

The role of seasonal trends in shaping tourist preferences for luxury resort: Big data approach

Luh Made Gunapria Hindu Rajeswari Pamungkas^{1*}, Putu Diah Sastri Pitanatri², Clearesta Adinda³

Bali Tourism Polytechnic, Bali, Indonesia^{1,2,3}

gunapriaa25@gmail.com¹, diahsastri@ppb.ac.id², clearestaadinda@ppb.ac.id³



Article History

Received on 20 May 2025

1st Revision on 9 June 2025

2nd Revision on 1 July 2025

Accepted on 24 July 2025

Abstract

Purpose: This study aims to examine seasonal patterns in tourist preferences for luxury resort stays in Bali, with a focus on how cultural backgrounds influence accommodation choices. The goal is to help resorts better understand guest behavior and optimize occupancy strategies.

Methodology/approach: The research analyzes monthly online review data from Tripadvisor for Bvlgari Resort Bali, a prominent luxury hotel. A time-series analysis using the ARIMA (Autoregressive Integrated Moving Average) model is applied to forecast occupancy trends. Prior to modeling, the data is tested for stationarity. In addition to forecasting, the study explores guest preferences by analyzing cultural characteristics inferred from reviews, categorizing them into collectivist and individualist orientations.

Results/findings: Findings reveal that occupancy trends do not strictly align with the hotel's predefined seasonal categories. Instead, they are shaped by global travel trends and cultural factors. Guests from collectivist cultures tend to prefer facilities that support group interaction and shared experiences, while those from individualist cultures prioritize privacy, exclusivity, and personalized services. The ARIMA model delivers accurate forecasting results, helping to predict future occupancy rates effectively.

Conclusion: IoT integration enhances the reliability of hospital-based PV systems. Tourist behavior is not solely dictated by conventional seasons but also by cultural expectations and travel motivations. Leveraging these insights allows hotels to better align operations, marketing, and pricing strategies with actual guest preferences.

Limitations: The study is limited to a single resort and uses data from one online review platform, which may not fully capture the diversity of all guests.

Contribution: This study contributes to tourism analytics, cross-cultural marketing, and hotel management by offering data-driven strategies to enhance occupancy performance.

Keywords: *ARIMA, Guest Preference, Occupancy, Seasonality, Time-Series*

How to Cite: Pamungkas, L. M. G. H. R., Pitanatri, P. D. S., Adinda, C. (2025). The role of seasonal trends in shaping tourist preferences for luxury resort: Big data approach. *Journal of Sustainable Tourism and Entrepreneurship*, 7(1), 107-123.

1. Introduction

Seasonality is a primary challenge in the tourism sector, significantly affecting stability and sustainability. This phenomenon leads to a concentration of tourist arrivals during specific periods, thereby impacting the spatial and temporal distribution of tourists throughout the year (Doran and Schofield, 2023). In the tourism industry, there are three types of seasonal patterns: off-peak, single-peak, and two-peak (Butler and Mao in Zvaigzne et al., and Dembovska (2022). Consequently, the

industry must adapt its facility and infrastructure management strategies to make them more efficient (Corluka, 2019).

Accommodation, particularly hotels, plays a crucial role in supporting tourism infrastructure and services, especially in Bali, by catering to tourists' needs (Pujiningrum & Sulistijanti, 2024). However, the hospitality sector also faces seasonal challenges, as room demand fluctuates due to trends, cycles, and random business movements. The primary patterns observed are peak seasons (high demand) and low seasons (off-seasons) (Maulana & Koesfardani, 2021).

The Bvlgari Resort Bali, a luxury hotel in Uluwatu, implements a strategy based on single-peak seasonality, dividing the year into low, shoulder, high, and peak seasons, as shown in Table 1. Variations in seasonal patterns significantly affect guests' length of stay, with occupancy rates often reflecting fluctuations in demand that influence revenue, transportation requirements, and employment in the tourism industry (Mitra, 2020).

Table 1. Categorization of seasonal patterns in Bvlgari Resort Bali

Seasonal Patterns	Month
Low Season	January, February, May, June
Shoulder Season	April, October, September, November
High Season	July, August
Peak Season	December

Source: Author's research (2025)

Nevertheless, Figure 1 reveals a significant discrepancy between the established seasonal categorization and actual average monthly occupancy trends. For instance, June is classified as a low season; however, occupancy levels show an increasing trend. Occupancy data for July and August peak during the high season, aligning with Central Statistics Agency data (2024), which reported a sharp increase in tourist visits in July to 625,665, representing a 20.11% rise. Conversely, October and November, categorized as the shoulder season, experience a steep decline in occupancy, approaching levels typical of the low season.

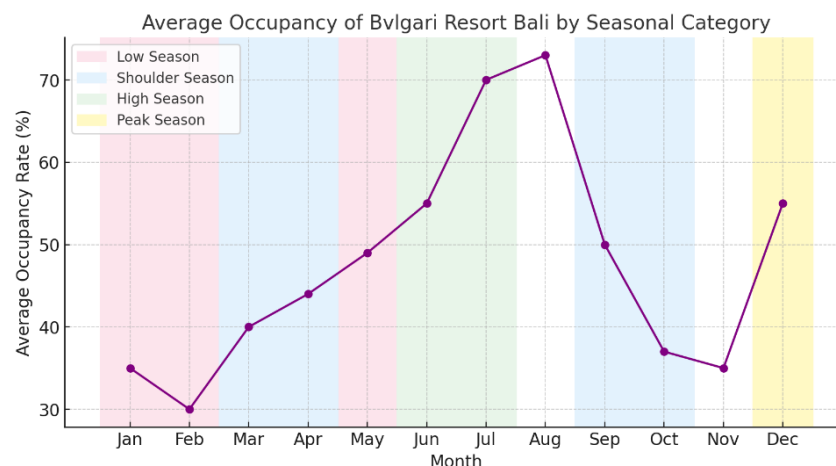


Figure 1. Monthly average occupancy at Bvlgari Resort Bali period 2022 – 2024 according to seasonal category
Source: Processed data (2025)

Therefore, understanding and managing seasonal patterns is essential for enhancing hotel efficiency and promoting tourism sustainability, particularly when serving a diverse range of guest preferences. Accordingly, scientific data analysis methods and mathematical models can be employed to identify trend patterns, seasonality, and factors influencing accommodation demand (Gudiato et al.). al. (Pujiningrum & Sulistijanti, 2024). These fluctuations stem from a complex interplay of seasonal patterns, cultural dynamics, and the preferences of tourists from various countries (Singgalen, 2025).

