ML-driven dynamic pricing models for revenue management of short-term accommodations

TYMUR ASTASHOV, University of Twente, The Netherlands

With the fast development of machine learning (ML) technologies, it is important to understand how they can help businesses optimize their revenue. In this thesis, the focus is on short-term accommodation, such as Airbnb provides. This research paper aims to analyze a dataset of 3551 Airbnb listings in Amsterdam to find what features of a listing influence its price the most and what external factors such as seasonality and events in the city have an effect on the price. Additionally, this paper describes what benefits ML techniques could bring to the pricing dynamics and market equilibrium of the short-term rental industry. The key findings revealed that numerous property features, including the accommodation capacity, location, and number of reviews, significantly impacted the pricing of a listing. Additionally, host attributes, such as their years of experience on the platform and superhost status, also played a crucial role in determining the price. In terms of external influences, seasonality was found to be much more influential on the average daily price of listings compared to the local events.

Additional Key Words and Phrases: Short-term accommodation, machine learning, price strategy, sharing economy

1 INTRODUCTION

Short-term accommodation has been a fast-growing market in the last decade. In 2022, 2.9 million people worldwide served as hosts on Airbnb [2]. Some researchers argue that it poses a threat to the hotel industry. Services like Airbnb offer various benefits compared to hotels, from exclusive and compiling amenities to lower average room prices [14]. It is worth mentioning that the platform's exponential expansion has increased competition for hosts, requiring more intelligent marketing techniques [7].

In the current market scenario, there are two types of hosts: casual and professional. In this article, a host is considered professional if they have more than one active listing. Professional hosts are more inclined towards using dynamic pricing strategies, while casual hosts are somewhat reluctant to do it as they find it complex and unclear [8]. This paper aims to find what features of a listing affect its pricing the most. Alongside this, the article provides an analysis of the cause-effect relationship between prices and external factors such as seasonality and local events.

It is crucial to consider pricing in the accommodation industry, as it is widely recognized as a key determinant of long-term success [9]. In the hotel industry, managers tend to maximize their profit not by controlling costs, but rather by implementing more sophisticated pricing strategies [4]. Due to the interlink between hotel and sharing economies, it is reasonable to suggest that the same principle also applies to Airbnb hosts.

2 RESEARCH QUESTION

This study addresses the following research question: *How can ML-driven dynamic pricing models enhance revenue management for hosts offering short-term accommodations?* This question is being answered by developing a tool for dynamic price adjustments which will be used to find a correlation between external factors that affect the price. Additionally, extensive literature exploration is performed to compare the results from data analysis with existing literature.

The main research question can be divided into three distinct sub-questions, each contributing to the overall answer:

(1) SQ 1: What key features of short-term accommodation listings significantly influence their pricing?
(2) SQ 2: How do external factors such as seasonality, competition, and local events affect the pricing of listings?
(3) SQ 3: What are the potential benefits and drawbacks of implementing ML-driven dynamic pricing models for revenue management in the short-term rental industry?

3 LITERATURE REVIEW

This section provides an overview of literature findings regarding factors which play a role in pricing policies and the benefits and drawbacks of using ML techniques in the sharing economy domain.

3.1 Factors affecting listing pricings

Researchers have different opinions on which features play a role in pricing decisions for short-term accommodations depending on their area of focus. In a sharing economy, it is important to ensure that a host's price meets the expectations of people looking for accommodation. In the following subsections, different features affecting the price will be discussed in detail.

3.1.1 Accomodation's features. Some researchers argue that general listing features, such as the number of bedrooms and bathrooms, type of property, and minimum amount of nights play a significant role [4]. Next to this, privacy, area, and distance to the city centre or tourist attractions seem to have a positive correlation with listing price per night [3].

Another important aspect of listings is photos. High-quality photos of the accommodation's interior and the host's profile picture play an important role in consumer decision-making when choosing an accommodation [15]. Due to the dataset containing only the primary image for each listing, photographs were not incorporated into the data analysis section of the research, as this is insufficient for conducting a comprehensive and consistent analysis.

A research, which applied ordinary least squares (OLS) together with quantile regression analysis, found that features such as air conditioning, free internet, and free parking were the primary factors influencing price [5]. In contrast, some general amenities, such as the kitchen, were not found to have a significant effect on the pricing, as most of the accommodations in the analyzed dataset (94%) had a kitchen available [5].

