

Article

Infrastructure, Governance, and Price Stability as Binding Constraints on Inbound Tourism to India

Bidyut Kumar Ghosh 

Manipal School of Commerce and Economics, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India; bidyutbwn@gmail.com or bidyut.ghosh@manipal.edu; Tel.: +91-96799980249

Abstract

India's inbound tourism potential has not been fully realised, even though it has rich heritage resources, expanding air networks, and sustained policy attention. To identify the key determinants of foreign tourism demand, this study applies a novel interpretable machine learning framework, XGBoost with SHAP and ALE plots, on a panel dataset of 61 source countries to India between 2002 and 2024. Using a gravity-based tourism demand model, the analysis uncovers nonlinearities and interactions in tourism demand based on origin-country income, India's hotel capacity, domestic and international aircraft movements, UNESCO heritage sites, mega-events, inflation, and governance indicators. Accumulated local effects (ALEs) and SHAP values were used as interpretable tools. The results show that source-country income and air connectivity are the most influential drivers of arrivals, while heritage sites and hotel rooms display clear saturation and diminishing returns, and governance and inflation exert only mild or nonlinear effects. Mega-events provide small and inconsistent short-run gains without strong persistence. The findings indicate that India's future tourism gains lie less in further capacity expansion and more in strengthening air connectivity, strategically targeting emerging middle-income markets, and upgrading quality, governance, and price stability to convert existing assets into sustained spatially dispersed arrivals.

Keywords: inbound tourism; gravity-based tourism demand; interpretable machine learning; XGBoost; SHAP values; accumulated local effects; air connectivity; UNESCO heritage sites; governance quality; India



Academic Editors: Hak-Seon Kim,
Jue Wang, Je Eun Yoo and Lewis Ting
On Cheung

Received: 8 December 2025

Revised: 14 January 2026

Accepted: 29 January 2026

Published: 21 February 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Tourism is one of the world's most dynamic and expanding economic sectors. It has multifaceted impacts across economic, social, and cultural aspects. Before the COVID-19 pandemic, the global tourist arrivals reached nearly 1.5 billion in 2019 ([OECD, 2020](https://doi.org/10.1016/j.tourhosp.2020.0057)). Improvements in transportation, digitalisation, visa liberalisation, and active global marketing have propelled sustained growth ([Buhalis & Law, 2008](#); [Gössling & Hall, 2006](#)).

Beyond its contributions to economic development, tourism promotes mutual understanding, cross-country cultural exchange, and destination branding ([Anholt, 2007](#); [Pike, 2016](#); [Telfer & Sharpley, 2015](#)). Especially for emerging economies, tourism catalyses economic diversification and entrepreneurship, which in turn attracts investment in the destination countries ([Ashley et al., 2007](#); [Torres, 2002](#)). Leading global tourist destinations such as Thailand, Malaysia, and Vietnam are standing proof of these tourism-based economic developments ([Dwyer et al., 2004](#); [Lim, 1997](#)).

India holds a tally of more than forty UNESCO World Heritage Sites across cultural and natural resources. With this, the substantial infrastructural growth across rail, road and civil aviation enables India to attract tourists globally for experiential and heritage travel (Government of India, 2020; Timothy & Boyd, 2006). However, India's share of the global tourist market remains poor compared to its regional peers. In 2019, while China, Thailand and Malaysia received 65.7 million, 39.8 million and 26.1 million international tourists, respectively, India attracted only 10.9 million (World Tourism Organization (UNWTO), 2021). This gap signals the need for a deeper understanding of underlying competitiveness and sectoral challenges.

Infrastructure, governance, cultural management, and creative policy frameworks are all intertwined with the growth of tourism (Hall et al., 2011; Prideaux, 2000). High standard destination quality, institutional capacity, event-based global visibility, and external diseconomies, such as global shocks and economic volatility, determine the growth of the tourism sector (Nunkoo & Ramkissoon, 2011; Wut et al., 2021). India launched an international tourism campaign, 'Incredible India,' in 2002 through streamlined e-visa programs and strategic mega-events to promote the country's rich history, diverse culture, spirituality, and varied geography. Despite measurable progress, the sector's vulnerability to external disruptions and persistent regional disparities underscores the need for robust empirical evidence to inform targeted and sustainable destination management strategies (World Tourism Organization (UNWTO), 2021).

Traditionally, econometric modelling, especially with the panel data method, was widely used to examine tourism demand factors such as income, infrastructure, price, and heritage assets (Crouch, 1994; Song et al., 2010). However, the empirical research on tourism in India is still dispersed and primarily descriptive. The use of sophisticated econometric techniques to capture intersectoral relationships, diminishing returns, and nonlinear behaviour is very much limited. When nonlinear or interaction terms are added, standard econometric models struggle to handle multicollinearity and model complex relationships (Wooldridge, 2010). Modern approaches—including ridge regression and tree-based machine learning—offer potential for greater explanatory power and managerial insight, yet they remain largely unexplored within Indian and emerging-economy contexts (Hastie et al., 2009; James et al., 2013).

This study addresses these critical empirical and methodological gaps by making use of an interpretable machine learning framework to identify the determinants of foreign tourist arrivals to India and their impact patterns. The study applies XGBoost regression with cooperative game theory-based Shapley additive explanation (SHAP) and accumulated local effect (ALE) plots to identify the nonlinear impacts, threshold effects, and complex interactions among the determinants of foreign tourists' arrivals. In this context, the analysis addresses four principal questions:

- What are the effects of heritage sites, hotel capacity, and source-country income on India's foreign tourist arrivals?
- How do nonlinearities, including diminishing returns and thresholds, shape these relationships?
- To what extent do governance and mega-events influence arrivals?
- Does interpretable machine learning (XGBoost) enhance explanatory robustness and policy relevance compared to conventional models?

Cross-country comparisons clarify India's development challenges and opportunities. Effective strategies in Thailand, Malaysia, and Singapore focus on integrated infrastructure, strategic marketing, and supportive governance to enhance tourism quality (Hall et al., 2011). In contrast, India's concentration of foreign tourists' arrivals in a few states and

institutional limitations underscore the policy imperative for greater inclusiveness and destination competitiveness (Ministry of Tourism, 2021).

Tourism demand modelling has changed its shape along with the growth of the sector and social change. Early economic studies used basic data analysis to show tourists' movement and employed cross-sectional and time-series econometric methods to explain tourists' inflow proportionally based on source-country income level and inversely with higher travelling costs, inflation and unfavourable exchange rates (Crouch, 1994; Lim, 1997; Song & Li, 2008).

Building on this intuition, panel and gravity-type models were increasingly used to explain tourism flows as functions of origin-country income, relative prices between origin and destination, and additional supply-side attributes such as purpose of visit and nationality (Cortés-Jiménez & Blake, 2011). The large-scale availability of panel data, especially in Western developed nations, facilitated the use of fixed- and random-effect models and to capture the heterogeneity across destinations and time dynamics (Dogan & Zhang, 2023; Song & Li, 2008).

However, these panel-data methods deploy straightforward linear approaches, unless specified exclusively, to capture the simple linear effects of the predictors on tourists' inflow. They fail to capture the nonlinear threshold-driven relationships that characterise tourism demand in complex real-world scenarios. It has been empirically found that visitors' reactions to economic factors, such as income growth, infrastructural development and other destination characteristics, often show saturation effects and diminishing marginal returns with respect to certain key threshold levels (Guan et al., 2022; Saayman & Botha, 2017; Wu et al., 2022).

