

Article

Factors Influencing Rental Investments in Paphos, Cyprus: Comparing Short- and Long-Term Rental Strategies

Sam Martin, Thomas Dimopoulos *  and Martha Katafygiotou 

Department of Real Estate, Neapolis University Paphos, 2 Danaï Avenue, 8042 Paphos, Cyprus; samgmartin@outlook.com (S.M.); m.katafygiotou@nup.ac.cy (M.K.)

* Correspondence: t.dimopoulos@nup.ac.cy

Abstract: Understanding the optimal strategy for a real-estate investment and how performance changes based on characteristics is crucial for optimising the achievable return. This is prominent in touristic areas such as Paphos, Cyprus, where there is no clear distinction as to whether short- or long-term approaches are optimal. This study aimed to develop a model for predicting the optimal rental strategy whilst assessing which model performed best and which property attributes impacted its return the greatest. Short-term data were collected from AirDNA and long-term data were manually collected from real-estate agents' websites. Furthermore, Random Forest, K-Nearest Neighbour, and Multiple Linear Regression models were created to predict the highest and best use for each property. Model accuracy varied between datasets, with the best-performing model for short-term properties being the Random Forest model (R-squared: 0.843), and the distance-based Multiple Linear Regression approach being the best for long-term properties (R-squared: 0.843). The study demonstrated that accurate models could be created to predict the optimal rental strategy with the number of bedrooms being the main driver for rental income, followed by luxury finishes and the presence of a pool. It was found that locational characteristics did not impact the returns significantly when assuming that the property was located within a touristic area.

Keywords: AirDNA; Airbnb; random forest; K-nearest neighbour; multiple linear regression; geographic information system



Citation: Martin, S.; Dimopoulos, T.; Katafygiotou, M. Factors Influencing Rental Investments in Paphos, Cyprus: Comparing Short- and Long-Term Rental Strategies. *Real Estate* **2024**, *1*, 136–157. <https://doi.org/10.3390/realestate1010007>

Academic Editor: Pierfrancesco De Paola

Received: 5 April 2024

Revised: 26 April 2024

Accepted: 8 May 2024

Published: 5 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Thanks to its natural beauty, international attractions, and Mediterranean climate, Cyprus sees a substantial amount of tourism, with the peak season of January 2022 to August 2022 seeing visitor spending on items such as travel, accommodation, and expenses reach EUR 1617 million, culminating in EUR 2439 million annually [1]. Although these 2022 figures are significant, the 2023 trends from January to May show an increase compared to the previous year of 34.2%, which reaffirms Cyprus' growing global appeal [2].

The Republic of Cyprus is home to four major cities, Nicosia, Larnaca, Limassol, and Paphos, and unlike larger European countries, each has its own distinct characteristics. Whilst cities such as Nicosia and Limassol boast the highest populations and commercial activity, Paphos is undoubtedly the home of tourism. It has been internationally recognised as such by winning numerous EU awards, the latest of which was the “2023 European Capital of Smart Tourism” award [3].

Paphos's population fluctuates considerably throughout the year due to tourism, and whilst there are around 95,400 permanent residents, Paphos Airport reported over 600,000 arrivals within the first four months of 2022 [4]. As such, a complex relationship between the long- and short-term rental market has been created, which sees many residents being outpriced from central and touristic areas. This issue is compounded by many investors optimising for a short-term strategy, which means that there is insufficient housing for residents. When considering the demand curve for long-term properties, this reduction

in available stock has caused a rise in rent, which at some point will surpass the short-term rental returns.

For investors, the balance between short- and long-term demand is crucial as the rental strategy of their property is directly impacted by the availability of each. Although this relationship occurs across the globe, cities such as Paphos with high levels of tourism are of most interest since the demand is high in relation to the overall size.

Currently, there are no publicly available studies regarding the Paphos rental market. This is concerning, since as Cyprus develops and expands its global reach, there is a requirement for clear, reliable, and informative studies to better advise investors. As such, this study aims to satisfy this by informing investors of the characteristics and trends present within the Paphos market, which may differ from other European areas.

Moreover, investors need to understand which rental strategy yields the highest return and which factors impact an investment's return the most. This study, therefore, identifies these factors whilst estimating which approach, be it short- or long-term renting, yields the highest return given a property's specific characteristics.

To achieve this, this study reviews a range of common statistical models, compares their results, and highlights any limitations they have in assessing the Paphos market. In practice, these models act as "estimators" that allow investors to input their property's characteristics, which in turn predicts the rental income for that property for each rental strategy.

Furthermore, this study will identify the model that most accurately predicts the return of rental properties when considering long- and short-term renting and suggest the highest and best use (HBU) for each. The various models must be reviewed based on their relative appropriateness to Paphos, as this directly impacts their predictive abilities and, as such, the accuracy of the information regarding their HBU. In addition, the factors that impact each rental strategy's performance will be identified so that advice can be provided to investors on how they can best manage their property to improve their return.

From this study, investors can expect to gain the tools needed to make informed decisions on their investment portfolio and apply them accordingly to optimise their rental portfolio.

2. Literature Review

2.1. Similar Studies

A study by Shokoohyar, Sobhani, and Sobhani into the optimal rental strategy based on the rate of return (RoR) between short-term and traditional residential rental investments was conducted for the city of Philadelphia [5]. It considered a range of different factors, including those relating to the property, neighbourhood, and location. A range of models were examined, with K-nearest neighbour (KNN), Random Forest (RF), and Multiple Linear Regression (MLR) analyses most accurately predicting the RoR. Moreover, they were able to define specific areas for which each rental strategy would be most suitable, with properties located centrally, close to historical attractions and nightlife, typically being more suited to short-term rental. Conversely, the suburbs were more suited to long-term residential rental. Whilst the study was relatively comprehensive, it only considered the area of Philadelphia, which is considerably different to that of Paphos.

Another study, this time on the city of Bologna, compared the performance of both short- and long-term residential properties, finding that 49% of short-term properties had an economic advantage over their long-term equivalent [6]. In the study, the authors also used MLR for the short-term rental data; however, for the long-term data, the city was split into 28 areas, in which the average return was taken for each using the national "Real Estate Market Observatory" due to limited public data. The comparisons made between the two were relatively fundamental and little interrogation was carried out regarding the factors affecting the rate of return; nevertheless, this study further supports the use of MLR and offers an alternative should data not be available for the Paphos market.

These two studies focused on touristic cities; however, neither of these cities were in a coastal area, as is the case with Paphos. For this reason, the study by Rodríguez-Pérez de Arenaza, Ángel Hierro, and Patiño was reviewed, which focused on the Andalusia area in Spain [7]. This study investigated how the rental prices of residential properties were impacted by the short-term holiday rental industry. Again, due to limited data on the market, the authors were unable to use a dynamic regression model and instead opted for a cross-sectional model. Due to it being fixed on a certain point in time, they repeated their analysis several times to account for touristic changes and found that short-term rental sites such as Airbnb increased the average residential rental price by 13.69%, creating an interesting dynamic between the two. The main limitation of this study relates to the approach taken to negate the absence of time-series data, which prevented the application of a more robust econometric analysis.

2.2. Multilinear Regression

One commonly used statistical technique for determining the correlation between data points is MLR analysis. Within real estate, this is known as “hedonic modelling”, wherein the aim is to estimate the price or value of a property based on its characteristics and location. These characteristics can include the number of bedrooms, the size of the internal area, and the location, to name a few. When creating an equation for MLR, the independent variables would be the property characteristics and the dependent variable would be the price, thus allowing for the price to be estimated based on a property with known characteristics. In our study, since multiple attributes are impacting the property prices, an MLR approach must be taken and will highlight which of the attributes has the largest impact on the property price [8].

The most common, and simplest, approach to MLR is the Ordinary Least-Squares (OLS) method. In this method, the analysis aims to find the line of best fit by minimising the sum of the squared residuals between the observed points and the predicted value of the model [9]. The accuracy of the model is then defined through an R-squared value, between 0 and 1, which displays how closely the model can describe the data. The higher the value, the better the fit. In addition, an adjusted R-squared value can also be used, and this is often more conservative, as it considers the number of independent variables in the model and penalises based on this number, which does not impact the model’s performance. The equation used to represent MLR can be seen below:

$$Y = b_0 + b_{1 \times 1} + b_{2 \times 2} + \dots + \epsilon, \quad (1)$$

where the following applies:

y—the property value (dependant variable);

b_0 —the y-intercept;

b—the slope coefficient for the explanatory variable;

x—the explanatory variable (independent variables);

ϵ —the model’s error term (residuals).

There are some assumptions that need to be made for this approach, such as that the relationship between the dependent and independent variables is linear, the variables have a normal distribution, and the variance is constant across all levels (homoscedasticity) [9].

One of the main issues with MLR is that it can oversimplify real-world scenarios as it is only capable of predicting linear relationships, when in practice, this may not be the case. It is also a stationary model and does not consider any time-series data, resulting in it having to be repeated periodically to maintain its relevance and accuracy [10].

With this in mind, there are still a large number of benefits of using MLR, such as variable importance, quick prediction and assumption validation, and also simplicity of interpretation. The limitations cannot be ignored and must be addressed during the analysis stage to mitigate their impact on the model.

2.3. Random Forest

Previous studies have utilised machine learning (ML) methods to predict the returns for an array of investments within an area, one of the most common of which is the Random Forest (RF) method. An RF consists of a collection of decision trees, with each tree being constructed independently using a random subset of training data and features. During this stage, a method called “Bootstrapping” is used, which randomly samples the training data with replacements, thus creating a collection of bootstrap samples. Overfitting is common within MLR, as discussed by Hawkins, whereby more independent variables are used than needed to predict the dependent variable, causing inaccuracies [11]. This is reduced within RFs, as random subsets of variables are selected at each node of the decision tree, hence improving the reliability of the model. Each decision tree can then be trained until a stopping criterion is met; this could be either the maximum tree depth or the minimum samples per node. Finally, the RF combines all of the decision tree predictions to make its final prediction; for regression purposes, the average/weighted prediction of the individual trees is used [12].

