

Enhancing the Airbnb Experience: Dynamic Pricing Strategy

Surekha Janrao¹, Vijay Kapure², Shudhodhan Bokefode³, Jayesh Sarwade⁴, Rohini Patil⁵,⁶Dr. Priya Karemore

¹ Department of Computer Engineering, K.J.Somaiya Institute of Technology, Sion, Mumbai-400022, India

² Department of Applied Science and Humanities, Xavier institute of engineering, Mahim, Mumbai, 400016, India

^{3,4} Department of Computer Engineering, Terna Engineering College, Nerul, Navi Mumbai-400706, India

⁵ Department of Information Technology, MIT School of Engineering and Science, Pune, 412101, India

⁶ Department of Information Technology, RSCOE, Pune, 412101, India

ARTICLE INFO

Received: 25 Oct 2024
Revised: 17 Nov 2024
Accepted: 18 Dec 2024

ABSTRACT

In the ever-changing landscape of the world's sharing economy, various platforms like Airbnb have revolutionized the way individuals browse through accommodations and monetize their properties. However, the challenge of efficiently matching guests with hosts while ensuring competitive pricing remains paramount. This paper presents methodologies that aim to optimize guest-host matching and pricing strategies to enhance the overall Airbnb experience. Leveraging real-world data, our approach combines advanced recommendation system and dynamic pricing algorithms to provide personalized recommendations for guests and maximize revenue for hosts. Through comprehensive evaluation and analysis, we demonstrate the effectiveness of our methodologies in improving guest satisfaction, increasing booking conversion rates, and enhancing revenue generation.

Keywords: Recommendation systems, Dynamic pricing, Optimization

INTRODUCTION

In recent years, the advent of online accommodation platforms has transformed the hospitality industry landscape, offering travellers a diverse range of lodging options while empowering individuals to monetize their properties. Among these platforms, Airbnb has emerged as a frontrunner, changing the way people seek accommodations and connect with hosts all over the world. However, with the proliferation of listings and the growing complexity of user preferences, optimizing guest-host matching and pricing strategies has become severely important to ensure a smooth and satisfying experience for both guests and hosts.

The success of Airbnb hinges on its ability to efficiently match guests with suitable hosts and determine optimal pricing strategies that balance affordability and competitiveness. Effective guest-host matching requires consideration of various factors, including location preferences, accommodation type, amenities, and host characteristics. Similarly, pricing optimization involves dynamically adjusting listing prices based on demand forecasting, market conditions, and competitor prices to maximize revenue for hosts while maintaining attractiveness to guests.

The paper [1] conducts a literature review highlighting the focus on dynamic pricing analysis for Amsterdam Airbnb hosts utilizing K-Means clustering, leveraging Airbnb data from March 4, 2021, to March 6, 2022, achieving a 96.21% accuracy rate. It aims to offer insights into marketing strategies, particularly dynamic pricing, and outlines the research method involving data acquisition and merging, while also referencing a previous study on dynamic pricing in the airline industry utilizing Support Vector Machine (SVM). The paper [2] examines dynamic pricing among Amsterdam Airbnb hosts, finding that frequent price adjustments correlate with higher revenue metrics, and hosts with more properties tend to perform better. It also discusses pricing strategies' social control aspect and hedonic models' impact on pricing, emphasizing the influence of location and property attributes. Additionally, it explores the tension between the sharing economy ethos and profit-driven practices of platforms like Airbnb, which it contends lead to the commercialization and commodification of residential areas. In [3] The paper explores various machine learning models, including Random Forests, Extra Trees, and XGBoost, for predicting Airbnb