There is a significant correlation between cultural background and variations in tourist destinations, types of activities, and places of origin (Velea et al., 2022), which extends to accommodation expectations, as evidenced by an in-depth analysis of guest feedback from diverse geographical regions (Apaza-Panca, Ramos, Ramos, Saico, & Apaza-Apaza, 2024). Cultural preferences manifest in distinct rating patterns: guests from collectivist societies tend to emphasize the importance of community spaces and group-oriented amenities, whereas guests from individualist cultures prioritize privacy and personalized experiences (Shen in Singgalen (2025)). Therefore, a hotel's market-sensing capabilities, specifically its ability to collect and utilize guest information, play a critical role in anticipating shifts in guest preferences.

TripAdvisor, the leading online accommodation review platform, offers benefits to tourists and presents opportunities for hotels (Ladhari & Michaud in Yilmaz, 2020). Baleiro (2023) confirmed the importance of online reviews in evaluating guest satisfaction; thus, text mining of these comments provides valuable insights into guest experiences. Although most previous studies have relied on static data analyses, such as topic modeling or sentiment analysis (Jain, Pamula, & Srivastava, 2021), longitudinal data can reveal the dynamics of accommodation feature importance across different seasons, long-term trends, or disruptive events (Teichert, González-Martel, Hernández, & Schweiggart, 2024). Time-series analysis is a valid tool for measuring secular trends, seasonal variations, and irregular changes (Burger in (Teichert et al., 2024)). The ARIMA (Autoregressive Integrated Moving Average) model has been used to project occupancy in various hotel classes in Bali (Indradewi & Parwita, 2022; Pramudita, 2020)

Several studies employing time-series analytics and the ARIMA method indicate that such data can effectively anticipate and predict occupancy trends Putri, Prabudhi, Putranto, Astasia, and Tjahjo (2024). Furthermore, the ARIMA model has been successfully applied to forecast occupancy using big data sources, such as Google Trends (Ayuningtyas & Wirawati, 2020; Rizalde, Mulyani, & Bachtiar, 2021). However, few studies have focused on analyzing TripAdvisor guest review data to identify seasonal patterns in guest preference.

This research gap primarily concerns the application of time-series analysis with the ARIMA method, combined with web scraping and Application Programming Interface (API) engineering, which offers a powerful opportunity to deepen the understanding of guest preferences through a detailed analysis of spatially and temporally authentic guest review data, specifically at the Bvlgari Resort Bali.

2. Literature Review

2.1. Seasonal

Yabancı (2023), defines seasonality as fluctuations in tourism demand occurring within specific periods each year. Similarly, Ekananda in Kristiyanti and Sumarno (2020) explained that seasonal patterns may be quarterly, semi-annual, or annual. The primary causes of seasonality are generally classified into two categories: natural seasonality, which relates to environmental and climatic variations, and institutionalized seasonality, which stems from social and policy-driven events (Butler in Yabancı (2023)). Furthermore, Butler and Mao, as cited in Zvaigzne et al. (2022), identified three main seasonal patterns in tourism: off-peak, single-peak, and two-peak.

Seasonal patterns profoundly influence tourism businesses by affecting staffing levels, inventory management, and promotional activities. Therefore, understanding these cycles is essential for optimizing revenue during peak periods and minimizing losses during the low season. Moreover, it facilitates better resource allocation and workforce planning and ensures consistent service quality throughout the year.

2.2. Guest Preferences

Venkatraman, cited in Kuncoro and Kusumawati (2021), posited that guest preferences for products or brands arise from a combination of product attributes (such as price and durability) and consumer characteristics (including goals, attitudes, and disposable income). Singgalen (2025) emphasized that analyzing the relationship between culture and guest preferences through global review data can reveal

distinctive patterns in accommodation choices, service expectations, and satisfaction levels among guests from diverse cultural backgrounds.

Guest preferences are dynamic and continuously evolve in response to market trends, lifestyle changes and global influences. Contemporary travelers increasingly prioritize personalization, sustainability, and convenience. Consequently, hotels that understand and adapt to these preferences can tailor their services to enhance guest satisfaction, foster loyalty, and achieve competitive differentiation through innovation.

2.3. *TripAdvisor*

Launched in 2000 by Stephen Kaufer, Tripadvisor serves as a digital guidebook that simplifies destination exploration for tourists (Bassett, 2024). Additionally, the analysis of scores and reviews enables hotels to better understand guest preferences, improve operations, and maximize occupancy rates. This research utilizes TripAdvisor's "Time of Year" feature to help hotels adjust pricing and service strategies according to visiting seasons, while the "Traveler Type" filter groups reviews by traveller categories, allowing hotels to customize services based on market segmentation.

As a trusted travel review platform, TripAdvisor functions both as a marketing channel and a feedback mechanism for hospitality providers. High ratings can increase bookings by building guest trust, while reviews promote transparency, enabling prospective guests to assess service quality and encouraging providers to continuously improve.

2.4. *Time-Series*

Hanke, as cited in Sorlury, Mongi, and Nainggolan (2022), defined time-series data as observations recorded sequentially over time. According to Athiyarath, Paul, and Krishnaswamy (2020), time-series data comprise systematic components—level, trend, and seasonal—and non-systematic components, or noise. Time-series patterns are generally categorized into four types: horizontal (random fluctuations), trend (long-term movements), seasonal (recurring variations within a year), and cyclical (long-term fluctuations influenced by economic and external factors).

Time-series analysis supports forecasting and strategic planning in hospitality by identifying historical trends that predict future demand. This approach aids decision-making related to pricing, staffing, and inventory management. Furthermore, visualizing time-series data helps stakeholders to monitor performance and prepare for potential disruptions, thereby enhancing operational readiness and competitiveness.

2.5. *ARIMA (Autoregressive Integrated Moving Average)*

Murbarani (2020) explains that ARIMA is a forecasting method that combines autoregressive (AR) and moving average (MA) components. It uses the past and present values of the dependent variable to generate short-term predictions without incorporating independent variables. ARIMA models include several variants, such as AR, MA, ARMA, and seasonal ARIMA (SARIMA). ARIMA is particularly effective for analyzing time-series data that exhibit trends and seasonality. Its integration component addresses non-stationary data, making it well-suited for real-world applications, such as tourism forecasting. In the hospitality sector, ARIMA facilitates predictions of occupancy rates, revenue, and guest flow, thereby supporting strategic adjustments and efficient resource utilization in the hospitality sector.

2.6. *Occupancy*

Ayuningtyas and Wirawati (2020) defined occupancy as the ratio of occupied to available rooms, expressed as a percentage over a given period. Moreover, the occupancy rate serves as a key performance indicator for hotels (Sugiarto in Jatmiko and Sandy, 2020), reflecting operational efficiency, marketing effectiveness, and pricing strategies.

Analyzing occupancy rates enables hotels to benchmark their performance against competitors and market standards. High occupancy indicates strong demand and effective operations, whereas low

occupancy may indicate service issues or misalignment with market needs. Continuous monitoring of occupancy also supports dynamic pricing strategies aimed at maximizing profits, while maintaining guest satisfaction.