Figure 1 below shows an overview of an RF, including the decision path, voting, and prediction, for a two-tree model. In practice, an RF is composed of hundreds of trees to tune the hyperparameters most effectively [13]. This topic will be analysed in depth to ensure that the most accurate model is created.

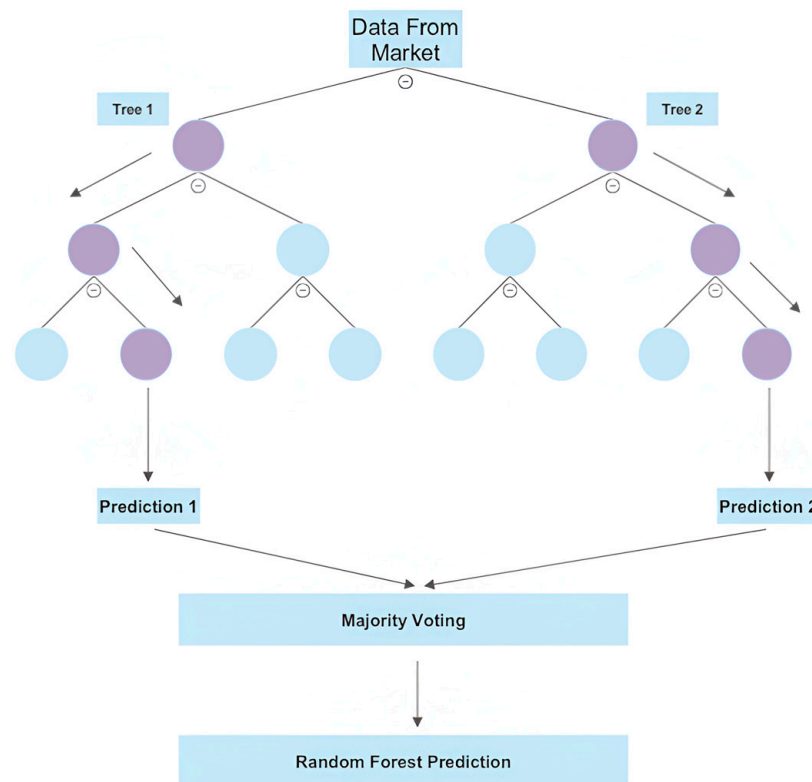


Figure 1. Random Forest overview.

Some of the main benefits of RFs are their high flexibility, accuracy, and ease of comprehension, as highlighted by Brieman in his exhaustive review of regression trees [14]. This is supported by Antipov and Pokryshevskaya’s study, which found that the RF predictions were supported by empirical data recorded directly from data on residential apartments within the Saint Petersburg area [15]. Moreover, Antipov and Pokryshevskaya found that the effectiveness of their RF performed better when the price per sqm was predicted followed by the calculation of the total price. This is believed to be linked to heteroscedasticity and some other problems inherent in real-estate data. They also found that factors such as the room area and the ground on which the apartment was situated

had little importance in the overall predictions. The accuracy of their model was between 0.80 and 0.91 depending on the coefficients used.

Mohapatra, Shreya, and Chinmay highlighted the importance of data pre-processing, parameter selection, and optimisation of the algorithm as key areas of focus to ensure the accuracy of the RF approach [16]. For the data available within the Cyprus market, data pre-processing should be closely managed and inaccuracies addressed before continuing with an RF analysis.

Finally, Potrawa and Tetereva's study investigating how a specific attribute of a property affects the property value found that an RF model outperformed MLR by 25% in RMSE and increased the R-squared value by 0.15 [17]. One major difference between our study and Potrawa and Tetereva's is the amount of available data: they had over 40,000 data points, which is much greater than what can be realistically achieved in the Paphos market.

2.4. K-Nearest Neighbour

Another ML method which has been used extensively in the real-estate industry is the KNN method. This approach is non-parametric, meaning it does not require the data to be collected following a specific distribution and uses these data to make predictions based on the similarity between the data points [18].

Using a training set of data, as mentioned previously, the algorithm makes a prediction for an "unknown" data point based on its proximity to all of the points in the training dataset [19]. Most commonly, the distance metric used is the Euclidean distance, which follows the equation below:

$$d = ((x_2 - x_1)^2 + (y_2 - y_1)^2)^{-2}, \quad (2)$$

where the following applies:

d —the linear distance between points;

x_i —the x-intercept for reaching a respective point;

y_i —the y-intercept for reaching a respective point.

Using the user-defined "k" value and calculated distances, the algorithm selects the closest "k" number of neighbours to consider for the analysis. When selecting the "k" value for the model, it is important to consider its magnitude, since low values yield better-localised results but can be highly impacted by noise, whereas large "k" values have less noise but provide less detailed predictions [20].

Depending on the prediction method, be it via majority voting or averaging, either the predicted class or value will be found, respectively, which is then assigned new data points as its final prediction.

KNN is among the simplest algorithms to use as it does not have many hyperparameters that need to be tuned, which makes it easy to implement and interpret. Unfortunately, this does come at a cost, which is its computational intensity with large datasets and high "k" values. Luckily, for the Cyprus market, there is limited information; so, in practice, this computational limitation will unlikely be met.

As mentioned before, KNN can complete both regression and classification predictions; therefore, for a real-estate application, it is versatile enough to solve a wide range of problems [20].

Finally, as with all of the methods discussed, data pre-processing is critical for accurate predictions using KNN, specifically in distance calculations. These are fundamental to the method and, as was shown by Nair and Kashyap's study, when advanced pre-processing is undertaken, the accuracy of the models can increase by as much as 0.18% [21].

2.5. Data Collection

The two main collection methods that we used to collect our data were AirDNA and manual gathering. AirDNA is an analytics platform focusing on short-term rental

data, which it collates from a range of rental platforms such as Airbnb, Vrbo, and Booking.com [22]. This platform provided most of the data, including the geographical information and proximity to POIs. For long-term properties, the data were collected through manual gathering from real-estate agents' websites and rental platforms, including Bazaraki and Facebook Marketplace. One concern was the accuracy and availability of these data, which could lead to data cleansing being required.

Various studies have investigated the effect of sample size on the accuracy of analyses, and depending on the quality of the data, the recommended sample ranges vary. Nguyen and Cripps found that MLR required around 500 data points to accurately create a model of relationships, whereas Benjamin, Guttery, and Sirmans found that 70 could yield good results [23,24]. Furthermore, Limsombunchai, Gan, and Lee were able to achieve an R-squared value of 0.75 with only 200 data points, which demonstrates that even smaller sample sizes can be used to effectively create MLR models [25]. Interestingly, one study that investigated developing markets with limited data found that MLR models outperformed both the k-NN and RF methods regarding their mean absolute percentage errors (MAPEs) [26]. The study used 318 data points from Szczecin in Poland, which is far less developed than Paphos and does not attract the same level of tourism. Based on these findings, in our study, a dataset of 200–300 properties was targeted for the long-term data and around 500+ properties were targeted for the short-term data due to the methods of data collection.

2.6. Enriching Data

One study on the Cyprus market aimed to understand whether the use of Artificial Intelligence (AI) and machine learning (ML) could supplement the use of Mass Appraisals (MAs) to achieve more accurate predictions [27]. It concluded that there were significant errors within the model used by the Department of Land and Surveys (DLS) and enriching it would help the reliability and accuracy of MAs. This could be executed using satellite imagery as well as geographical locations relative to key value-influencing areas such as hospitals and schools. As such, locational characteristics can significantly impact the return rates for short- and long-term rentals; so, for this reason, a combination approach incorporating satellite imagery may be beneficial for improving the results of the aforementioned study.

For this project, the use of GIS spatial data, coordinates, and features was invaluable as it helped enrich our analysis, allowing for variables outside of just the property characteristics to be analysed. Din, Hoesli, and Bender state that even when GIS data are incomplete, they should be used where available to improve the accuracy of hedonic pricing models [28].

3. Materials and Methods

3.1. Short-Term Rental Data

As previously mentioned, the short-term rental data were collected through AirDNA with consent from AirDNA and Axia Valuers, to whom the data belongs. To ensure the data were clean and specific to the research, thorough data cleansing was carried out to remove outliers and properties that were not within the scope of the project. From the initial 9584 properties, only 825 satisfied all of the qualities that were relevant to the study; this cleansing has been discussed further below.

3.1.1. Locations Outside of the Paphos District

By using the longitudinal and latitudinal coordinates as the bounding box of Paphos, all of the properties located outside of this area were removed, with the limitation of assuming the Paphos district as a rectangular shape. Most properties were not located near the district borders, and so their impact was deemed of low significance. In total, 5248 were removed due to being located outside of the Paphos district.

3.1.2. Duration of Operation

This research focused on one fixed point in time, so only data from 2022 were considered and each property needed to have been operating for the entire year in order to not negatively skew the equivalent monthly rate. As a result, 2249 properties were found to have not operated for the entirety of 2022 and so were removed.

3.1.3. Outliers and Ghost Properties

Through analysis of the data, it was found that many properties were not active, had unrealistic prices, or were so-called “ghost” properties. These were removed as they caused significant inaccuracies, as supported by Fleischer, Ert, and Bar-Nahum [29]. This was carried out through outlier removal (q-score) and resulted in a further 942 properties being removed.

3.1.4. Missing Data

Properties with missing or partial data were removed due to the impracticality of manually imputing property characteristics; this accounted for another 320 properties being removed.

