

DOI: https://doi.org/10.48009/1_iis_2026_111

Interpretable AI for hotel occupancy forecasting in Taiwan's tourist hotels: lessons learned from multiple periods of crisis and market Transition

Jenq-Foung Yao, *Georgia College & State University*, jf.yao@gcsu.edu

Yu-Hsiang (John) Huang, *Georgia College & State University*, john.huang@gcsu.edu

Daniel Wu, *Georgia College & State University*, Daniel.wu@gcsu.edu

Cheng-Ying Yang, *University of Taipei*, cyang@utapei.edu.tw

Abstract

The economic downturn (e.g., the COVID-19 pandemic) caused unprecedented disruption to the global hospitality industry, fundamentally reshaping hotel occupancy patterns and demand structures. This study examines how hotel occupancy dynamics in Taiwan evolved across the pre-COVID (2017–2019), COVID (2020–2022), and post-COVID recovery (2023–2024) periods using machine learning and explainable artificial intelligence (AI) techniques. We analyze shifts in the distribution of occupancy categories, changes in the relative importance of demand-side predictors—particularly domestic and international visitors—and the role of operational and employment-related features in supporting occupancy resilience. Using proprietary Taiwanese hotel data, this study classifies occupancy levels into three classes: low (<50%), medium (50–75%), and high (>75%). Based on these classifications, separate Gradient Boosting Classifier models are trained for each period, and SHAP values are then applied to quantify global and class-specific feature contributions. Results reveal a dramatic collapse of high occupancy during the economic downturn, a corresponding surge in low occupancy, and partial normalization during recovery. International visitors strongly predict occupancy before and after COVID-19 but lose influence during the economic downturn, when domestic visitors become the primary driver. Employment-related features, especially food and beverage and room staffing—grow in importance during recovery, indicating their role in signaling operational readiness and resilience. Model accuracy peaks in the pre-COVID period (86%) and declines thereafter, reflecting structural instability and evolving demand relationships. Overall, this study demonstrates the value of interpretable AI in capturing crisis-induced structural changes in hospitality markets and offers actionable insights for hotel managers and policymakers working to support sustainable recovery through integrated tourism and workforce strategies.

Keywords: AI, Machine Learning, forecasting, Occupancy, Hotel, Crisis

Introduction

The hospitality industry is a critical pillar of Taiwan's tourism-driven economy, contributing substantially to employment, regional development, and service-sector growth. Prior to 2020, Taiwan's hotel industry benefited from stable international tourism demand, with average occupancy rates for tourist hotels consistently exceeding mid-60 percent levels (Chang et al., 2021). This relative stability enabled hotels to optimize operations, staffing, and pricing strategies around predictable seasonal patterns.

The outbreak of the COVID-19 pandemic fundamentally disrupted this equilibrium (Lee & Chen, 2022), (Wang & Chen, 2023). Border closures, quarantine requirements, and global travel restrictions caused international arrivals to collapse almost overnight, driving hotel occupancy rates to historic lows between 2020 and 2022 (Chen et al., 2020), (Chen & Chen, 2021). In response to the economic downturn, the government implemented domestic travel subsidies and containment policies that partially offset demand losses, but occupancy patterns diverged sharply from pre-pandemic norms (Filimonau et al., 2020). As restrictions eased in 2023–2024, the hotel sector entered a recovery phase characterized by uneven demand, labor constraints, and gradual reentry of international visitors. Importantly, emerging evidence suggests that post-COVID recovery has not simply reverted to pre-pandemic structures, but instead reflects altered relationships between demand sources, staffing, pricing, and operational scale.

While existing hospitality research has documented the impacts of pandemic-related economic downturns using descriptive statistics, surveys, and econometric models (Chen et al., 2020), (Hao et al., 2020), (Wang & Chen, 2023), fewer studies have leveraged machine learning-driven AI techniques to examine how predictors of hotel occupancy evolved across distinct crisis phases. Moreover, many predictive studies prioritize accuracy without adequately addressing interpretability, limiting their usefulness for managerial decision-making and policy formulation (Chen et al., 2022). Understanding *why* models make certain predictions is especially important during periods of structural disruption, when historical relationships may no longer hold.

To address these gaps, this study applies a Gradient Boosting Classifier combined with SHAP (SHapley Additive exPlanations) to analyze Taiwanese hotel data across three distinct periods: pre-COVID (2017–2019), during the economic downturn caused by COVID-19 (2020–2022), and post-COVID recovery (2023–2024). Rather than predicting continuous occupancy rates, we classify occupancy into low, medium, and high categories to reflect operationally meaningful thresholds used by hotel managers. By integrating explainable AI techniques, we move beyond prediction to identify how the importance and direction of key features—such as visitor composition, employment structure, pricing, and hotel size—shifted across crisis phases.

Building on prior work that examined pre-COVID conditions (Huang et al., 2023) and pre- versus during-COVID comparisons (Yao et al., 2026), this study extends the literature by explicitly modeling three temporal regimes and systematically comparing feature importance and SHAP values across them. In doing so, it contributes both methodologically and substantively to hospitality analytics and crisis resilience research.