Mega-events like international festivals and global sports events certainly provide a temporary surge in international tourist arrivals in the host country. However, long-term impacts depend on factors such as infrastructure readiness, governance quality, and long-term destination management (Black & Van Der Westhuizen, 2004; Gursoy et al., 2017; Matheson, 2006).

These nonlinear and threshold effects can be captured directly by linear models through polynomial and interaction terms. But this approach faces a severe multicollinearity problem, which eventually leads to unstable coefficient estimates (Wooldridge, 2010). Multicollinearity is particularly problematic in tourism demand modelling since macroeconomic variables such as income, prices, and infrastructure often interact in complex ways (Song et al., 2023; Wasiaturrahma & Rohmawati, 2021).

Tree-based machine learning and penalised regression techniques can stabilise coefficient estimates and enhance out-of-sample prediction. In fact, they have been widely adopted in economic modelling but remain underutilised in tourism demand studies in emerging economies (Hastie et al., 2009; James et al., 2013). This methodological gap limits the robustness and applicability of many empirical findings.

India is no exception as empirical research on tourism demand has focused primarily on classical econometric and time-series approaches. In recent times, Shah et al. (2022) used the augmented gravity-based model on a panel fixed-effect setup and showed that the importance of origin-country income, exchange rates, distance, and cultural proximity is shaping inbound tourism. In another state-level panel analysis, a similar set of features, such as infrastructure, accessibility, and amenities, significantly affected foreign tourist arrivals but with greater heterogeneity among the states (Suresh et al., 2015). Time-series and cointegration methods elucidate both short-run and long-run impacts of macroeconomic variables on tourism demand (Devi et al., 2024; Sharma et al., 2023). Using linear model frameworks, these studies provide valuable insights, but none of these works focus on the more complex relationships and threshold values. This indicates both the availabil-

ity of rich panel data and the methodological opportunities for more robust and flexible modelling of tourism demand in India. In fact, recent research highlights the growing use of artificial intelligence and machine learning methods for predicting tourism inflow and hospitality-sector revenue guidance (Henriques & Pereira, 2024). Recent research also highlights the importance of digital infrastructure and ICT in tourism demand (Gössling, 2021; Xia et al., 2024).

Finally, over the last decade, a growing number of researchers have identified the importance of governance and administrative quality as an important driver of tourism demand. Seminal works by Hall (2011) and Bramwell and Lane (2011) mention the importance of collaborative and multi-level governance in developing the resilient tourism sector. Later, this was empirically observed both across OECD countries and emerging countries (OECD, 2022; Rasoolimanesh et al., 2024). These works underscore how governance, infrastructure investment, and digital transformation together can strengthen the resilience and competitiveness of the tourism sector.

After synthesising, we identify the following three key gaps:

- Limited nonlinear and interaction modelling of economic, infrastructural, mega-events, and governance quality in shaping tourism demand in India;
- Underuse of machine learning methods for resolving highly correlated multi-dimensional modelling in tourism demand;
- Integration of governance and institutional quality alongside traditional drivers in the policy-making context.

This study addresses these gaps through a nonlinear interpretable machine learning framework to analyse India's foreign tourist arrivals quantitatively and comparatively.

2. Materials and Methods

This study conceptualises foreign tourist arrivals as outcomes of multifaceted determinants spanning economic, infrastructural, cultural, event, and institutional domains. The framework is built on the gravity-based tourism-demand theory (Morley et al., 2014). Among the economic variables, per capita income in origin countries primarily determines travel affordability and demand. However, its marginal returns diminish beyond certain thresholds due to market saturation and diversification (Nguyen, 2021). In a fully specified gravity model, prices would ideally be represented by bilateral relative prices or real exchange rates between each source country and India. In a gravity model, prices would ideally be represented by real CPI or real exchange rates (Morley et al., 2014). However, due to a lack of reliable and consistent all-source-country CPI, we have treated India's consumer price inflation as a destination-side proxy for relative prices.

From the infrastructure side, hosting tourists in quality hotels and offering end-to-end transportation are the most important variables to predict tourism demand. We have proxied these two variables by the total number of hotel rooms at the macro level and aircraft flown both internationally and domestically. Cultural and heritage assets such as UNESCO sites, festivals, and historical landmarks act as a stimulus to attract foreign tourists. India has a rich heritage of culture and geographical landscapes to count on. Although these assets generate destination uniqueness, they are subject to the risk of overcrowding effects, which may diminish marginal benefits at high levels (Kordej-De Villa & Šulc, 2021; Kutlu et al., 2024). We have considered UNESCO World Heritage Sites due to continuous additions of sites over the years and a high global ranking (current 6th).

India did host several mega-events over the years, including sports, political and religious events. All these mega-events are considered both in terms of short-term and event-lag-wise. Finally, the recent literature also strongly considers the institutional quality in terms of governance, safety, and regulatory efficiency. We have conceptualised that gov-

ernance quality will have an impact on the international tourism inflow. Conceptually, we hypothesise nonlinear effects (diminishing returns; thresholds) and synergistic interactions.

The dataset is structured as panel data indexed by source country c (61 countries) over years t (from 2002 to 2024). While the dependent variable (tourist inflow) varies by country-year, several key regressors (e.g., infrastructure, heritage assets, CPI growth, institutional quality, and events) vary only by year and apply uniformly across source countries. The dependent variable is log-transformed for stabilising variance and interpreting coefficients as elasticities.

The log-transformed real Gross Domestic Product (GDP) per capita of the source country (in constant 2015 US dollars) was taken as a demand-side proxy for tourism inflow to India. Due to data limitations, the price competitiveness is proxied by India's Consumer Price Index (CPI), which acts as an indicator of inflation in the destination market. As mentioned earlier, the destination infrastructure is captured by two proxy variables—the total number of classified and approved hotel rooms available at the macro level and the counts of both international and domestic aircraft movements. Cultural appeal is measured by the number of UNESCO heritage sites.

To account for temporary shocks of mega-events, we have considered a binary variable indicating major event hosting (such as the Commonwealth Games, Cricket World Cup, or G20 Summit) in a year. We have also considered a lagged event indicator to capture the persistent impacts of the event beyond the year of occurrence.

Institutional quality is measured by three governance metrics—Rule of Law (RL), Political Stability (PV), and Government Effectiveness (GE). These variables capture legal, political, and administrative dimensions that may influence international travellers' perceptions and decisions about the destination. A simple average of these three variables ($pvgerl$) was used in the study. All this information was gathered from the Reserve Bank of India database, OECD database and World Bank's Worldwide Governance Indicators.

Building on the gravity-based tourism demand model, foreign tourist arrivals from origin i to India in year t can be viewed as

$$\ln(FTA_{ct}) = \beta_0 + \beta_1 \ln(GDP_{PCct}) + \beta_2 CPI_t + \beta_3 \ln(Rooms_t) + \beta_4 \ln(Airdom_t) + \beta_5 \ln(Airint_t) + \beta_6 \ln(UNESCO_t) + \beta_7 (Event_t) + \beta_8 (Event_{t-1}) + \beta_9 \ln(pvgerl_t) + \mu_c + \epsilon_{ct}$$

The variables used in the above gravity-based tourism demand equation, along with their definitions, measurement units, and sources, are summarised in Table 1.

We initially employed classical panel data regression models, including pooled ordinary least squares (OLS), fixed-effect (FE), and random-effect (RE) estimations. With the objective to account for nonlinearities and interaction effects, we extended the baseline linear model by incorporating polynomial terms and interaction variables. Model selection tests (F-test, Breusch–Pagan LM test, and Hausman test) indicated that the fixed-effect model was the preferred specification given the unobserved country-specific heterogeneity.

