Co więcej, wystarczająca liczba zdjęć i elastyczna polityka odwołania rezerwacji mogą zrównoważyć niedogodności lokalizacyjne i dostarczyć wyższych przychodów w porównaniu z centralnymi lokalizacjami.

Ograniczenia/implikacje badawcze: ograniczeniem tego projektu jest to, że badanie obejmowało tylko jedno miasto na przestrzeni 12 miesięcy, co wprawdzie umożliwia odzwierciedlenie rocznej sezonowości, jednak może zawierać zdarzenia wyjątkowe. Powtórzenie analizy na danych z innych lokalizacji (miejskich lub nie) lub innych okresów stanowi rekomendację dla przyszłych badań.

Oryginalność/wartość: projekt wzbogaca nurt literatury opisującej determinanty przychodów na trzy sposoby. Po pierwsze, oferuje zastosowanie metod uczenia maszynowego do określenia cech lokalu oraz cech i zachowań gospodarzy, które sprzyjają osiąganiu wysokich wyników. Po drugie, w odróżnieniu od większości istniejących badań, ten projekt definiuje wyniki jako przychody, a nie jako cenę. Po trzecie, budując na hedonicznej teorii cen, projekt ten rzuca światło na sposoby zarządzania kompleksową ofertą prywatnych pokoi oferowanych w usługach noclegowych typu peer-to-peer, w odróżnieniu od całych domów/mieszkań, na których badacze koncentrowali się do tej pory.

Słowa kluczowe: usługi noclegowe typu peer-to-peer, uczenie maszynowe, przychody, determinanty wyników, Airbnb.

1. Introduction

In the recently developing stream of research devoted to the performance determinants of peer-to-peer accommodation (P2PA), the researchers studied an array of potential performance determinants that covered the features of the property, the features and behaviours of the host, and the reputation factors (Sainaghi, 2021). Usually, the *property type* determined the performance in such a way that entire home/apartment properties generated higher revenue and prices than private or shared rooms. As the vast majority of the P2PA offer is represented by entire homes/apartments, the conclusions referred mostly to them. In times of inflation and economic turbulence, the private room offer,

https://doi.org/10.7172/2956-7602.106.2

being more affordable than the entire home/ apartment, might gain importance, especially among cost-conscious tourists. Thus, this research aims to concentrate on the private room offers that have thus far been underrepresented, yet their performance might be shaped by other factors than is the case with offers of entire homes/apartments. The research question was set as follows:

RQ: What are the performance determinants of private rooms in P2PA?

The available research on P2PA performance determinants, built on the hedonic price theory, predominantly used regression methods to identify factors determining performance, most often defined as price. However, regression modelling has a rigid set of assumptions that are imposed before learning from the data, which might not be possible to fulfil (Camatti et al., 2024). Linearity and multicollinearity issues have turned researchers' attention to more advanced methods, exploring many new techniques based on data mining that are non-parametric in nature (Chong, 2010), with special focus on artificial intelligence techniques (neural networks, decision trees, random forest) (Panahandeh et al., 2025). Following this path, our research focused on machine learning methods that make minimal demands on model structure and assumptions, thus are applicable on broad variety of datasets. Namely, clustering and artificial neural network algorithm were employed to identify the P2PA performance determinants. On top of minimum demands on data, the methods have the advantage of possessing tools to identify predictor importance.

This study adds threefold to the P2PA performance determinants stream. First, using an Airbnb dataset for Warsaw, Poland, it offers an application of machine learning methods to identify property and host features and behaviours that deliver high performance. This methodology is especially useful for datasets with highly skewed variables, facing normality issues, where linear regression modelling is not an option. Second in contrast to the majority of available research, it defines performance as revenue. Third, building on the hedonic price theory, it sheds new light on how to manage a complex offer for private rooms in the urban settings of P2PA, in contrast to entire homes/apartments that the research has been focused on so far. The proposed methods carry minimal demands on datasets and are easily replicable to other settings or geographies, thus are useful for researchers and practitioners.

As with other small and medium enterprises in the hospitality sector, the properties featured on P2PA platforms are usually managed by regular people that have limited access to knowledge, analytical tools, or resources (Floričić, 2016; Gibbs, et al., 2018b), this study constitutes valuable input for their strive to improve business performance and competitiveness.

2. Literature Review

2.1. P2PA Performance Determinants

Apart from *property type*, an array of P2PA performance determinants were identified, such as *size* (*no. of bedrooms, no. of bathrooms, maximum no. of guests*; positive relation to performance) (Wang & Nicolau, 2017; Sainaghi, Abrate & Mauri, 2021; Chapman, Mohammad & Villegas, 2023; Wang, 2023; Camatti et al., 2024; Panahandeh et al., 2025); *location* (*distance to: the city centre; tourist attractions; transportation; shopping*; negative relationship) (Perez-Sanchez et al., 2018; Benítez-Aurioles, 2018a; Sainaghi et al., 2021; Gyódi & Nawaro, 2021; Jang & Kim, 2022; Sainaghi & Chica-Olmo, 2022; Chapman et al., 2023; Panahandeh et al., 2025); *amenities* (*Internet; elevator; parking*; positive relationship) (Chattopadhyay & Mitra, 2019); *host professionalism* (*single-/multi-unit; Superhost*; *amateur/experienced; part-/full-time;* positive relationship) (Chen & Xie, 2017; Deboosere et al., 2019; Berentsen et al., 2019; Xie, Heo & Mao, 2021; Abrate, Sainaghi & Mauri, 2022; Panahandeh et al., 2025); *management* (*cancellation policies, minimum stay, dynamic pricing, response rate, smoking/pets allowed*; mixed relationship) (Wang & Nicolau, 2017; Benítez-Aurioles, 2018b; Sainaghi et al., 2021); *advertising* (*no. of photos; no. of reviews; overall rating*; mixed relationship) (Abrate & Viglia, 2019; Xie et al., 2021).

The performance was most often operationalised as *price* (Chen & Xie, 2017; Wang & Nicolau, 2017; Benítez-Aurioles, 2018b; Gibbs et al., 2018a; Magno, Cassia & Ugolini, 2018; Perez-Sanchez et al., 2018; Chattopadhyay & Mitra, 2019; Sainaghi, Abrate & Mauri, 2021; Casamatta et al., 2022; Ferreira et al., 2023; Wang, 2023; Camatti et al., 2024; Panahandeh et al., 2025), while *revenue* was discussed in only 19% of cases (Sainaghi, 2021).

As for research methods, the existing research included hedonic price modelling (Wang & Nicolau, 2017; Gibbs et al., 2018a); multilevel hedonic analysis (Deboosere et al., 2019); ordinary least squares (Gunter & Onder, 2017); multivariate regression (Perez-Sanchez et al., 2018); random effect models and quantile regressions (Falk, Larpin & Scaglione, 2019); random forest, decision trees and ordinary least squares (Chattopadhyay & Mitra, 2019); geographically weighted regression (Zhang et al., 2017) as well as advanced machine learning techniques such as ensemble frameworks (Ghosh, Jana & Abedin, 2023). Islam et al. (2022) employed methods such as XGBoost and Latent Dirichlet Allocation. Next, neural networks were utilised by Kalehbasti, Nikolenko and Rezaei (2021) in comparison to linear regression, tree-based models, support-vector regression, and K-means clustering. Further comparative studies exploring the effectiveness of conventional approaches (linear and generalised linear models), and artificial intelligence techniques (random forest, decision trees, neural networks) are also available from Cai, Zhou and Scott (2019), Kokasih and Paramita (2020), Chapman et al. (2023), Wang (2023), Kanakaris and Karacapilidis (2023), Yang (2024), Camatti, et al. (2024), and Panahandeh et al. (2025). In summary, the artificial intelligence techniques proved to deliver better predictive accuracy vs.

https://doi.org/10.7172/2956-7602.106.2

conventional approaches. The limited predictive accuracy of the linear regression may be attributed to its inability to capture the non-linear relationships and interactions among these variables (Yang, 2024).