rental prices in New York City, noting their comparable performance. Additionally, other studies emphasize the effectiveness of decision tree-based models, particularly Random Forests, in accurately predicting rental prices and addressing challenges in establishing reliable pricing for rental properties in the real estate industry. Furthermore, machine learning algorithms, such as the RIPPER algorithm and a combination of Random Forests and Recurrent Neural Networks, have shown promise in predicting housing prices accurately by leveraging diverse data types emphasizes the focus on pricing strategies of new Airbnb hosts, highlighting the necessity of dynamic pricing and the challenges they encounter in determining optimal prices[4]. It aims to validate hypotheses concerning the impact of various factors on pricing decisions, including listing characteristics, pandemic effects, performance enhancements, and rental policies. Previous research underscores the influence of factors like property attributes, host credentials, ratings, amenities, and customer feedback on pricing in the sharing economy. Additionally, the study stresses the significance of city-level analysis for accurately identifying price determinants for both new and experienced hosts within platforms like Airbnb. The study delves into the intricacies of pricing strategies within the realm of Airbnb accommodations, shedding light on the pivotal roles played by host experience and market demand dynamics. It intricately examines how hosts adeptly navigate the pricing landscape, dynamically adjusting rates based on their accumulated experience and the prevailing market demand for specific booking dates. Moreover, the study uncovers a fascinating trend: the inverse relationship between the number of reviews received by an accommodation and its pricing. This intriguing finding suggests that an unusually high number of reviews may trigger guest skepticism, prompting hosts to lower prices in response. Furthermore, the research underscores the significant advantage that professional hosts enjoy over their non-professional counterparts in commanding higher prices. This disparity underscores the vital role of experience and expertise, particularly in marketing and price management, in shaping effective pricing strategies within the Airbnb ecosystem. Additionally, the study sheds light on hosts' astute responsiveness to market demand levels, providing further evidence of sophisticated revenue management practices permeating the shared accommodation sector. This keen adaptation to market dynamics underscores the nuanced approach hosts employ to optimize their pricing strategies and maximize revenue potential in a competitive landscape [5]. The study highlights the extensive exploration of dynamic pricing across various industries, including retail, airlines, travel, and hotel revenue management. Traditional approaches in retail focus on demand distribution or polynomial functions tied to price and product expiry, while recent advancements in hotel revenue management involve stochastic models like Markov Chain Monte Carlo (MCMC) [6].

In [7] authors introduced, Journey Ranker, a sophisticated multi-task deep learning model crafted to enhance the Airbnb search journey, ensuring a harmonious balance between guest and host preferences. Comprising four modules designed with distinct responsibilities, Journey Ranker offers versatility for adaptation to diverse use cases beyond Airbnb's search ranking realm. By leveraging intermediate guest actions as guiding milestones and contextual cues such as guest state and search queries, the model aims to facilitate successful bookings seamlessly.

In [8] author extensively references pertinent studies in the realm of dynamic pricing, encompassing various methodologies such as linear stochastic bandits, demand estimation with machine learning, model combination, and feature-based dynamic pricing. Additionally, it incorporates insights from research on the pricing dynamics of networked hospitality exchange platforms like Airbnb. Moreover, the paper introduces a pricing strategy model deployed in production to aid Airbnb hosts in setting prices more effectively, which underwent rigorous evaluation both offline and online. Results indicated its superiority over a direct max-rev pricing strategy; although the authors acknowledge the on-going necessity for refining demand curve estimation to further refine pricing strategies.

In [9] the study underscores its focus on dynamic pricing and matching in two-sided queues with strategic servers, particularly in online marketplaces like ridesharing, with the goal of maximizing profit while minimizing expected waiting times. The authors devise a versatile framework to analyse diverse strategic behaviours among servers, encapsulating their actions within a tailored cost function adaptable to various scenarios. They introduce a novel probabilistic fluid problem as an infinite-dimensional optimization program, offering an upper bound on attainable profit.

In [10] the authors of the study delve into the investigation of intertemporal price discrimination's implementation and its impact on revenue within Airbnb listings, addressing a notable gap in the literature regarding revenue management techniques applied in non-traditional accommodations and pricing strategies on peer-to-peer platforms. It situates the study within existing research on dynamic pricing strategies in both traditional accommodation industries and peer-to-peer marketplaces, as well as within the realm of estimating heterogeneous treatment effects.