3.1.2 Host's features and reviews. Some researchers analyse the price aspect from the perspective of social features. They revealed that the responsiveness of hosts and the number of reviews play a major role in accommodation pricing [11]. Furthermore, reviews are a complex topic, with various studies describing them in different ways. On one hand, it's logical that higher star ratings correlate with higher prices. On the other hand, due to the abundance of high ratings, it is unclear how much of a competitive advantage high ratings actually provide to hosts [7]. Renters seem to pay more attention to the amount of reviews than to a star rating, therefore, according to the literature, the amount has more importance than the actual value of the reviews [11]. It is also worth mentioning that a superhost status is influential on the pricing of listings. Researchers discovered that hosts with superhost status tend to have more reviews, higher ratings, and customers who are more willing to spend money [13]. Airbnb awards its superhost label to hosts that meet specific criteria. The following requirements have to be met in order to get a superhost badge:

- Minimum of 10 bookings annually
- No cancellations (except extraordinary circumstances)
- At least 80& of 5-star ratings
- 90% response rate

3.1.3 External factors. Seasonality has an enormous impact on the prices of accommodations. One study, similar to this one, investigated the listings in Corsica, France. The study concluded that for the same listing, the price during the season can be 19.8% higher than off-season [3]. The same study found that the property is expected to have a higher price if it offers greater privacy and area [3]. Location to the city centre and tourist attractions has a similar effect on pricing. Another study analyzed the effect of 8 sporting events on room pricing in Finnish Lapland. In general, room prices increase by an average of 14% during the event time [6]. The results discovered that only multi-day mega-events lead to a significant price increase. The Levi FIS Ski competition showed the biggest price effect, at 63.5%, as it is the most-known event in the region. However, one important finding is that the pricing effect of sports events is limited to the event period. There are no significant positive price effects for the two nights following the sporting event; there are even considerable negative pricing implications. One of the objectives of this paper was to identify if the competition affects the pricing, however, the literature suggested that modelling competition is highly complex due to the extensive variety in accommodation listings [7]. Researchers only focus on the competition between Airbnb and hotels and not on Airbnb itself. Therefore, this part was excluded from the analysis due to the aforementioned reasons.

3.2 Benefits and drawbacks of utilizing ML-driven pricing models

Pricing is undoubtedly one of the most significant benchmarks for business practices in the hospitality industry. Hosts are continuously looking for ways to optimize their business to maximize their profits. It is vitally important to set a reasonable price for the listings from the perspectives of demand and available listings in the area. The hosts' ability to determine their own prices requires them to summarise the qualities of their listings, provide a relevant estimate of demand, and, last, have a thorough understanding of the market in which they operate [3]. This could be challenging, as hosts might not have enough time to spend on analyzing similar listings and events in the area, therefore it would be highly beneficial to automate this process with ML techniques. Literature suggests that even though the price demanded by professional hosts does not differ on average from the price asked by casual hosts, the former receive a higher level of revenue on average [3]. This could be because professional hosts engage in dynamic pricing strategies, while casual hosts are reluctant to do so. Furthermore, automated pricing allows hosts to adapt their prices to market conditions quickly, which leads to the optimization of pricing strategies [12]. To sum up, the main benefit of implementing ML techniques is generating a higher revenue through educated price variability.

On the negative side, concerns such as unfairness and decreasing customer loyalty could rise due to automated pricing strategies[12]. In practice, this means that if a client checks the price of accommodation and decides to take some time before booking, but in the meantime, a major event is announced, automated pricing techniques would increase the price. This could lead to customer frustration and a subsequent decrease in customer loyalty. Also, it is challenging to build an ML model which takes into account all the relevant factors for the analysis, and if the model makes a mistake, it could lead to suboptimal pricing decisions and potential revenue loss. Literature finds that a one-fits-all model for price prediction could not take into account different listing features and therefore make less correct price predictions [12]. From the practical perspective, in the current state of the art, making a tool for price prediction requires advanced analytical skills, which could be a barrier for some hosts who may lack these skills. Therefore, they tend to either buy an out-of-the-box solution or just use fixed pricing for their listings.

4 MODELS USED

4.1 RandomForestRegressor

The RandomForestRegressor algorithm is commonly used to predict continuous outcomes. A random collection of raw data is used to generate each of the several decision trees that make up a RandomForestRegressor model. As a result, the decision tree's architecture aims to improve the model's usefulness using various data sets. Prediction models therefore use marginal forests to determine the average or weighted outcomes of every decision tree. The benefit of RandomForestRegressor is that it uses highly advanced non-linear processing and data collection techniques to accept, stabilise, and lower risk.

4.2 XGBoostRegressor

XGBoostRegressor is a powerful machine-learning algorithm used for regression tasks. The main distinctive feature of it is the gradient gradient-boosting framework, which improves prediction accuracy by sequentially building and optimizing decision trees. This strategy produces models that are simple and easy to interpret. XGBoost is highly scalable, and capable of handling substantial datasets because of its optimised memory utilisation and computational efficiency.