However, the addition of polynomial and interaction terms introduced substantial multicollinearity and resulted in unstable coefficient estimates and inflated standard errors. To mitigate this problem, we have tried applying all types of regularisation (L1, L2 and elastic net). However, due to strong multicollinearity, none of these regularisation methods produced the optimal desired results, including nonlinear and interaction terms. Due to space constraints, these results are not included in the main manuscript but are available from the authors on request.

Table 1. Definitions and aspects of variables.

Variable	Description	Data Source	Unit	Transformation
$\ln(FTA_{ct})$	Foreign tourist arrivals to India from the source country c in year t	Ministry of Tourism, India; UNWTO	Number of persons	Natural logarithm
$\ln(GDP_{PCct})$	Real GDP per capita of source country c (2015 USD)	World Bank	USD	Natural logarithm
CPI_t	Consumer Price Index growth rate in India	Reserve Bank of India	Percentage	Level
$\ln(Rooms_t)$	Total classified and approved hotel rooms in India	Ministry of Tourism, India	Number of rooms	Natural logarithm
$\ln(Airint_t)$	International aircraft movements in India	Ministry of Civil Aviation	Number of flights	Natural logarithm
$\ln(Airdom_t)$	Domestic aircraft movements in India	Ministry of Civil Aviation	Number of flights	Natural logarithm
$\ln(UNESCO_t)$	Number of UNESCO World Heritage Sites in India	UNESCO World Heritage Centre	Count	Natural logarithm
$Event_t$	Indicator for hosting a major international event	Author compilation	Dummy (0/1)	Level
$Event_{t-1}$	One-year lag of the mega-event indicator	Author compilation	Dummy (0/1)	Level
$\ln(pvgerl_t)$	Average of Political Stability, Government Effectiveness, and Rule of Law indices	World Bank (WGI)	Standardised index	Level

Source: author's compilation based on secondary data.

Consequently, this study employed Extreme Gradient Boosting (XGBoost), a machine learning ensemble method capable of intrinsically modelling complex nonlinearities and interactions without requiring pre-specification of polynomial terms or interaction variables. The model uses a squared-error loss function with a learning rate of 0.1, a maximum tree depth of eight, subsampling and column-subsampling rates of 0.8, four parallel threads, and a fixed random seed of 123, allowing up to 500 boosting iterations with early stopping after 20 rounds without improvement in root mean squared error. To respect the panel structure and avoid temporal leakage, custom time–entity cross-validation folds are constructed by sorting each source country's observations by year and splitting them into three contiguous blocks. These are then rotated as training and validation sets. This time-aware cross-validation procedure ensures that reported performance reflects genuine out-of-sample prediction across countries and years and that the accumulated local effect (ALE) plot-based interpretation is based on a fully specified and reproducible XGBoost training design. Shapley additive explanations (SHAP) values and ALE plots were used to achieve model interpretation.

3. Results

3.1. Summary Statistics of Key Variables

The key variables used in the analysis include both source-country-level time-variant variables and India-level time-variant and time-invariant aggregates. Table 2 presents descriptive statistics for the variables at the macro level.

Table 2. Statistical description of the variables included in the model.

Variable	Mean	Median	Std. Dev.	Min	Max	Nature
Foreign tourist arrivals	100,341	34,360	218,948	530	2,577,727	Source country–year variant
GDP per capita	22,325	12,979	21,433	338	98,776	Source country–year variant
CPI growth	0.06	0.05	0.02	0.03	0.11	India-level time variant
Hotel rooms	99,045	99,328	6976	86,106	117,815	India-level time variant
Heritage sites	33.52	32	6.53	23	45	India-level time variant
PVGERL	0.01	−0.02	0.09	−0.11	0.19	India-level time variant
Aircraft movement (domestic)	662.69	542.1	210.89	490.6	1063.1	India-level time variant
Aircraft movement (international)	105.93	97.6	45.11	47.2	208.5	India-level time variant

Source: author’s estimates based on secondary data.

The foreign tourist arrivals average 100,341 per source country–year, with a median of 34,360 and a range extending from 530 to over 2.5 million arrivals. This clearly reflects the heterogeneity in tourism flows across diverse source countries and over time. Time-similar heterogeneity is observed in the economic conditions of the source countries, where the GDP per capita of the source countries shows substantial variations both among the countries and over time. With an average inflation rate of 6%, the data shows mild inflation, which in turn proves to be modestly price-competitive. A mean accommodation capacity of 99,045 rooms, with modest variability, reflects steady growth over time. There is a wide gap in aircraft movement between the domestic and international levels. In recent years, due to the Government of India’s flagship initiatives (UDAN, launched in 2016), there has been better growth in domestic aircraft movement compared with international movement. The World Bank’s World Governance Indicators (in standardised form) hover around zero with minor fluctuations. This suggests relative governance stability over the years. Thus, there are wide variations both on the demand and supply sides of international tourist inflow to India.

3.2. Trend of Foreign Tourists Arriving in India

Figure 1 depicts the inbound foreign tourist arrivals to India from multiple source countries over the period 2002–2024. We have standardised the inbound tourist inflow from the respective source country to compare the growth dynamics, ignoring the volume of arrivals. The plot reveals three dominant features: (1) a general upward trend in arrivals for most countries from 2002 to 2019; (2) a sharp and synchronised contraction in 2020–2021 due to the COVID-19 pandemic; and (3) a marked post-pandemic rebound starting in 2022. This pattern indicates the resilience of international tourism demand and persistent cross-country heterogeneity in shock exposure and recovery.

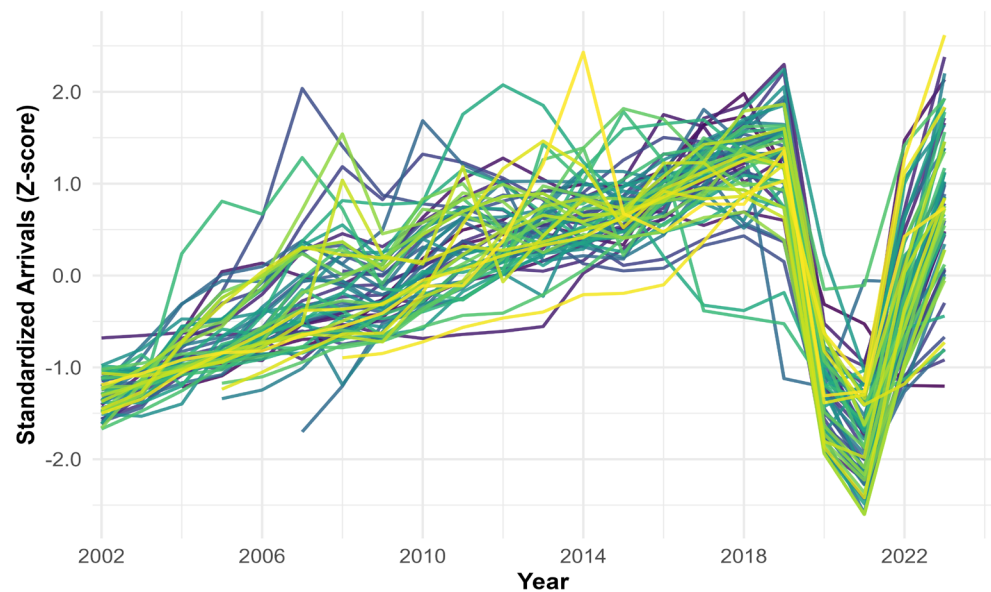


Figure 1. Tourist arrivals over time in India. Different colours is used for different source countries. There are 61 line (no of source countries). Source: author's estimates based on secondary data.