OBJECTIVES

This research paper aims to address the challenges associated with guest-host matching and pricing optimization on the Airbnb platform. Leveraging real-world data and innovative methodologies, we propose strategies to enhance the Airbnb experience for users. By optimizing guest-host matching and pricing strategies, we aim to improve guest satisfaction, increase booking conversion rates, and maximize revenue generation for hosts. The paper is structured as follows: in the subsequent sections, we present the methodology used to develop and evaluate our proposed strategies, followed by a comprehensive analysis of the results obtained. We discuss the implications of our findings and provide recommendations for practitioners and stakeholders in the hospitality industry. Through this research, we endeavor to contribute to the advancement of online accommodation platforms and the broader hospitality ecosystem.

METHODS

We adopt a systematic approach to develop and evaluate our methodologies, comprising several key steps. Firstly, we collect and preprocess comprehensive datasets from Airbnb, including information about listings, hosts, and guest reviews. We then define guest preferences and filter potential hosts based on compatibility criteria such as neighbourhood preferences, room type, and price range. Utilizing selected features, we calculate the similarity between guests and hosts and generate personalized recommendations for guests. Additionally, we develop dynamic pricing algorithms that adjust listing prices in real-time based on demand forecasting and market dynamics.

3.1 Dataset Description

The dataset utilized in this research, titled "New York City Airbnb Dataset," is sourced from Kaggle, a platform renowned for data science competitions and datasets. It offers an extensive collection of information pertaining to Airbnb listings and host data within the vibrant metropolis of New York City. Providing a comprehensive snapshot of the city's short-term rental market, this dataset encompasses various facets of property listings, host characteristics, geographical attributes, pricing details, and review metrics. Each entry in the dataset is characterized by a unique identifier, encompassing features such as the listing's title, host ID, host name, borough (neighbourhood group), specific neighbourhood, geographical coordinates (latitude and longitude), room type, nightly price, minimum nights required for booking, total number of reviews, date of the last review, average reviews per month, number of listings managed by the host, and availability within the next 365 days.

The dataset's scope spans Airbnb listings across New York City's five boroughs: Manhattan, Brooklyn, Queens, Bronx, and Staten Island. It represents a snapshot of the short-term rental market during a specific timeframe, enabling analysis of trends and patterns within the city's dynamic hospitality landscape. While the dataset provides valuable insights into Airbnb activity within the city, it is not without limitations. Potential issues such as missing values, outliers, and inconsistencies necessitate preprocessing steps to ensure data quality and reliability. Additionally, being a snapshot of listings during a specific timeframe, it may not reflect real-time changes or account for seasonal fluctuations in demand.

Despite these limitations, the dataset serves as a valuable resource for researchers, analysts, and

Industry practitioners seeking to delve into Airbnb trends, explore market dynamics, and develop strategies for optimizing guest-host matching and pricing. By leveraging the rich trove of data contained within this dataset, researchers can gain a deeper understanding of the New York City short-term rental market and derive actionable insights to enhance the Airbnb experience for both guests and hosts.