4.3 CatBoostRegressor

CatBoostRegressor is the best-performing model in terms of R^2 and Mean Absolute Error (MAE) in comparison to RandomForestRegressor and XGBoostRegressor. It has several distinct features, including automatic handling of categorical variables and efficient treatment of missing values. However, it may require more processing resources and take longer to train than other gradient-boosting techniques.

4.4 Why autoregressive moving average models were not used?

There are numerous models which are using autoregressive moving averages (ARMA), which are used for time series forecasting. ARMA is a powerful tool for accurate predictions, blending the strengths of both autoregressive and moving average models to analyze and forecast time-dependent data effectively. Some more advanced models, such as seasonal autoregressive integrated moving average exogenous (SARIMAX), support adding external information to determine how it affects the target value. However, the particularity of all ARMA models was that they required each timestamp in the dataset to be unique to maintain the integrity of the temporal structure and ensure accurate modelling of the time series data. This was not possible with the analyzed dataset, as it contained prices for each date for each listing, therefore it was not possible to make timestamps unique.

4.5 Performance

The measurements of the performance of the three models are described in Table 1. The performance of the models with the chosen dataset is not necessarily high when it comes to out-of-sample data. R^2 ranges from 26% to 47%, which means that only those amounts of variation in listing prices can be explained by the selected features for the analysis, while the rest is due to some other factors. Root Mean Squared Error (RMSE) describes how far predictions of the model are from the actual outcomes. It is the square root of the average of the squared differences, which in the used models range from 116.7 to 136. Mean Absolute Error (MAE) measures the differences between predicted and actual values by taking the absolute value of each difference and then calculating the average of these absolute differences, which was measured to be from 75.8 to 85.3. There are several possible reasons for these measurements:

- The dataset contained information only for one year of listings, which might not be sufficient to provide good predictive capabilities. Although numerous datasets are available online, this research required a dataset with daily prices for listings. Unfortunately, a free dataset with daily prices over a significantly longer period was not found. Theoretically, if the model were used on a larger dataset, its performance could improve, but this could not be tested. Alternatively, the dataset could include listings from other cities but Amsterdam, however, it would significantly increase the complexity of the analysis, considering the numerous differences between cities.
- The dataset might not include all possible features that significantly affect the price, which were out of scope of this research.
- The trend in the out-of-sample data may have changed, as it was taken from a dataset scraped in a different month.
- The models could be overfitted, as there are numerous features but data only from one year, which could lead to low performance. Overfitting occurs when the model fits the noise in the training data rather than the underlying relationships. This issue was addressed by decreasing the number of features, but no performance improvement was observed.

The models might not be suitable for maximizing the profit by predicting the most fitting price in the future, as their MAE are relatively high, however, they are still useful for analyzing how different features affect the pricing of listings in a specific year.

Regressor	Data Type	R^2	RMSE	MAE
RandomForest	In-sample data	0.99	24.5	7.1
	Out-of-sample data	0.26	136.0	85.3
XGBoost	In-sample data	0.46	119.0	82.2
	Out-of-sample data	0.37	127.8	87.6
CatBoost	In-sample data	0.98	24.2	11.1
	Out-of-sample data	0.47	116.7	75.8

Table 1. Performance metrics of different regressors

5 ANALYSIS OF DATA

In this section, the paper describes findings gained from data analysis and their comparison to the literature. The dataset was analyzed using several models mentioned in the previous section.

5.1 Dataset choice

During the search for a dataset, an InsideAirbnb website was discovered [1]. It had data on Airbnb listings in many cities around the world for free. However, the downside was that the data is free only for the last year, therefore, if a researcher needs to get historical data, a data request has to be created, which makes the data not free anymore, the pricing for a single city varies from \$300 to \$500. Consequently, it majorly restricts the research, as the data is freely available only for one year. Therefore, this paper focuses on a dataset which was freely available and published on 06 June 2023. The initial dataset contained 82 columns and 1,276,768 rows (containing price and features for every 3551 listing for a calendar year). This dataset was chosen as Amsterdam has a high density of Airbnb listings per district and numerous events are happening there, which could potentially affect the prices of listings.