3.1.5. Calculated Data

Due to inaccuracies in how AirDNA calculates revenue data for properties, the estimated monthly revenue was calculated using the average daily rate and the average occupancy rate of all properties within the study. Whilst this is not necessarily the most accurate approach since it does not consider the occupancy variations between each property, the method used by AirDNA heavily skews the performance of properties that are rented for long periods and are not easily identifiable through data cleansing, preventing deeper analyses from taking place. This is supported by Agarwal, Koch, and McNab, who also found an upward bias in the metrics published by AirDNA [30]. The AirDNA report assumes that the properties are held for a long period; thus, averaging over 12 months for all properties is a reasonable assumption, especially in heavily touristic areas such as Paphos, where the demand is high.

3.2. Long-Term Rental Data

Long-term data needed to be collected manually through traditional means, which involved collecting data from platforms such as Bazaraki, Facebook Marketplace, and the agents’ websites. As these data are publicly available, no consent or approvals were required; however, all data that could be considered sensitive were redacted.

While this step ensured that all of the relevant data were collected, it meant that the process was extremely time-consuming. The main benefit of this approach was that because the data were vetted as they were recorded, very little data cleansing was required, and the data were almost immediately ready for analysis.

The main limitation was the fact that the prices advertised were the asking prices, not the agreed rental prices. Throughout the data collection, it was evident that some values were highly inflated, so they were ignored; however, others were only marginally inflated and therefore remained. Although this marginal inflation may only be a few hundred euros (EUR), since it was apparent in many properties, it is likely that it impacted the final HBU decision.

Over 400 properties were reviewed, with only 203 properties being used due to inflated prices or missing data. Properties with inflated prices were identified through knowledge of the Paphos market since most were easily identifiable, and these accounted for 93 properties. The remaining 106 were removed due to partial or missing data.

As with the short-term data, the removed properties showed no correlation so their removal should have had little impact on the model’s ability to make accurate predictions. Although this is a considerably small amount compared to the short-term rentals, this figure still falls within the recommended range found during the literature review.

3.3. Geospatial Data

Using QGIS, an open-source GIS application, the road network of the Paphos district and all properties and Points of Interest (POIs) were mapped. With this information, the distances across the network were interpolated for each property and POI to define the distances to each. The POIs included beaches, golf courses, hospitals, the central business district, and Paphos Airport, to name a few. These data were then reintroduced into the raw dataset for use within the regression models.

Heat maps of the monthly returns of both long- and short-term properties showed some interesting trends regarding their geospatial distribution. The long-term map, as shown in Figure 2, shows that the highest-returning properties were found outside of the city in areas known for their luxurious nature. On the other hand, in Paphos, there was a range of different returns, but generally, these remained lower; this is most likely linked to the property type and size being more apartment-focused, whilst the suburbs had more houses/villas.

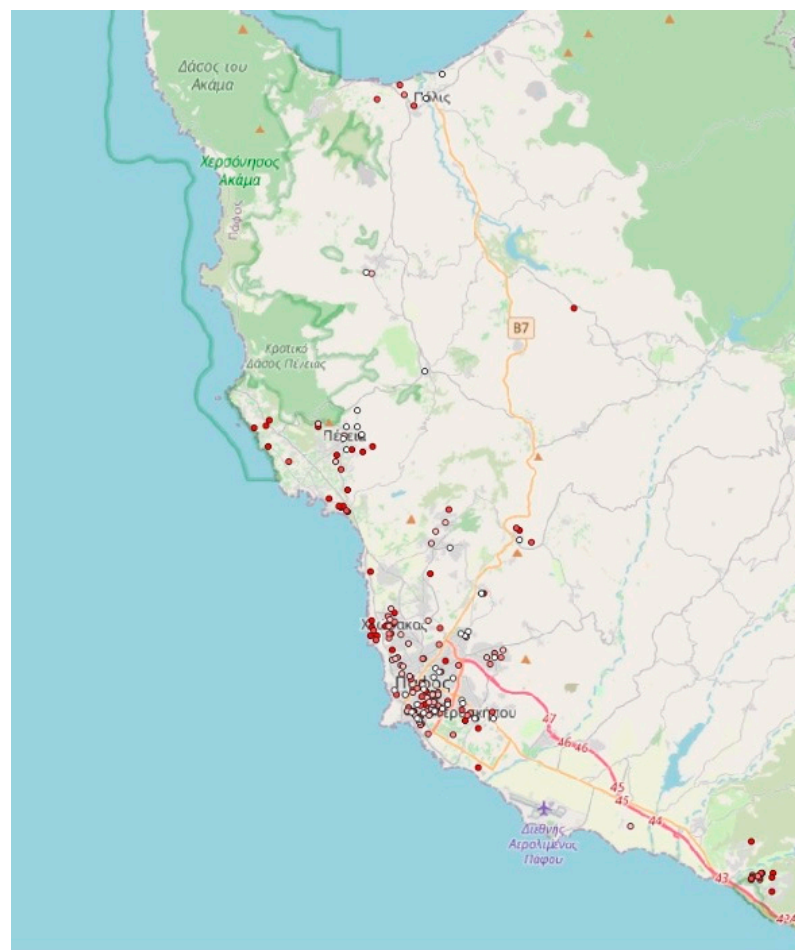


Figure 2. Price heatmap for the long-term properties.

A similar trend can be seen for the short-term rentals, as per Figure 3; however, the extent is much more severe, with the aforementioned areas showing considerably higher values than their Paphos counterparts.

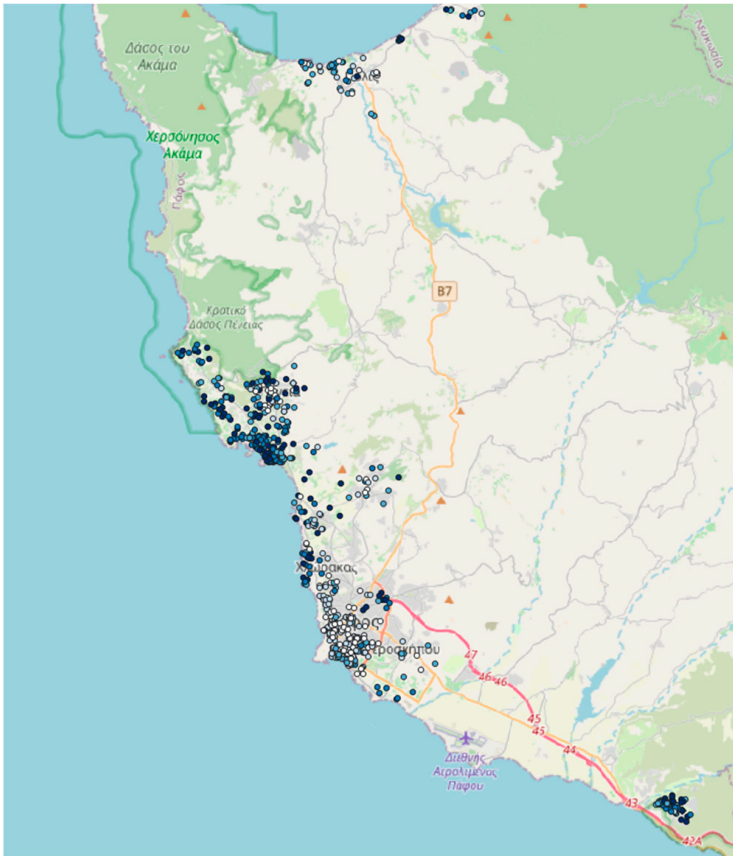


Figure 3. Price heatmap for the short-term properties.

3.4. Regression Analysis

3.4.1. Data Preparation

The first approach used for the prediction of the alternative strategy was MLR, with the property characteristics being a primary focus to allow investors to manage their property accordingly. Correlations between the attributes, known as multicollinearity, were addressed before the regression process as they severely affect the stability and interpretability of the model. This is in part due to them becoming hypersensitive to minor changes in the data, leading to large variations in the predicted results, but they can also cause other variables to become insignificant in their presence. A correlation matrix was created to highlight any variables that had a high correlation, which is commonly defined as follows:

$$x > 0.5 \quad x < -0.5, \tag{3}$$

Several variables were found to have a level of correlation; these can be seen in Table 1, and their relative correlation to other variables, in addition to whether it remained in the analysis or not, is shown.

Table 1. Regression correlation summary.

Independent Variable	Correlation
Property Type	Yes, high correlation with bedrooms and bathrooms.
Latitude	Regression model-dependent, removed when using GIS data.
Longitude	Regression model-dependent, removed when using GIS data.

Table 1. Cont.

Independent Variable	Correlation
Finish Quality	No correlation.
Bedrooms	No, property type removed instead.
Bathrooms	Yes, removed due to correlation with bedrooms.
Parking	No correlation.
Pool	No correlation.
Dist. to Nearest Golf Course	Yes, correlation with lots of GIS variables.
Dist. to Nearest Beach	No, other variables removed due to their importance to this study.
Dist. to Nearest Hospital	Yes, correlation with lots of GIS variables.
Dist. to CBD	Yes, correlation with lots of GIS variables.
Dist. to Old Town	Yes, correlation with lots of GIS variables.
Dist. to Airport	Yes, correlation with lots of GIS variables.
Dist. to Harbour	Yes, correlation with lots of GIS variables.
Dist. to Sea Caves	Yes, correlation with lots of GIS variables.
Dist. to Coral Bay	No, other variables removed due to their importance to this study.
Dist. To Tombs of Kings	Yes, correlation with lots of GIS variables.
Dist. To Aphrodite Hills	Yes, correlation with lots of GIS variables.
Dist. To ISOP	Yes, correlation with lots of GIS variables.
Dist. To Polis	No, other variables removed due to their importance to this study.