2.2. Clustering Methods

Clustering methods aim at identifying clusters of objects that are as similar as possible to other objects within a cluster, and as different as possible to the objects in other clusters (Wierzbiński, 2009; Yu, Dong & Yao, 2022). Among the three major types of clustering algorithms ([1] hierarchical clustering, [2] non-hierarchical clustering, and [3] two-step clustering), the lattermost is considered the most versatile (Chong, 2010). This is because it has several desirable features that other clustering methods do not carry. Besides being capable of automatically returning the number of clusters and scalability that allows the efficient analysis of thousands of observations (Chiu et al., 2001), the algorithm is reasonably robust against the violation of variables' normal distribution assumptions, and thus renders assumption checking unnecessary (Chong, 2010).

2.3. Artificial Neural Networks

Artificial neural networks are very popular for modelling non-linear problems and for the prediction of the output values for given input parameters from their training values. Traditionally-used linear regression has a rigid model structure and set of assumptions that are imposed before learning from the data. By contrast, the neural networks make minimal demands on model structure and assumptions. Thus, a neural network can approximate a wide range of statistical models without requiring to hypothesise in advance certain relationships between the variables. Instead, the form of the relationships is determined during the learning process. If a linear relationship between the dependent and independent variables is appropriate, the results of the neural network should closely approximate those of the linear regression model. If a nonlinear relationship is more appropriate, the neural network will automatically approximate the 'correct' model structure (IBM, 2024).

Multilayer perceptron (MLP) is one of the most commonly used types of artificial neural networks; it utilises backpropagation for training (a supervised learning technique) (Gorjani et al., 2021). MLP is composed of multiple layers, including an input layer, hidden layers, and an output layer, where each layer contains a set of perception elements, known as 'neurons'. The multilayer perceptron algorithm can be used to build a prediction system or forecasting model (Chan et al., 2023). The results confirm claims made by many researchers stating the superiority of neural network over statistical models in prediction tasks (Ismail et al., 2015; Pirmohammadi & Mast, 2020).

3. Data and Methods

3.1. Data

The source of the data was AirDNA, the worldwide provider of property-level data of Airbnb and Vrbo (AirDNA, 2023). The dataset covered 9,599 Airbnb properties based in Warsaw, Poland, that delivered any revenue in the 12-months from March 2022 to February 2023, of which private rooms accounted for 1,614 properties. After several exclusions covering *minimum stay* >31, *no. of bedrooms* >1 and outliers for Annual Revenue (AR) and Average Daily Rate (ADR), the final dataset included 1,444 properties.

The list of variables and their measurement is displayed in Table 1.

Table 1
The variables

<table>
<tr><th>Type</th><th>Group</th><th>Variable</th><th>Abbrevia-tion</th><th>Measurement</th><th>Source</th></tr>
<tr><td rowspan="4">Performance</td><td rowspan="4">Revenue</td><td>Annual Revenue</td><td>AR</td><td>USD</td><td>AirDNA</td></tr>
<tr><td>Median Annual Revenue</td><td>MAR</td><td>1=low, 2=high</td><td>AirDNA</td></tr>
<tr><td>Tertile Annual Revenue</td><td>TAR</td><td>1=low, 2=med, 3=high</td><td>AirDNA</td></tr>
<tr><td>Quartile Annual Revenue</td><td>QAR</td><td>1=low, 2=lower med, 3=upper med, 4=high</td><td>AirDNA</td></tr>
<tr><td rowspan="5">Property</td><td rowspan="3">Size</td><td>Maximum no. of guests</td><td>Max Guests</td><td>1; 2; 3; ...; 8</td><td>AirDNA</td></tr>
<tr><td>No. of bedrooms</td><td>Bedrooms</td><td>0; 1</td><td>AirDNA</td></tr>
<tr><td>No. of bathrooms</td><td>Bathrooms</td><td>0; 1; 1,5; 2; 2,5; ...; 6</td><td>AirDNA</td></tr>
<tr><td rowspan="2">Location</td><td>Distance to Central Railway Station</td><td>CRS</td><td rowspan="2">latitude/longitude orthodromic distance (km)</td><td rowspan="2">AirDNA</td></tr>
<tr><td>Distance to Old Town Square</td><td>OTS</td></tr>
<tr><td rowspan="6">Host</td><td rowspan="3">Manage-ment</td><td>Cancellation Policy</td><td>CP</td><td>0; 1; 2; 3</td><td>AirDNA</td></tr>
<tr><td>Minimum Stay</td><td>Min Stay</td><td>1; 2; 3; ...; 31</td><td>AirDNA</td></tr>
<tr><td>Pets Allowed</td><td>Pets</td><td>0; 1</td><td>AirDNA</td></tr>
<tr><td rowspan="2">Profes-siona-lism</td><td>Superhost</td><td>Superhost</td><td>0; 1</td><td>AirDNA</td></tr>
<tr><td>No. of days blocked</td><td>Days Blocked</td><td>0; 1; 2; ...; 364</td><td>AirDNA</td></tr>
<tr><td>Adverti-sing</td><td>No. of photos</td><td>Photos</td><td>1; 2; 2; ...; 92</td><td>AirDNA</td></tr>
</table>

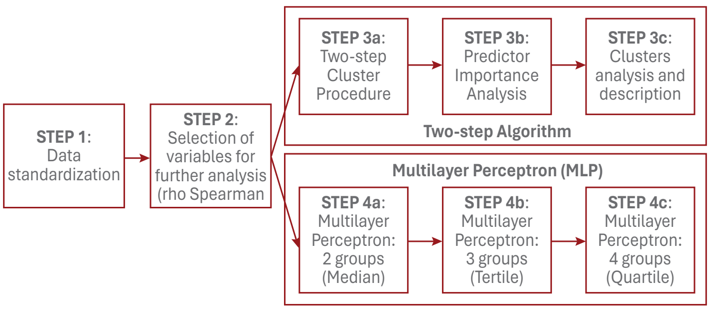
https://doi.org/10.7172/2956-7602.106.2

As for the details, three of above-mentioned variables need additional explanation. Airbnb provides their operators (hosts) with four standard cancellation policies (*CPs*) (Airbnb, 2023a), which define cancellation periods and the terms of refund, and vary from flexible conditions (coded = 0) to strict (coded = 4). In case of minimum-length-of-stay, the host can set the property's minimum rent period for 1 or more day. For this study, the *minimum stay* variable was developed as number of days set by the host, from 1 to 31. Finally, the *Superhost* status is a set of certain conditions to be met by the host that covers high activity rate, high response rate, high overall rating, and low cancellation rate (Airbnb, 2023b). As the majority of the variables had far from normal distribution, the application of regression models was not an option. Therefore, alternative approaches of machine learning were employed.