3. Research Methodology

3.1. Design & Strategy

The research design was systematically developed to identify the key elements necessary for analyzing seasonal patterns in tourist preferences (Saunders, Lewis, & Thornhill, 2023). This study adopts a quantitative, data-driven approach that emphasizes objectivity and leverages numerical data to identify and predict behavioral trends. A quantitative design is particularly suitable when the aim is to measure variables and uncover patterns using statistical or computational methods (Creswell and Creswell, 2017). Unlike qualitative research, quantitative research prioritizes outcomes and the generalizability of results by utilizing standardized instruments to ensure replicability.

Importantly, this study addresses a research gap in the application of big data to analyze seasonal tourist preferences by utilizing the ARIMA time-series model on guest reviews from TripAdvisor. While previous studies have employed ARIMA and tools such as Google Trends to predict hotel occupancy (Ayuningtyas & Wirawati, 2020; Putri et al., 2024; Rizalde et al., 2021), few have focused on user-generated reviews to uncover these temporal patterns. To fill this gap, this study uses web scraping via the TripAdvisor API to collect monthly review data for the Bvlgari Resort Bali and applies ARIMA to identify seasonal trends. Ultimately, the goal is to provide data-driven insights that support more effective pricing, marketing, and resource planning in the luxury hospitality industry.

3.2. Data Collection & Preparation

The primary dataset comprises online guest reviews sourced from TripAdvisor covering the period from January 2022 to December 2024. Data were collected using a web scraping process that leveraged TripAdvisor's publicly accessible Application Programming Interface (API). This method enables the automated extraction of large-scale textual content. Following data extraction, a custom scripting process was employed to transform and structure the raw data, which were then stored in a database and exported into Excel format for analysis (Sirisuriya, 2023). This approach ensured a consistent and analyzable dataset that was suitable for time-series modeling.

To ensure the reliability of the analysis, the raw dataset underwent rigorous data cleaning. This included addressing missing values, standardizing date formats, and filtering out irrelevant content from the data. Python, a high-level, open-source programming language, was used for these tasks because of its robust data manipulation libraries and flexibility in handling unstructured data. The versatility of Python makes it an ideal tool for preparing large datasets for time-series analysis (Nongthombam & Sharma, 2021).

3.3. Data Analysis & Theoretical Framework

The cleaned dataset was subjected to time-series analysis to identify patterns, such as trends, seasonal fluctuations, and irregular events. The ARIMA model was applied because of its ability to manage non-stationary time-series data by removing trend and seasonal components through differencing. The model then isolates the signal from the noise, producing accurate forecasts of future occupancy rates. These insights provide practical value in guiding pricing strategies, marketing efforts, and resource allocation in luxury hospitality operations (Kontopoulou, Panagopoulos, Kakkos, & Matsopoulos, 2023).

Building upon the time-series forecasting process, the conceptual framework in this study integrates empirical data and theoretical insights to explain the underlying factors influencing occupancy patterns at the Bvlgari Resort Bali. By grounding the research in a data-driven model while incorporating established theories of consumer behavior and tourism demand, the framework supports a systematic analysis of observed patterns. In this quantitative approach, the framework not only guides hypothesis testing but also informs the interpretation of ARIMA-generated forecasts, linking statistical outputs to actionable insights for decision-making in luxury resort management (Syahputri et al., 2023).

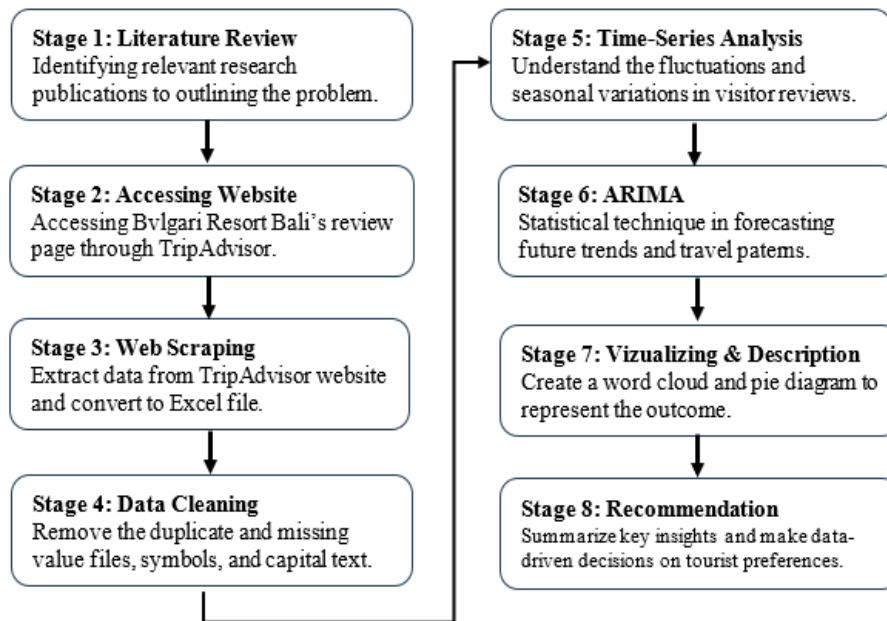


Figure 2. Theoretical framework
Source: Adapted from (Bawana, Mansor, & Noordin, 2024)

3.3.1 Model of the study

3.3.1.1 ARIMA

The ARIMA model is defined by three parameters as follows:

- p (autoregressive order),
- d (degree of differencing),
- q (moving average order).

Once these components are identified, the model is estimated using the historical data. The general form of the ARIMA(p, d, q) model used in this study is expressed as follows:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

This formulation consists of the following:

- Autoregressive (AR) terms: capturing the influence of past values of the series,
- Moving Average (MA) terms: capturing the impact of past error terms,
- A constant term c ,
- The current error term ε_t .

The model summary provides the estimated coefficients for these components along with their statistical significance, which is typically evaluated using p-values and confidence intervals. A well-fitting model is expected to show statistically significant parameters and low residual errors. Residual diagnostics were performed to validate the adequacy of the model. The residuals, defined as the difference between the actual and predicted values, were computed as follows:

$$e_t = Y_t - \hat{Y}_t$$

3.3.1.2 Augmented Dickey-Fuller (ADF)

Before applying time-series forecasting models, such as ARIMA, it is essential to determine whether the series is stationary. Stationarity implies that the statistical properties of a series, such as the mean and variance, remain constant over time. This assumption is crucial for the validity of several forecasting techniques. The Augmented Dickey-Fuller (ADF) test was employed to assess stationarity. The ADF test builds upon the basic Dickey-Fuller test by incorporating lagged differences of the

dependent variable to account for serial correlation in the error terms, thereby improving its robustness. The regression model used in the ADF test is expressed as

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \epsilon_t$$

Where:

- ΔX_t is the first difference of the series,
- t is the time trend,
- γ represents the coefficient for the lagged level of the series,
- δ_i are the coefficients for the lagged differences, and
- ϵ_t is the error term.