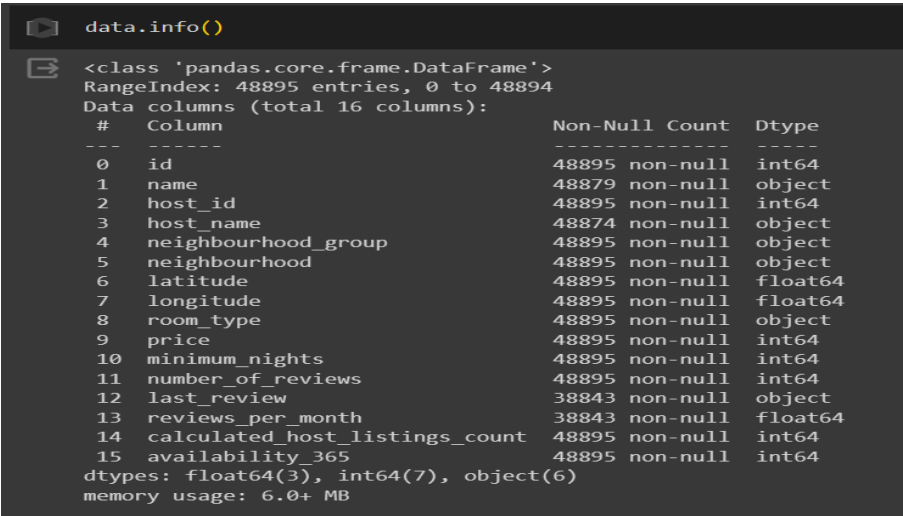


Figure 1. Dataset Description

3.2 Statistical Analysis

3.2.1 Descriptive Analysis

Descriptive analysis involves summarizing and visualizing the characteristics of the dataset to gain insights into its distribution, central tendency, and variability. This type of analysis provides a snapshot of the data's current state without making predictions or prescribing actions. Descriptive statistics such as mean, median, standard deviation, and quartiles are commonly used to summarize numerical features. Visualization techniques such as histograms, box plots, and scatter plots help visualize the distributions and relationships between variables.

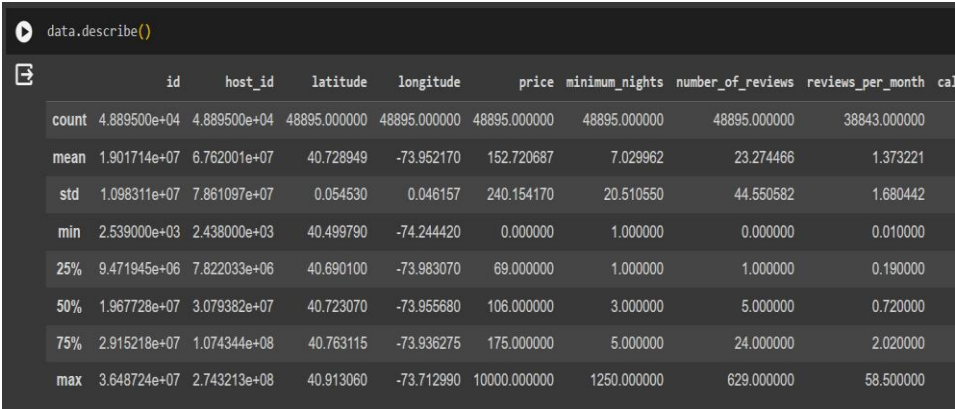


Figure 2. Statistics of Data

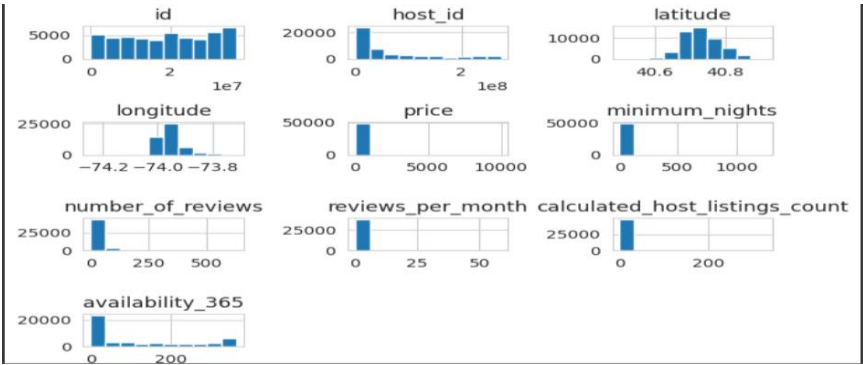


Figure 3. Descriptive Statistics

3.2.2 Predictive Analysis

Predictive statistics involves forecasting future trends, outcomes, or behaviours based on historical data. It aims to provide insights into what might happen in the future. Predictive analytics can be leveraged to build models predicting prices, reviews per month, and availability for Airbnb listings, allowing users to anticipate and plan for their stays more effectively.