3.2. Methods

The analysis was delivered in several steps (Figure 1). First, as advocated by previous research (e.g., Sharaf Addin et al., 2022), the data were standardised, as they were expressed in different measurement units (Table 1). Second, variables significantly correlated to AR (Spearman's Rho) were selected to be applied in further analysis. Then, two procedures were carried out, namely Two-step Cluster Algorithm, and Multilayer Perceptron (MLP) using IBM SPSS. In Two-step Cluster Algorithm (Steps 3a, 3b, 3c), the IBM SPSS procedure was run, then the Predictor Importance analysis was made, and the detailed clusters' analysis and profiling were made, based on selected variables by property. In the MLP stream, the procedures were repeatedly run for three scenarios: (4a) assigning the properties to 2 groups (Median, low/high AR); (4b) assigning the properties to 3 groups (Tertile, low/medium/high AR); (4c) assigning the properties to 4 groups (Quartile, low/lower medium/upper medium/ high AR). As with each rerun of the MLP algorithm, the system randomly assigns cases to the Training and Testing groups, rerunning leads to different results each time. Therefore, usually several attempts to run the MLP should be taken in order to check the stability of the solution.

Figure 1
Analysis procedure



4. Results

4.1. Two-step Clustering Algorithm

Eight variables were implemented into the Two-step clustering procedure (marked in grey in Table 2): seven variables exhibiting significant relations to AR, and AR was introduced into the Evaluation field.

The Silhouette measure for the model with 6 clusters was 0.3, which suggests the quality of the model was fair (Figure 2). The ratio of sizes between the largest and the smallest cluster was 3.42, which is acceptable.

As far as predictor importance was concerned, the *distance to* variables (OTS and CRS) turned out to be the most important in shaping the clusters (Figure 3), followed by *CP*, *Max Guests*, and *no. of days blocked*. The least important, yet still with importance above 0.25, were *no. of bathrooms* and *no. of photos*.

Final clusters' centres and short descriptions are displayed in Table 3. To serve the study objective, detailed depiction will cover only the highest revenue clusters (3, 2, and 4).

Cluster 3. 2 bathrooms, many photos (11% of properties). This cluster was characterised by the availability of 2 *bathrooms* and many *photos* (avg. of 19.4). In the case of room renting, the shared bathroom can become an annoying bottleneck for guests, especially when there are several rooms rented in the same property or the bathroom is shared with the host. Apparently, it is also important for the guests to see the pictures of not only the bedroom, but also the shared areas (e.g., kitchen, bathroom, common room), which might be the reason why the properties with higher *no. of photos* fell into this high AR cluster.

Cluster 2. 4-person (8.2% of properties). This cluster included predominantly 4-person rooms or bigger (68% and 15% respectively), leading to the conclusion that higher room capacity was the factor connected to above-average AR. However, this cluster was the smallest of all, suggesting that it was the least popular room configuration in the market.

Cluster 4. Best location, flexible CP (21.7% of properties). This cluster was defined by the best *location*, including properties within a 1 km (33%) and 2 km (31%) radius from CRS. Also, this was the cluster with the most flexible CP and the fewest *days blocked* (on average only 43.6 *days blocked*). Despite the rooms being considerably smaller (70%: 2-person) than those in Cluster 2, the AR was on average only 11% lower.

All of the best revenue performing clusters were very active throughout the year, with the lowest average *no. of days blocked*.

Table 2
Correlation table (Spearman's Rho)

	Variable	2	3	4	5	6	7	8	9	10	11	12	13	14
1	C. Railway Station (CRS)	.737**	-.083**	0.009	0.005	.173**	.116**	-.292**	-.088**	0.016	..226**	-.115**	-0.046	-.242**
		0.000	0.002	0.740	0.846	0.000	0.000	0.000	0.001	0.548	0.000	0.000	0.080	0.000
2	Old Town Square (OTS)	1.000	-.078**	.087**	0.016	.187**	.094**	-.259**	-0.012	.112**	.174**	-.097**	-0.008	-.188**
			0.003	0.001	0.541	0.000	0.000	0.000	0.656	0.000	0.000	0.000	0.759	0.000
3	Bedrooms	-.078**	1.000	-0.021	-0.024	0.039	-0.025	0.025	-0.014	-0.025	0.010	-0.002	0.003	0.009
		0.003		0.425	0.353	0.140	0.341	0.344	0.584	0.342	0.692	0.935	0.907	0.733
4	Bathrooms	.087**	-0.021	1.000	-0.023	.143**	0.046	-.089**	.148**	.181**	-.071**	.089**	-.091**	.098**
		0.001	0.425		0.393	0.000	0.081	0.001	0.000	0.000	0.007	0.001	0.001	0.000
5	Max Guests	0.016	-0.024	-0.023	1.000	-0.010	-.180**	-0.027	-.177**	.056*	.180**	.228**	.456**	-.218**
		0.541	0.353	0.393		0.714	0.000	0.303	0.000	0.032	0.000	0.000	0.000	0.000
6	Airbnb Superhost	.187**	0.039	.143**	-0.010	1.000	.106**	-.139**	0.047	.176**	-0.039	0.035	-.185**	.064*
		0.000	0.140	0.000	0.714		0.000	0.000	0.072	0.000	0.141	0.178	0.000	0.015
7	No. of days blocked	.094**	-0.025	0.046	-.180**	.106**	1.000	-0.015	.168**	.127**	-0.033	-.330**	-.321**	.188**
		0.000	0.341	0.081	0.000	0.000		0.574	0.000	0.000	0.209	0.000	0.000	0.000
8	Cancellation Policy (CP)	-.259**	0.025	-.089**	-0.027	-.139**	-0.015	1.000	-0.015	-.074**	-.071**	-.085**	-0.012	.109**
		0.000	0.344	0.001	0.303	0.000	0.574		0.559	0.005	0.007	0.001	0.638	0.000
9	Minimum Stay	-0.012	-0.014	.148**	-.177**	0.047	.168**	-0.015	1.000	.163**	-.142**	-0.006	-.147**	.238**
		0.656	0.584	0.000	0.000	0.072	0.000	0.559		0.000	0.000	0.829	0.000	0.000

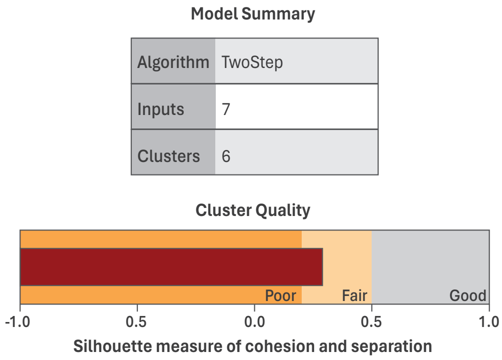
Table 2 (cont.)

	Variable	2	3	4	5	6	7	8	9	10	11	12	13	14
10	No. of Photos	.112**	-0.025	.181**	.056*	.176**	.127**	-.074**	.163**	1.000	.057*	.129**	.130**	0.027
		0.000	0.342	0.000	0.032	0.000	0.000	0.005	0.000		0.030	0.000	0.000	0.302
11	Pets Allowed	.174**	0.010	-.071**	.180**	-0.039	-0.033	-.071**	-.142**	.057*	1.000	-0.006	.377**	-.200**
		0.000	0.692	0.007	0.000	0.141	0.209	0.007	0.000	0.030		0.814	0.000	0.000
12	Annual Revenue (AR)	-.097**	-0.002	.089**	.228**	0.035	-.330**	-.085**	-0.006	.129**	-0.006	1.000	.270**	.256**
		0.000	0.935	0.001	0.000	0.178	0.000	0.001	0.829	0.000	0.814		0.000	0.000
13	Average Daily Rate	-0.008	0.003	-.091**	.456**	-.185**	-.321**	-0.012	-.147**	.130**	.377**	.270**	1.000	-.455**
		0.759	0.907	0.001	0.000	0.000	0.000	0.638	0.000	0.000	0.000	0.000		0.000
14	Occupancy Rate	-.188**	0.009	.098**	-.218**	.064*	.188**	.109**	.238**	0.027	-.200**	.256**	-.455**	1.000
		0.000	0.733	0.000	0.000	0.015	0.000	0.000	0.000	0.302	0.000	0.000	0.000	