3.3.1.2 Autocorrelation Function (ACF)

The autocorrelation Function (ACF) is a statistical tool in time-series analysis that measures the correlation between a time series and its past values at different time lags. It helps determine the extent to which current observations in the series are influenced by historical observations, providing insight into underlying patterns such as seasonality, trends, and cyclic behavior. Mathematically, the ACF at lag k , denoted by ρ_k , is defined as follows:

$$\rho_k = \frac{E[(Y_t - \mu)(Y_{t-k} - \mu)]}{\sigma^2}$$

Where:

- Y_t the value of the time series at time t ,
- μ is the mean of the series,
- σ^2 is the variance, and
- E denotes the expected value or mean of a product.

3.3.1.3 Autocorrelation Function

The autocorrelation Function (ACF) measures the total correlation between a time series and its lagged values, and the Partial Autocorrelation Function (PACF) provides a more refined perspective. The PACF isolates the direct correlation between an observation and its lagged counterpart after accounting for the effects of all shorter lags. This function is especially important for identifying the presence and order of autoregressive (AR) components in a time-series model. Mathematically, the PACF at lag k , denoted ϕ_{kk} , is defined as:

$$\phi_{kk} = \text{Corr}(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})$$

Where:

- Y_t the value of the time series at time t ,
- μ is the mean of the series,
- σ^2 is the variance, and
- E denotes the expected value or mean of a product.

3.3.1.4 ARIMA model residual plot and residual density plot

The density distribution of the residuals was plotted to test whether the residuals followed a normal distribution, which is a key assumption of the ARIMA models. The normal distribution of the residuals was evaluated using the probability density function:

$$f(e) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(e-\mu)^2}{2\sigma^2}}$$

Where:

- μ is the mean (ideally 0),
- σ^2 is the variance of the residuals,
- $f(e)$ is the probability of observing specific residual values.

3.3.1.5 Linear regression

The analysis of the relationship between the monthly average occupancy and the number of online reviews for the Bvlgari Resort Bali, as published on TripAdvisor, revealed a strong positive linear correlation. This relationship was modeled using a simple linear regression equation of the following form:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Where:

- Y represents **monthly occupancy rate (%)**,
- X denotes the **number of online reviews**,
- $\beta_0 = 23.17$ is the **intercept** of the regression line,
- $\beta_1 = 1.08$ is the **regression coefficient (slope)**,
- ϵ is the **error term** that captures the unexplained variation.

4. Results and Discussion

4.1. Results

4.1.1. Web scraping API (Artificial Processing Intelligence)

Web scraping refers to the automated retrieval of data using web scraper software that systematically loads and extracts information from websites based on user-defined requirements (Sirisuriya, 2023). Furthermore, APIs enable software systems to interact and build applications that leverage the data or functionalities of other systems (Raatikainen et al., 2021). The initial approach in this study involved accessing the TripAdvisor website to extract online reviews from guests who stayed at the Bvlgari Resort Bali between 2022 and 2024. The page URL was copied and entered into the API software, which then processed the data through serverless program configuration in the cloud, requiring no additional manual input. The automated extraction process rendered the review pages and parsed the HTML (HyperText Markup Language), providing an interface for software services available over the network. This process facilitated the publication of relevant open datasets in Excel format, which are easy to analyze.

Figure 3. Dataset online reviews of Bvlgari Resort Bali on Tripadvisor

Source: Processed data by API (2025)

4.1.2. Data Cleaning

According to Nongthombam and Sharma (2021), data cleaning involves scanning and removing inconsistent or duplicate entries. Upon examining the extracted Excel data using Python, several columns related to TripAdvisor reviews were identified, including user information, ratings, and review text. The cleaned data frame retained the columns 'text', 'title,' and 'user,' which are essential for the analysis. However, the 'user' column was recorded as 'user/name', necessitating an adjustment to the list of key columns to reflect the actual structure of the cleaned data frame. The next step involved deleting any rows with missing values in the 'user/name', 'text', or 'title' columns before finalizing the data-cleaning process. After these adjustments, data cleansing was successfully completed, resulting in a final dataset comprising 71 rows and 44 columns.

4.1.3. Time Series Analysis

A fundamental concept in time-series analysis is that current observations X_t are influenced by one or more lagged observations, such as X_{t-1} , X_{t-2} , and so forth. This temporal dependence underpins many forecasting models, including autoregressive models that leverage historical patterns to predict future values.

For this study, the dataset included a column labeled “publishedDate”, which recorded the date each online review was posted. This column is critical for time-series analysis because it serves as a temporal index anchoring each observation to a specific point in time. Before the analysis, the ‘publishedDate’ column was converted to a datetime format to ensure compatibility with Python’s time-series functions, such as resampling and indexing. Once converted, the dataset was transformed into a deep-index data frame to enable more sophisticated time-based operations.

Subsequently, the data were aggregated on a monthly basis, producing a new time series that represents the number of online reviews per month. Monthly aggregation is particularly effective for identifying broad patterns such as seasonality, cyclical trends, or long-term growth or decline. This level of temporal granularity strikes a balance between detail and the reduction of noise from the daily fluctuations. The resulting monthly review counts were visualized to provide a clearer understanding of the trends over the study period.

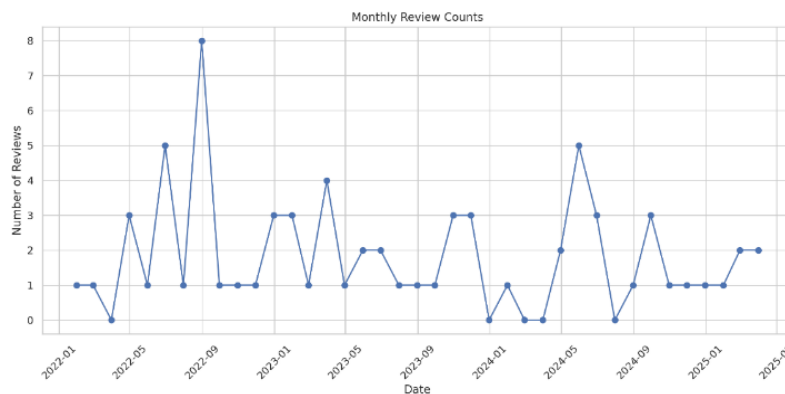


Figure 4. Fluctuations in the number of online reviews from 2022 – 2025

Source: Processed data by Python (2025)

In this study, the Augmented Dickey-Fuller (ADF) test produced an ADF statistic of approximately -4.05, with an associated p-value of 0.0012. Since the p-value is less than the commonly used significance level of 0.05, the null hypothesis that the series contains a unit root is rejected. This result confirms that the time series data are stationary. Given the stationarity of the data (i.e., the order of differencing $d=0$), the time series is suitable for modeling using the Autoregressive Integrated Moving Average (ARIMA) approach. The absence of a unit root simplifies the modeling process, as no further differencing is required, and ARIMA can proceed directly using the original (or first-differenced, if necessary) data.