Accordingly, this study addresses the following research questions:

- RQ1: How does the distribution of hotel occupancy categories (low, medium, high) differ across the pre-COVID, during-COVID, and post-COVID recovery periods in Taiwan?
- RQ2: How do the relative importance and SHAP-based contributions of key predictors—particularly domestic and international visitor counts—change across these three periods?
- RQ3: Which operational and employment-related features are most strongly associated with occupancy resilience during the post-COVID recovery phase?

By answering these questions, this research provides interpretable, data-driven insights into how Taiwan's hotel industry adapted to an unprecedented global shock and which factors may support sustained recovery in the post-pandemic era.

Data Source

The data are sourced from the Tourism Administration, Ministry of Transportation and Communications, Taiwan (Tourism Administration, 2026). The dataset consists of three Excel files: *TW 47 hotel data_2017–2019_reduced-revenue.xlsx* (141 records), *TW 47 hotel data_2020–2022_reduced-revenue.xlsx* (141 records), and *TW 47 hotel data_2023–2024_reduced-revenue.xlsx* (95 records). Each file excludes those that went out of business during the study period. The variables captured include Location (urban vs. regional), Management Type (I: Domestic; II: Chain), Average Room Rate, Occupancy Rate, Domestic and International Visitor Counts, Number of Rooms, and Employee Counts (Room, F&B, Other). Occupancy rates were preprocessed and categorized into three classes: low (<50%), medium (50%–75%), and high (>75%).

We excluded revenue-related variables—such as room revenue, food and beverage revenue, and other revenues—from the set of predictors. Although these variables are highly correlated with occupancy rates and contain useful information, they are directly tied to the outcome variable. Including them would therefore undermine the purpose of this study, which aims to build models that predict occupancy rather than rely on variables that are essentially proxies for it.

Methods

Gradient Boosting Classifier

A Gradient Boosting Classifier is an ensemble machine learning algorithm that improves predictive accuracy by building models sequentially, with each new model (typically a decision tree) correcting the errors (residuals) of the combined ensemble of previous models. It minimizes a loss function, such as log-loss, using gradient descent to create a strong, highly accurate classifier from many weak learners, making it robust for complex, tabular data.

Gradient Boosting begins with a simple foundation: an initial prediction generated by a very basic model, often just the average of the target values. From this modest starting point, the algorithm enters a carefully structured, iterative learning process designed to progressively improve predictive accuracy.

At each iteration, the current model's errors—known as residuals—are calculated. These residuals represent the aspects of the data that the model has not yet captured well. A new weak learner, typically a shallow decision tree, is then trained specifically to predict these residuals. By focusing on what the model previously got wrong, each added tree helps fine-tune the ensemble's predictive ability.

This approach is inherently sequential. Unlike Random Forests, which train many trees in parallel and aggregate their predictions, Gradient Boosting builds its trees one at a time. Every new tree is crafted to correct the shortcomings of the trees that came before it. This orderly, step-by-step structure enables the model to steadily reduce error as it grows.

At the heart of this process lies loss function optimization. Gradient Boosting uses gradient descent to minimize the loss between predicted and actual values. Each new model is added in the direction that most reduces this loss, ensuring that the ensemble improves with every iteration.

To keep the model from overfitting and becoming too tightly tailored to the training data, Gradient Boosting applies a technique called shrinkage, or learning rate adjustment. This scales down the contribution of each newly added tree, making updates more conservative and allowing the model to generalize more effectively.

Through this repeated cycle of error calculation, weak learner training, model updating, and controlled refinement, Gradient Boosting constructs a powerful ensemble capable of capturing complex relationships in data—one small, purposeful step at a time.

SHAP Methodology

SHAP (SHapley Additive exPlanations) is a unified framework for interpreting machine learning model predictions. Introduced by Scott Lundberg and Su-In Lee (Lundberg & Lee, 2017), it applies principles from cooperative game theory to quantify how each feature contributes to a specific prediction (local explanation) while also enabling an overall assessment of feature importance across the entire dataset (global explanation). SHAP has become a widely used tool in explainable AI (XAI), especially for complex models such as gradient boosting machines, neural networks, and random forests, where understanding the rationale behind predictions is essential for trust, model validation, and regulatory compliance.

At its core, SHAP assigns value to each feature that represents its marginal contribution to the difference between the model's prediction for a given instance and the model's average (baseline) prediction. These contributions are additive: the sum of all SHAP values for a single observation equals the deviation of that observation's prediction from the baseline. This additivity makes SHAP both intuitive and mathematically rigorous, providing consistent and transparent explanations for model outputs.

Applications and Visualizations

SHAP values enable both local and global interpretations:

- **Local Explanations:** For a single prediction, a force plot shows how features "push" the output from baseline to final value (positive/negative directions).
- **Global Insights:** Aggregate via mean absolute SHAP ($|\text{SHAP}|$) for feature importance, or beeswarm/summary plots to show distribution and direction (e.g., high values of a feature increasing predictions).