3.3. XGBoost Model Performance

An XGBoost regression model is estimated to predict log (foreign tourist arrivals) with origin-year observations linked to a set of already specified predictors. Among the regressors, source-country income is log-transformed along with the destination country's hotel rooms, heritage sites, and aircraft flown domestically and internationally. Other regressors (event, event count, CPI growth and pvgerl) are kept at their level form. We have dropped the lagged values of the tourist inflow to have an autoregressive-temporal-effect-free estimation for checking the real predictive power of the regressors. Even without the lagged dependent variable, the model performed well—an out-of-sample root mean squared error (RMSE) of approximately 0.752 and an R-squared (R^2) of 0.59. So, although the lagged outcomes contribute to improved fit, the model robustly explains variations in tourist arrivals through economic, infrastructure, event, and governance features alone.

3.4. SHAP Summary Plot: Global Feature Importance and Value Effects

To further explore the pattern through which key regressors influence foreign tourist arrivals, we leveraged Shapley Additive explanation (SHAP) values to assess the feature values' impact on model output. The SHAP summary plot in Figure 2 globally ranks all predictors' contributions to the model's predictive output for inbound tourism. The figure shows not only which drivers are most influential but also how the features modulate their predictive power on tourism inflow. The SHAP plot uses a colour gradient from blue to red ("low-high") to reveal not just the rank of importance of the regressors but also the directionality and conditionality of each feature's effect. Features that move from blue (low value; negative effect) to red (high value; positive effect) are policy levers with direct, scalable benefits. Conversely, negative movement of high values (e.g., inflation) highlights risk points and the necessity for economic stability. Air traffic movement (both international and domestic) is the next most vital determinant. This affirms that reductions in spatial frictions (lowering time, cost, and uncertainty of travel) robustly expand India's effective tourist catchment area. The positive SHAP value range for both international aircraft count ("Aircraft flown intl") and domestic count ("Aircraft flow dms") empirically supports the hypothesis that boosting international air connectivity and robust domestic networks generates nontrivial systemic gains in foreign tourist arrivals in India.

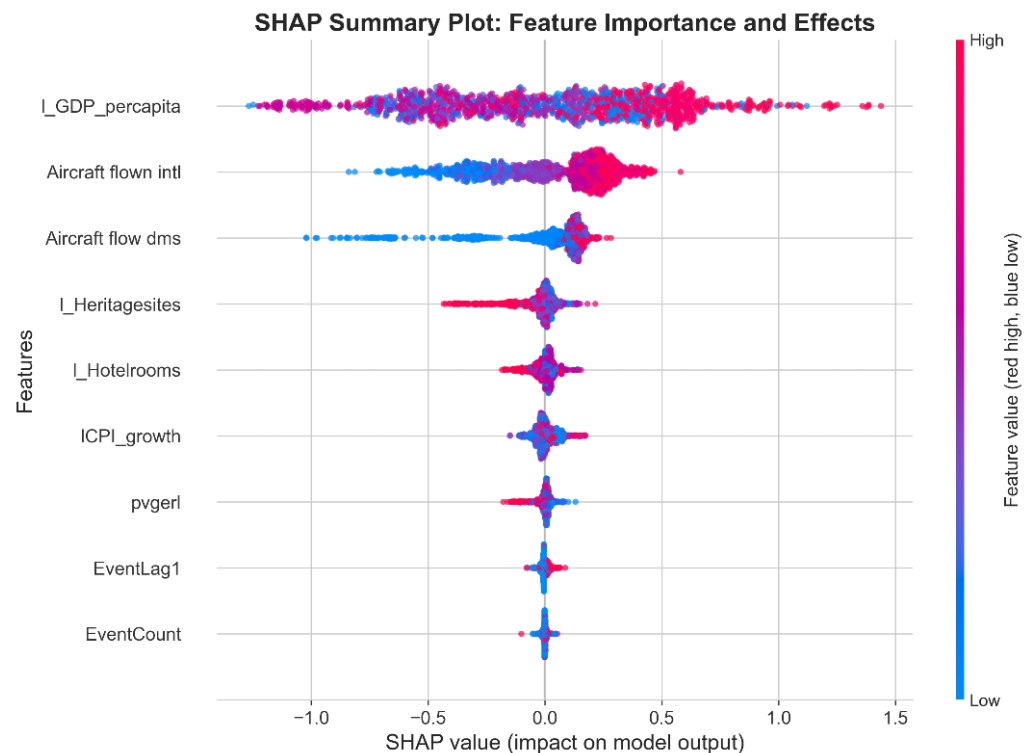


Figure 2. SHAP summary plot: feature importance and effects. Source: Author's estimates based on secondary data.

The prominence of log heritage sites reinforces the growing literature on differentiation and experience-driven tourism. However, most of the SHAP values are clustered near zero on the positive side. This implies that the gains in tourism inflow are limited compared to the earlier two important regressors. The same is true for log of hotel rooms. More hotel rooms to accommodate more tourists boosts potential inflow; however, the effect size is limited compared to income and civil aviation infrastructure impact. The SHAP values for governance (pvgerl) are tightly clustered near zero, slightly on the positive side, implying that governance has a modest direct marginal effect on tourist inflows—but not at the scale of income and aviation infrastructure.

The SHAP values of CPI growth are clustered around zero, with most of them skewed towards negative values. This essentially implies that higher CPI growth values show a negative SHAP association, consistent with price sensitivity in model inbound tourism prediction. Finally, the near-zero clustered SHAP values of mega-events and their lag impact indicate that hosting a major event produces at best a small and inconsistent short-run boost to tourist inflows, with no strong systematic impacts over the subsequent time periods.

3.5. ALE Plots: Feature Interactions and Economic Intuition

One limitation of SHAP values is their sensitivity to collinearity among regressors. As our model exhibits strong collinearity, we apply an ALE plot to get unbiased model-consistent interpretations under strong multicollinearity and nonlinear interactions. Figure 3 presents one-dimensional accumulated local effect (ALE) plots for all the key determinants of international tourists' inflow to India. The ALE profile for the log per capita GDP of the source countries displays a nonlinear and partially cyclical trajectory. At lower levels of per capita GDP of source countries, an increase generates strong positive marginal gains in predicted tourist arrivals to the destination. However, this momentum is lost substantially beyond a mid-income threshold; it exhibits brief reversals and

risers again at the highest income levels. There can be multiple reasons for this. Initially, the outbound tourism market matures at a certain income level, and then, at a further higher income level, outbound travel diversifies and thereby distributes the demand across multiple destinations.