4.1.4. ARIMA

4.1.4.1 Autocorrelation Function (AF)

The Autocorrelation Function (ACF) equation captures the normalized covariance between values that are kk periods apart, providing a standardized measure of the strength and direction of the correlation for each lag kk . The ACF values range from -1 to 1, where values closer to 1 or -1 indicate strong positive or negative correlation, respectively, and values near zero suggest little to no correlation. In the context of a Moving Average process of order qq , or $MA(qq)$, the ACF is vital for model identification. Theoretical properties of an $MA(qq)$ process suggest that the ACF will exhibit non-zero autocorrelations only up to lag qq , after which it will drop to zero. This characteristic allows researchers to estimate the order qq by observing the point at which autocorrelation becomes insignificant.

However, as shown in Figure 5, the empirical ACF plot of the time-series data reveals a gradual decline in the autocorrelation values across successive lags. Rather than dropping off sharply to zero after a specific lag (as expected in a simple MA process), the ACF values tail off slowly. This behavior indicates that the time series exhibits memory or influence extending across multiple lags, which is characteristic of a model with an MA component where $q > 0$.

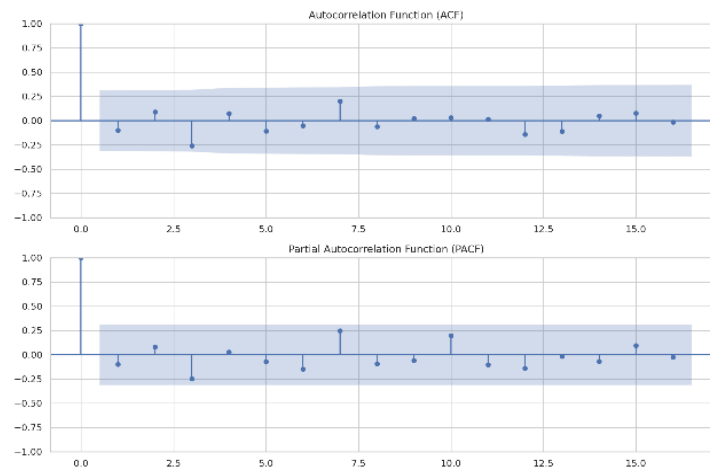


Figure 5. ACF plot
Source: Processed data by Phyton (2025)

The slow decay of the ACF values indicates persistent stochastic dependencies, suggesting that a simple low-order MA model may be inadequate and that multiple MA terms may be necessary. This pattern also hints at a potential mixed ARMA or ARIMA process, requiring both autoregressive and moving average components to capture the dynamics of the data. To confirm and refine the model structure, further diagnostics, such as the Partial Autocorrelation Function (PACF) and model fitting criteria, are essential.

4.1.4.2 Partial Autocorrelation Function (PACF)

This expression measures the correlation between Y_t , and Y_{t-k} , conditional on all intermediate lags from Y_{t-1} , through Y_{t-k+1} . By removing indirect influences, the PACF provides a clearer view of the direct relationship between current and specific lagged values, revealing the structure of the autoregressive (AR) process.

For a pure autoregressive model of order pp ($AR(p)$), the PACF will display significant partial correlations up to lag pp and will cut off to zero afterward. This behavior is a defining characteristic of AR processes and is essential for determining the appropriate value of pp in time-series modeling.

Figure 6 shows that the PACF plot for the dataset reveals a sharp cutoff after the first lag, indicating that only the first lag exhibits a statistically significant partial autocorrelation, whereas subsequent lags show little to no significant influence. This pattern strongly suggests the presence of an $AR(1)$ process, in which the current value of the time series is primarily influenced by its immediately preceding value.

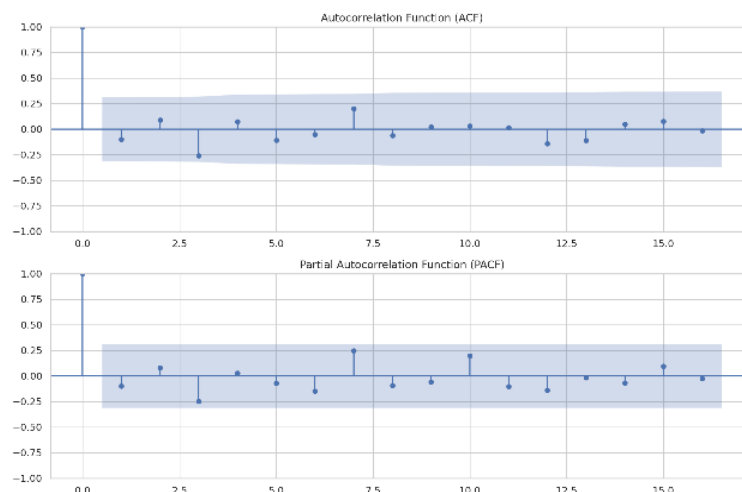


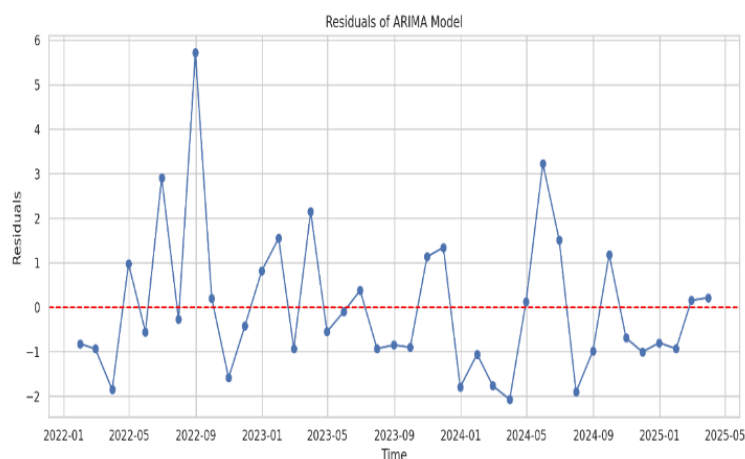
Figure 6. PACF plot
Source: Processed data by Phyton (2025)

The cutoff observed in the PACF supports the inclusion of an autoregressive component with order $p = 1$ in the time-series model. When used alongside ACF analysis, which indicated the presence of a moving average component ($q > 0$), the PACF findings suggest that the appropriate model structure could be an ARIMA(p,d,q) model with $p = 1, d = 0$ (as established by the ADF test confirming stationarity), and a suitable value for q based on the ACF.

4.1.4.3 Residual Plot and Density Plot of Residuals

A residual plot was used to assess whether the residuals behaved like white noise, that is, fluctuating randomly around zero, with no apparent patterns or autocorrelation. In this study, the residuals were mostly centered around zero, indicating that the ARIMA model effectively captured the general trend in the data. However, the presence of several sharp spikes suggests potential outliers or latent patterns that may not have been fully addressed by this model.