Class-Specific: In multi-class problems (like our occupancy categories: low, medium, high), SHAP computes values per class, revealing differential impacts (e.g., international visitors strongly boosting "high" occupancy pre-COVID). We used mean $|\text{SHAP}|$ to compare feature influences across periods, showing shifts like the drop in "Visitors Count Int" during 2020–2022. This highlighted pandemic effects, where domestic visitors became dominant for low occupancy predictions.

All in all, a Gradient Boosting Classifier (from scikit-learn) was trained separately for each period, using features excluding occupancy rate. Hyperparameters were default, with 5-fold cross-validation for accuracy. SHAP values were computed to explain predictions, providing mean absolute SHAP for global and class-specific insights.

Results

Class Distribution of Occupancy Categories Across Time Periods

The occupancy rates were categorized as follows: Low: $< 50\%$, Medium: $50\%–75\%$, High: $> 75\%$. The distributions (Table 1) show clear shifts over time, likely influenced by external factors like the COVID-19 pandemic:

Table 1 Occupancy Rate Distributions

Period	Low (%)	Medium (%)	High (%)
2017–2019	15.60	48.23	36.17
2020–2022	67.38	27.66	4.96
2023–2024	24.47	47.87	27.66

- 2017–2019 (Pre-COVID): Balanced toward higher occupancy, with medium and high dominating (over 84% combined). This suggests a robust hotel industry with strong demand.
- 2020–2022 (COVID Impact): Sharp increase in low occupancy (over 67%), with high occupancy nearly vanishing. This reflects travel restrictions, lockdowns, and reduced tourism.
- 2023–2024 (Recovery): Partial rebound, with medium occupancy returning to pre-COVID levels, but low occupancy remains elevated compared to pre-COVID, and high occupancy is lower. This indicates ongoing recovery but not yet full return to pre-pandemic normals.

Model Performance (Cross-Validation Accuracy)

A Gradient Boosting Classifier was trained for each period using the following features: Location, Mgt (encoded as 0 for domestic, 1 for chain), Average Room Rate, Visitors Count Dom, Visitors Count Int, Rooms Count, Room Emp, F&B Emp, Other Emp. 5-fold cross-validation was used to assess accuracy. The analysis results are presented in Tables 2 and 3.

Table 2 Analysis Results of Gradient Boosting Classifier

Period	Macro F1	Weighted F1	Accuracy
2017–2019	0.858	0.861	0.86
2020–2022	0.652	0.765	0.77
2023–2024	0.648	0.656	0.66

Table 3 More Analysis Results of Gradient Boosting Classifier

	2017-2019			2020-2022			2023-2024		
Class	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
High	0.81	0.87	0.84	0.5	0.5	0.5	0.56	0.62	0.59
Low	0.86	0.86	0.86	0.83	0.86	0.85	0.8	0.57	0.67
Medium	0.9	0.86	0.88	0.64	0.58	0.61	0.67	0.71	0.69

Cross-validation accuracies were 86% (2017–2019), 77% (2020–2022), and 66% (2023–2024). Lower post-2019 performance suggests increased variability from pandemic effects. The model performed best during the 2017–2019 period, achieving its highest Macro and Weighted F1 scores. During these years,

occupancy patterns appear to have been more stable and predictable, allowing the model to capture trends more effectively (see Table 2).

Performance declined noticeably between 2020 and 2022. The Macro F1 score dropped to 0.65, and predictions for the “High occupancy” category were particularly weak, with an F1 score of only 0.50 (See Table 3). This decline likely reflects the significant disruptions caused by COVID-19, which dramatically altered typical patterns in the tourism and hospitality sectors.

Model performance in 2023–2024 remained relatively unstable. Accuracy reached its lowest point at 0.66, and while the model showed balanced performance across occupancy classes, overall predictive strength was still only moderate. These results suggest that the industry may still be in a post-pandemic adjustment phase, with occupancy patterns continuing to fluctuate.

Finally, the gap between Macro and Weighted F1 scores highlights an important aspect of the dataset. The higher Weighted F1 score indicates strong performance on the majority class—low occupancy—while the lower Macro F1 score shows that the model struggles with minority classes such as high occupancy. This imbalance underscores the challenges the model faces when predicting less common occupancy outcomes.