** $p < .001$; * $p < .005$; $N = 1{,}444$

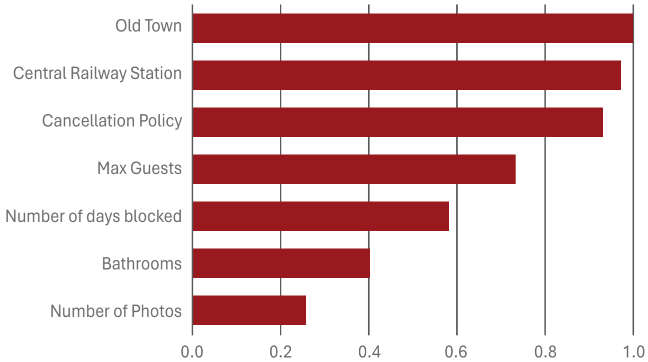
Source: IBM, SPSS based on AirDNA.

Figure 2
Two-step clustering Model Summary



Source: IBM, SPSS based on AirDNA.

Figure 3
Predictor Importance



Source: IBM, SPSS based on AirDNA.

Table 3
Final cluster characteristics

	Cluster	# / %	AR	OTS	CRS	CP	Max Guests	Blocked Days	Bathrms	Photos
3	2–3 person, **2 bathrooms, many photos, active**	159	high	medium	medium	moderate	2–3	low	2	high
		11.0%	5,274.97	3.71	2.80	0.98	2.13	51.74	1.90	19.43
2	**4-person, active**	119	high	medium	medium	moderate	4	low	1	low
		8.2%	5,209.74	3.98	3.43	1.20	4.13	50.79	1.05	8.46
4	1–2 person, **central location, flex/ moderate CP, active**	314	high	close	close	flexible/ moderate	1–2	low	1	low
		21.7%	4,678.68	2.67	2.04	0.44	1.79	43.57	1.07	8.66
5	1–2 person, **distant, flex CP**	262	medium	distant	distant	flexible/ moderate	1–2	medium	1	medium
		18.1%	2,446.30	9.79	8.59	0.49	1.95	73.16	1.13	10.76
1	1–2 person, **strict CP, central location**	407	medium	close	close	firm/strict	1-–2	medium	1	low
		28.2%	2,166.75	2.54	2.02	2.69	1.79	79.44	1.05	8.23
6	1–2 person, **high blocked days, flex CP**	183	low	medium	medium	flexible/ moderate	1–2	high	1	medium
		12.7%	1,278.57	4.35	3.79	0.55	1.71	258.61	1.22	12.57

Source: Author's elaboration based on AirDNA.

https://doi.org/10.7172/2956-7602.106.2

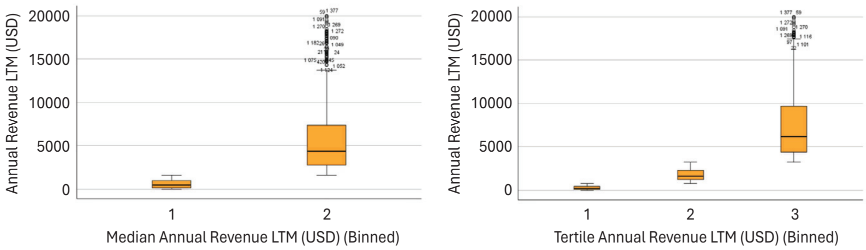
4.2. Multilayer Perceptron (MLP)

The variable used for assigning properties to two groups (Median Annual Revenue) was created by using the Median, while the three groups variable (Tertile Annual Revenue) was created using the Tertile. Similarly, the four groups variable was created using the Quartile (Quartile Annual Revenue), however, as the accuracy of the model was not satisfactory, only Median and Tertile procedures will be further described. The details of the groups are depicted in Table 4 and visualised in Figure 4. Because the AR variable was highly skewed (1.960; se=.064), the ranges of the clusters in both solutions were not equal.

Table 4
Median and Tertile group split. Descriptive statistics

Descriptives	Median:	Cluster	Tertile:	Cluster	
Annual Revenue	1	2	1	2	3
N	722	722	482	481	481
Mean	597	5.892	292	1.770	7.677
Std. Deviation	495	4.330	235	676	4.299
Min	7	1.635	7	792	3.249
Max	1.634	19.979	789	3.229	19.979
Range	1.627	18.344	782	2.437	16.730

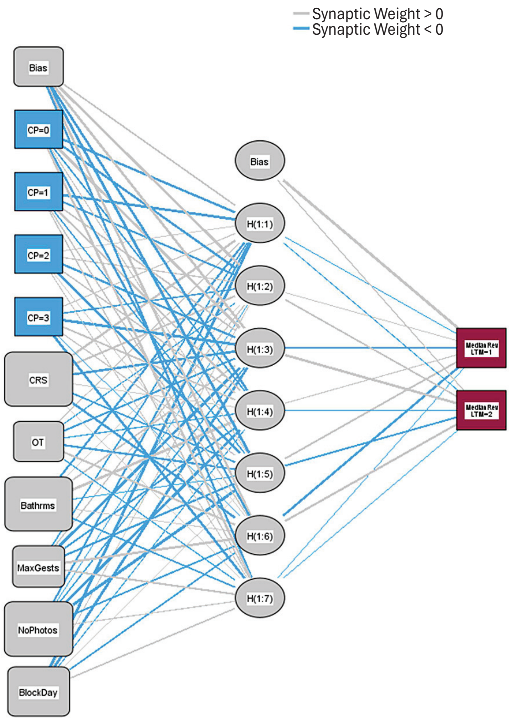
Figure 4
AR variable distribution across Median and Tertile-created groups



The MLP diagrams are shown in Figure 5. The MLP diagram for the 2-group solution (Median) comprised 10 input variables, seven neurons in the hidden layer, and two output levels (low and high AR), while for the 3-group solution (Tertile) it comprised 10 input layers, eight neurons, and three output levels (low, medium, and high). The input variables included CP (4 levels), *distance to* CRS, *distance to* OTS, *no. of bathrooms*, *maximum no. of guests*, *no. of photos*, and *no. of days blocked*. Of the 1,444 properties, approx. 70% were used to train the network, and 30% were used to test the network.

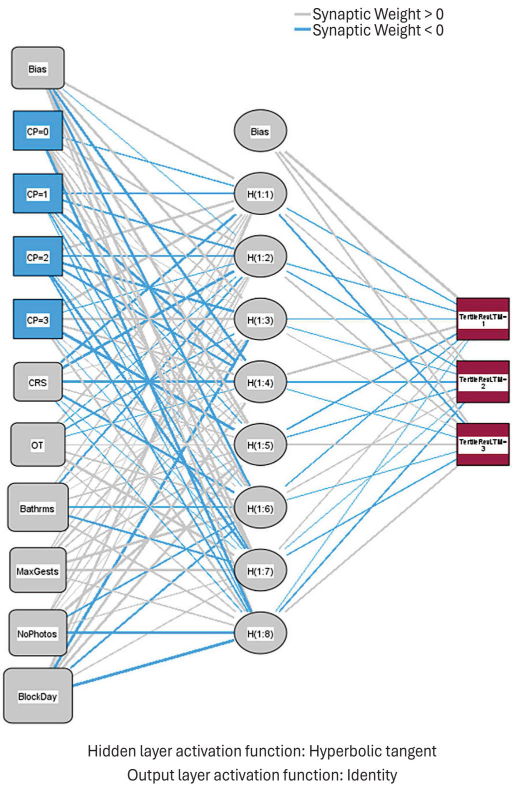
Figure 5
MLP diagrams for Median (Model 1) and Tertile (Model 3) model exemplifications

Model 1. (Median)



Hidden layer activation function: Hyperbolic tangent
Output layer activation function: Identity

Model 3. (Tertile)



Source: SPSS Multilayer Perceptron.