As shown in Figure 7, the residual distribution was not perfectly symmetrical, indicating a degree of skewness. Although this deviation was not severe, it suggests that the model could benefit from further refinement or transformation of the input data. Residual fluctuations around zero indicate that the model has captured the main pattern, but some extreme peaks may be indicative of outliers or patterns that have not yet been fully captured. Meanwhile, the residual distribution is not completely symmetrical (skewness), whereas it should ideally approximate a normal distribution (a symmetrical bell shape).



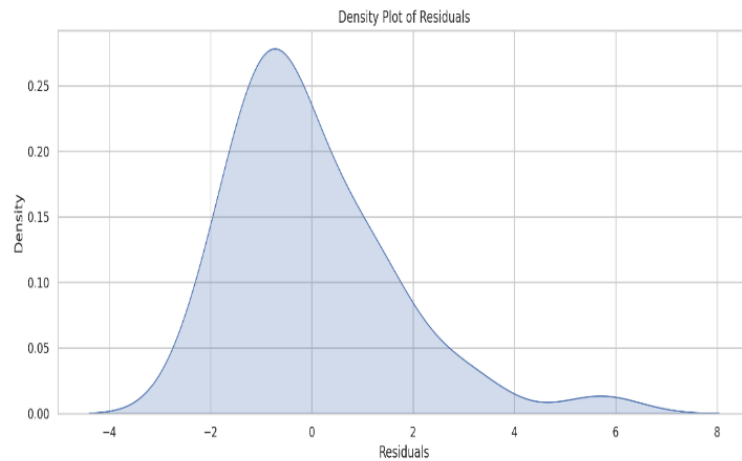


Figure 7. ARIMA model residual plot and residual density plot
Source: Processed data by Phyton (2025)

Once the ARIMA components are determined and validated, the model is used to generate the forecasts. Figure 8 presents a visualization of the forecast results, showing the predicted values alongside the historical data. The forecast is accompanied by confidence intervals, which provide a range of values within which future observations are expected to fall within. This visualization is critical for understanding the expected trends in online review counts and aids decision-makers in resource planning and strategic forecasting.

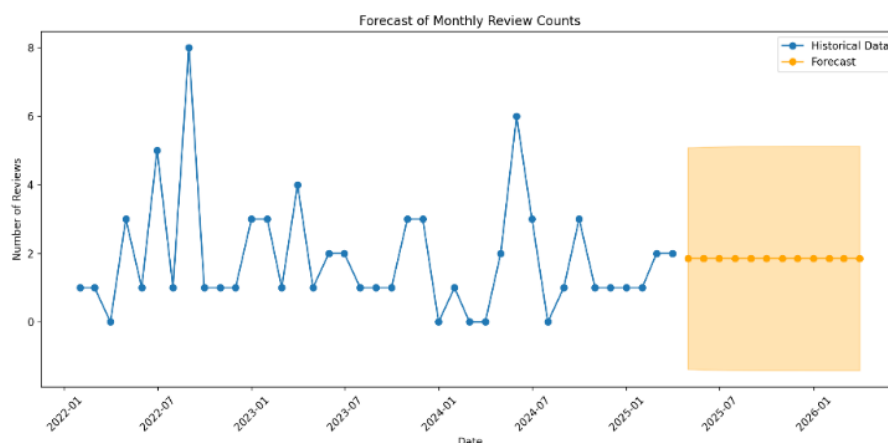


Figure 8. Plot *forecast* number of monthly online reviews of Bvlgari Resort Bali
Source: Processed data by Phyton (2025)

4.2. Discussion

Based on Figure 9, each additional online review is associated with an estimated 1.08% increase in the monthly occupancy rate. This demonstrates the measurable and economically meaningful impact of online engagement on hotel performance. The R-squared value of 94.5% indicates that a very high proportion of the variability in occupancy rates can be explained by the fluctuations in the number of reviews. In statistical terms, this high R-squared value suggests an excellent model fit and implies that the model captures the majority of the patterns in the data. Additionally, the p-value of the regression coefficient is below 0.05, confirming that the relationship between the number of online reviews and occupancy is significant.

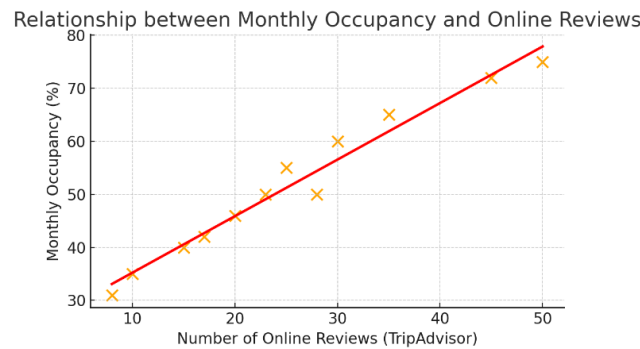


Figure 9. Scatter plot of monthly average occupancy and number of online reviews of Bvlgari Resort Bali on TripAdvisor
Sources: Processed data by Phytion (2025)

Figure 10 complements the regression model by illustrating the monthly trends of both occupancy rates and online reviews throughout the year. A clear seasonal pattern was observed. Occupancy and review volume both rise steadily from January, peaking in July and August, and then taper off sharply in September and October before recovering slightly in December. This synchronized pattern between occupancy and review counts supports the interpretation that online reviews reflect guest satisfaction and signal market demand. During peak tourism seasons, more guests stay at the resort and are likely to leave reviews, suggesting that the number of online reviews may serve as a leading or coincidental indicator of demand.

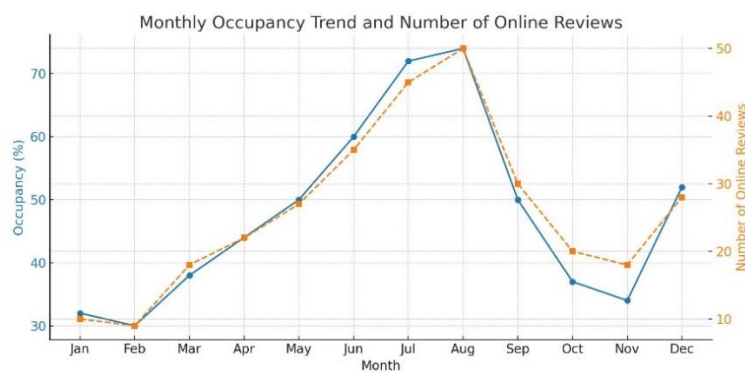


Figure 10. Time-series plot train occupancy monthly and number of online reviews of Bvlgari Resort Bali on TripAdvisor
Sources: Processed data by Phytion (2025)

The analysis of the Bvlgari Resort Bali's occupancy trends and guest feedback revealed a strong positive relationship between the number of online reviews and monthly occupancy rates, with a linear regression model showing that each additional review is associated with a 1.08% increase in occupancy. This model demonstrated high predictive accuracy, with an R-squared value of 94.5% and statistically significant results ($p\text{-value} < 0.05$).