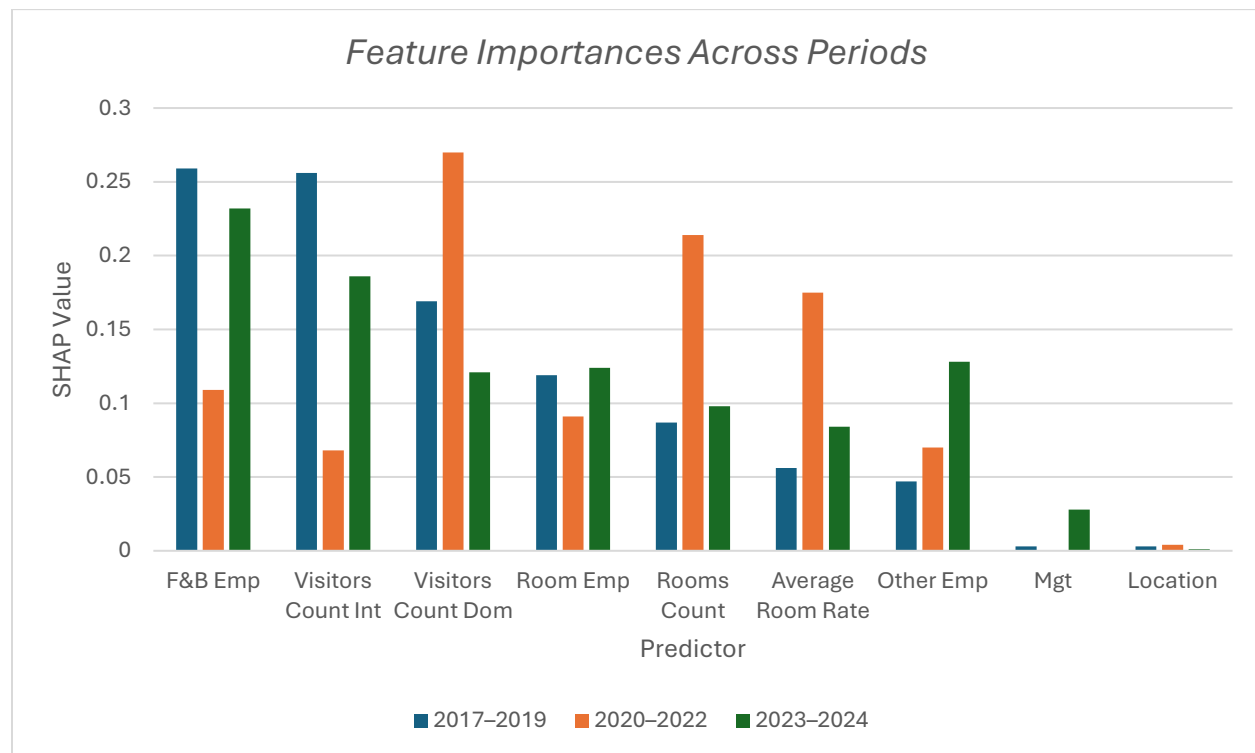
Feature Importances Across Time Periods

Feature importances from the fitted models highlight which variables most strongly predict occupancy categories. Values are normalized (sum to 1.0); higher values indicate greater influence (See Table 4 and Figure 1.)

Table 4. Feature Importance

Feature	2017–2019	2020–2022	2023–2024
F&B Emp	0.259	0.109	0.232
Visitors Count Int	0.256	0.068	0.186
Visitors Count Dom	0.169	0.270	0.121
Room Emp	0.119	0.091	0.124
Rooms Count	0.087	0.214	0.098
Average Room Rate	0.056	0.175	0.084
Other Emp	0.047	0.070	0.128
Mgt	0.003	0.000	0.028
Location	0.003	0.004	0.001

Figure 1 Feature Importances



From 2017 to 2019, occupancy levels were driven primarily by international tourism and food-and-beverage activity. During this period, F&B employment and international visitor counts consistently ranked as the most influential factors, underscoring how strongly hotels depended on global travelers and robust service operations to maintain high occupancy.

The landscape changed dramatically between 2020 and 2022, which was the economic downturn caused by the pandemic. As international travel collapsed, international visitor counts lost nearly all predictive power. Instead, domestic visitor volume became the dominant driver, supported by structural factors such as the number of available rooms and the increasing importance of the average room rate. This shift reflects a market that relied heavily on local demand and operational scale during the pandemic's restrictions and uncertainty.

By 2023 and 2024, F&B employment and international visitor counts reemerged as top predictors, indicating a gradual return to pre-pandemic patterns. However, their influence was now accompanied by more balanced contributions from other employment-related variables, such as other hotel staff and room employees. Management type also gained slight relevance, possibly highlighting differences in how chain-managed and domestically managed properties approached recovery.

Across all periods, employment-related features—F&B staff, room staff, and other hotel employees—remained consistently significant, suggesting that workforce capacity is closely tied to occupancy resilience. Visitor-type importance shifted according to global circumstances: international visitors dominated in stable periods, domestic visitors during the COVID disruption. Room count and average room rate became especially influential during the pandemic, when structural attributes mattered more than travel patterns. Throughout the entire timeline, management style and location played only minor roles.

Overall, the patterns reveal the pandemic's disruptive force: a collapse of high-occupancy models reliant on international travel from 2020 to 2022, followed by an uneven but notable return to earlier dynamics by 2023 and 2024.

Occupancy Distribution Shifts

Pre-COVID (2017–2019), distributions favored medium (48.23%) and high (36.17%) occupancy, indicating strong demand. During the economic downturn caused by the pandemic (2020–2022), low occupancy dominated (67.38%), with high nearly absent (4.96%), aligning with reported drops to 20% in April 2020 (Chen et al., 2020). Recovery (2023–2024) showed partial normalization: low (24.47%), medium (47.87%), high (27.66%).