5.2 Data pre-processing

The data was cleansed by deleting columns that were not useful for the research. In total, only 18 columns were used for the analysis. Their overview is available in Appendix A.2. Considering that some columns had textual values, which are not suitable for the models often used for machine learning in the accommodation sector, they had to be transformed. Here is an overview of the steps taken:

(1) True/False values were transformed to 1/0 so that the models could make use of them. This was applied to `host_is_superhost` and `instant_bookable` columns.
(2) Date of booking was transformed from "text" to "DateTime" format and split into three columns: `day_of_month`, `month`, and `year`.
(3) `Price` was transformed from "text" to "float64" format after deleting the dollar sign at the beginning.
(4) `Neighbourhoods_cleansed` initially had 22 possible values, but they were categorized into 6 groups: East-South, East, South, North, Centrum, and West. It was required for one-hot encoding to be performed. The overview of the distribution is available in Appendix A.1.
(5) For the `room_type` column, 3 dummies were created: Entire home/apt, Private room, and Shared Room. And they were also encoded using a one-hot technique.
(6) The column `host_since` contained information on when the host joined the platform, however, it was not particularly useful for the model, so it was converted to a column `years_on_platform`, which shows a rounded number of years on the platform
(7) Information about the bathrooms, which was initially contained in a column with strings was converted into two columns: the number of bathrooms in the accommodation and a boolean-type column which represents if the bathroom(s) is/are private or not
(8) Listings priced above $1500 or with more than 1000 reviews were excluded from the analysis, as these values were considered extreme.

The data was normalized by using a `StandardScaler`, which standardizes the features by removing the mean and scaling to unit variance. Specifically, each value of numerical columns is transformed according to the formula:

$$z = \frac{(x - \mu)}{\sigma}$$

Note: x is the original value, μ is the mean of the training samples, σ is the standard deviation of the training samples, and z is the standardized value.

This transformation ensures that the data has a mean of 0 and a standard deviation of 1, which helps in improving the performance of machine learning models by bringing all features to the same scale.

5.3 Internal features affecting accommodation pricing

From the feature importance graph (Figure 1), it is possible to determine which features have the most effect on the pricing of accommodation. Based on the figure, a mixture of accommodation features and reviews plays a crucial role in listing pricing. The most important features of a property are the unit's capacity, minimum nights per booking, and the number of bathrooms. From the perspective of reviews, the most influential figures are the amount of reviews per month, the number of overall reviews, and the review score. Interestingly, the actual score for reviews matters 15.4% less than the amount of monthly reviews a listing gets. Among other features, specific attributes considering the host, play a role, including

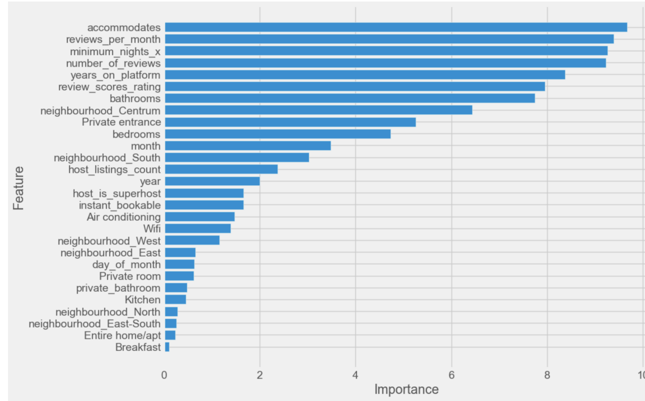

Fig. 1. Feature importance chart of CatBoostRegressor Model

the duration of the host's tenure on the platform, superhost status, and the number of listings managed. Tenure on the platform holds more considerable importance, approximately three and a half times greater than that of superhost status and the number of listings. This observation aligns with the expectation that hosts with more experience often charge higher prices compared to newly registered proprietors.

By analyzing the heatmap graph, it is possible to find out about correlations between different features (see Appendix B). The most valuable column for this paper is the last one, about the price. The highest positive correlation (0.34-0.44) is observed between price and the capacity to accommodate people, number of bedrooms, and bathrooms. Also, it is worth mentioning that having a private room is highly positively correlated to having a private bathroom, their correlation coefficient is 0.69. On the contrary, there is a negative relationship concerning having a private bathroom in an entire apartment/house. This may be because hosts often do not emphasize bathroom privacy when describing an entire unit. A similar correlation is found between the number of total reviews and reviews per month, which entails that listings with a continuous inflow of new reviews tend to have a high total amount of reviews. In terms of amenities, it is noticeable that there is a strong correlation between the presence of a kitchen and a listing being an entire unit, whereas the correlation is the opposite for kitchens and private rooms. Usually, when working with heatmaps, researchers aim to reduce the presence of highly correlated independent variables, a phenomenon known as multicollinearity. However, since this paper focuses on tree models, multicollinearity is not a significant concern. Decision trees inherently handle multicollinearity due to their distinctive structure and methodology [10].