However, further insights from guest reviews and cultural preferences suggest that relying solely on numerical review counts may overlook important behavioral factors that affect hotel performance. A word cloud generated from TripAdvisor reviews highlights key themes such as "experience," "hospitality," "luxury," "service," and "privacy," reflecting the emotional and cultural drivers of satisfaction.



Figure 11. *Word-cloud* Bvlgari Resort Bali guest reviews on TripAdvisor
Sources: Processed data by Phyton (2025)

Guests from collectivist cultures (55%) tended to value social interaction and communal experiences, whereas those from individualist cultures (45%) prioritized privacy and exclusivity, as shown in Figure 12. If these cultural distinctions are not incorporated into the predictive model, it may lead to inaccurate forecasts and residual patterns that violate the regression assumptions. Therefore, incorporating cultural preferences into the model could enhance its robustness and relevance. Strategically, the resort can benefit from segmenting guest experiences by offering private, serene services for individualists and group-oriented, culturally immersive activities for collectivists. Moreover, culturally tailored marketing and review solicitation strategies can further improve guest satisfaction and online engagement, ultimately strengthening resort occupancy performance and market positioning.

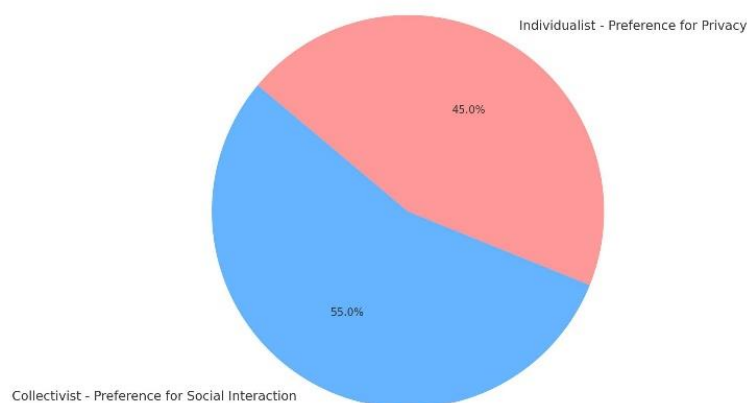


Figure 12. Bvlgari Resort Bali guest preference pie diagram on Tripadvisor
Sources: Processed data by Phyton (2025)

5. Conclusion

5.1. Conclusion

The results of the analysis demonstrate that seasonal patterns significantly affect the occupancy levels at the Bvlgari Resort Bali. Through the application of time-series analysis using the ARIMA model, we found that occupancy variations do not always align with the predefined seasonal categorizations set by the hotel. This suggests that additional elements, such as evolving tourist travel trends and changing guest preferences, shape occupancy patterns. Moreover, an analysis of online reviews on TripAdvisor revealed that guest preferences are also influenced by cultural backgrounds. Guests from collectivist cultures tend to prioritize community-related amenities and private experience. The ARIMA model proved effective in forecasting future occupancy trends, supported by graphs showing review trends that aligned with occupancy rates and reflected forecast uncertainty.

5.2. Limitation

This study offers valuable insights; however, several limitations should be acknowledged in this regard. Key variables that may influence occupancy levels, such as macroeconomic factors, political

conditions, and global travel disruptions, have not been fully considered. Furthermore, the heavy reliance on online reviews may introduce bias as not all guests share their feedback on digital platforms. Focusing on a single resort also limits the generalizability of the findings to broader hospitality contexts. Therefore, future research should incorporate a wider range of external factors and include multiple properties across different hotel categories. This approach will provide a more comprehensive perspective and enhance the relevance of the findings for the hospitality industry.

5.3. Suggestion

Based on the findings and limitations, it is recommended that the Bvlgari Resort Bali adopt a data-driven marketing strategy focusing on peak occupancy months. Dynamic pricing models should be employed to match demand fluctuations, and promotional events should be planned during the predicted low-occupancy periods. Additionally, tailoring offerings to suit cultural preferences and utilizing big data analytics in strategic decision-making will enhance the competitiveness and operational sustainability of the hospitality industry.

References

- Apaza-Panca, C., Ramos, K., Ramos, A., Saico, C., & Apaza-Apaza, S. (2024). Quality Of Service Of Accommodation Establishments As A Factor In Tourism Competitiveness. *International Journal of Innovative Research and Scientific Studies*, 7(3), 1128-1139. doi:<https://doi.org/10.53894/ijirss.v7i3.3100>
- Athiyarath, S., Paul, M., & Krishnaswamy, S. (2020). A Comparative Study And Analysis Of Time Series Forecasting Techniques. *SN Computer Science*, 1(3), 175. doi:<https://doi.org/10.1007/s42979-020-00180-5>
- Ayuningtyas, I., & Wirawati, I. (2020). Nowcasting Tingkat Penghunian Kamar Hotel Menggunakan Google Trends. *Seminar Nasional Official Statistics*, 2020(1), 338-343. doi:<https://doi.org/10.34123/semnasoffstat.v2020i1.636>
- Baleiro, R. (2023). Understanding Visitors' Experiences At Portuguese Literary Museums: An Analysis Of Tripadvisor Reviews. *European Journal of Tourism Research*, 33, 3305-3305. doi:<https://doi.org/10.54055/ejtr.v33i.2839>
- Bassett, K. (2024). TripAdvisor as a 'geo-pastoral technology'. *Tourism Geographies*, 26(4), 672-686. doi:<https://doi.org/10.1080/14616688.2023.2275734>
- Bawana, T. A., Mansor, F., & Noordin, K. (2024). Gauging Customer Sentiment Regarding Indonesian Islamic Digital Banks. *AL-IKTISAB: Journal of Islamic Economic Law*, 8(1), 101-118. doi:<https://doi.org/10.21111/aliktisab.v8i2.12838>
- Corluka, G. (2019). Tourism seasonality—an overview. *Journal of business paradigms*, 4(1), 21-43.
- Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, And Mixed Methods Approaches*: Sage publications.
- Doran, A., & Schofield, P. (2023). Tourism and the seasons. *Tourism: A temporal analysis*, 29-42. doi:<https://doi.org/10.23912/9781911635840-5446>
- Indradewi, I. G. A. A. D. M., Ni Putu Sri, & Parwita, W. G. S. (2022). Peramalan Tingkat Penghunian Kamar Berdasarkan Kelas Hotel Di Bali Menggunakan Metode ARIMA Forecasting Room Occupancy Rates Based on Hotel Class in Bali Using the ARIMA Method. *Journal of Computing Engineering, System and Science*, 7(2), 325-339. doi:<https://doi.org/10.24114/cess.v7i2.33959>
- Jain, P. K., Pamula, R., & Srivastava, G. (2021). A Systematic Literature Review On Machine Learning Applications For Consumer Sentiment Analysis Using Online Reviews. *Computer science review*, 41, 100413. doi:<https://doi.org/10.1016/j.cosrev.2021.100413>
- Jatmiko, H., & Sandy, S. R. O. (2020). Faktor–Faktor Yang Mempengaruhi Tingkat Hunian Kamar Pada Hotel Di Kota Jember. *Sadar Wisata: Jurnal Pariwisata*, 3(1), 32-40. doi:<https://doi.org/10.32528/sw.v3i1.3371>
- Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A Review Of Arima Vs. Machine Learning Approaches For Time Series Forecasting In Data Driven Networks. *Future Internet*, 15(8), 255. doi:<https://doi.org/10.3390/fi15080255>
- Kristiyanti, D. A., & Sumarno, Y. (2020). Penerapan Metode Multiplicative Decomposition (Seasonal) Untuk Peramalan Persediaan Barang Pada Pt. Agrinusa Jaya Santosa. *Jurnal SISKOM-KB*