When performing our regression analysis, categorical variables needed to be removed as they were another cause of multicollinearity in the model. For this, dummy variables were created using a binary format, where 1 represents true and 0 represents false. In this analysis, dummy variables were taken for the finish quality, parking, and pools. In addition, a technique called “dummy variable exclusion” was followed, which sees one item from each category being removed, thus creating none-perfect multicollinearity without affecting the overall fit or performance of the model. In equation form, the “dummy variable exclusion” technique is as follows:

$$Y = \beta_0 + \beta_1 \times \text{Dummy_A} + \beta_2 \times \text{Dummy_B} + \beta_3 \times \text{Dummy_C}, \quad (4)$$

Taking logarithms of values often improves regression analyses by reducing heteroskedasticity; although, this was found to not be beneficial to the overall performance of our model, since the R-squared value remained unchanged.

3.4.2. Regression Models

Three regression models—standard, geospatial, and distance-based spatial regression—were carried out on the dataset, with the variables being adjusted based on their nature and purpose. A summary of each one and its variables can be seen in Tables 2–4.

Table 2. Standard regression descriptive characteristics.

	Long-Term Data			Short-Term Data		
	Coefficient	Std Error	p-Value	Coefficient	Std Error	p-Value
Intercept	−239.5670	594.7145	0.6875	−1925.1626	231.6099	0.0000
Quality_Luxury	2772.9211	306.2989	0.0000	2753.5876	164.0873	0.0000
Quality_Good	741.5633	211.9946	0.0006	1807.0815	163.4928	0.0000
Quality_Standard	119.1536	198.8557	0.5497	950.0772	146.0243	0.0000
Bedrooms	613.5142	71.2152	0.0000	1564.2426	41.6072	0.0000
Parking_Yes	−44.8237	577.0686	0.9382	−237.3274	175.6139	0.1769
Pool_Yes	1256.7473	176.1012	0.0000	377.9773	140.8164	0.0074
Pool_Communal	138.4187	132.5471	0.2976	258.6321	195.4374	0.1861

Table 3. Geospatial regression descriptive characteristics.

	Long-Term Data			Short-Term Data		
	Coefficient	Std Error	p-Value	Coefficient	Std Error	p-Value
Intercept	12,072.9884	57,834.6681	0.8349	−201,152.1098	40,900.6610	0.0000
Longitude	−1210.8226	944.8119	0.2015	−1038.4441	605.9397	0.0870
Latitude	920.7146	1080.5340	0.3952	7258.9539	814.3556	0.0000
Quality_Luxury	2713.3189	308.5177	0.0000	2845.4105	153.7311	0.0000
Quality_Good	664.6562	214.8312	0.0023	1792.2597	152.4968	0.0000
Quality_Standard	108.8621	197.8807	0.5829	859.3971	136.4362	0.0000
Bedrooms	615.4072	70.8811	0.0000	1629.2139	39.5664	0.0000
Parking_Yes	−61.9064	574.2631	0.9143	−280.3170	163.8273	0.0875
Pool_Yes	1309.2286	181.0859	0.0000	340.6442	133.3233	0.0108
Pool_Communal	110.7387	132.5807	0.4046	11.7053	183.7805	0.9492

Table 4. Distance-based regression descriptive characteristics.

	Long-Term Data			Short-Term Data		
	Coefficient	Std Error	p-Value	Coefficient	Std Error	p-Value
Intercept	−540.9797	616.8445	0.3816	−3380.4834	263.5197	0.0000
Quality_Luxury	2568.4478	306.6768	0.0000	2839.6123	156.2421	0.0000
Quality_Good	561.3841	215.1968	0.0098	1831.8435	155.3048	0.0000
Quality_Standard	47.1591	196.1454	0.8103	879.0517	138.3407	0.0000
Bedrooms	635.0916	70.4320	0.0000	1630.5508	40.5600	0.0000
Parking_Yes	4.8095	566.3701	0.9932	−283.5373	165.8014	0.0876
Pool_Yes	1320.2718	175.8417	0.0000	412.1615	137.3391	0.0028
Pool_Communal	60.9690	131.8932	0.6444	60.9744	190.0521	0.7484
Nearest_Beach	−0.0527	0.0225	0.0203	0.0904	0.0387	0.0197
Dist_Coral_Bay	0.0132	0.0090	0.1434	0.0286	0.0055	0.0000
Dist_Polis	0.0105	0.0078	0.1798	0.0311	0.0047	0.0000

It is important to note that these results were achieved after numerous optimisations involving data cleansing, logarithmic tests, and adjustments to the chosen independent variables. Since long- and short-term rentals were to be compared as part of this study, the variables needed to be matched to ensure that they could be applied to the opposing dataset without impacting the integrity of the analysis.

Based on these results, it is found that the property quality and the number of bedrooms consistently influenced the achievable rental income for both long- and short-term properties, whilst amenities such as parking or pool access had mixed significance. The distance-based regression results also show that the short-term income was more greatly affected by the property's proximity to the touristic areas when compared to the long-term income.

3.5. Random Forest Models

Machine learning can be utilized through various techniques; for this study, Python was used, along with scikit-learn (sklearn). Due to the characteristics of RFs, the entire dataset was used so that the model was able to capitalise on all descriptive characteristics, which strengthened its prediction of returns.

As with MLR, dummy variables were created since categorical ones are not handled well in RFs. Other than this, the data were not modified as logarithms and scaling did not impact the performance of the model.

Unlike MLR, because the entire dataset could be used for both long- and short-term rentals, only two models were created: one for each rental strategy. These datasets were then separated into a “training” and “testing” split automatically using the sklearn module, with the test size equalling 20% of the total values. In addition, various sizes of Random Forests were tested, and it was found that 100 trees performed the best in both situations.

Due to its black-box nature, it is not possible to define an equation for the RF like that defined in the case with MLR; however, Figures 4 and 5 show the relative feature importance for both the long- and short-term rentals.

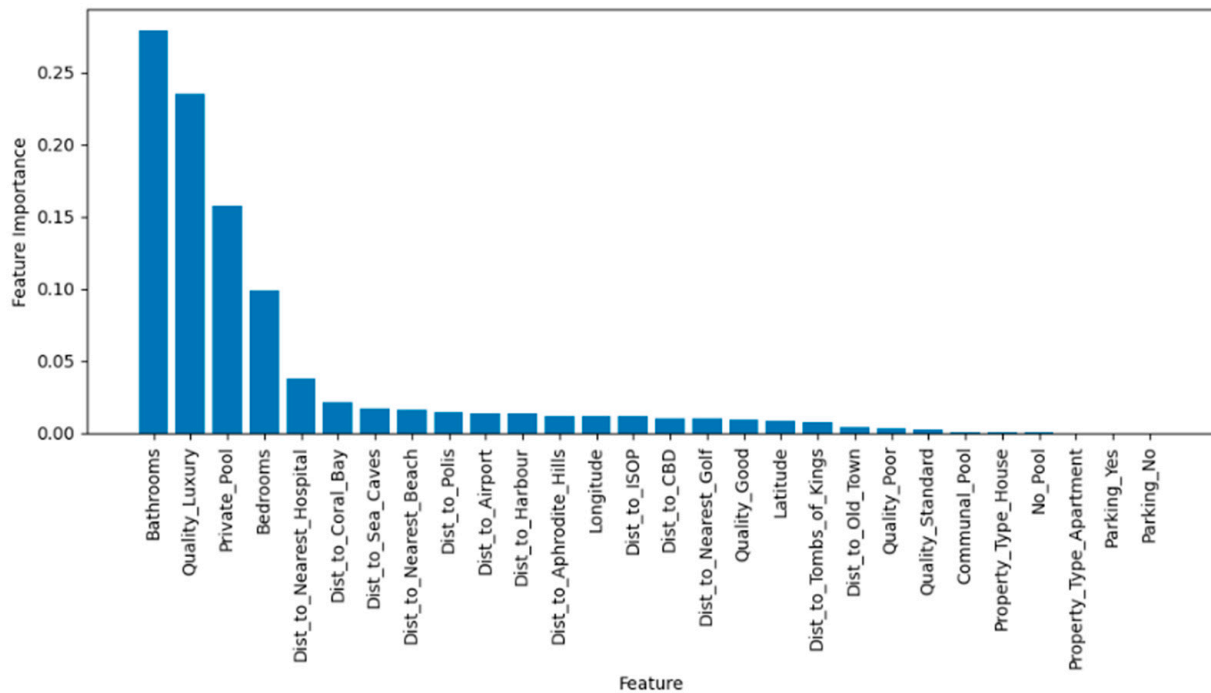


Figure 4. Random Forest feature importance (long-term predictions).