As indicated earlier, rerunning MLPs returns different results, as each time the assignment of the items to training and testing groups is random, therefore it is advisable to run the MLP several times to check the stability of the model. MLPs run for low/high (Median split) AR solution returned models with incorrect prediction stable levels within 30% to 35%; the best performing models summaries are shown in Table 5. The Training Models returned Incorrect predictions of 29.9% for Model 1 and 33.7% for Model 2, while Testing Models returned Incorrect Predictions of, respectively, 31.1% and 31.5%. MLPs run for low/ medium/ high (Tertile split) AR solution (Model 3 and Model 4), returned considerably higher Percent of Incorrect Predictions (from 45.9% to 48.5%), which would make the models not appropriate for further usage.

Table 5
Model Summary

Model Summary		Model 1[b] (Median)	Model 2[b] (Median)	Model 3[c] (Tertile)	Model 4[c] (Tertile)
Training	Sum of Squares Error	204.625	217.703	293.509	288.553
	Percent Incorrect Predictions	29.9%	33.7%	47.7%	45.9%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]			
	Training Time	0:00:00,09	0:00:00,09	0:00:00,05	0:00:00,06
Testing	Sum of Squares Error	89.699	81.844	128.054	127.233
	Percent Incorrect Predictions	31.1%	31.5%	47.6%	48.5%

[a] Error computations are based on the testing sample.
[b] Dependent Variable: Median Annual Revenue (Binned).
[c] Dependent Variable: Tertile Annual Revenue (Binned).

Source: SPSS Multilayer Perceptron.

The Classification Table (Table 6) indicated slightly higher Percent Predicted Correct for High AR properties (75.8% Predicted Correct for Model 1, and 75.1% for Model 2), which could indicate somewhat higher model accuracy in predicting high AR properties.

Overall Percent Predicted Correct was similar across Training and Testing environments, 70.1% and 68.9% for Model 1, and 66.3% and 68.5% for Model 2. Equally, Models 3 and 4 (Table 6) scored higher on High AR Percent Predicted Correct (73.9%; 74.3% for Model 3; and 71.1%; 68.3% for Model 4, Training and Testing samples respectively), which could indicate higher model accuracy in predicting High AR properties.

In MLPs, the Area Under the ROC Curve indicates how optimal the model is. In this study, we adopted the following criteria to interpret the area under the ROC curve: (A) 0.90–1 = excellent; (B) 0.80–0.90 = good; (C) 0.70–0.80 = fair; (D) 0.60–0.70 = poor; (E) 0.50–0.60 = fail (Aryadoust & Goh, 2014). In our case, for Median solution, both models indicated fair levels (Table 7), attaining .749 for Model 1 and .734 for Model 2, while for Models 3 and 4, the predictions of Medium AR grouping have just crossed the poor level border (.600 for Model 3, and .608 for Model 4), but the Low and High AR groups results fell well within the limits of fair (.738 and .748 for Low AR group; .767 and .771 for High AR group).

https://doi.org/10.7172/2956-7602.106.2

Table 6
Classification table

Sample		Model 1[a] (Median)			Model 2[a] (Median)			Model 3[b] (Tertile)				Model 4[b] (Tertile)			
		Predicted			Predicted			Predicted				Predicted			
		1	2	% Correct	1	2	% Correct	1	2	3	% Correct	1	2	3	% Correct
Training	1	353	163	68.4%	324	186	63.5%	196	52	84	59.0%	184	66	75	56.6%
	2	142	361	71.8%	161	360	69.1%	114	80	142	23.8%	96	125	131	35.5%
	3	–	–	–	–	–	–	51	38	252	73.9%	38	59	239	71.1%
	Overall %	48.6%	51.4%	70.1%	47.0%	53.0%	66.3%	35.8%	16.8%	47.4%	52.3%	31.4%	24.7%	43.9%	54.1%
Testing	1	127	79	61.7%	132	80	62.3%	85	26	39	56.7%	83	40	34	52.9%
	2	53	166	75.8%	50	151	75.1%	45	39	61	26.9%	42	40	47	31.0%
	3	–	–	–	–	–	–	18	18	104	74.3%	19	27	99	68.3%
	Overall %	42.4%	57.6%	68.9%	44.1%	55.9%	68.5%	34.0%	19.1%	46.9%	52.4%	33.4%	24.8%	41.8%	51.5%

[a] Dependent Variable: Median Annual Revenue (USD) (Binned).
[b] Dependent Variable: Tertile Annual Revenue (USD) (Binned).

Source: SPSS Multilayer Perceptron.

Table 7
Area Under the ROC Curve

	Model 1 (Median)	**Model 2 (Median)**	**Model 3 (Tertile)**	**Model 4 (Tertile)**
Group	Area	Area	Area	Area
1	0.749	0.734	0.738	0.748
2	0.749	0.734	0.600	0.608
3	–	–	0.767	0.771

Source: SPSS Multilayer Perceptron.

Finally, a normalised Importance Index for each independent variable was calculated, which indicates the weight of each independent variable in predicting the dependent variable, and ranges from 0% to 100%. Higher indices indicate the higher contribution of the variable to predicting or classifying the dependent variable (Aryadoust & Goh, 2014). In our case, the importance of the independent variables differed across the four models (Table 8), which was further depicted in Figure 6. In both Median models, the *distance to* CRS, *no. of days blocked* and *no. of photos* scored high (above 70%), CP and *distance to* OTS scored lower (46–57%), while the role of *no. of bathrooms* differed considerably (97.7% in Model 1, vs. 41.9% in Model 2). However, as both models delivered similar Area in ROC Curve results (0.749; 0.734) and Percent correct (68.9%; 68.5%), we could assume an important role (above 40%) of the *no. of bathrooms* variable in predicting the AR. In both Tertile models, the most important variable was *no. of blocked days* (100% in both cases). As with the Median models, CP scored low (33.3%; 39.7%), while the role of *distance to* CRS scored considerably lower than was the case in Median models (35.8%; 38.1%), which could indicate a lower importance of CRS in grouping the properties into Tertile groups.

In summary, the MLP models of Median and Tertile delivered satisfactory results in predicting the AR, especially in high AR ranges. Further attempts to run the MLP models for the Quartile split of properties according to their AR, when they delivered satisfactory performance on high AR properties, they failed to perform on medium ranges (2nd and 3rd quartiles), as well as on total model performance, therefore the detailed analysis will not be provided here.