- (Sistem Komputer dan Kecerdasan Buatan), 3(2), 45-51. doi:<https://doi.org/10.47970/siskom-kb.v3i2.145>
- Kuncoro, H., & Kusumawati, N. (2021). A Study Of Customer Preference, Customer Perceived Value, Sales Promotion, And Social Media Marketing Towards Purchase Decision Of Sleeping Product In Generation Z. *Adv. Int. J. Business, Entrep. SMEs*, 3(9), 265-276. doi:<https://doi.org/10.35631/aijbes.39018>
- Maulana, A., & Koesfardani, C. F. P. P. (2021). Pola Musiman Kunjungan Wisatawan Mancanegara Ke Bali. *Jurnal Kepariwisata Indonesia: Jurnal Penelitian Dan Pengembangan Kepariwisata Indonesia*, 14(2), 73-90. doi:<https://doi.org/10.47608/jki.v14i22020.73-90>
- Mitra, S. K. (2020). Estimating The Duration Of Different Seasons And Their Impact On Hotel Room Prices. *International Journal of Hospitality Management*, 90, 102604. doi:<https://doi.org/10.1016/j.ijhm.2020.102604>
- Murbarani, N. (2020). Pemodelan Tingkat Hunian Kamar (Occupancy) Hotel Grand Inna Malioboro Yogyakarta Berdasarkan Pendekatan Analisis Intervensi.
- Nongthombam, K., & Sharma, D. (2021). Data Analysis Using Python. *Int. J. Eng. Res. Technol*, 10(07), 463-468. doi:<https://doi.org/10.17577/IJERTV10IS070241>
- Pramudita, A. (2020). Memperkirakan Tingkat Penghuni Hotel Menggunakan Analisis Arima Dengan Aplikasi Minitab. *EDUSAINTEK*, 4.
- Pujiningrum, R. D., & Sulistijanti, W. (2024). Peramalan Tingkat Penghunian Kamar Hotel Bintang Lima Provinsi Bali Menggunakan Metode ARIMA dan Fuzzy Time Series Lee. *JURNAL SOSIAL EKONOMI DAN HUMANIORA*, 10(2), 158-166. doi:<https://doi.org/10.29303/jseh.v10i2.520>
- Putri, C. R., Prabudhi, S. A., Putranto, F. G. F., Astasia, A., & Tjahjo, T. W. (2024). Forecasting Room Occupancy Rates in Batu City: Implicationsfor Government Policy Using ARIMA Method. *PANGRIPTA*, 7(1), 37-48. doi:<https://doi.org/10.58411/gzftfs54>
- Raatikainen, M., Kettunen, E., Salonen, A., Komssi, M., Mikkonen, T., & Lehtonen, T. (2021). State Of The Practice In Application Programming Interfaces (Apis): A Case Study. *European Conference on Software Architecture*, 191-206. doi:https://doi.org/10.1007/978-3-030-86044-8_14
- Rizalde, F. A., Mulyani, S., & Bachtiar, N. (2021). Forecasting Hotel Occupancy Rate In Riau Province Using Arima And Arimax. *Proceedings of The International Conference on Data Science and Official Statistics*, 2021(1), 578-589. doi:<https://doi.org/10.34123/icdsos.v2021i1.199>
- Saunders, M. N. K., Lewis, P., & Thornhill, A. (2023). Research Methods for Business Students.
- Singgalen, Y. A. (2025). Tourist Preferences at Hotel and Resort Based on Review Data. *Journal of Business and Economics Research (JBE)*, 6(1), 179-190. doi:<https://doi.org/10.47065/jbe.v6i1.6844>
- Sirisuriya, S. D. S. (2023). Importance Of Web Scraping As A Data Source For Machine Learning Algorithms-Review. *2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS)*, 134-139. doi:<https://doi.org/10.1109/iciis58898.2023.10253502>
- Sorlury, F. N., Mongi, C. E., & Nainggolan, N. (2022). Penggunaan Model Autoregressive Integrated Moving Average (ARIMA) Untuk Meramalkan Nilai Tukar Petani (NTP) di Provinsi Sulawesi Utara. *d\Cartesian: Jurnal Matematika dan Aplikasi*, 11(1), 59-66. doi:<https://doi.org/10.35799/dc.11.1.2022.38744>
- Syahputri, A. Z., Della Fallenia, F., & Syafitri, R. (2023). Kerangka berfikir penelitian kuantitatif. *TARBIYAH: Journal of Educational Science and Teaching*, 2(1), 160-166. doi:<https://doi.org/10.1342/tarbiyah.v2i1.25>
- Teichert, T., González-Martel, C., Hernández, J. M., & Schweiggart, N. (2024). Dynamics In Accommodation Feature Preferences: Exploring The Use Of Time Series Analysis Of Online Reviews For Decomposing Temporal Effects. *International Journal of Contemporary Hospitality Management*, 36(7), 2521-2541. doi:<https://doi.org/10.1108/ijchm-03-2023-0279>
- Velea, L., Gallo, A., Bojariu, R., Irimescu, A., Craciunescu, V., & Puiu, S. (2022). Holiday Climate Index: Urban—Application For Urban And Rural Areas In Romania. *Atmosphere*, 13(9), 1519. doi:<https://doi.org/10.3390/atmos13091519>
- Yabancı, O. (2023). Managing Seasonality In Tourism. *International Journal of Geography and Geography Education*(50), 353-369. doi:<https://doi.org/10.32003/igge.1299610>

Zvaigzne, A., Litavniece, L., & Dembovska, I. (2022). Tourism Seasonality: The Causes And Effects. *Worldwide Hospitality and Tourism Themes*, 14(5), 421-430. doi:<https://doi.org/10.1108/whatt-07-2022-0080>