The CatBoostRegressor model was also analyzed for the SHapley Additive exPlanations (SHAP) values. SHAP is a method based on game theory that helps to explain the results produced by any machine learning model. The results for SHAP analysis are provided in Figure 2. This figure shows the 20 most influential features for listing pricing. The list is sorted by the amount of influence, so the capacity of the unit, whether it is located in the city centre and the amount of bathrooms are the most influential on the price. The

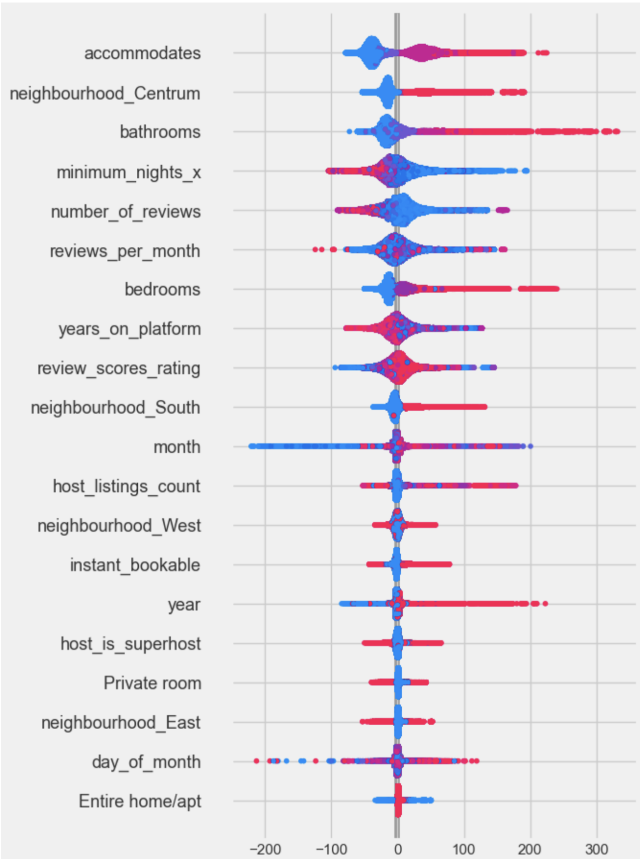

Fig. 2. SHAP Values for CatBoostRegressor Model

colour gradient represents feature value, so red means higher value and blue means lower value. For instance, in terms of the number of people, a listing can accommodate, a predominantly red colour signalises that properties accommodating more people tend to have higher predicted outcomes. On the contrary, the month has both positive and negative influence on the price as it has a lot of both red and blue, meaning that for some months the price is increasing, while for others it is sharply decreasing. Overall, this graph is useful to analyze what features are the most influential on the price and specifically in what way.

5.4 External features affecting accommodation pricing

There are several things to consider when it comes to the external effects of prices. Firstly, it is noticeable from the figure that prices are highly dependent on the day of the week (Figure 3). They tend to accelerate by approximately 8% when the weekend comes. Secondly, there is a major seasonal effect noticed in the average prices. During all other seasons, except Spring, average prices fluctuate between $287 to $342. However, during spring the average prices surge to a maximum of $387 per night on King's Day. In examining the impact of special occasions on average prices, the data does not indicate a consistent correlation between specific events and elevated average prices, with the notable exceptions of King's Day, New Year's Day and the Tulip Festival. These three events are outliers, suggesting that their special significance for incoming travellers may drive price increases. However, overall, the data reveals that seasonality and the day of the week have a more substantial influence on average prices. This suggests that periodic factors play a more critical role in price fluctuations than singular events. This aligns with the findings from the literature, as a dataset of Airbnb listings from Corsica, France, revealed that for the same listing, the price during the season can be 19.8% higher than off-season [3]. The research on listings in Finish Lapland examined that the increase in pricing during the events is limited to their duration and afterwards the prices tend to decrease [6]. This tendency is not observed in the analyzed dataset, however, a comparable pattern related to seasonality is noticed. After the tulip season in Amsterdam finishes (marked by the end of Keukenhof), there is a 7.5% drop in the average price.

6 LIMITATIONS

The main limitation of this research paper is focused on the data availability. The analysed dataset provides information only for 12 consequent months. It was not possible to find another free dataset with dates and prices, as there were no freely available datasets available for a significantly longer duration of time. Another limitation was the analysed city, as the dataset contained information only on listings in Amsterdam. Even though Amsterdam was found to be a good start for the analysis, as it has a great tourist influx, numerous events, and a great variety of properties; adding more cities could potentially improve the prediction benchmarks of the models.