SHAP Values Analysis Across Time Periods

SHAP values provide a detailed explanation of how each feature contributes to the model's predictions for the occupancy categories (low, medium, high). Unlike feature importances (which show global contribution), SHAP values are local but aggregated here as mean absolute values (mean |SHAP|) to indicate the average magnitude of impact per feature per class. Higher values mean the feature has a stronger influence on predicting that class. More details on SHAP values across time periods can be found in Tables 5–7.

The classes are ordered as: high (label 0), low (label 1), medium (label 2). We also computed an overall mean |SHAP| per feature (averaged across classes) for a high-level view. Results are compared across periods, highlighting shifts (e.g., due to COVID impacts). Feature importances from the models are included for reference, as they align closely but lack class-specific nuance.

Mean Absolute SHAP Values and Feature Importances

Table 5: 2017–2019 SHAP values

Feature	High	Low	Medium	Overall Mean	Importance
Average Room Rate	0.047	0.008	0.050	0.035	0.056
F&B Emp	0.119	0.023	0.107	0.083	0.259
Location	0.002	0.002	0.003	0.003	0.003
Mgt	0.001	0.005	0.004	0.003	0.003
Other Emp	0.055	0.008	0.053	0.039	0.047
Room Emp	0.061	0.010	0.062	0.045	0.119
Rooms Count	0.096	0.016	0.097	0.070	0.087
Visitors count Dom	0.067	0.062	0.108	0.079	0.169
Visitors Count Int	0.211	0.115	0.194	0.173	0.256

Table 6: 2020–2022 SHAP values

Feature	High	Low	Medium	Overall Mean	Importance
Average Room Rate	0.046	0.087	0.066	0.066	0.175

F&B Emp	0.011	0.072	0.069	0.051	0.109
Location	0.011	0.003	0.011	0.008	0.004
Mgt	0.000	0.000	0.000	0.000	0.000
Other Emp	0.013	0.020	0.026	0.020	0.070
Room Emp	0.006	0.029	0.027	0.021	0.091
Rooms Count	0.080	0.129	0.087	0.099	0.214
Visitors count Dom	0.099	0.259	0.194	0.184	0.270
Visitors Count Int	0.007	0.050	0.052	0.036	0.068

Table 7: 2023–2024 SHAP values

Feature	High	Low	Medium	Overall Mean	Importance
Average Room Rate	0.040	0.014	0.046	0.033	0.084
F&B Emp	0.047	0.155	0.144	0.115	0.232
Location	0.001	0.001	0.001	0.001	0.001
Mgt	0.028	0.006	0.022	0.019	0.028
Other Emp	0.128	0.011	0.133	0.091	0.128
Room Emp	0.092	0.023	0.095	0.070	0.124
Rooms Count	0.024	0.026	0.037	0.029	0.098
Visitors count Dom	0.037	0.049	0.084	0.057	0.121
Visitors Count Int	0.225	0.072	0.188	0.162	0.186

A summary of the overall mean absolute SHAP values for the Gradient Boosting Classifier across the three time periods is shown in Table 8.

Table 8: Overall mean SHAP values (Extracted from Table 5, 6, 7)

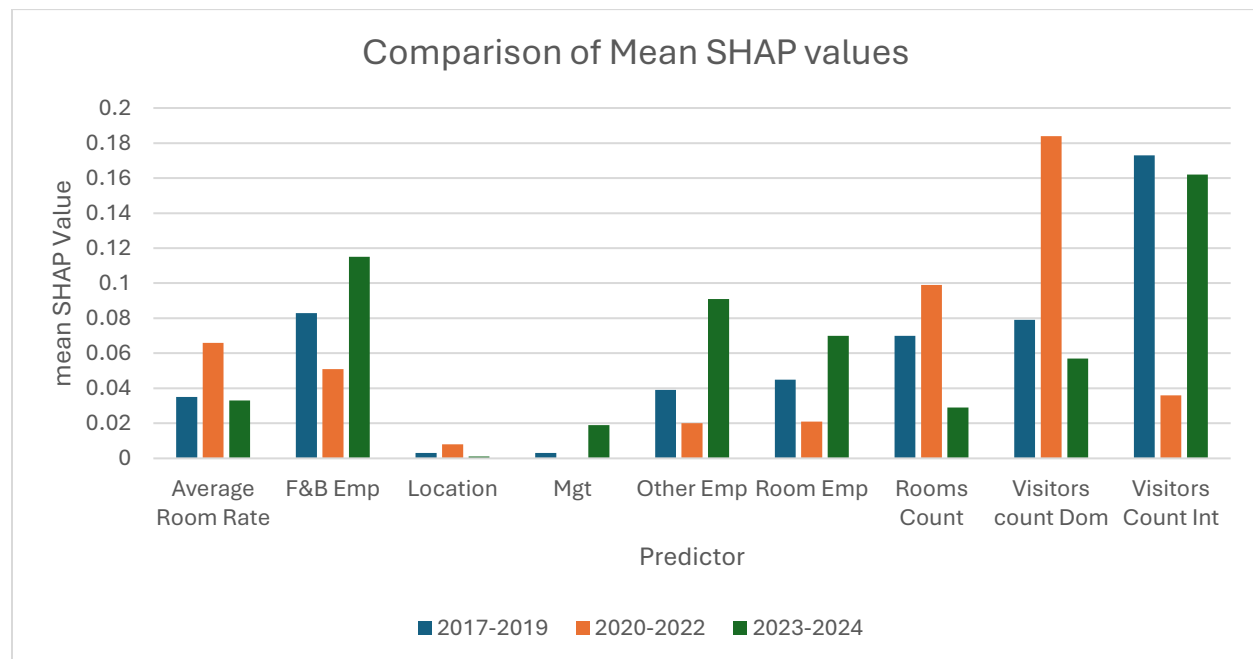
Feature	Overall mean		
	2017-2019	2020-2022	2023-2024
Average Room Rate	0.035	0.066	0.033
F&B Emp	0.083	0.051	0.115
Location	0.003	0.008	0.001
Mgt	0.003	0.000	0.019