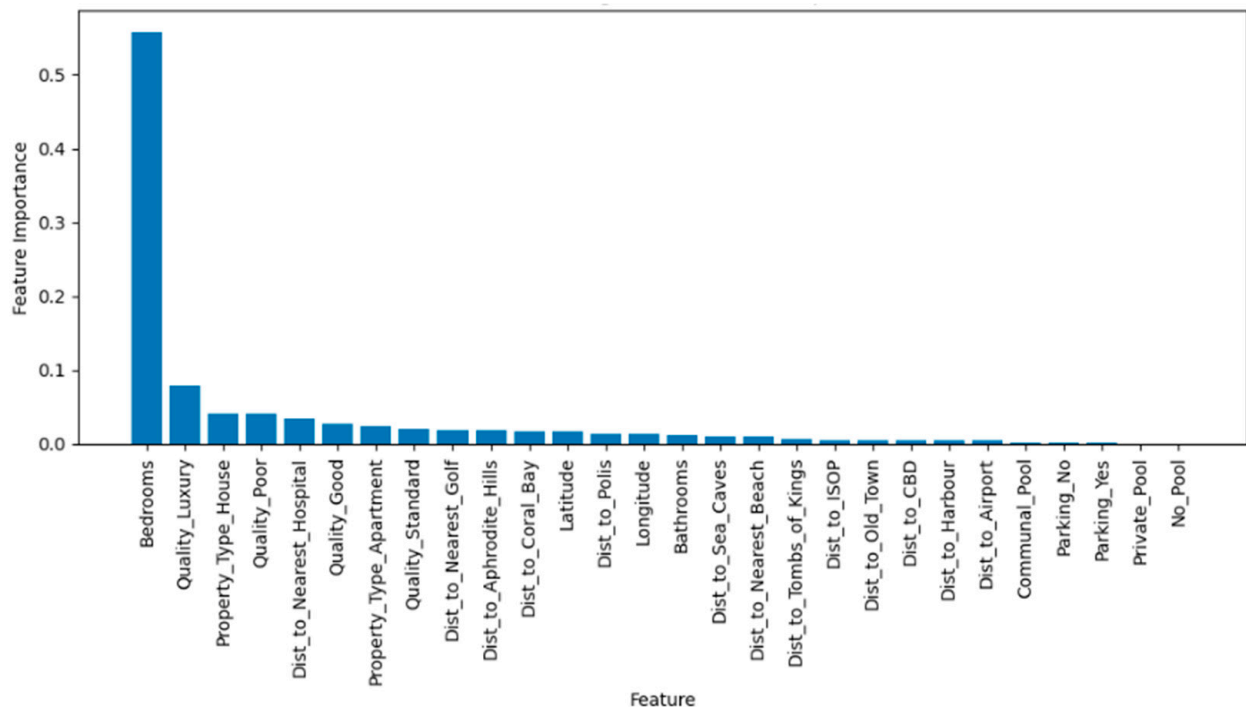


Figure 5. Random Forest feature importance (short-term predictions).

3.6. K-Nearest Neighbour

Like the RF, the KNN approach was carried out using sklearn on the Python platform and data were pre-processed similarly. There were several differences for KNN, the first being the conversion of longitude and latitude values to Cartesian coordinates, which is

necessary due to the geometry of the Earth's surface being curved and not flat. Secondly, due to the model making predictions based on the distances between points, the data needed to be scaled so that larger scales did not disproportionately affect the results, and this helped to ensure that each feature was treated with equal importance.

Optimisation of the model was primarily carried out by adjusting the number of neighbours used for the predictions. To find this optimal value, a balance between variance and bias needed to be chosen; this was aided using the MSE, along with sklearn's integrated cross-valuation analysis. The MSE values for training and validation for both the long- and short-term rentals can be seen in Figures 6 and 7, respectively.

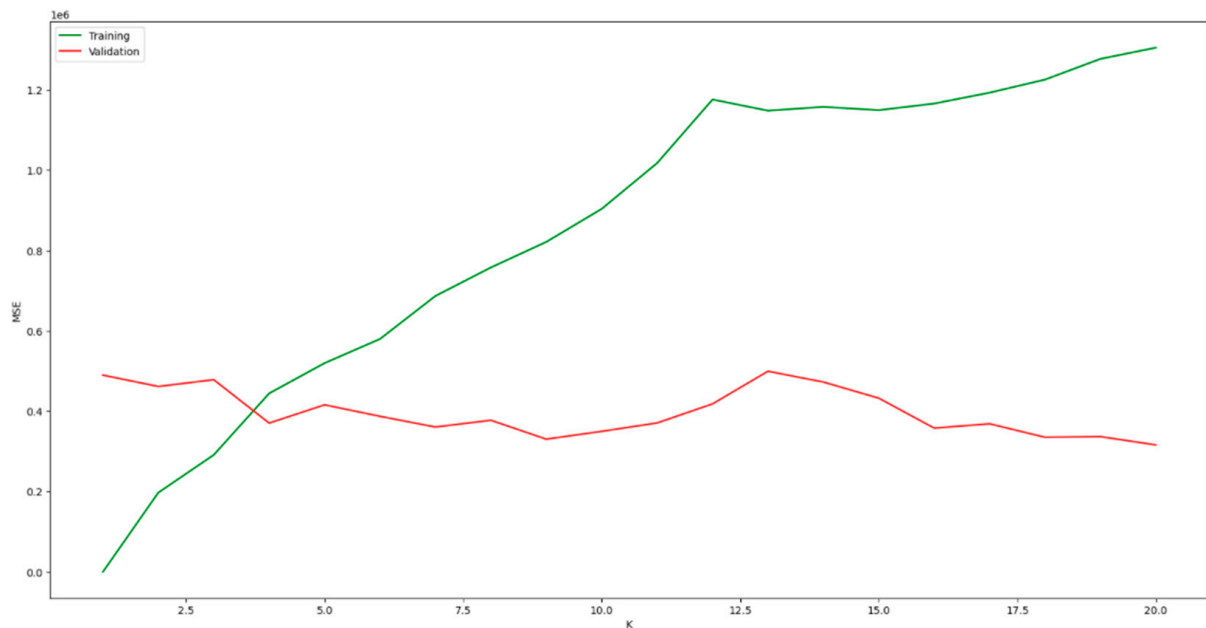


Figure 6. KNN long-term MSE variation.

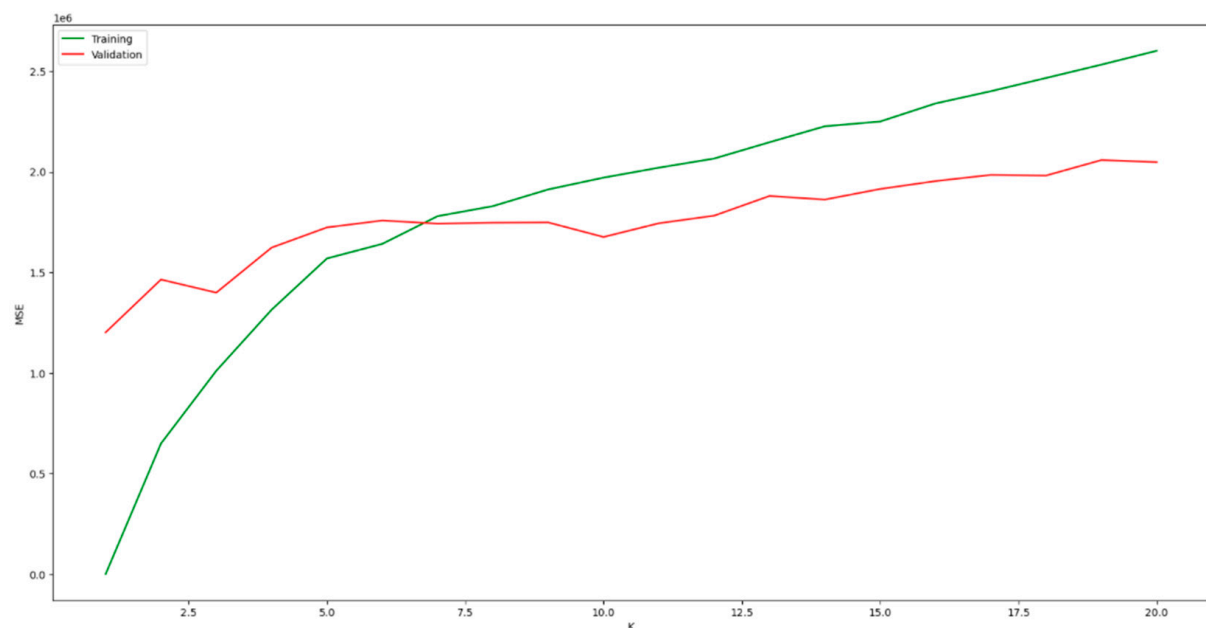


Figure 7. KNN short-term MSE variation.

From this information, a “k” value of 5 was chosen for the long-term rentals and a value of 7 was chosen for the short-term rentals. Due to the methodology of KNN, the

weighting of each attribute is not defined, and instead the average of the nearest neighbours is used to predict the estimated returns.

4. Results

4.1. Optimal Model Selection

With multiple approaches being taken, identifying the best-fitting model was crucial to ensure that the predicted returns for the properties were accurate. For this, the R-squared value, also known as the coefficient of determination, was found and compared for each model. The R-squared value is a measure of “goodness-to-fit”, where 0 does not explain any variability and 1 explains the model perfectly. The R-squared value for each method can be seen in Table 5.

Table 5. Comparison of model performance (R-squared values).

	Long-Term Predictions	Short-Term Predictions
Random Forest	0.560	0.843
KNN	0.518	0.752
MLR (Distance-Based)	0.803	0.739
MLR (with Lat/Long)	0.796	0.745
MLR (no Geo)	0.791	0.706

From these data, the distance-based MLR was chosen for the long-term predictions and the RF model was chosen for the short-term predictions, since these returned the highest R-squared values of 0.803 and 0.843, respectively. These values show that in both cases, over 80% of the variation in the rental incomes can be explained by the property characteristics used in the analysis. Based on these models, investigations into the alternative strategy were conducted based on the estimates they provided.

4.2. Descriptive Characteristics

For each property, the HBU was identified based on the predictions and a summary of their characteristics can be seen in Table 6.

Table 6. Characteristics of the predicted data.

Short-Term Data (Predicted)		Long-Term Data (Predicted)	
Mean	EUR 4136	Mean	EUR 3082
Median	EUR 3333	Median	EUR 3077
Mode	EUR 1010	Mode	EUR 5108
Std. Dev	EUR 3032	Std. Dev	EUR 1239
Range	EUR 20,666	Range	EUR 5630
Minimum	EUR 628	Minimum	EUR 869
Maximum	EUR 21,294	Maximum	EUR 6500
Count	601	Count	427

From the total collected data points, around 58.5% of the properties were found to achieve a higher return if they were rented in the short term compared to the long term. Moreover, on average, the return for these properties was around 25.5% higher than that of their long-term counterparts; however, the standard deviation between them was over 59.1% higher.

4.3. Incorrect Rental Strategy

Hypothetically, it is possible for property owners to incorrectly market their property according to an opposing strategy, thus yielding lower returns. The costs of this can be seen in Figure 8.