7 CONCLUSION

This paper provided an overview of factors, both internal and external, which affect the pricing of short-term accommodation. These factors were analyzed both from the perspective of literature and a practical perspective by examining a dataset of Airbnb listings in Amsterdam. The main findings included that numerous property features, such as the number of people it can serve, location, and amount of bathrooms and reviews were highly influential towards the pricing of a particular listing. The same implies for host features, such as the number of years they have been using the platform and a superhost status. A simple app with user interface was created to test the models with users. The models, trained on the dataset, did not provide a significantly sound prediction performance for revenue management due to several potential reasons, such as low availability of data, absence of additional potentially valuable features affecting prices, or overfitting. The models were additionally cross-validated on a dataset of 18 months, but the performance did not improve, potentially due to the fact that this increase in duration was not significant. However, the models were useful for providing an analysis of factors affecting listing prices for a specific period of 12 months in Amsterdam. In theory, ML techniques could be highly beneficial for revenue management, but a broader selection of datasets is needed, so that the predictions are precise and in line with ongoing trends. In contrast, they could negatively impact customer loyalty and cause unfairness in the market of the sharing economy.

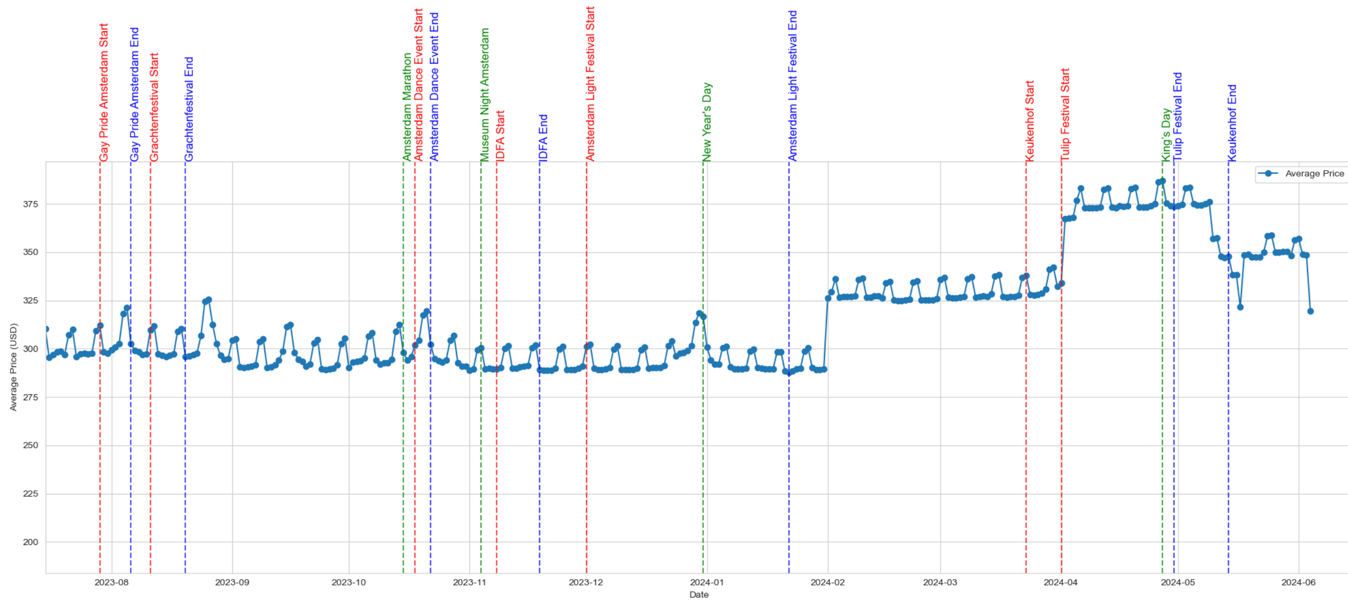

Fig. 3. Average Prices With Events in Amsterdam

8 FUTURE WORK

This research provides a great basis for the analysis of the effects of different features on listing pricing. However, it would be beneficial to include more data (longer period of analysis, additional cities). Also, a set of other models, which might provide more accurate results, such as ARMA could be used for the analysis, if a suitable dataset is found. From the perspective of the app, it could be tested on user experience features and improved upon the findings.

9 AI STATEMENT

During the preparation of this work, the author used ChatGPT in order to paraphrase complex sentences, refine code, and support creation of Graphical User Interface, as well as Grammarly to correct grammatical mistakes in the text. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the work.

REFERENCES

[1] [n. d.]. Get the Data — insideairbnb.com. https://insideairbnb.com/get-the-data/. [Accessed 30-06-2024].