Other Emp	0.039	0.020	0.091
Room Emp	0.045	0.021	0.07
Rooms Count	0.07	0.099	0.029
Visitors count Dom	0.079	0.184	0.057
Visitors Count Int	0.173	0.036	0.162

Feature Importances and SHAP Analysis

The analysis of feature importances and SHAP values showed clear shifts over time. Before and after COVID-19, international visitors were the strongest contributors to predictions, with overall SHAP means of 0.173 and 0.162. During the pandemic, however, their influence dropped sharply to 0.036, while domestic visitors became the dominant factor, reaching a SHAP mean of 0.184. As the sector began to recover, employment-related variables—such as food-and-beverage employment—grew in importance, reflected in a SHAP mean of 0.115. This rise suggests that staffing levels played a meaningful role in shaping operational resilience during the recovery period.

Figure 2 Mean Absolute SHAP Values Across Periods



Bar chart of SHAP values (Figure 2) illustrate these trends, with international visitors showing a U-shaped recovery pattern. Lines for key features like Visitors Count Int (high pre/post, low during), Visitors Count Dom (peak during), and F&B Emp (rising in recovery). Class-specific SHAP highlighted international visitors' stronger push toward high occupancy pre-COVID (0.211) vs. domestic for low during COVID (0.259).

Key Insights from SHAP Analysis

From 2017 to 2019—before the onset of the COVID-19 pandemic—international visitors played the dominant role in predicting occupancy across all categories. Their influence highlights the strength of global

tourism during this period. Hotels achieving high occupancy were strongly associated with employment-related features, particularly food and beverage staffing, suggesting that service capacity and quality were essential to supporting peak performance. While international visitors drove most of the variation, domestic visitor numbers were more relevant for medium and low occupancy levels, indicating a more balanced demand base in those segments.

The pandemic years of 2020 to 2022 brought a dramatic shift. With international travel curtailed, domestic visitors became the primary driver of occupancy. International visitors fell sharply in importance across all classes. During this disruption, operational and structural factors—especially the number of available rooms and the average room rate—gained influence. These features reflect the reality of operating in a low-demand environment, where pricing strategy and the scale of the property played an outsized role. When high occupancy did occur during this period, it was largely supported by domestic demand and by the underlying infrastructure of the hotel.

In the recovery years of 2023 and 2024, international visitors regained their status as the leading predictor of occupancy, marking a return to pre-pandemic dynamics. At the same time, employment features such as F&B staff and room staff surged in importance, particularly for low and medium occupancy levels. This pattern suggests that staffing and operational readiness were fundamental to stabilizing occupancy as travel resumed. Hotels with stronger workforce capacity were better positioned to capture returning demand and maintain consistent performance.

Across all periods, several overarching themes emerge. The roles of domestic and international visitors fluctuate dramatically in response to external conditions, especially the pandemic, while employment and structural features such as room count consistently provide resilience. SHAP value breakdowns reveal meaningful class-specific patterns—such as the especially strong effect of international visitors on high occupancy rates—while the global feature importances capture the broader performance landscape over time.

Contrary to traditional hospitality theories that emphasize the competitive advantage of prime locations or international chain management (Yang et al., 2014, Ivanova & Ivanov, 2016), this study found that Management Type and Location played only minor roles across all three periods. This suggests that the COVID-19 shock was a universal crisis in Taiwan. The collapse of international demand was so profound that it bypassed traditional structural advantages. Whether a hotel was a domestic brand or a chain, or whether it was located in an urban or regional area, the primary determinant of occupancy shifted instead to broader demand sources and operational capacity.

Overall, the results underscore the pandemic's disruptive effect on the hospitality industry and the gradual path back toward pre-COVID drivers of hotel occupancy.

Discussion

This study provides empirical evidence of how the economic downturn, exemplified by the COVID-19 pandemic, reshaped hotel occupancy dynamics in Taiwan, revealing both short-term disruptions and longer-term structural adjustments. By combining Gradient Boosting classification with SHAP-based explainability, the analysis uncovers not only changes in predictive accuracy across periods, but also meaningful shifts in the underlying drivers of hotel occupancy.