The histogram in figure reveals that a significant concentration of values lies below \$200. Specifically, the majority of reservations, totalling 47,994 (98.15%), fall within the range of 0 to \$58,823, as highlighted in the frequency distribution table. This visualization enhances our comprehension of the price distribution, emphasizing the prevalence of values within a more affordable price range.

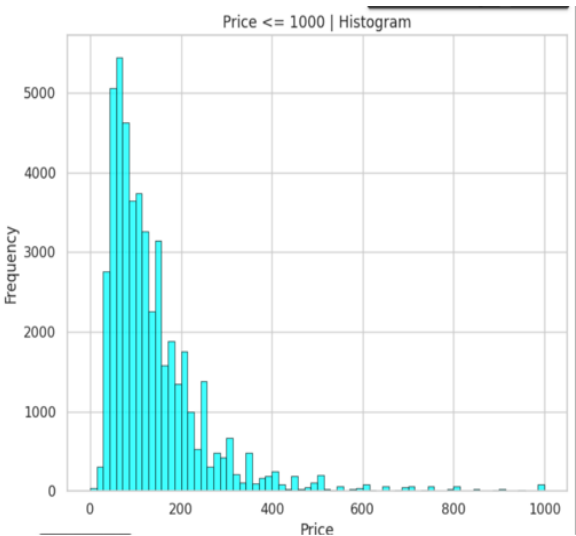


Figure 4. Histogram for price and frequency

Among the myriad questions arising from our analysis, one crucial query finds an answer: "Which neighbourhood stands out as the most sought-after by customers for advertisements and accommodation reservations on the Airbnb website among the 221 neighbourhoods considered?" The below histogram gives the most popular neighbourhoods.

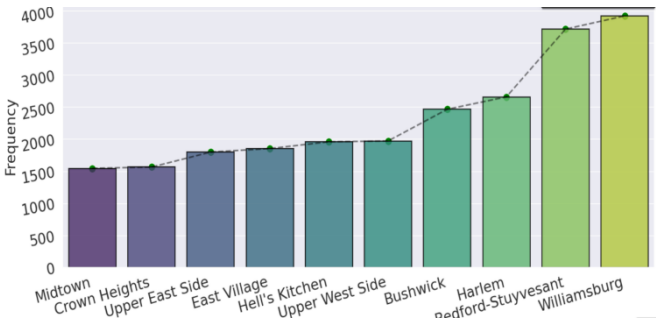


Figure 5. Descriptive Statistics

3.3 Data Preprocessing

Data preprocessing plays a crucial role in ensuring the quality and usability of the dataset for analysis. In this section, we outline the steps followed to address missing values, outliers, and inconsistencies, and to prepare the data for further analysis.

3.3.1 Handling Missing Values

Missing values are a common issue in real-world datasets and can adversely affect the integrity of analyses. To mitigate this, we first identify and quantify missing values across all features in the dataset. For categorical features such as 'name' and 'host_name', missing values are handled by imputing the most frequent value. Numerical features such as 'reviews_per_month' and 'last_review' are imputed using appropriate statistical measures such as mean or median. For features with a significant number of missing values, we evaluate the feasibility of imputation versus removal based on the impact on data quality and analysis outcomes. Handling missing values in a dataset is crucial for ensuring the accuracy and reliability of any analysis or model built upon it. When dealing with missing values in an Airbnb dataset, we want to employ various strategies depending on the nature of the data and the specific context of your analysis. There's no one-size-fits-all solution for handling missing values.



Figure 6. Dataset Pre-processing

3.3.2 Outlier Detection and Treatment

Outliers, or data points that deviate significantly from the rest of the distribution, can distort analysis results and model performance. We employ various techniques to detect outliers in numerical features such as 'price', 'minimum_nights', and 'number_of_reviews'. Visualization tools such as box plots and scatter plots are utilized to identify anomalous data points. Outliers are then treated using methods such as winsorization, where extreme values are replaced with less extreme values based on predefined thresholds, or through transformation techniques such as log transformation.