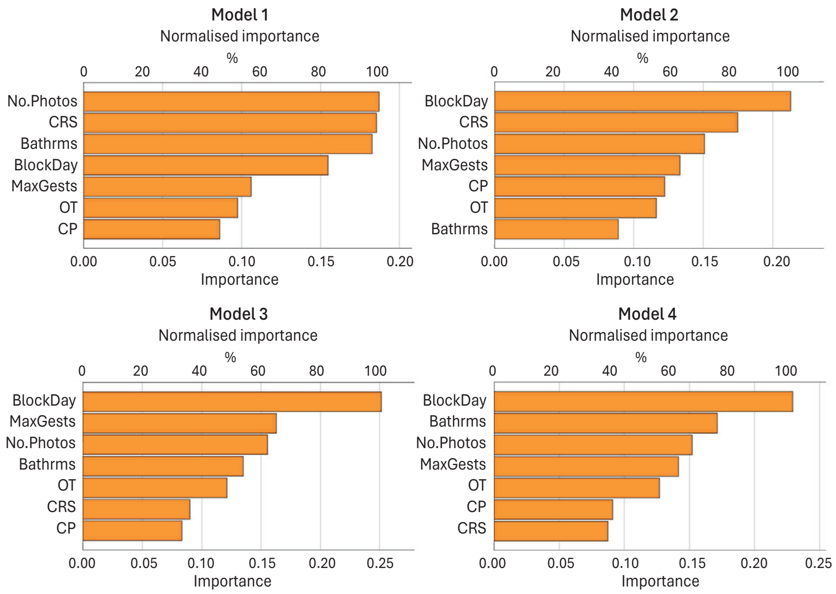
https://doi.org/10.7172/2956-7602.106.2

Table 8
Independent Variable Importance

	Model 1 (Median)		Model 2 (Median)		Model 3 (Tertile)		Model 4 (Tertile)	
	Importance	Normalised Importance	Importance	Normalised Importance	Importance	Normalised Importance	Importance	Normalised Importance
CP	0.086	46.0%	0.122	57.4%	0.084	33.3%	0.091	39.7%
CRS	0.185	99.1%	0.175	82.1%	0.090	35.8%	0.087	38.1%
OTS	0.098	52.1%	0.116	54.7%	0.121	48.4%	0.127	55.5%
Bathrooms	0.183	97.7%	0.089	41.9%	0.135	53.8%	0.171	74.7%
Max Guests	0.106	56.8%	0.133	62.6%	0.163	64.9%	0.142	61.8%
No. Photos	0.187	100.0%	0.151	71.0%	0.155	61.8%	0.152	66.4%
Days Blocked	0.155	82.7%	0.213	100.0%	0.251	100.0%	0.229	100.0%

Source: SPSS Multilayer Perceptron.

Figure 6
Independent Variables Normalized Importance



Source: SPSS Multilayer Perceptron.

5. Discussion

The clusters with the highest *annual revenue* were characterised by either the best *location* (4), a high *maximum-no-of-guests* (2) or more than 1 *bathroom* availability (3). This is in line with the previous research that confirmed a negative relationship between performance and *location* (defined as *distance to*) and a positive relationship between performance and *size* (*maximum no. of-guests, no. of bathrooms*) (Wang & Nicolau, 2017; Benítez-Aurioles, 2018a; Sainaghi, 2021; Sainaghi et al., 2021; Jang & Kim, 2022). However, in the case of previous research mainly covering entire house/ apartment property type, *no-of-bathrooms* has usually played secondary importance vs. *maximum no. of guests*, and *no. of bedrooms*.

This study employed more than one reference point for the *location* variable. In addition to *city centre* (OTS) (Wang & Nicolau, 2017; Gibbs et al., 2018a; Benítez-Aurioles, 2018a; Sainaghi et al., 2021), this study proposed a *transportation hub* (CRS) as a second reference point. In every cluster the average *distance to* CRS was smaller than to OTS, which indicated that the transportation hub was more the centre of gravity for the *location* factor of Airbnb private rooms than the historical city-centre / main tourist spot.

In terms of *CP*, among the highest revenue clusters (2, 3), the moderate *CP* prevailed, while the strictest *CP* cluster (1) delivered almost the lowest revenues. This partly contradicted the findings of Sainaghi et al. (2021), who advocated that changing from flexible to moderate or strict *CP* positively affected revenue; while Wang and Nicolau (2017) and Benítez-Aurioles (2018b) found that flexible *CP* had negative impact on price.

The *minimum stay* variable has usually favoured 1-day *minimum stay* properties in terms of house/apartment performance (Sainaghi et al., 2021). This study did not confirm this pattern, as the *minimum stay* variable did not reveal any significant relation to *revenue* (*p=.829*). This likely was due to the fact that the vast majority of the private rooms offered 1- or 2-day *minimum stay* (67% and 16% respectively), so it was not identified as a differentiating factor.

The market was dominated by 2-person (62%) and 1-person (22%) rooms. Among six analysed clusters, four offered mostly 1–2 person rooms. The interesting finding among clusters 5 and 1 was that the *location* (defined as *distance to*) was not a negative decisive factor as far as *revenue* was concerned. In fact, it was the opposite. Cluster 5, which included the most distant locations, thanks to higher *no. of photos* and more flexible *CP*, was able to generate higher *revenues* than was the case with properties located much closer to CRS or OTS (Cluster 1). The cluster with the lowest *revenue* (6), although, in comparison to other 1–2 people clusters, occupied medium quality *locations*, limited their *revenue* potential by being available only for a limited number of days (on average 258.61 *days blocked*).

As for *advertising* variable, the *no. of photos* should be a minimum of 9–11, the better to exhibit not only the bedroom but also the shared areas (kitchen, bathroom, common room, etc.).

The neural network analysis confirmed the overall findings of clustering analysis. However, some additional points could be stressed. First, the results further indicated that location variables could be of lesser importance than is assumed with the correlation analysis. Especially, the *distance to* OTS repeatedly fell below the average Normalised Importance of the employed variables, while the *distance to* CRS presented a mixed picture depending on the dependent variable definition (dichotomic vs. 3-level variable). The reason for it could be that the private rooms are considered by the Guests as the affordable option of accommodation. As such, this type of accommodation addresses the needs of the most cost-conscious tourists. In search for the best option to spend several nights in the tourist-destination city, cost-conscious tourists might be prepared to offset the price with the location, provided other features fulfil their expectations. Second, *no. of photos* consistently maintains high importance in shaping the annual revenue, whether it's the AR variable definition or the model version. Again, for the Guests it is evidently important to see not only the bedroom, but also the shared areas, so the properties providing a high *no. of photos* can expect to achieve higher revenues. Third, as long as the accuracy of the models is concerned, the neural networks seem to more accurately assign the properties of high revenue. This could be the result of the AR variable distribution, being very much skewed, suggesting the vast majority of the properties achieve low and very low AR. This is, however, good news in terms of the objective of this study: to define the high performance determinants of private rooms in P2PA. Namely, the determinants applied into the models would be more accurate in describing the high-revenue properties.

6. Conclusions, Limitations, and Future Research Recommendations

Building on the hedonic price theory, which regards the price as the sum of customer valuations of a certain set of product (or service) characteristics (Lancaster, 1966; Chen & Xie, 2017), this research exploited and enhanced the existing research on *P2PA performance determinants*, centred around property or host characteristics and behaviours that impact the performance. First, we employed *clustering* and *neural network* methods to define combinations of property features and host behaviours that return high revenues. These methods carry low demands on normality and multicollinearity of data, and therefore can be used on highly differentiated datasets. Second, unlike available studies that mostly used *price* as performance indicator, this study employed *annual revenue*, that encompasses *price* and *occupancy rate* effects. Third, it focused on *private rooms*, which allowed to depict some features specific for this property type.