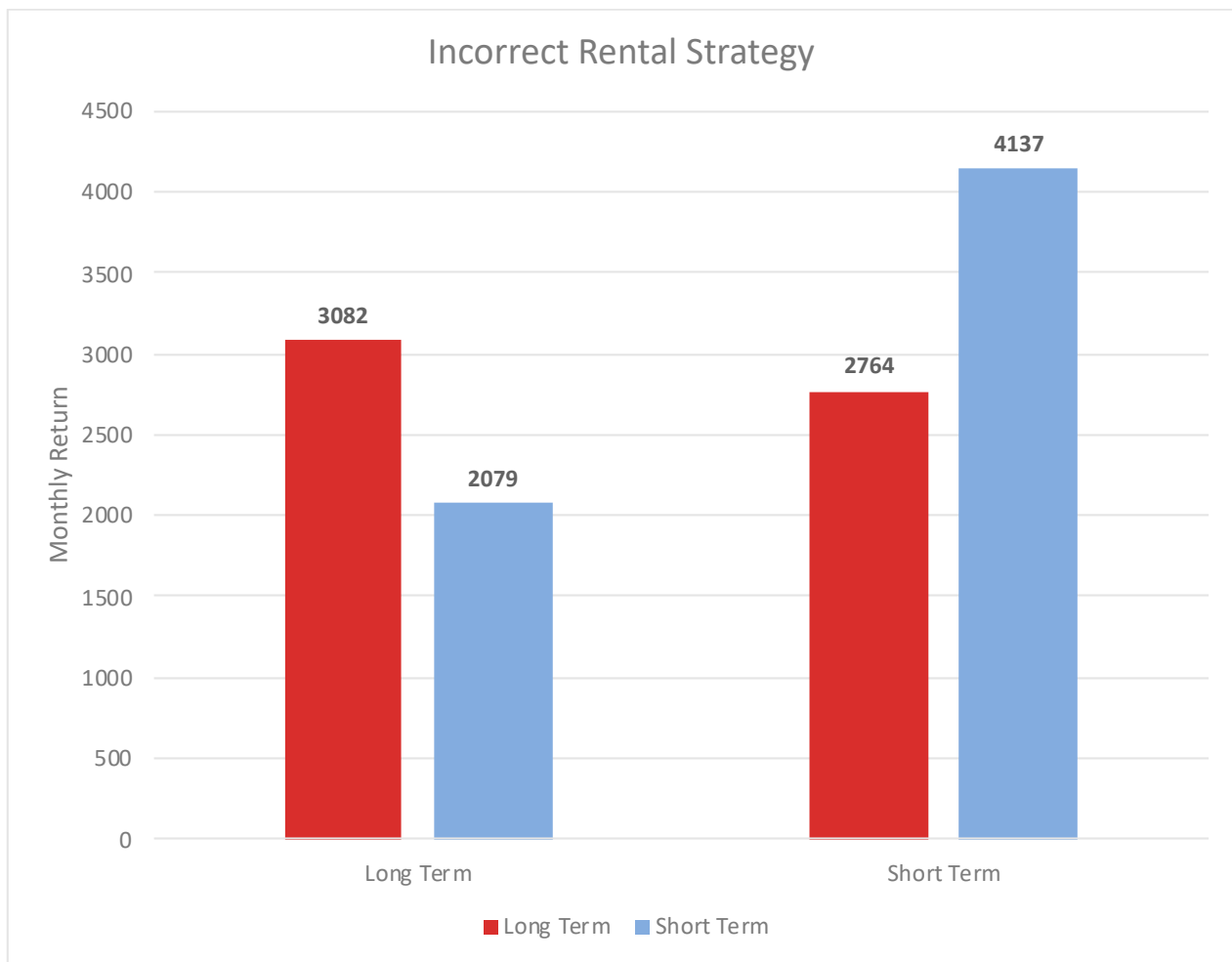


Figure 8. Costs of following the incorrect rental strategy.

Where the HBU is long-term rental, if short-term rental is chosen, then the owner would lose 1003 EUR/month, earning only 67.5% of the property's potential. On the other hand, if the HBU is short-term rental and the property is rented out long-term, the owner would lose 1373 EUR/month and earn around 66.8% of its potential.

4.4. Locational Trends

By plotting the HBU for each property in QGIS, the individual points for each approach, red for the long term and blue for the short term, could be displayed, highlighting any potential trends. Shown in Figure 9, the distribution can be seen as relatively consistent across the city, with short-term properties tending to be closer to the sea/beach due to the higher proportion of touristic amenities there such as restaurants and shops. Interestingly, it is shown that a short-term approach outperforms the long-term approach in some villages that are located away from touristic activity.

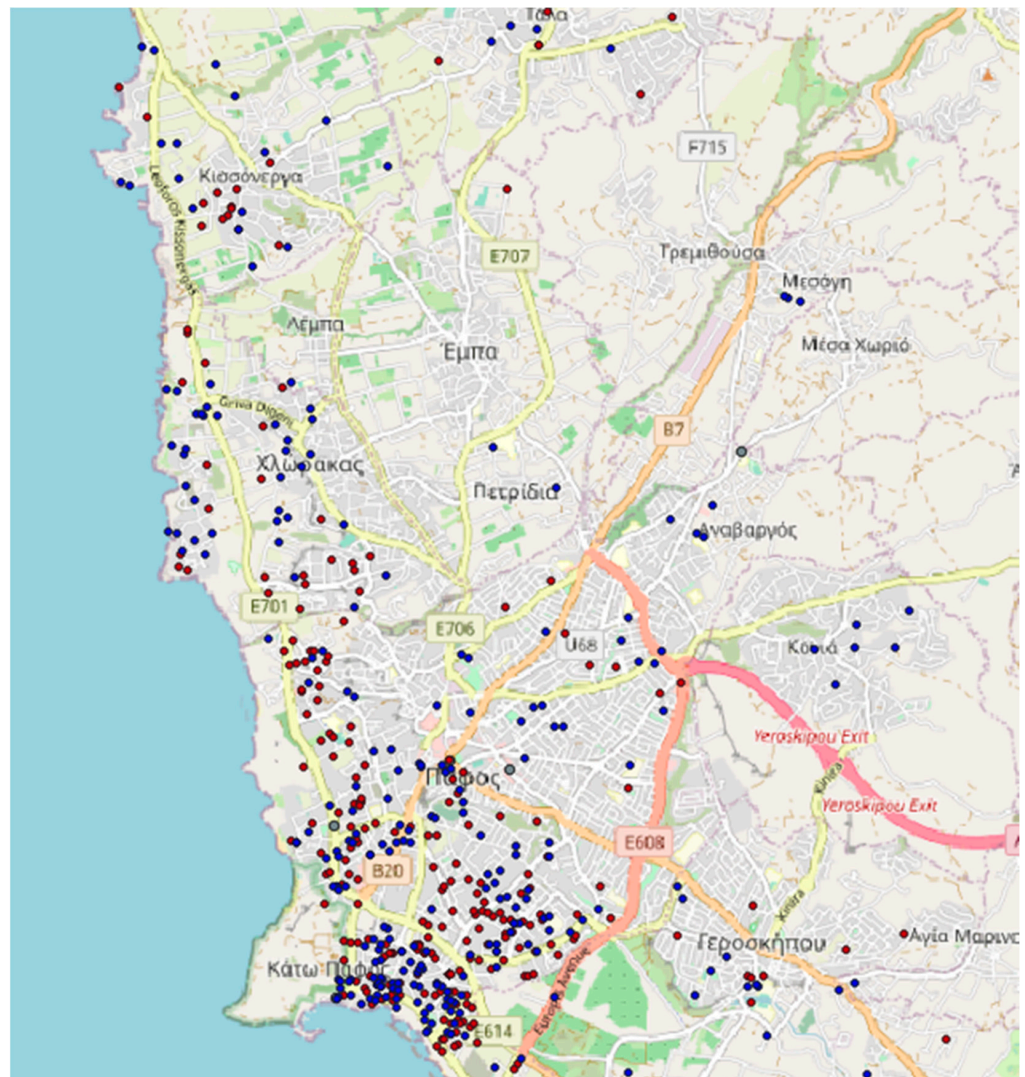


Figure 9. Long-term (Red) and short-term (Blue) HBU properties (Paphos).

4.5. Variable Weighting

Figures 4 and 5 highlight the relative importance of different variables regarding a specific rental approach. As such, it was found that for short-term predictions, the number of bedrooms was by far the most significant feature, with the RF predicting an importance of 0.526. The next most important factor was a luxury finish, at only 0.098, thus making the number of bedrooms over five times more important when predicting short-term returns.

For the long-term predictions, a luxury finish was found to be the most important, having an MLR coefficient of 2568; this was followed by a private pool, with 1320, and finally the bedrooms, at 635. It is important to note that the first two features were only applied singularly, whereas the bedroom importance feature was applied multiple times depending on the number of bedrooms each property held. For all data points recorded, the average number of bedrooms per property was 2.47, so the importance of bedrooms averages at around 1568, putting it behind the luxury finish variable as the feature of highest importance for long-term rentals.

4.6. Sensitivity Analysis

Since assumptions were made regarding the occupancy rates of short-term properties, two additional scenarios were briefly reviewed to assess how they would impact the choice of strategy. These two scenarios took the original occupancy rate (approximately 65%) and then increased and reduced it by 5%, as shown in Table 7.

Table 7. Sensitivity analysis.