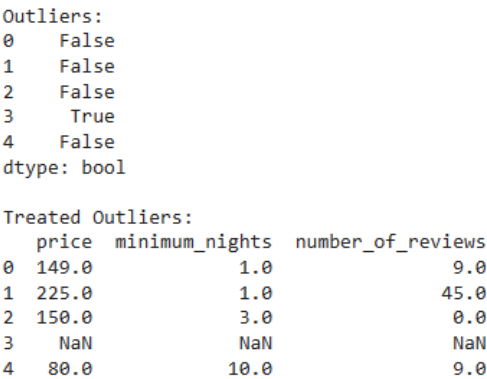


Figure 7. Outlier Detection

3.3.3 Feature Engineering

Feature engineering involves creating new features or transforming existing ones to enhance the predictive power of the dataset. In this step, we derive new variables based on domain knowledge and insights gained from exploratory data analysis. For example, we create a new feature 'host_experience' by aggregating the total number of reviews received by each host, providing a measure of their experience and reputation. Additionally, we encode categorical variables such as 'neighbourhood_group' and 'room_type' using one-hot encoding to convert them into numerical representations suitable for modeling.

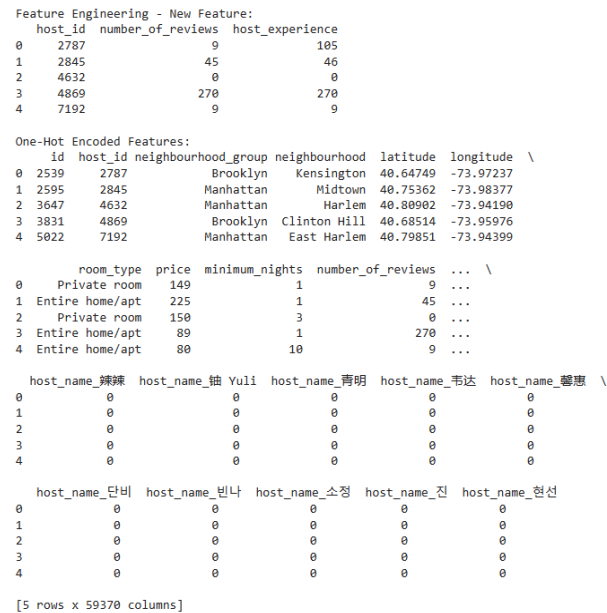


Figure 8. Feature Engineering

3.3.4 Data Cleaning and Standardization

Data cleaning involves identifying and correcting errors or inconsistencies in the dataset to ensure uniformity and accuracy. This includes standardizing formats for date variables such as 'last_review', converting them into a consistent format for analysis. Furthermore, we standardize numerical features such as 'latitude' and 'longitude' to a common scale to facilitate comparisons and modelling.. Finally, we perform data validation checks to verify the integrity of the dataset and ensure that it adheres to predefined quality standards.

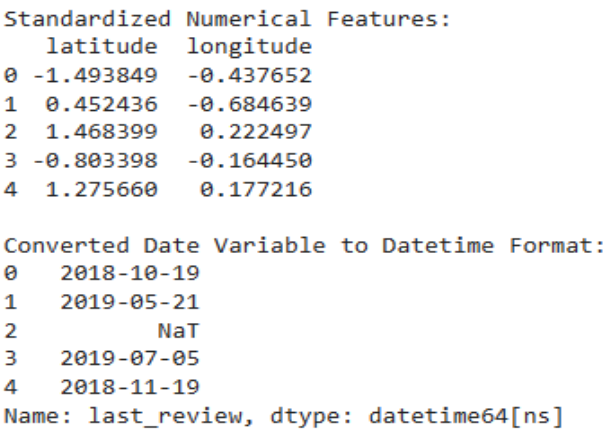


Figure 9. Standardization

3.3.5 Data Splitting

To facilitate model training and evaluation, we split the preprocessed dataset into training and testing sets. The training set is used to train predictive models, while the testing set is reserved for evaluating model performance on unseen data. We ensure temporal consistency in the data splitting process, maintaining the chronological order of observations to avoid data leakage and ensure the reliability of model evaluation results.