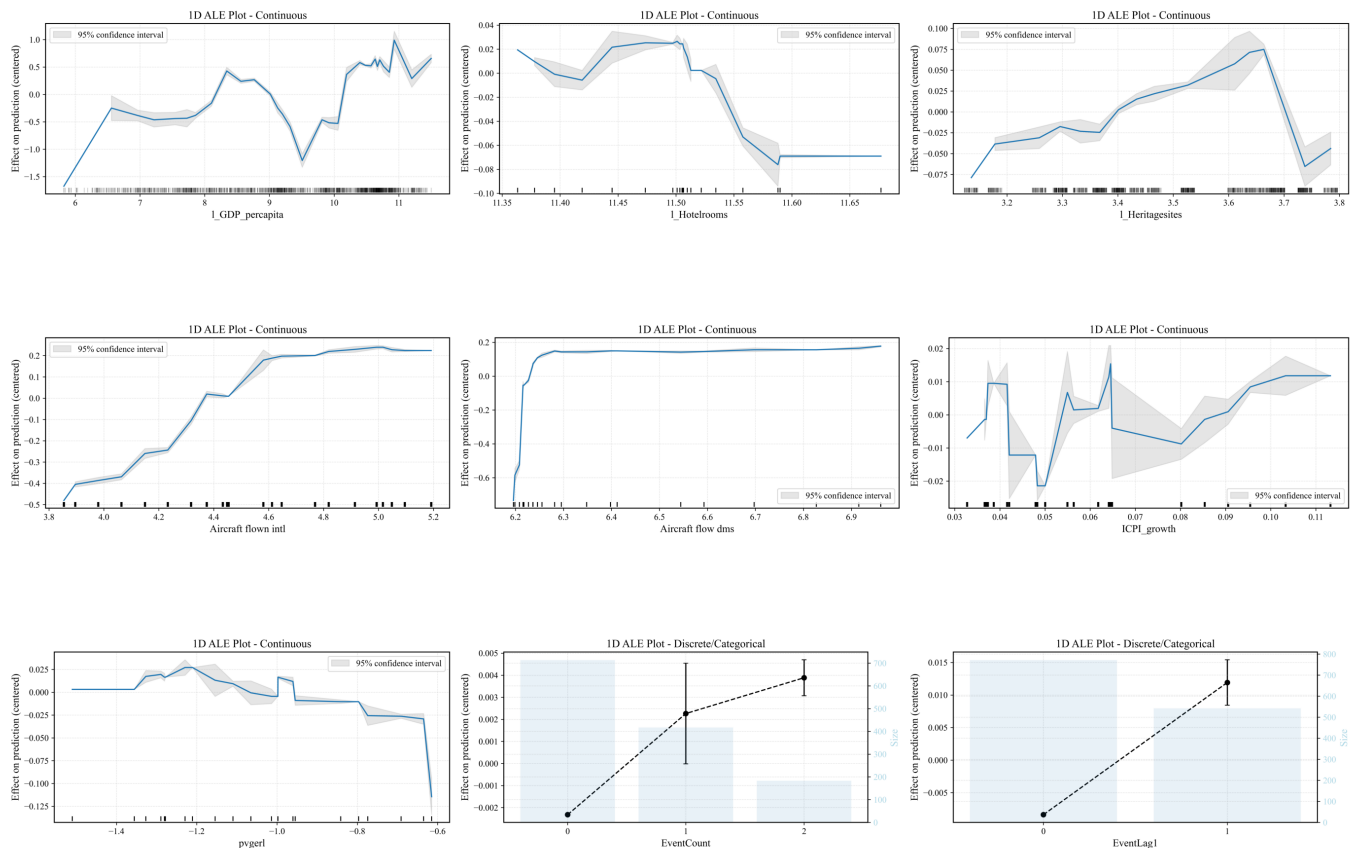


Figure 3. One-dimensional ALE plot of top features. Source: author's estimates based on secondary data.

The ALE curve for the log of heritage sites shows a strictly concave and diminishing return pattern. In the beginning, developing a few heritage sites gives a huge boost to tourism inflow. But beyond the threshold level, the addition of new sites brings in fewer extra visitors. This clearly indicates that India already ranks 6th in the UNESCO heritage list and has hit the maximum carrying capacity. The ALE plot shows diminishing then flat effects beyond the 32 heritage sites, suggesting saturation in this model's learned patterns. Now what matters is managing the service quality at the heritage sites and their worldwide marketing to keep tourists coming back.

Among the infrastructure-related regressors, the log of hotel rooms clearly indicates that India has reached the threshold level. When the number of hotel rooms was low, any addition to it made tourists stay and yielded a positive impact. As the hotel rooms increased further, the extra rooms helped, but their incremental benefits became weaker, and eventually they started to fall. In this situation of saturation, what really matters is the expansion of air connectivity, which emerges as the second most important driver behind the income of the source countries. The ALE plots of both intentional and domestic air connectivity have shown steep rising slopes and then flattening at high levels, without any signs of south direction. This is indicative of a positive impact but with diminishing returns—no signs of downward impact at higher levels.

Institutional quality exhibits a nonlinear pattern and eventually a negative relationship with the inflow of foreign tourists. At low to moderate values of institutional quality, the effect on the model prediction was low to slightly positive, which implies that a very small incremental gain is associated with it in terms of attracting foreign visitors. After a mid-threshold level, the ALE curve for institutional quality moves slowly in a downward direction. This is possibly due to India's relatively low score in institutional quality, with limited variation and much behind the OECD benchmark level. Due to this, the marginal effects of small changes in governance quality may not necessarily reflect international tourism gains. Finally, the inflationary situation in India has a mild negative impact on inbound tourism.

Overall, the ALE plots show that, along with source countries' economic development, India's rich collection of heritage sites and growing air connectivity are pulling foreign visitors in a nonlinear fashion, while the impacts of domestic inflation and institutional quality are mild to negative. Also, most of the key drivers, except air connectivity, have shown a saturated status.

3.6. Finding Feature Interactions

We have utilised the pairwise SHAP values for finding possible interactions among the features, and the top interacting features based on their aggregate mean absolute values are reported in Table 3 below.

Table 3. SHAP-based feature interaction strength.

Feature 1	Feature 2	Interaction Strength
l_GDP_percapita	Aircraft flown intl	0.058254
l_Heritagesites	l_GDP_percapita	0.029339
l_GDP_percapita	Aircraft flow DMS	0.023133
l_GDP_percapita	ICPI_growth	0.017775
l_Heritagesites	Aircraft flow dms	0.017162
l_Heritagesites	Aircraft flown intl	0.015683
Aircraft flown intl	Aircraft flow dms	0.015478
l_Hotelrooms	l_GDP_percapita	0.014072
ICPI_growth	Aircraft flown intl	0.01375
l_GDP_percapita	pvgerl	0.011266

Source: author's estimates on the basis of secondary data.

The magnitudes of interactions clearly show that the interactions are mostly anchored around the source countries' economic affluence (l_GDP_percapita). This confirms that the economic affluence of the source countries does not operate in isolation; rather, it amplifies the tourism inflow to India through interacting with supply-side factors. Taking the top four pairwise interaction factors anchoring around the economic affluence of sources, we further estimate the interaction mechanism using a 2D ALE plot, as shown in Figure 4 below:

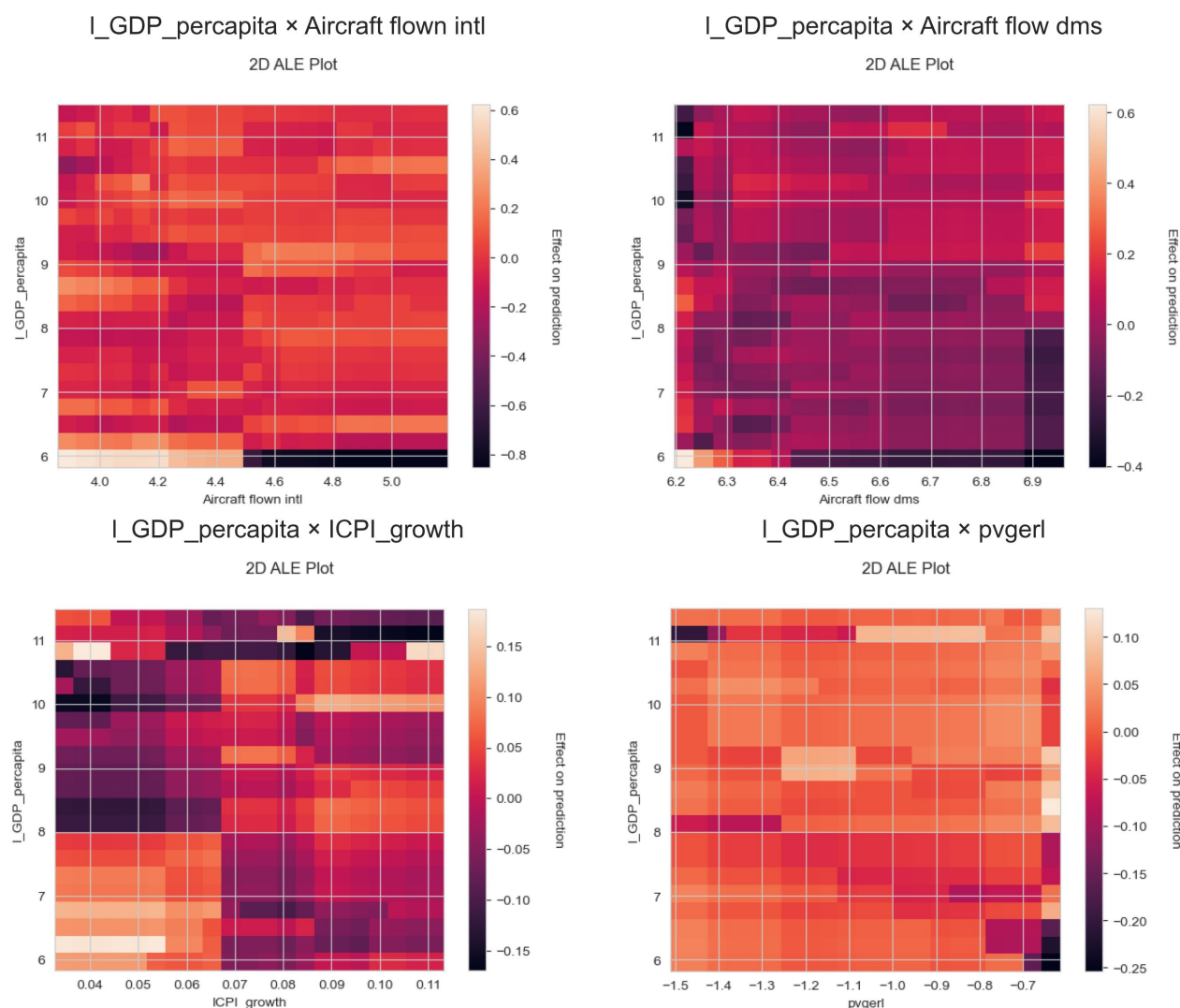


Figure 4. Two-dimensional ALE plots for feature interaction. Source: author's estimates on the basis of secondary data.

To understand the 2D ALE plot, we must know the colour gradient. The colour gradient reflects how different combinations of the two features increase (lighter colours) or decrease (darker colours) the model's prediction relative to its average. The strongest interactions are between source countries' economic affluence and international air connectivity to India. The 2D ALE plot clearly shows that international connectivity amplifies foreign tourists' inflow to India from middle-income countries. At the highest income level and highest international connectivity, the tourist inflow to India diminishes. This is probably due to saturation of demand or the most affluent foreign tourists diversifying their travel patterns to other destinations. The same is true for domestic air connectivity as it strengthens foreign tourists up to a certain threshold level and, after that, serves internal mobility. The interaction between the economic affluence of the source country and India's inflation reveals a nonlinear price elasticity of foreign tourists' demand. Tourists from low- and middle-income countries show high sensitivity to India's domestic inflation. On the other hand, tourists from high-income countries show muted and heterogeneous responses due to their relatively high tolerance of higher inflation and destination diversification at higher income levels. However, this was not sufficient to compensate for the loss arising from high-price-sensitive tourists from low- and middle-income countries. Finally, the interaction plot of institutional quality in India and source countries' economic affluence shows

that economically rich tourists are more sensitive to institutional quality than budget-driven low- and middle-income foreign tourists.

3.7. SHAP-Based Elasticities: Translating Model Interpretability into Economic Magnitudes

To provide policymakers and researchers with actionable insights, we translate the SHAP values (marginal effects) into pseudo-elasticities for estimating the percentage change in foreign tourist arrivals associated with a one percent change in key predictors. Table 4 presents these pseudo-elasticities along with 95% bootstrap confidence intervals.

Table 4. SHAP-based pseudo-elasticity estimate.

Feature	Absolute Elasticity Magnitude	Absolute Elasticity Magnitude (95% CI)	Signed Elasticity	Signed Elasticity (95% CI)
ICPI_growth	0.495	[0.029, 0.032]	0.003	[0.000, 0.005]
l_GDP_percapita	0.047	[0.419, 0.449]	−0.004	[−0.032, 0.025]
EventLag1	0.027	[0.026, 0.028]	0.001	[−0.001, 0.003]
pvgerl	−0.016	[−0.017, −0.015]	0	[−0.001, 0.002]
l_Heritatesites	0.014	[0.047, 0.054]	−0.024	[−0.029, −0.020]
EventCount	0.008	[0.008, 0.009]	−0.001	[−0.002, −0.000]
l_Hotelrooms	0.003	[0.030, 0.033]	0	[−0.003, 0.002]
Aircraft flown intl	0.002	[0.002, 0.002]	0	[−0.000, 0.000]
Aircraft flow dms	0	[0.000, 0.000]	0	[0.000, 0.000]

Source: author's estimates on the basis of secondary data.

We have estimated two complementary SHAP-based pseudo-elasticity measures. First, we compute the average absolute SHAP values to measure the impact magnitude of the predictors on foreign tourists' arrivals to India. Secondly, the signed elasticities are estimated from the average SHAP values for the continuous predictors and for level predictors, SHAP values by the predictor mean. These second-class estimates provide the net average direction of the marginal effect over the sample range.

Based on the mean absolute elasticity, inflation in India, source countries' per capita income, India's heritage sites, and hosting of mega-events by India exert the largest average influence on tourists' arrivals in India. The signed elasticities confirm the visual plots—the SHAP summary plot and 1D ALE plot. Adding more heritage sites exhibits a small but statistically significant negative elasticity. This further confirms our earlier observation in the 1D ALE plot.

Inflation in India shows a very small positive elasticity (with a narrow confidence interval). So, this implies a mild positive SHAP contribution towards tourist arrivals in India when inflation is at a moderate level. However, when inflation reaches a higher level, it will certainly have a negative impact on tourists' arrivals in India. For hosting mega-events and air connectivity infrastructure, the signed elasticities are estimated to be near zero despite the substantial absolute mean SHAP elasticities. This is due to the fact that their marginal elasticities are positive in some value ranges and negative in others, which eventually cancel each other out.

4. Discussion

The log GDP per capita of the source country stands out as the overwhelmingly dominant factor. This validates the economic gravity model expectations and the central factor of international tourism demand, that is, travellers' ability and willingness to pay increases with rising affluence (see [Song & Li, 2008](#); [Stabler et al., 2009](#)). The horizontal

spread and concentration of high-value (red) points on the positive SHAP side imply disproportionately strong marginal impacts at higher income levels. Tourism being a luxury good, rising income levels boost tourism sentiments, reflecting a positive income elasticity since both are measured on a logarithmic scale.