In terms of the *location* variable, measured as *distance to*, the interesting finding was that among 1–2 person rooms, the most distant rooms delivered on average 12.9% more *revenue* than rooms that were centrally located. This was thanks to the flexible *CP*, sufficient *no. of photos* and free parking, which would likely attract guests arriving by car. Further, following neural networks method results, higher importance seemed to be given to nearness of the transportation hub than nearness to the main tourist attraction. Also, the results indicated that the *location* could be of lesser importance than to *size* or more than 1 *bathroom* availability. Second, the availability of more than 1 *bathroom* proved to be an important *revenue* building factor for *private rooms*, as it could ease the annoying bottlenecks in multi-room properties. Third, neither *minimum stay* nor *Superhost* variables were related to the *revenue* performance – one probable explanation could be the high percentage of properties offering no *minimum stay* restrictions and low percentage of rooms run by *Superhosts* in the analysed dataset. Furthermore, according to neural networks methods, *no. of photos* consistently appeared among the top 3 factors for normalised importance, stressing the value of presenting the visuals of shared areas, and even of the surroundings. Finally, as a majority of the rooms in high *revenue* clusters offered moderate *CP*, it could be an indication of this *CP* to deliver proper balance between the interests of the hosts and those of the guests.

https://doi.org/10.7172/2956-7602.106.2

As for practical implications, this study brings value to current and future hosts, as well as to P2PA platforms (such as Airbnb), by offering a model of a set of private room features and behaviours designed to deliver the highest *revenue*. For current hosts it outlines how to improve the existing offer; switching from strict or firm to moderate *CP* and displaying at least 9 to 11 *photos* would be the fastest way to do so, whereas rearranging the offer by making one more *bathroom*[s] available or fitting 1 more person in the room would be more demanding. For future hosts, to maximise the *annual revenue*, several options could be pursued, of which the 2–3 person room within a 3 km radius from CRS would be the most promising. For distant locations, availability of either the large room (4-person), or 2 or more *bathrooms* option would be recommended, and the availability of *free-parking* should be highlighted.

The limitations of this study are that it covered only one destination during a 12-month period, which while encompassing yearly seasonality might include some exceptional events. However, implementation of the proposed methodology is fast, simple, and suitable for different data types (Akbar, Liu & Latif, 2020), which makes it easy to replicate for other destinations (urban or non-urban) and timeframes. These would be potentially valuable future research recommendations. Further, in terms of the *location* variable, this study employed a mainstream solution of orthodromic distance to two reference points (transportation hub, and city centre), while more advanced spatial analysis would be recommended for future research, taking into account access to municipal transportation (e.g., an underground/subway/metro), or natural barriers (e.g., a river). Also, this study concentrated mainly on the *property* and *host* features and behaviours that were previously proved to impact P2PA performance. However, in the case of *private rooms*, there might be other specific factors that could influence the price, which offer interesting future research options, e.g., private bathroom availability, or whether the host does or does not live in the rented house/apartment.

Funding

This work was supported by the Kozminski University Warsaw research support fund.

Declaration of Conflicting Interests

The author declare that there is no conflict of interest related to the research paper. The study was conducted without any financial or personal relationships that could be perceived as potentially influencing the research or its outcomes.

Declaration about the scope of AI utilization

The author did not use an AI tool in the preparation of the article.

References

Abrate, G., Sainaghi, R., & Mauri, A.G. (2022). Dynamic pricing in Airbnb: Individual versus professional hosts. *Journal of Business Research*, *141*(December 2021), 191–199. https://doi.org/10.1016/j.jbusres.2021.12.012

Abrate, G., & Viglia, G. (2019). Personal or Product Reputation? Optimizing Revenues in the Sharing Economy. *Journal of Travel Research*, *58*(1), 136–148. https://doi.org/10.1177/0047287517741998

Airbnb (2023a). *Choose the right cancellation policy for you*. Retrieved from Choose the right cancellation policy for you – Resource Centre – Airbnb (2.10.2023).

Airbnb (2023b). *How to become a Superhost*. https://www.airbnb.co.uk/help/article/829/ (26.10.2023).

AirDNA (2023). *About AirDNA*. https://www.airdna.co/about (2.10.2023).

Akbar, Z., Liu, J., & Latif, Z. (2020). Discovering knowledge by comparing silhouettes using K-means clustering for customer segmentation. *International Journal of Knowledge Management*, *16*(3), 70–88. https://doi.org/10.4018/IJKM.2020070105

Aryadoust, V., & Goh, C. (2014). Predicting Listening Item Difficulty with Language Complexity Measures: A Comparative Data Mining Study. *CaMLA Working Papers*, 2014-02.

Benítez-Aurioles, B. (2018a). The role of distance in the peer-to-peer market for tourist accommodation. *Tourism Economics*, *24*(3), 237–250. https://doi.org/10.1177/1354816617726211

Benítez-Aurioles, B. (2018b). Why are flexible booking policies priced negatively?. *Tourism Management*, *67*, 312–325. https://doi.org/10.1016/j.tourman.2018.02.008

Berentsen, A., Breu, M.R., & Waller, C.J. (2019). What is the Value of Being a Superhost? FRB St. Louis Working Paper No. 2019-19. https://doi.org/10.20955/wp.2019.019

Cai, Y., Zhou, Y., & Scott, N. (2019) Price determinants of Airbnb listings: evidence from Hong Kong. *Tourism Analysis*, *24*, 227–242. https://doi.org/10.3727/108354219X15525055915554

Camatti, N., di Tollo, G., Filograsso, G., & Ghilardi, S. (2024). Predicting Airbnb pricing: a comparative analysis of artificial intelligence and traditional approaches. *Computational Management Science*, *21*(30). https://doi.org/10.1007/s10287-024-00511-4

Casamatta, G., Giannoni, S., Brunstein, D., & Jouve, J. (2022). Host type and pricing on Airbnb: Seasonality and perceived market power. *Tourism Management*, 88, 104433. https://doi.org/10.1016/j.tourman.2021.104433

Chapman, S., Mohammad, S., & Villegas, K. (2023). *Predicting Listing Prices In Dynamic Short Term Rental Markets Using Machine Learning Models*. Cornell University. https://arxiv.org/abs/2308.06929 (8.01.2025).

Chan, K.Y., Abu-Salih, B., Qaddoura, R., Al-Zoubi, A.M., Palade, V., Pham, D.-S., Del Ser, J., & Muhammad, K. (2023). Deep neural networks in the cloud: Review, applications, challenges and research directions. *Neurocomputing*, *545*, 126327. https://doi.org/10.1016/j.neucom.2023.126327

Chattopadhyay, M., & Mitra, S.K. (2019). Do airbnb host listing attributes influence room pricing homogenously?. *International Journal of Hospitality Management*, *81*, 54–64. https://doi.org/10.1016/j.ijhm.2019.03.008

Chen, Y., & Xie, K.L. (2017). Consumer valuation of Airbnb listings: a hedonic pricing approach. *International Journal of Contemporary Hospitality Management*, *29*(9), 2405–2424. https://doi.org/10.1108/IJCHM-10-2016-0606

Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). *A robust and scalable clustering algorithm for mixed type attributes in large database environment*. Proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco. CA 263–268. https://doi.org/10.1145/502512.502549

Chong, H.Y. (2010). Exploratory data analysis in the context of data mining and resampling. *International Journal of Psychological Research*, *3*(1), 9–22. https://doi.org/10.21500/20112084.819

Deboosere, R., Kerrigan, D. J., Wachsmuth, D., & El-Geneidy, A. (2019). Location, location and professionalization: A multilevel hedonic analysis of airbnb listing prices and revenue. *Regional Studies, Regional Science*, *6*(1), 143–156. https://doi.org/10.1080/21681376.2019.1592699

Falk, M., Larpin, B., & Scaglione, M. (2019). The role of specific attributes in determining prices of Airbnb listings in rural and urban locations. *International Journal of Hospitality Management*, 83, 132–140. https://doi.org/10.1016/j.ijhm.2019.04.023