3.4 Pricing Optimization

The dynamic pricing strategy was implemented using a linear function that adjusts the listing prices of Airbnb properties in real-time based on perceived demand. The methodology involved pre-processing the dataset to normalize the demand and price features, allowing for fair comparison and effective modelling. The demand for each property was calculated as the difference between the total number of days in a year and the availability of the property throughout the year. This normalized demand feature, along with the normalized price, was used as input to the linear function, which incorporated parameters such as slope and intercept to dynamically adjust the listing prices. The dynamic pricing algorithm was then applied to predict the optimal listing prices for Airbnb properties, aiming to maximize revenue while responding to fluctuations in demand. This methodology ensures that pricing decisions are data-driven and responsive to market dynamics, enabling hosts to optimize their pricing strategy in real-time.

The linear function is defined as:

$$\text{new_price} = \text{current_price} + \alpha \times \text{demand_norm}$$

Where:

- new_price is the adjusted price.
- current_price is the current price of the listing.
- demand_norm is the normalized demand for the listing.
- α is the slope of the linear function, representing the rate of change of price with respect to demand.



Figure 10. Demand vs. Original Price

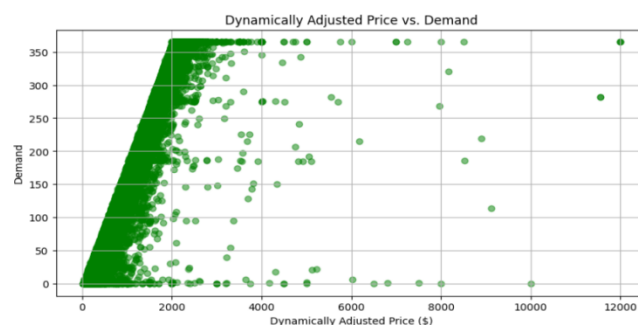


Figure 11. Dynamically adjusted Price vs. Demand

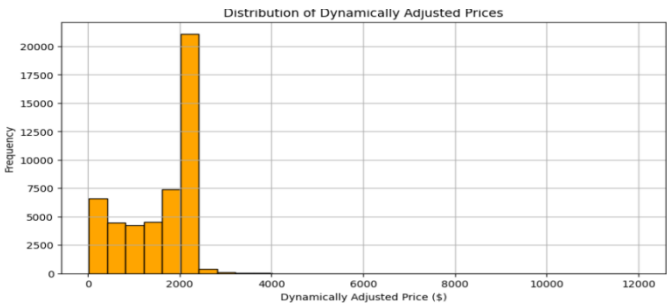


Figure 12. Dynamically adjusted Prices

The performance of the dynamic pricing strategy was evaluated alongside several regression algorithms, including linear regression, decision tree regression, random forest regression, gradient boosting regression, and support vector regression. Evaluation metrics such as RMSE, MAE, and R-squared were used to assess the predictive accuracy and goodness-of-fit of each model. The results revealed that the dynamic pricing strategy yielded a RMSE of 1569.46, MAE of 1396.57, and R-squared of -54.68, indicating suboptimal performance compared to other regression algorithms. In contrast, the gradient boosting regression model demonstrated superior performance with a RMSE of 3.65, MAE of 0.86, and R-squared of 0.999, suggesting its effectiveness in predicting optimal listing prices for Airbnb properties. These findings highlight the importance of model selection and algorithmic sophistication in achieving accurate and reliable dynamic pricing strategies for maximizing revenue in the hospitality industry.