Air traffic movement, both international and domestic, ranks as the next most vital determinant after source-country income. This aligns with prior findings that enhanced air connectivity reduces spatial frictions and expands destination accessibility, as seen in panel analyses of Asian tourism flows (Shah et al., 2022). The positive SHAP range for both aircraft counts supports systematic gains from international links and domestic networks. However, the ALE plots reveal saturated returns at high levels. This is a contrasting finding from the linear assumptions in classical gravity models (Crouch, 1994).

Accommodation infrastructure (hotel rooms) shows limited positive SHAP clustering around zero. This is also indicative of saturation effects beyond threshold values. Regarding the growing networks of hotel accommodation, we differ from the early infrastructure studies assuming constant returns (Suresh et al., 2015). In the current study, hotel capacity reaches a plateau where further expansion yields falling benefits.

Similarly, heritage assets also show a saturation condition (limited positive SHAP clustering near zero). The concave ALE curve for UNESCO sites confirms diminishing marginal returns for India (sixth globally), echoing overcrowding risks in heritage tourism (Kordej-De Villa & Šulc, 2021; Kutlu et al., 2024).

Governance quality (pvgerl) exhibits modest positive but near-zero SHAP values, with ALE turning negative post-threshold—suggesting limited appeal at India's scores. This partially supports governance as a resilience factor (Hall, 2011; Rasoolimanesh et al., 2024) but highlights its secondary role versus economic drivers, unlike the OECD emphasis (OECD, 2022). Inflation (CPI growth) skews negative in SHAP, with a mild overall impact, aligning with price sensitivity in emerging markets (Nguyen, 2021) yet showing nonlinearity that is absent in linear panels (Song et al., 2023). Mega-events yield near-zero SHAP effects, which is clearly indicative of inconsistent short-run boosts without long-term persistence—tempering optimistic legacy views (Gursoy et al., 2017; Matheson, 2006) and underscoring infrastructure preconditions (Black & Van Der Westhuizen, 2004).

Methodologically, the setup of XGBoost plus SHAP/ALE resolves both multicollinearity (arising from the inclusion of polynomial extensions of fixed-effect models) (Wooldridge, 2010; Wasiaturrehman & Rohmawati, 2021) and captures nonlinearities and interaction effects (e.g., income–air connectivity) better than traditional linear econometric approaches (Saayman & Botha, 2017; James et al., 2013). Interactions centred on source affluence amplify supply-side effects for middle-income markets, informing targeted policies over broad strategies.

The current study's XGBoost-SHAP/ALE framework confirms the foundational gravity findings, with the income of source countries and India's air connectivity predominantly predicting India's inbound tourism demand. The study also reveals key novelties in terms of nonlinear thresholds absent in linear panels. The key novelties include heritage/hotel saturation, prioritising upgrades over expansion (contralinear positives; aligns qualitatively with overcrowding risks) and air connectivity interaction with mid-income GDP per capita, suggesting untapped emerging markets beyond proximity effects. Mild CPI effects validate the price-stability requirement, but nonlinearity refines targeting. Low governance scores and limited variation are possible reasons for not having clear predictive power in tourism demand modelling. However, the potential endogeneity between tourist arrivals and supply variables (air transport and hotels) remains unaddressed; future work should employ instrumental variables or causal ML approaches.

5. Conclusions

This study uses interpretable machine learning methods to provide a data-driven perspective on the determinants of foreign tourist arrivals to India. It tries to reveal the interplay between economic, infrastructure, governance, and event-driven factors for India's tourism demand in the global market. The empirical findings show that the economic affluence of source countries is the single most influential driver of arrivals. However, this alone cannot determine demand. It is an interaction between domestic inflation, connectivity and quality of governance. Also, in most cases, the predictors show diminishing marginal returns, indicating market maturity levels. This convex response indicates the continuing necessity of tailoring marketing and market development efforts towards emerging middle-income segments, which seems to be more promising for India since highly affluent foreign tourists are diversifying their destination portfolios.

Air connectivity emerges as a crucial infrastructure pillar. Both international and domestic aircraft movements feature among the top predictors. Both reveal strong and mostly linear positive relationships with inbound tourism, with no sign of downward movement. So, direct international links with more middle-income foreign cities to Indian airports and systemically bridging the domestic flight network will help in the greater dispersal of arrivals.

The study also reaffirms the power of cultural and heritage assets. However, due to saturation, any further addition to the list will not provide a substantial gain to India's tourism demand. Rather, the maintenance of the sites' service quality and effective world-wide marketing to the specific target audience will be strategically more crucial. The same strategy applies to another infrastructure ingredient, that is, hotel rooms, where a clear saturation state is visible. Thus, instead of aggressive expansion of hotel networks, the focus should be on qualitative improvements.

Institutional quality exhibits a nonlinear pattern and eventually a negative relationship with the inflow of foreign tourists. There is ample room for improvement in this field. As India scores at relatively lower levels, only significant jumps in institutional quality will attract sensitive foreign tourists. Finally, the inflationary situation in India has a mild negative impact on inbound tourism, especially for low-middle-income foreign tourists, due to increased difficulty in economic affordability. However, as long as India's long-term inflation remains under the RBI's suggestive band, there is no significant worry from this side.

Finally, event strategies are empirically validated for their enduring tourism stimulus. Mega-events not only provide immediate influxes but also produce persistent gains in subsequent years. However, finding support for small impacts on tourism demand, judicious selection, preparation, and legacy planning for major events is of utmost importance. Such mega-events must be aligned with sustainable practices and destination branding goals.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The entire manuscript was conceptualised, analysed and written by the author. However, GenAI (Grammarly 14.1271.0) has been used for final proofreading and fine-tuning of language.

Conflicts of Interest: The author declares no conflicts of interest.