Consistent with crisis management and tourism resilience theory, the results demonstrate a clear demand substitution effect during the pandemic period. Prior to the COVID-19 outbreak, international visitor counts

were the dominant predictor of high occupancy, reflecting Taiwan's reliance on global tourism flows. During the 2020–2022 economic downturn, this pattern collapsed, and domestic visitor counts emerged as the primary driver of occupancy outcomes. This finding aligns with crisis adaptation frameworks, which posit that tourism systems temporarily reorient toward internal markets when external demand is constrained. The SHAP analysis further refines this insight by showing that domestic visitors were particularly influential in distinguishing low-occupancy outcomes during the pandemic, underscoring the limited but stabilizing role of domestic travel under restrictive conditions.

From an organizational resilience perspective, employment-related variables—especially food and beverage, room, and other staff counts—exhibited consistent importance across all periods and increased in relative prominence during the recovery phase. Rather than implying causality, these associations suggest that staffing levels function as indicators of operational readiness and service capacity. In the post-COVID context, where labor shortages and service quality concerns are widely reported, the growing influence of employment features may reflect hotels' ability to scale operations in response to returning demand. This interpretation is consistent with resilience theory, which emphasizes adaptive capacity and resource flexibility as key determinants of recovery performance.

Structural characteristics, such as room count and average room rate, became more influential during the pandemic-related economic downturn, as demand was suppressed and pricing strategies played a greater role in occupancy outcomes. This pattern aligns with information asymmetry and pricing literature, which suggests that during periods of uncertainty, consumers respond more strongly to price signals and capacity constraints. As international travel resumed, these structural features declined in importance relative to demand and staffing variables, indicating a partial normalization of market dynamics rather than a complete return to pre-pandemic conditions.

Importantly, the divergence between feature importances and class-specific SHAP values highlights the value of explainable machine learning in hospitality research. While traditional feature importance metrics identify globally influential variables, SHAP analysis reveals how specific features differentially affect low, medium, and high occupancy outcomes across time. This granularity provides insights that would be difficult to obtain using conventional regression approaches, particularly under non-stationary conditions such as a global pandemic (Yang & Chen 2009).

Several limitations should be acknowledged. The dataset excludes hotels that exited the market during the study period, which may bias the results toward more resilient establishments. While these insights are invaluable for understanding how the sector stabilizes, they may not fully capture the extreme vulnerabilities that led to market exits during the height of the pandemic. Additionally, class imbalance, particularly during the economic downturn caused by COVID-19, likely contributed to reduced predictive accuracy. Finally, SHAP values reflect model-based associations rather than causal relationships, and findings should be interpreted accordingly.

Despite these limitations, the results offer practical implications for policymakers and hotel managers. Strategies that support international tourism recovery remain essential for restoring high occupancy levels, while sustained investment in workforce stability may enhance resilience during future shocks. More broadly, this study demonstrates that interpretable machine learning can serve as a powerful tool for understanding structural change in hospitality markets, bridging predictive performance with actionable insight.

Conclusion

This study examined how hotel occupancy dynamics in Taiwan evolved before, during, and after the COVID-19 pandemic by applying a Gradient Boosting Classifier with SHAP-based explainability to longitudinal hotel data spanning 2017–2024, capturing the economic downturn. By framing the analysis around three distinct temporal regimes, the study moves beyond static assessments of pandemic impact and provides a structured, interpretable account of how occupancy drivers shifted across crisis phases.

Addressing RQ1, the results demonstrate pronounced changes in the distribution of occupancy categories over time. Prior to COVID-19, medium and high occupancy levels dominated, reflecting stable international tourism demand. During the pandemic, low occupancy became the prevailing outcome, with high occupancy nearly disappearing. The post-COVID recovery period showed partial normalization, with medium occupancy returning to near pre-pandemic levels, while high occupancy remained below historical benchmarks—indicating an ongoing, uneven recovery rather than a full structural reset.

In response to RQ2, the analysis revealed substantial temporal shifts in the relative importance and SHAP-based contributions of key demand-side predictors. International visitor counts were the dominant driver of occupancy outcomes in both the pre-COVID and recovery periods but lost nearly all predictive influence during the pandemic. Conversely, domestic visitor counts became the most influential predictor during the 2020–2022 economic downturn, highlighting a temporary substitution effect under travel restrictions. The recovery phase exhibited a rebalancing of these influences, suggesting that international tourism remains essential for achieving high occupancy levels in Taiwan’s hotel sector.

Regarding RQ3, employment-related and operational features emerged as central indicators of occupancy resilience during the recovery period. Food and beverage staff, room employees, and other employment categories increased in relative importance after COVID-19, reflecting the role of operational capacity and service readiness in stabilizing occupancy as demand returned. These associations do not imply causal effects but indicate that staffing levels serve as meaningful signals of a hotel’s ability to adapt and scale in a post-crisis environment.