Table 1. Comparison of metrics

Algorithm	RMSE	MAE	R-squared
Linear Regression	5.43e-14	3.67e-14	1.0
Dyanamic pricing strategy	1569.46	1396.57	-54.68
Decision Tree Regression	5.62	0.11	0.999
Random Forest Regression	4.51	0.08	0.999
Gradient Boosting Regression	3.65	0.86	0.999
Support Vector Regression	200.44	71.77	0.092

CONCLUSION

The study delved into exploring pricing optimization strategies for Airbnb rental listings through the utilization of various algorithms. While linear regression demonstrated a near-perfect fit to the data, its potential application in real-world scenarios may be limited due to its simplistic nature. Despite the conceptual appeal of dynamic pricing strategy, its predictive performance was lackluster, indicating the necessity for further refinement. Machine learning algorithms, including decision tree regression, random forest regression, and gradient boosting regression, showcased promising outcomes with minimal errors and high R-squared values. However, support vector regression, while providing reasonable predictions, exhibited higher errors compared to its counterparts. The selection of an appropriate algorithm hinges on factors such as dataset characteristics, computational resources, interpretability requirements, and the specific objectives of the pricing strategy. Future research endeavours could concentrate on fine-tuning dynamic pricing models, incorporating additional features, and experimenting with

alternative algorithms to bolster predictive accuracy and augment decision-making within the Airbnb rental market.

REFERENCES

- [1] Itrianingsih, D. A. Rahayu and F. R. Zazila, "Dynamic Pricing Analytic of Airbnb Amsterdam Using K-Means Clustering," 2022 Seventh International Conference on Informatics and Computing (ICIC), Denpasar, Bali, Indonesia, 2022, pp. 01-07, doi: 10.1109/ICIC56845.2022.10006966.
- [2] Oskam, J., van der Rest, JP. & Telkamp, B. What's mine is yours—but at what price? Dynamic pricing behavior as an indicator of Airbnb host professionalization. *J Revenue Pricing Manag* 17, 311–328 (2018). <https://doi.org/10.1057/s41272-018-00157-3>.
- [3] A. Lektorov, E. Abdelfattah and S. Joshi, "Airbnb Rental Price Prediction Using Machine Learning Models," 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp.0339-0344,doi: 10.1109/CCWC5344.2023.1009.
- [4] V. Toader, O. R. Bode, and R. V. Rus, "Pricing Strategies of Porto's Airbnb New Listings," in Proceedings of the 9th European Conference on Social Media, Cluj-Napoca, Romania, 2020.
- [5] Magno, F., Cassia, F. and Ugolini, M.M. (2018), "Accommodation prices on Airbnb: effects of host experience and market demand", *The TQM Journal*, Vol. 30 No. 5, pp. 608-620. <https://doi.org/10.1108/TQM-12-2017-0164>.
- [6] I. Singh, "Dynamic Pricing using Reinforcement Learning in Hospitality Industry," 2022 IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 2022, pp. 1-6, doi: 10.1109/IBSSC56953.2022.10037523.
- [7] C. H. Tan et al., "Optimizing Airbnb Search Journey with Multi-task Learning," in Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), Aug. 6–10, 2023, Long Beach, CA, USA, and New York, NY, USA: ACM, 2023, pp. 10. [Online]. Available: <https://doi.org/10.1145/3580305.3599881>
- [8] P. Ye et al., "Customized Regression Model for Airbnb Dynamic Pricing," in KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Aug. 19–23, 2018, London, United Kingdom, New York, NY, USA: ACM, 2018, pp. 9. [Online]. Available: <https://doi.org/10.1145/3219819.3219830>.
- [9] S. M. Varma, F. Castro, and S. T. Maguluri, "Dynamic Pricing and Matching for Two-Sided Markets with Strategic Servers," in Abstract Proceedings of the 2021 ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '21), New York, NY, USA: Association for Computing Machinery, 2021, pp. 61–62. [Online]. Available: <https://doi.org/10.1145/3410220.3456272>
- [10] V. Leoni and W. Nilsson, "Dynamic pricing and revenues of Airbnb listings: Estimating heterogeneous causal effects," *Int. J. Hosp. Manag.*, vol. 95, p. 102914, 2021. [Online]. Available: <https://doi.org/10.1016/j.ijhm.2021.102914>