References

- Anholt, S. (2007). *Competitive identity: The new brand management for nations, cities and regions*. Palgrave Macmillan.
- Ashley, C., De Brine, P., Lehr, A., & Wilde, H. (2007). *The role of the tourism sector in expanding economic opportunity (Corporate social responsibility initiative)*. Harvard Kennedy School. Available online: <https://ashleyinsight.co.uk/wp-content/uploads/2020/04/tourism-in-expanding-economic-opportunity.pdf> (accessed on 29 November 2025).
- Black, D., & Van Der Westhuizen, J. (2004). The allure of global games for “semi-peripheral” polities and spaces: A research agenda. *Third World Quarterly*, 25(7), 1195–1214. [\[CrossRef\]](#)
- Bramwell, B., & Lane, B. (2011). Critical research on the governance of tourism and sustainability. *Journal of Sustainable Tourism*, 19(4–5), 411–421. [\[CrossRef\]](#)
- Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism Management*, 29(4), 609–623. [\[CrossRef\]](#)
- Cortés-Jiménez, I., & Blake, A. (2011). Tourism demand modeling by purpose of visit and nationality. *Journal of Travel Research*, 50(4), 408–416. [\[CrossRef\]](#)
- Crouch, G. I. (1994). The study of international tourism demand: A survey of practice. *Journal of Travel Research*, 32(4), 41–55. [\[CrossRef\]](#)
- Devi, R., Agrawal, A., Dhar, J., & Misra, A. K. (2024). Forecasting of Indian tourism industry using modeling approach. *MethodsX*, 12, 102723. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dogan, E., & Zhang, X. (2023). A nonparametric panel data model for examining the contribution of tourism to economic growth. *Economic Modelling*, 128, 106487. [\[CrossRef\]](#)
- Dwyer, L., Forsyth, P., & Spurr, R. (2004). Evaluating tourism’s economic effects: New and old approaches. *Tourism Management*, 25(3), 307–317. [\[CrossRef\]](#)
- Government of India. (2020). *India tourism statistics (Tourism statistics)*. Ministry of Tourism, Government of India.
- Gössling, S. (2021). Tourism, technology and ICT: A critical review of affordances and concessions. *Journal of Sustainable Tourism*, 29(5), 733–750. [\[CrossRef\]](#)
- Gössling, S., & Hall, C. M. (2006). Uncertainties in predicting tourist flows under scenarios of climate change. *Climatic Change*, 79(3–4), 163–173. [\[CrossRef\]](#)
- Guan, B., Silva, E. S., Hassani, H., & Heravi, S. (2022). Forecasting tourism growth with state-dependent models. *Annals of Tourism Research*, 94, 103385. [\[CrossRef\]](#)
- Gursoy, D., Yolal, M., Ribeiro, M. A., & Panosso Netto, A. (2017). Impact of trust on local residents’ mega-event perceptions and their support. *Journal of Travel Research*, 56(3), 393–406. [\[CrossRef\]](#)
- Hall, C. M. (2011). A typology of governance and its implications for tourism policy analysis. *Journal of Sustainable Tourism*, 19(4–5), 437–457. [\[CrossRef\]](#)
- Hall, C. M., Gössling, S., & Scott, D. (2011). Tourism and sustainability: An introduction. In *The routledge handbook of tourism and sustainability*. Routledge.
- Hastie, T., Tibshirani, R., & Friedman, F. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Henriques, H., & Pereira, L. N. (2024). Hotel demand forecasting models and methods using artificial intelligence: A systematic literature review. *Tourism & Management Studies*, 20(3), 39–51. [\[CrossRef\]](#)
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.
- Kordej-De Villa, Ž., & Šulc, I. (2021). Cultural heritage, tourism and the UN sustainable development goals: The case of Croatia. In M. B. Andreucci, A. Marvuglia, M. Baltov, & P. Hansen (Eds.), *Rethinking sustainability towards a regenerative economy* (Vol. 15, pp. 341–358). Springer International Publishing. [\[CrossRef\]](#)
- Kutlu, D., Zambak, M., Soykan, S., Kasalak, M. A., & Aktaş Çimen, Z. (2024). The influence of world heritage sites on tourism dynamics in the EU 27 nations. *Sustainability*, 16(20), 9090. [\[CrossRef\]](#)
- Lim, C. (1997). Review of international tourism demand models. *Annals of Tourism Research*, 24(4), 835–849. [\[CrossRef\]](#)
- Matheson, V. (2006). *Mega-events: The effect of the world’s biggest sporting events on local, regional, and national economies*. College of the Holy Cross, Department of Economics. Available online: https://hcapps.holycross.edu/hcs/RePEc/hcx/HC0610-Matheson_MegaEvents.pdf (accessed on 30 November 2025).
- Ministry of Tourism, Government of India. (2021). *India tourism statistics 2021*. Available online: <https://tourism.gov.in/sites/default/files/2022-09/India%20Tourism%20Statistics%202021%20%281%29.pdf> (accessed on 30 November 2025).
- Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: Theory and use. *Annals of Tourism Research*, 48, 1–10. [\[CrossRef\]](#)
- Nguyen, Q. H. (2021). Tourism demand elasticities by income and prices of international market regions: Evidence using Vietnam’s data. *Economies*, 10(1), 1. [\[CrossRef\]](#)
- Nunkoo, R., & Ramkissoon, H. (2011). Developing a community support model for tourism. *Annals of Tourism Research*, 38(3), 964–988. [\[CrossRef\]](#)

- OECD. (2020). *OECD tourism trends and policies 2020*. Organisation for Economic Co-operation and Development. Available online: https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/03/oecd-tourism-trends-and-policies-2020_7cfc4549/6b47b985-en.pdf (accessed on 30 March 2025).
- OECD. (2022). *OECD tourism trends and policies 2022*. OECD Publishing. [CrossRef]
- Pike, S. (2016). *Destination marketing essentials* (2nd ed.). Routledge.
- Prideaux, B. (2000). The role of the transport system in destination development. *Tourism Management*, 21(1), 53–63. [CrossRef]
- Rasoolimanesh, S. M., Wang, M., Ragavan, N. A., & Poulain, J.-P. (2024). Tourism governance towards sustainability: A review and a metagovernance model. In K. A. Boluk, F. Higgins-Desbiolles, & M. Akhoundoghli (Eds.), *The Elgar companion to tourism and the sustainable development goals* (pp. 260–283). Edward Elgar Publishing. [CrossRef]
- Saayman, A., & Botha, I. (2017). Non-linear models for tourism demand forecasting. *Tourism Economics*, 23(3), 594–613. [CrossRef]
- Shah, I. A., Nengroo, T. A., & Haq, I. U. (2022). Determinants of international tourism demand in India: An augmented gravity model approach. *Studia Universitatis Vasile Goldiș Arad, Seria Științe Economice*, 32(3), 102–115. [CrossRef]
- Sharma, M., Mohapatra, G., & Giri, A. K. (2023). The determinants of inbound tourism demand in India: New evidence from ARDL co-integration approach. *Business Perspectives and Research*, 227853372211485. [CrossRef]
- Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management*, 29(2), 203–220. [CrossRef]
- Song, H., Li, G., Witt, S. F., & Fei, B. (2010). Tourism demand modelling and forecasting: How should demand be measured? *Tourism Economics*, 16(1), 63–81. [CrossRef]
- Song, H., Qiu, R. T. R., & Park, J. (2023). Progress in tourism demand research: Theory and empirics. *Tourism Management*, 94, 104655. [CrossRef]
- Stabler, M. J., Papatheodorou, A., & Sinclair, M. T. (2009). *The economics of tourism* (1st ed.). Routledge. [CrossRef]
- Suresh, M., Bid, K., Sangeetha, G., & Krishnan, N. (2015, December 10–12). *Inbound international tourism development in India: A panel data analysis on its affecting factors*. IEEE International Conference on Computational Intelligence and Computing Research, Madurai, India. Available online: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84965072205&partnerID=40&md5=c78f7a2b853471055ca3ac5b7713a9a0> (accessed on 30 September 2025). [CrossRef]
- Telfer, D. J., & Sharpley, R. (2015). *Tourism and development in the developing world* (2nd ed.). Routledge. [CrossRef]
- Timothy, D. J., & Boyd, S. W. (2006). Heritage tourism in the 21st Century: Valued traditions and new perspectives. *Journal of Heritage Tourism*, 1(1), 1–16. [CrossRef]
- Torres, R. (2002). Cancun's tourism development from a Fordist spectrum of analysis. *Tourist Studies*, 2(1), 87–116. [CrossRef]
- Wasiaturrehman, W., & Rohmawati, H. (2021). Multicollinearity in tourism demand model: Evidence from Indonesia. *Economics Development Analysis Journal*, 10(1), 54–69. [CrossRef]
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.
- World Tourism Organization (UNWTO). (2021). *International tourism highlights* (2020 ed.). World Tourism Organization (UNWTO). [CrossRef]
- Wu, Y., Jia, Z., & Yu, T. (2022). Tourism and green development: Analysis of linear and non-linear effects. *International Journal of Environmental Research and Public Health*, 19(23), 15907. [CrossRef]
- Wut, T. M., Xu, J., & Wong, S. (2021). Crisis management research (1985–2020) in the hospitality and tourism industry: A review and research agenda. *Tourism Management*, 85, 104307. [CrossRef]
- Xia, H., Muskat, B., Karl, M., Li, G., & Law, R. (2024). Destination competitiveness research over the past three decades: A computational literature review using topic modelling. *Journal of Travel & Tourism Marketing*, 41(5), 726–742. [CrossRef]

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