Collectively, the findings contribute to hospitality and tourism research by demonstrating the value of explainable machine learning for analyzing structural change under extreme disruption. Methodologically, the integration of Gradient Boosting with SHAP enables both accurate classification and transparent interpretation in a non-stationary context. Substantively, the results reinforce theories of demand substitution, organizational resilience, and adaptive capacity within tourism systems.

From a practical perspective, the study suggests that sustained recovery in Taiwan’s hotel industry depends on both the revival of international tourism and the stabilization of the hospitality workforce. Policymakers may benefit from aligning tourism promotion with labor-support initiatives, while hotel managers can leverage staffing and service readiness as strategic levers during periods of uncertainty.

Future research could extend this framework by incorporating macroeconomic indicators, policy variables, and firm-level financial data, while also exploring causal inference methods to complement explainable machine learning approaches. In addition, further investigation is warranted into the relationship between staffing levels and occupancy rates, particularly to better understand the operational dynamics that influence hotel performance during periods of disruption and recovery. More broadly, the framework presented in this study offers a transferable blueprint for analyzing hospitality resilience and adaptive capacity in the face of future global shocks.

References

- Chang, C. L., McAleer, M., & Wong, W. K. (2021). Effectiveness of the talent cultivation training program for industry transformation in Taiwan during the COVID 19 pandemic. *International Journal of Environmental Research and Public Health*, 18(14), 7291. <https://doi.org/10.3390/ijerph18147291>
- Chen, C. L., & Chen, M. H. (2021). Hospitality industry employees' intention to stay in their job after the COVID 19 pandemic. *Administrative Sciences*, 11(4), 144. <https://doi.org/10.3390/admsci11040144>
- Chen, C. M., Chen, S. H., & Lee, H. T. (2020). The impact and recovering strategies of the COVID 19 pandemic: Lessons from Taiwan's hospitality industry. *Cogent Social Sciences*, 6(1), 1829806. <https://doi.org/10.1080/23311886.2020.1829806>
- Chen, M. H., Jang, S. C., & Peng, N. (2022). How do hotel characteristics moderate the impact of COVID 19 on hotel performance? Evidence from Taiwan. *Current Issues in Tourism*, 25(8), 1192–1197. <https://doi.org/10.1080/13683500.2021.1908965>
- Filimonau, V., Derqui, B., & Matute, J. (2020). The COVID 19 pandemic and organizational commitment of senior hotel managers. *International Journal of Hospitality Management*, 91, 102659. <https://doi.org/10.1016/j.ijhm.2020.102659>
- Hao, F., Xiao, Q., & Chon, K. (2020). COVID 19 and China's hotel industry: Impacts, a disaster management framework, and post pandemic agenda. *International Journal of Hospitality Management*, 90, 102636. <https://doi.org/10.1016/j.ijhm.2020.102636>
- Huang, Y. H. J., Yao, J. F. and Yang, C. Y. (2023). Enhancing Management of Diverse Taiwanese Hotels Through Machine Learning and Data Mining, *IEEE 3rd International Conference on Social Sciences and Intelligence Management (SSIM)*, Taichung, Taiwan, pp. 149 152, doi: 10.1109/SSIM59263.2023.10468732.
- Ivanova, M., & Ivanov, S. (2016). Hotel chains' conceptual models. In *The Routledge handbook of hotel chain management* (1st ed., pp.69 82). Routledge.
- Lee, C. K., & Chen, M. H. (2022). The impact of COVID 19 on the tourism and hospitality industry: Evidence from international stock markets. *Annals of Tourism Research Empirical Insights*, 3(2), 100054. <https://doi.org/10.1016/j.annale.2022.100054>
- Lundberg S. & Lee, S I (2017). A Unified Approach to Interpreting Model Predictions, Cornell University, <https://arxiv.org/pdf/1705.07874v2>
- Tourism Administration, Taiwan Ministry of Transportation and Communications (MOTC). (2020–2024). *Tourist hotel operating statistics and occupancy rates (monthly/annual reports)*. Retrieved February 4, 2026, from <https://stat.taiwan.net.tw/>
- Wang, C. H., & Chen, C. M. (2023). Evaluating information asymmetry effects on hotel pricing: A comparative analysis before and during the COVID 19 pandemic in Taiwan's market. *Humanities and Social Sciences Communications*, 10, 258. <https://doi.org/10.1057/s41599-023-02582-1>

- Yang, H. Y., & Chen, K. H. (2009). A general equilibrium analysis of the economic impact of a tourism crisis: A case study of the SARS epidemic in Taiwan. *Journal of Policy Research in Tourism, Leisure and Events*, 1(1), 37–60. <https://doi.org/10.1080/19407960902738313>
- Yang, Y., Luo, H., & Law, R. (2014). Theoretical, empirical, and operational models in hotel location research. *International Journal of Hospitality Management*, 36, 209–220. <https://doi.org/10.1016/j.ijhm.2013.09.004>
- Yao, J. F., Huang, Y. H. J., Wu, Daniel, Yang, C.Y. (2026). AI Driven Hotel Resource Optimization Before and During COVID 19, ICCE TW 2026, Taiwan