[2] Nicola Camatti, Giacomo di Tollo, Gianni Filograsso, and Sara Ghilardi. 2024. Predicting Airbnb pricing: a comparative analysis of artificial intelligence and traditional approaches. *Computational Management Science* 21, 1 (May 2024), 30. https://doi.org/10.1007/s10287-024-00511-4

[3] Georges Casamatta, Sauveur Giannoni, Daniel Brunstein, and Johan Jouve. 2022. Host type and pricing on Airbnb: Seasonality and perceived market power. *Tourism Management* 88 (Feb. 2022), 104433. https://doi.org/10.1016/j.tourman.2021.104433

[4] Manojit Chattopadhyay and Subrata Kumar Mitra. 2020. What Airbnb Host Listings Influence Peer-to-Peer Tourist Accommodation Price? *Journal of Hospitality & Tourism Research* 44, 4 (May 2020), 597–623. https://doi.org/10.1177/1096348020910211 Publisher: SAGE Publications Inc.

[5] Gábor Dudás, Tamás Kovalcsik, György Vida, Lajos Boros, and Gyula Nagy. 2020. Price determinants of Airbnb listing prices in Lake Balaton Touristic Region. *European Journal of Tourism Research* 24 (Feb. 2020), 2410.

[6] Martin Falk and Markku Vieru. 2020. Short-term hotel room price effects of sporting events. *Tourism Economics* 27 (Jan. 2020), 135481662090195. https://doi.org/10.1177/1354816620901953

[7] Chris Gibbs, Daniel Guttentag, Ulrike Gretzel, Jym Morton, and Alasdair Goodwill. 2017. Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing* 35 (April 2017), 1–11. https://doi.org/10.1080/10548408.2017.1308292

[8] Chris Gibbs, Daniel Guttentag, Ulrike Gretzel, Lan Yao, and Jym Morton. 2017. Use of dynamic pricing strategies by Airbnb hosts. *International Journal of Contemporary Hospitality Management* 30 (Nov. 2017), 00–00. https://doi.org/10.1108/IJCHM-09-2016-0540

[9] Wei-Ting Hung, Jui-Kou Shang, and Fei-Ching Wang. 2010. Pricing determinants in the hotel industry: Quantile regression analysis. *International Journal of Hospitality Management* 29, 3 (Sept. 2010), 378–384. https://doi.org/10.1016/j.ijhm.2009.09.001

[10] S. B. Kotsiantis. 2013. Decision trees: a recent overview. *Artificial Intelligence Review* 39, 4 (April 2013), 261–283. https://doi.org/10.1007/s10462-011-9272-4

[11] Donghun Lee, Woochang Hyun, Jeongwoo Ryu, Woo Jung Lee, Wonjong Rhee, and Bongwon Suh. 2015. An Analysis of Social Features Associated with Room Sales of Airbnb. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW'15 Companion)*. Association for Computing Machinery, New York, NY, USA, 219–222. https://doi.org/10.1145/2685553.2699011

[12] Veronica Leoni and William Nilsson. 2021. Dynamic pricing and revenues of Airbnb listings: Estimating heterogeneous causal effects. *International Journal of Hospitality Management* 95 (May 2021), 102914. https://doi.org/10.1016/j.ijhm.2021.102914

[13] Sai Liang, Markus Schuckert, Rob Law, and Chih-Chien Chen. 2017. Be a "Superhost": The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management* 60 (June 2017), 454–465. https://doi.org/10.1016/j.tourman.2017.01.007

[14] Troy Lorde, Jadon Jacob, and Quinn Weekes. 2019. Price-setting behavior in a tourism sharing economy accommodation market: A hedonic price analysis of AirBnB hosts in the caribbean. *Tourism Management Perspectives* 30 (April 2019), 251–261. https://doi.org/10.1016/j.tmp.2019.03.006

[15] Richard TR Qiu, Daisy Xuefeng Fan, and Anyu Liu. 2018. Exploring the Booking Determinants of the Airbnb Properties: An Example of the Listings of London. In *Information and Communication Technologies in Tourism 2018*, Brigitte Stangl and Juho Pesonen (Eds.). Springer International Publishing, Cham, 44–51. https://doi.org/10.1007/978-3-319-72923-7_4