Short-Term Data (−5% Occupancy)		Short-Term Data (+5% Occupancy)	
Mean	EUR 3929	Mean	EUR 4323
Median	EUR 3283	Median	EUR 3452
Mode	EUR 1216	Mode	EUR 1441
Std. Dev	EUR 2898	Std. Dev	EUR 3229
Range	EUR 19,077	Range	EUR 22,257
Minimum	EUR 579	Minimum	EUR 676
Maximum	EUR 19,657	Maximum	EUR 22,933

For the pessimistic scenario (−5% occupancy), 534 properties (51.9%) performed better when rented short-term compared to 494 performing better in the long term, whereas for the optimistic scenario (+5% occupancy), 669 properties (65.1%) performed better when rented short-term compared to 359 in the long term. When compared to the original scenario, the number of properties where the HBU was higher for short-term rental reduced by 67 in the pessimistic scenario but increased by 68 in the optimistic scenario. As a percentage change, this was −10.9% and 11.1%, respectively. Concerning the average rental income, the pessimistic scenario saw a reduction of EUR 207 and around −5.01%, whereas the optimistic saw an increase of EUR 187 and 4.53%. The rest of the characteristics varied by similar degrees.

5. Discussion

5.1. Optimal Strategy

During this study, ten predictive models for forecasting rental returns were created: five for long-term rentals and five for short-term rentals. When comparing these models to find the most suitable one, R-squared values were used as the determining factors for identifying the best-fitting model, as was the case for several similar studies such as Shokoohyar, Sobhani, and Sobhani's [5].

In doing so, some clear differences between the long- and short-term models were found, which led to two different approaches being required for predicting the alternative strategy. Firstly, both ML methods for long-term rental were found to be significantly lower than that of the MLR approaches, returning R-squared values of 0.560 and 0.518 for the RF and KNN, respectively. While ML models generally perform better than MLR since they can identify complex relationships and do not suffer from overfitting or multicollinearity, where data are limited, they can be outperformed simply due to them having insufficient data to make precise predictions. These results are also supported by Gnat and Doszyn's findings, which were discussed previously [26]. Due to the size of the Paphos market and the accuracy provided regarding long-term rental properties, acquiring sufficient data to achieve higher-performing ML models would be difficult, since through extensive market research, almost 50% of the properties advertised were found to be unsuitable. Therefore, the predictions of long-term returns were made through the MLR approach.

Conversely, where more data were available, both ML models outperformed their equivalent MLR methods, as can be seen for the short-term rentals. It is also interesting to note that even with 825 properties, the ML models were able to achieve a maximum R-squared value of 0.843, which, when compared to similar studies, shows that they performed well considering the amount of data. This suggests that the data were relatively predictable compared to similar studies, following a similar trend and having a high level of data quality. Moreover, it suggests that even when using a data-scraping approach to collect the data, a model can still return good results when extensive data cleansing takes place.

When comparing the MLR between the long- and short-term predictors, strangely, the long-term model outperforms the short-term one in every approach, even when it contained significantly less data. This was most likely due to the use of manual data collection versus data scraping, which resulted in the quality of the data being much

higher and inaccuracies such as outliers being more effectively removed. The drawback is the considerable time investment required to manually collect data, with only around 20 properties being collected per hour, and the risk of human error involved. In many applications, this would not be an appropriate approach, but for smaller or low-budget projects, it demonstrates that accurate models are still achievable. This also highlights the potential drawback of using AirDNA data, be it related to the data collection approach or the processing that is carried out to publish these data concisely. For this reason, special care should be taken when relying on AirDNA data and a good understanding of how these data are processed is needed to avoid misinterpretations. Regardless, these results show that both data quality and volume are crucial factors when deciding whether MLR or ML methods are used and so should be a focus when developing similar models.

Finally, when comparing the three MLR methods, every model containing geographical data, be it geospatial or distance-based, outperformed the equivalent nongeographical model, albeit to varying degrees. This supports the findings of Din, Hoesli, and Bender, who found that whenever geographical data are available, they should be used to improve model performance [28]. In this study, the long-term regression was improved by 1.5% and the short-term by 5.2% just by introducing geographical data. For the short-term prediction models, the extent to which the geographical data were beneficial was considerably greater, which may be in part due to these models generally performing worse than their long-term equivalents, but which was most probably due to the quality of the data. As such, this suggests that where the data quality is poorer, geographical data can more significantly increase a model's accuracy compared to an already well-performing model. These findings also support the conclusion that AirDNA's data may present issues regarding their accuracy, so care should be taken when using these data.

5.2. Variable Weighting

For the long-term properties, it was found that the most significant factor to consider was the overall quality of the property's finish, with a luxury finish holding the highest weight of all property attributes. This is partly expected due to the large amounts of international buyers who can afford a higher price tag thanks to relatively high wages compared to rental prices. Improving the finish quality can be achieved with varying degrees of ease depending on the specific property and its age or structure; so, targeting this must be carefully assessed taking into consideration the ROI.

Pools also considerably increased the achievable returns, especially when a private pool was available; this is thanks to the hot Mediterranean climate, which demands these types of amenities. Consequently, investors should consider the practicality of installing these at their properties where possible, as this will increase their returns and add value to the properties should they wish to sell. As with renovation, the construction of a pool is not simple and has associated issues, including acquiring permits, losing income during construction, and additional costs for landscaping before/after the work.

Adding bedrooms to a property also increases returns but would most likely require an extension; however, the conversion of spare rooms within the property could be an option. Most studies also support this, including those by Limsombunchai, Gan, and Lee and Shokoohyar, Sobhani, and Sobhani, which is unsurprising since the number of bedrooms correlates strongly with the property's size [5,25].

For all of the variables, the conversion of existing properties is most likely not cost-effective unless the property requires it. For new investors, when looking for a property to invest in, they should consider each of these and how they can ensure that they are maximised to increase returns. The purchase price for acquiring a property that maximises these characteristics will be higher than the price for a property without this maximisation, so the ROI should be calculated to identify the best property accordingly.

The variable weight for short-term properties on the other hand was significantly weighted towards the number of bedrooms when compared to all other attributes, including a luxury finish. This is understandable since the main driving factor for short-term renters

choosing a property is the number of people that the property can sleep and their choice is less about the finish quality. In addition, Airbnb is not aimed at luxury travellers, so there are much fewer luxury properties compared to the normal quality expected from short-term rentals. Therefore, for short-term rentals, the focus should be on sleeping as many people as possible whilst maintaining a quality that still ensures the property is marketable and attracts guests. As a result, investors should seek to convert spare rooms into bedrooms to maximise their ROI where possible.

5.3. Locational Importance

It was found that the location of the property had a relatively low importance when compared to the other features such as the number of bedrooms and the finish quality according to the data. This differs from many other studies' findings, including Kim, Kwon, and Choi's, which found that rental income was directly related to the property's proximity to favourable and less favourable areas [10]. A potential reason for this finding is that when looking at the location of properties in our study, they were predominantly located within touristic areas of Paphos, close to the coast, due to the nature of the city. Interestingly, it also shows that regardless of the location, both long- and short-term approaches can be viable assuming that the property is in or near touristic areas, and even suburban locations can see higher returns for short-term rentals. When looking at Paphos, this can be attributed to the large volume of tourists, with those seeking larger cheaper properties opting for the suburbs compared to the centre when on a tighter budget. This range in demand means that there is a large area of influence for short-term rentals, so assuming that the property has the desired characteristics, the location may have less of an impact. As such, this gives investors some level of flexibility to seek properties outside of the highly desirable areas, which may have overinflated prices where the demand outweighs the supply.

5.4. Incorrect Strategy

Assuming the model is accurately able to predict the optimal rental strategy for a property, it shows that many investors should consider adjusting their approach from short- to long-term renting. This is evidenced by the 224 short-term properties finding higher returns if they were rented as long-term (of course this may be slightly more since some long-term properties may also have the incorrect strategy). This stands as a stark reminder to investors to continually review their investment strategy as economic trends such as rising interest rates and reduced construction can significantly impact the dynamic of the market.

It was also found that renting a property out using the incorrect strategy would result in losing between 66.8% and 67.5% of income depending on which strategy returned the highest. As an investor, this difference is significant and can make or break an investment, which clearly shows the importance of thorough portfolio management.

5.5. Sensitivity Analysis

When comparing the optimistic and pessimistic scenarios, the overall trends were consistent with the scale of the variation. This is evident from the average returns increasing and decreasing by similar magnitudes compared to the occupancy rate, which is expected since these variables were used to define the models. Nevertheless, it is noteworthy that even with a 5% change in occupancy rate, a minimum of a 10% shift between the HBU strategy was seen in favour of the alternative strategy. This reaffirms the importance of the occupancy rate for high returns but also shows that in uncertain touristic conditions, a more reliable approach, in the form of long-term rental, could be most advantageous. Furthermore, investors should seek higher ROIs for short-term rentals to account for the increased risk that they may face due to fluctuations in vacancy.

6. Conclusions

In summary, it was found that with a larger dataset of 825 properties, ML methods such as the RF and KNN techniques significantly outperformed standard MLR approaches. Where larger datasets could not be found, the opposite occurred, whereby the MLR models outperformed ML. In these cases, the incorporation of geographical information in the form of longitude and latitude or spatial distances improved the performance of the models, and their incorporation should be used whenever possible.

The data collection methods used had both benefits and drawbacks, including the high achievable volume from data scraping or the accuracy of manual collection. It is, therefore, important to assess the aspects needing improvement in future studies and apply improvements accordingly. In general, where lower amounts of data are required, a manual approach should be taken as it ensures that the data collected are accurate and reduces the risks of misinterpretation, which can be common for data scraping. Where large amounts of data or data over a large period of time are required, data scraping, when data cleansing is performed well, can still achieve good results, especially when used in combination with ML methods.

Several key variables were found to impact the returns of properties, regardless of the strategy used: these included the number of bedrooms, the finish quality, and, for long-term properties, whether the property had a pool. These variables are difficult to change once the property is constructed without significant financial and time investments, so as part of the purchasing process, investors need to pay special attention to these factors so that they can capitalise on their investment.