Ferreira, J.P., Bijen Saha, B., Carrero, G.C., Kim, J., & Court, C. (2023). Impacts of red tide in peer-to-peer accommodations: A multi-regional input-output model. *Tourism Economics*, *29*(3), 812–834. https://doi.org/10.1177/13548166211068276

Floričić, T. (2016). Specialized SMEs in the Hotel Industry and Market Positioning in Croatia. *European Management Studies (previously: Problemy Zarządzania – Management Issues)*, *14*(3(62)), 151–171. https://doi.org/10.7172/1644-9584.62.9

Ghosh, I., Jana, R. K., & Abedin, M. Z. (2023). An ensemble machine learning framework for Airbnb rental price modeling without using amenity-driven features. *International Journal of Contemporary Hospitality Management*, *35*(10), 3592–3611. https://doi.org/10.1108/IJCHM-05-2022-0562

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018a). 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

Gibbs, C., Guttentag, D., Gretzel, U., Yao, L., & Morton, J. (2018b). Use of dynamic pricing strategies by Airbnb hosts. *International Journal of Contemporary Hospitality Management*, *30*(1), 2–20. https://doi.org/10.1108/IJCHM-09-2016-0540

Gorjani, O. M., Byrtus, R., Dohnal, J., Bilik, P., Koziorek, J., & Martinek, R. (2021). Human Activity Classification Using Multilayer Perceptron. *Sensors*, *21*(18), 6207. https://doi.org/10.3390/s21186207

Gunter, U., & Onder, I. (2017). Determinants of Airbnb demand in Vienna and their implications for the traditional accommodation industry. *Tourism Economics*, *24*(3), 270–293. https://doi.org/10.1177/1354816617731196

Gyódi, K., & Nawaro, Ł. (2021). Determinants of Airbnb prices in European cities: A spatial econometrics approach. *Tourism Management*, *86*. https://doi.org/10.1016/j.tourman.2021.104319

IBM. (2024). *What Is a Neural Network?*. https://www.ibm.com/docs/en/spss-statistics/saas?topic=networks-what-is-neural-network (6.11.2024).

Islam, D., Li, B., Islam, K. S., Ahasan, R., Mia, M. R., & Haque, E. (2022). Airbnb rental price modeling based on Latent Dirichlet Allocation and MESF-XGBoost composite model. *Machine Learning With Applications*, *7*(100208). https://doi.org/10.1016/j.mlwa.2021.100208

Ismail, M.R., Awang, M.K., Rahman, N., & Makhtar, M. (2015). A multi-layer perceptron approach for customer churn prediction. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(7), 213–222. https://doi.org/10.14257/ijmue.2015.10.7.22

Jang, S., & Kim, J. (2022). Remedying Airbnb COVID-19 disruption through tourism clusters and community resilience. *Journal of Business Research*, *139*, 529–542. https://doi.org/10.1016/j.jbusres.2021.10.015

Kalehbasti, P.R., Nikolenko, L., & Rezaei, H. (2021). Airbnb price prediction using machine learning and sentiment analysis. In A. Holzinger, P. Kieseberg, A.M. Tjoa, E. Weippl (Eds.), *Machine Learning and Knowledge Extraction* (p. 173–184). Springer International Publishing. https://doi.org/10.1007/978-3-030-84060-0_11

Kanakaris, N., & Karacapilidis, N. (2023). Predicting prices of Airbnb listings via Graph Neural Networks and Document Embeddings: The case of the island of Santorini. *Procedia Computer Science, 219*, 705–712. https://doi.org/10.1016/j.procs.2023.01.342

Kokasih, M., & Paramita, A. (2020). Property Rental Price Prediction Using the Extreme Gradient Boosting Algorithm. *International Journal of Informatics and Information Systems*, *3*(2), 54–59. https://doi.org/10.47738/ijiis.v3i2.65

Lancaster, K.J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*, 132–157. https://doi.org/10.1086/259131

Magno, F., Cassia, F., & Ugolini, M.M. (2018). Accommodation prices on Airbnb: effects of host experience and market demand. *TQM Journal*, *30*(5), 608–620. https://doi.org/10.1108/TQM-12-2017-0164

Panahandeh, A., Rabiei-Dastjerdi, H., Goktas, P., & McArdle, G. (2025). Answering new urban questions: Using eXplainable AI-driven analysis to identify determinants of Airbnb price in Dublin. *Expert Systems with Applications*. *260*, 125360, https://doi.org/10.1016/j.eswa.2024.125360

Perez-Sanchez, R., Serrano-Estrada, L., Marti, P., & Mora-Garcia, R.T. (2018). The what, where, and why of airbnb price determinants. *Sustainability (Switzerland)*, *10*(12). https://doi.org/10.3390/su10124596

Pirmohammadi, G., & Mast, M. (2020). 'Comparison of Artificial Neural Network and SPSS Model in Predicting Customers Churn of Iran's Insurance Industry'. *International Journal of Computer Applications*, *176*(32), 14–21. https://doi.org/10.5120/ijca2020920345

Sainaghi, R. (2021). Determinants of price and revenue for peer-to-peer hosts. The state of the art. *International Journal of Contemporary Hospitality Management*, *33*(2), 557–586. https://doi.org/10.1108/IJCHM-08-2020-0884

Sainaghi, R., Abrate, G., & Mauri, A. (2021). Price and RevPAR determinants of Airbnb listings: Convergent and divergent evidence. *International Journal of Hospitality Management*, *92*, 102709. https://doi.org/10.1016/j.ijhm.2020.102709

Sainaghi, R., & Chica-Olmo, J. (2022). The effects of location before and during COVID-19: Impacts on revenue of Airbnb listings in Milan (Italy). *Annals of Tourism Research*, *96*, 103464. https://doi.org/10.1016/j.annals.2022.103464

Sharaf Addin, E.H., Admodisastro, N.I., Mohd Ashri, S.N.S., Kamaruddin, A., & Chew, Y.C. (2022). Customer Mobile Behavioral Segmentation and Analysis in Telecom Using Machine Learning. *Applied Artificial Intelligence*, *36*(1). https://doi.org/10.1080/08839514.2021.2009223

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

Wang, H. (2023). Predicting Airbnb Listing Price with Different models. *Highlights in Science, Engineering and Technology*, *47*, 79–86. https://doi.org/10.54097/hset.v47i.8169

Wierzbiński, J. (2009). Limitations of k-means cluster analysis in creating typologies. *European Management Studies (previously: Problemy Zarządzania – Management Issues)*, *7*(4(26)), 232–233.

Xie, K., Heo, C.Y., & Mao, Z.E. (2021). Do professional hosts matter? Evidence from multi-listing and full-time hosts in Airbnb. *Journal of Hospitality and Tourism Management*, *47*, 413–421. https://doi.org/10.1016/j.jhtm.2021.04.016

Yang, Y. (2024). Predicting US Airbnb Listing Prices by Machine Learning Models. *Highlights in Business, Economics and Management*, *24*, 1408–1417. https://doi.org/10.54097/m187nw17

Yu, D., Dong, S., & Yao, S. (2022). Improvement of K-Means Algorithm and Its Application in Air Passenger Grouping. *Computational Intelligence and Neuroscience*, 2022. https://doi.org/10.1155/2022/3958423

Zhang, Z., Chen, R. J. C., Lee, H., & Ling, Y. (2017). Key factors affecting the price of Airbnb listings: A geographically weighted approach. *Sustainability*, *9*(9), 1635. https://doi.org/10.3390/su9091635