A HEADINGS IN APPENDICES

A.1 Appendix A.1

Column Name	**Description**
accommodates	Number of guests the listing can accommodate
bedrooms	Number of bedrooms in the listing
bathrooms	Number of bathrooms in the listing
private_bathroom	Indicates if the bathroom is private (1) or shared (0)
host_is_superhost	Indicates if the host is a superhost (1) or not (0)
host_listings_count	Total number of listings by the host
minimum_nights_x	Minimum number of nights required for a stay
number_of_reviews	Total number of reviews the listing has received
review_scores_rating	Average rating score from reviews
instant_bookable	Indicates if the listing is instantly bookable (1) or not (0)
reviews_per_month	Average number of reviews per month
Entire home/apt	Indicates if the listing is an entire home/apartment (1) or not (0)
Private room	Indicates if the listing is a private room (1) or not (0)
neighbourhood_Centrum	Indicates if the listing is in the Centrum neighbourhood (1) or not (0)
neighbourhood_East	Indicates if the listing is in the East neighbourhood (1) or not (0)
neighbourhood_East-South	Indicates if the listing is in the East-South neighbourhood (1) or not (0)
neighbourhood_North	Indicates if the listing is in the North neighbourhood (1) or not (0)
neighbourhood_South	Indicates if the listing is in the South neighbourhood (1) or not (0)
neighbourhood_West	Indicates if the listing is in the West neighbourhood (1) or not (0)
day_of_month	Day of the month when the data was recorded
month	Month when the data was recorded
year	Year when the data was recorded
years_on_platform	Number of years the listing has been on the platform
Kitchen	Indicates if the listing has a kitchen (1) or not (0)
Wifi	Indicates if the listing has wifi (1) or not (0)
Private entrance	Indicates if the listing has a private entrance (1) or not (0)
Air conditioning	Indicates if the listing has air conditioning (1) or not (0)
Breakfast	Indicates if the listing offers breakfast (1) or not (0)

Table 2. Description of each column in the dataset

A.2 Appendix A.2

East-South	**East**	**South**
Bijlmer-Centrum Bijlmer-Oost Gaasperdam - Driemond	Oostelijk Havengebied - Indische Buurt Noord-Oost Oud-Oost Watergraafsmeer IJburg - Zeeburgereiland	De Pijp - Rivierenbuurt Buitenveldert - Zuidas Zuid
North	**West**	**Centrum**
Oud-Noord	Bos en Lommer De Aker - Nieuw Sloten De Baarsjes - Oud-West Noord-West Osdorp Slotervaart Westerpark Geuzenveld - Slotermeer	Centrum-Oost Centrum-West

Table 3. Neighborhoods in Amsterdam categorized by region

B APPENDIX B

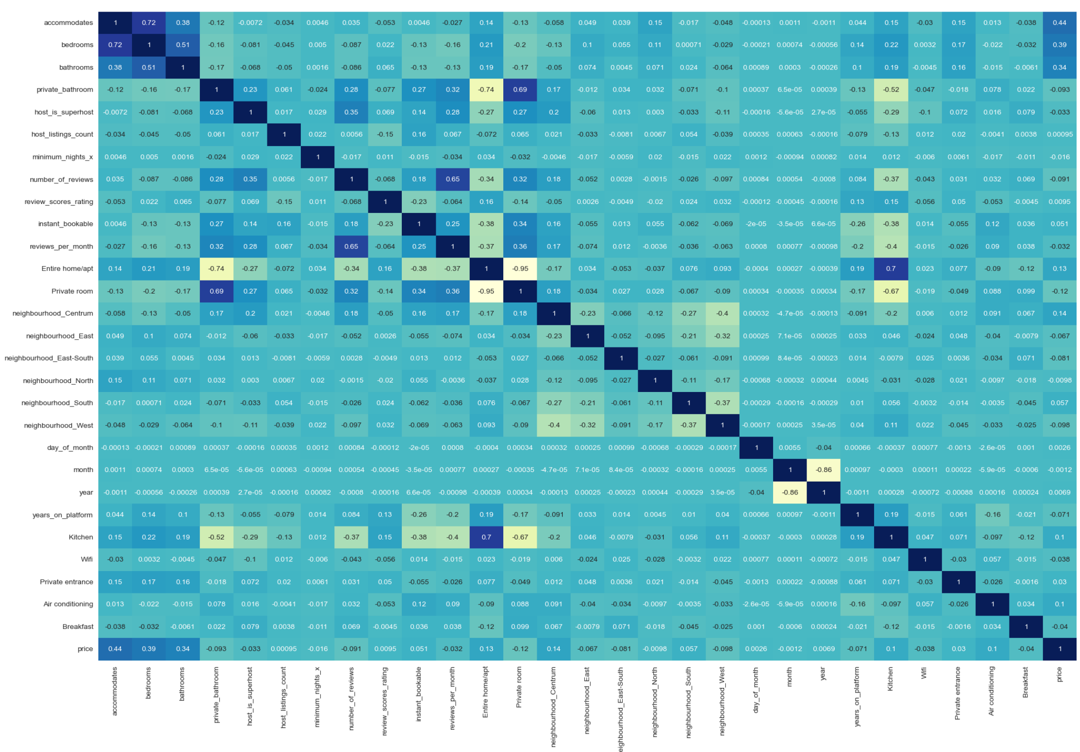

Fig. 4. Heatmap for dataset features