This study also found that many investors had the incorrect rental strategy, which significantly impacted their returns. This emphasises the importance of good portfolio management, be it understanding economic changes or being flexible to changes in order to capitalise on current demands. Moreover, investors should always review their situation so that they can plan accordingly through more beneficial lease contracts and marketing materials to enable seamless changes.

In examining the impact of the Airbnb model on the local real-estate market in Paphos, Cyprus, our findings indicate a significant transformation in housing dynamics, particularly affecting affordability for local residents. The advent of short-term rentals has appreciably escalated property values and rental rates, primarily due to the shift by property owners towards short-term tourism-oriented rental strategies, which tend to offer higher returns than traditional long-term leases. This transition has effectively reduced the availability of affordable housing for locals, placing upward pressure on both rent and property prices. The economic disparity between the incomes of local residents and the elevated housing costs, exacerbated by the global appeal of Airbnb-style accommodations, underscores a growing market disequilibrium. Consequently, this trend warrants a critical re-evaluation of housing policies to safeguard the residential stability of the local community against the burgeoning short-term rental market. Such measures are essential to ensure that the development of tourist-centric rental properties does not inadvertently marginalise the local populace by rendering housing unaffordable.

Limitations and Future Research

Due to limited data, the volume of long-term rental data used was relatively small, meaning that the ML models were not suitable. Throughout the literature review, it was found that more accurate results could be achieved through these methods, so it can be assumed that with more data the results would be more accurate. To achieve this, additional research needs to be conducted to create a tool that can automatically carry out data scraping on the market through various sources to capture large volumes of data at one time. Of course, the data quality must remain high, or the benefit would be offset by the inaccuracies caused by poor-quality data.

Another limitation was the fact that the AirDNA data had large amounts of “ghost” properties and properties that were clearly overpriced or incorrectly marketed. The main

reason for this was that property owners did not input the data properly into the platform. As a result, significant time was spent on cleaning the data, but without manually checking each data point, it is hard to reliably say that all outliers were removed. Although studies into the accuracy of AirDNA data found that these data could generally be relied upon, they also found that AirDNA tended to overestimate the potential returns [29]. Within this study, this overestimation was not considered, nor was the fact that the long-term rentals showed advertised and not agreed prices, so there is inherent uncertainty surrounding the extent to which these impacted the predicted returns. For this reason, future research tracking the variations in these advertised values compared to the agreed prices would be interesting and would paint a clearer picture for investors on how to market their properties. This approach could also be applied retrospectively to this study, assuming economic conditions have not changed too significantly.

Another figure that was not considered was the operational costs associated with both short- and long-term rentals. Whilst these figures vary greatly depending on the property, in general, short-term operational costs are much higher than long-term costs, which will therefore impact profitability. Additional research should be conducted in this area, and again, the results can be applied retrospectively.

As discussed previously, the same occupancy rate was assumed for all short-term rental properties based on the average occupancy rates of all properties in the study. This assumption will have skewed the data for certain properties, but due to the unreliability of the AirDNA data regarding the revenue, it was the only option for estimations of the HBU. That being said, additional research into the short-term rentals in the Paphos area could shed light on the true occupancy rate, whilst additional analyses of AirDNA revenue calculations may help better define potential revenues for the Paphos region.

Finally, as all real-estate professionals know, the age of a property significantly impacts the returns that one can expect, regardless of the strategy. Unfortunately, these data are not published on short-term rental platforms, and so were left from our analysis.

Author Contributions: Conceptualisation, S.M. and T.D.; methodology, S.M. and T.D.; software, S.M.; validation, S.M.; formal analysis, S.M.; investigation, S.M. and T.D.; resources, S.M. and T.D.; data curation, S.M.; writing—original draft preparation, S.M.; writing—review and editing, S.M. and T.D.; visualisation, S.M.; supervision, T.D. and M.K.; project administration, S.M. and T.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Due to privacy reasons, the dataset and generated data are unavailable to the public.

Acknowledgments: We would like to express our gratitude to all those who have contributed to the completion of this work, including members of Neapolis University, Paphos for providing the necessary resources and fostering an environment conducive to academic growth. We want to extend our thanks to the AXIA Chartered Surveyors for supplying the AirDNA data, which enabled us to carry out this research and to AirDNA for allowing the information to be shared. Special thanks also to REAL GEOSOLUTIONS LTD who sponsored the publication expenses.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cystat. Available online: www.cystat.gov.cy/en/Announcement?id=65400 (accessed on 31 October 2023).
2. Cystat. Available online: www.cystat.gov.cy/en/PressRelease?id=67917 (accessed on 31 July 2023).
3. European Commission. Available online: www.smart-tourism-capital.ec.europa.eu/pafos-winner-2023-competition_en (accessed on 9 August 2023).
4. Cyprus Mail. Available online: www.cyprus-mail.com/2022/05/18/paphos-airport-sees-600000-passengers-in-first-four-months-of-2022 (accessed on 18 May 2022).

5. Shokoohyar, S.; Sobhani, A.; Sobhani, A. Determinants of rental strategy: Short-term vs. long-term rental strategy. *Int. J. Contemp. Hosp. Manag.* **2020**, *32*, 3873–3894. [CrossRef]
6. Manganello, B.; Tataranna, S.; De Paola, P. A Comparison of Short-Term and Long-Term Rental Market in an Italian City. *Comput. Sci. Its Appl.* **2020**, *12251*, 884–899.
7. Rodríguez-Pérez de Arenaza, D.; Ángel Hierro, L.; Patiño, D. Airbnb, sun-and-beach tourism and residential rental prices. *Curr. Issues Tour.* **2019**, *25*, 3261–3278. [CrossRef]
8. Monson, M. Valuation Using Hedonic Pricing Models. *Cornell Real Estate Rev.* **2009**, *7*, 62–73.
9. Ordinary Least Squares Method: Concepts & Examples. Available online: <https://vitalflux.com/ordinary-least-squares-method-concepts-examples> (accessed on 24 April 2024).
10. Kim, H.; Kwon, Y.; Choi, Y. Assessing the Impact of Public Rental Housing on the Housing Prices in Proximity: Based on the Regional and Local Level of Price Prediction Models Using Long Short-Term Memory (LSTM). *Sustainability* **2020**, *12*, 7520. [CrossRef]
11. Hawkins, D. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [CrossRef] [PubMed]
12. Liaw, A.; Wiener, M. Classification and Regression by Random Forest. *R News* **2002**, *23*, 18–22.
13. Crunching the Data. Available online: <https://crunchingthedata.com/number-of-trees-in-random-forests> (accessed on 3 July 2023).
14. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*, 1st ed.; Chapman and Hall/CRC: New York, NY, USA, 1984.
15. Antipov, E.; Pokryshevskaya, E. A Mass appraisal of residential apartments: An application of Random Forest for valuation and a CART-based approach for model diagnostics. *Munich Pers. RePEc* **2010**, *1*, 1772–1778. [CrossRef]
16. Mohapatra, N.; Shreya, K.; Chinmay, A. Optimization of the Random Forest. *Data Eng. Commun. Technol.* **2020**, *37*, 201–208.
17. Potrawa, T.; Tetereva, A. How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. *J. Bus. Res.* **2022**, *144*, 50–65. [CrossRef]
18. IBM. Available online: <https://www.ibm.com/docs/en/db2woc?topic=procedures-statistics-parametric-nonparametric> (accessed on 24 April 2024).
19. Zouhal, L.; Denoeux, T. An evidence-theoretic k-NN rule with parameter optimization. *IEEE Trans. Syst. Man Cybern.* **1998**, *28*, 263–271. [CrossRef]
20. Yazdani, M. Machine Learning, Deep Learning, and Hedonic Methods for Real Estate Price Prediction. Master's Thesis, Colorado University, Boulder, Colorado, 15 October 2021.
21. Nair, P.; Kashyap, I. Optimization of kNN Classifier Using Hybrid Preprocessing Model for Handling Imbalanced Data. *J. Eng. Res. Technol.* **2019**, *12*, 697–704.
22. AirDNA. Available online: <https://www.airdna.co/vacation-rental-data> (accessed on 8 July 2023).
23. Nguyen, N.; Cripps, A. Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. *J. Real Estate Res.* **2020**, *22*, 313–336. [CrossRef]
24. Benjamin, J.D.; Guttery, R.S.; Sirmans, C. Mass Appraisal: An Introduction to Multiple Regression Analysis for Real Estate Valuation. *J. Real Estate Pract. Educ.* **2020**, *7*, 65–77. [CrossRef]
25. Limsombunchai, V.; Gan, C.; Lee, M. House Price Prediction: Hedonic Price Model vs. Artificial Neural Network. *Am. J. Appl. Sci.* **2004**, *1*, 193–201. [CrossRef]
26. Gnat, S.; Doszyń, M. Parametric and Non-parametric Methods in Mass Appraisal on Poorly Developed Real Estate Markets. *Eur. Res. Stud. J.* **2020**, *23*, 1230–1245. [CrossRef]
27. Dimopoulos, T.; Bakas, N. Sensitivity Analysis of Machine Learning Models for the Mass Appraisal of Real Estate. *Remote Sens. Appl. Geoinf.* **2019**, *11*, 3047. [CrossRef]
28. Din, A.; Hoesli, M.; Bender, A. Environmental Variables and Real Estate Prices. *Urban Stud.* **2001**, *38*, 1989–2000. [CrossRef]
29. Fleischer, A.; Ert, E.; Bar-Nahum, Z. The Role of Trust Indicators in a Digital Platform: A Differentiated Goods Approach in an Airbnb Market. *J. Travel Res.* **2021**, *61*, 1173–1186. [CrossRef]
30. Agarwal, V.; Koch, J.V.; McNab, R.M. Differing Views of Lodging Reality: Airdna, STR, and Airbnb. *Cornell Hosp. Q.* **2018**, *60*, 193–199. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.