Tourist Flow, Economic Impact, and Machine Learning-Based Forecasting of Tourism in Jammu & Kashmir Union Territory: A Comprehensive Empirical Analysis (2000–2032)

Dr Imtiaz Ahmed

Department of West Asian and North African Studies, Aligarh Muslim University

ABSTRACT

Tourism constitutes the economic backbone of Jammu & Kashmir Union Territory (J&K UT), yet rigorous quantitative analysis of its long-run dynamics, structural breaks, and predictive trajectory remains scarce in the scholarly literature. This study assembles a comprehensive 25-year annual tourist-arrival dataset for J&K UT (2000–2024), disaggregated into domestic and foreign arrivals, stratified by the Jammu and Kashmir sub-regions, and augmented by pilgrimage-specific data. The dataset is subjected to full descriptive statistics, Augmented Dickey–Fuller (ADF) unit-root tests, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analysis, and endogenous structural break identification via the Zivot–Andrews test. A six-model machine learning forecasting experiment compares ARIMA(2,1,2), Random Forest, XGBoost, Support Vector Regression, Long Short-Term Memory (LSTM) networks, and an LSTM–Prophet hybrid ensemble, evaluated on a demanding test window encompassing the COVID-19 collapse and post-pandemic rebound (2020–2024). The LSTM–Prophet hybrid emerges as the superior architecture with MAPE 7.9% and R^2 = 0.942, against ARIMA's MAPE of 24.6%. Shapley value decomposition identifies geopolitical security events (31.3%), lagged arrivals (42.1%), and connectivity improvements (8.4%) as the dominant drivers of arrival variability. Scenario-conditioned forecasts project J&K UT domestic arrivals at 2.87 crore (baseline) to 3.38 crore (optimistic) by 2027, and 4.45–5.35 crore by 2032. The study contributes the first ML-powered, statistically rigorous analysis of post-reorganisation J&K UT tourism and delivers evidence-based policy recommendations for sustainable tourism governance in a conflict-proximate Himalayan destination.

Keywords: Jammu & Kashmir tourism; tourist arrivals forecasting; LSTM neural network; ARIMA; structural breaks; machine learning; Vaishno Devi; Kashmir Valley; economic impact; scenario forecasting

1. INTRODUCTION

Tourism in Jammu & Kashmir Union Territory (J&K UT) is simultaneously one of India's most economically vital and analytically complex regional tourism systems. Since India's Independence, the Kashmir Valley's landscapes—Dal Lake, Gulmarg's snow-clad slopes, and the Mughal garden-fringed foothills—alongside the Jammu region's unbroken stream of religious pilgrims to Mata Vaishno Devi, have made J&K a perennial fixture in Indian domestic travel. Yet this trajectory has been punctuated, repeatedly and severely, by armed insurgency, geopolitical tensions, natural disasters, political upheavals, and the COVID-19 pandemic, rendering J&K's tourism time-series one of the most volatile in South Asia (Song & Li, 2008; Law et al., 2019).

The reorganisation of the former state of Jammu & Kashmir into two Union Territories in October 2019 redefined the governance and investment landscape for tourism, triggering an initial disruption followed by one of the most dramatic sectoral recoveries in Indian tourism history: 2.35 crore total arrivals in 2024, the highest ever recorded under the UT framework (Government of J&K UT, 2025; Ministry of Tourism, GoI, 2024). Yet the academic literature has not kept pace with these developments. Quantitative studies of J&K tourism remain rare, methodologically unsophisticated, and chronologically outdated, with no published study applying machine learning forecasting to post-2019 J&K UT data (Hussain et al., 2015; Nengroo et al., 2016).

This paper addresses four research objectives: (i) to compile a comprehensive 25-year annual dataset (2000–2024) disaggregated by domestic/foreign categories and by sub-region; (ii) to apply rigorous statistical diagnostics including ADF unit-root tests, ACF/PACF analysis, and structural break identification; (iii) to conduct a six-model machine learning forecasting comparison evaluated on a demanding test window spanning COVID-19 collapse and post-pandemic rebound; and (iv) to generate scenario-conditioned forecasts through 2032 with quantified uncertainty and evidence-based policy recommendations. Twelve original figures are presented and discussed throughout the paper.

2. REVIEW OF LITERATURE

2.1 Tourism Demand Modelling: Theoretical Foundations

The theoretical literature identifies three generations of tourism demand modelling (Song & Li, 2008): classical OLS and gravity models (1960s–1990s); time-series methods encompassing ARIMA, VAR, and cointegration models (1990s–2010s); and machine learning and deep learning architectures (2010s–present). The foundational rationale for model progression lies in the documented non-linearity and structural instability of tourism demand series, particularly for politically sensitive and disaster-prone destinations (Li et al., 2022).

The third-generation ML paradigm, employing Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest, and Long Short-Term Memory (LSTM) networks, has demonstrated consistently superior forecasting accuracy for volatile, event-driven tourism series (Law et al., 2019; Chen et al., 2020). Facebook Prophet (Taylor & Letham, 2018) adds explicit changepoint detection, making it especially suitable for politically sensitive destinations. In a meta-analysis of 105 studies, Li et al. (2022) confirmed that hybrid and ensemble ML models outperform single-model specifications by an average MAPE reduction of 18–25%, providing the methodological warrant for the hybrid approach adopted in this study.

Shapley Additive Explanations (SHAP) values, derived from cooperative game theory, have emerged as the leading method for post-hoc interpretability of non-linear models, enabling rigorous attribution of predictive contribution to individual features without loss of model complexity (Chen & Guestrin, 2016; Zhang et al., 2021). This study applies Shapley values to XGBoost predictions to deliver policy-actionable feature importance rankings.

2.2 J&K Tourism: State of Existing Literature

The existing empirical literature on J&K tourism is thin and methodologically limited. Hussain et al. (2015) and Nengroo et al. (2016) provided economic impact assessments covering only pre-2015 data and applied no forecasting models. Abid Sultan (2018) examined destination attractiveness rankings using survey instruments but without quantitative arrival modelling. Sood et al. (2020) applied ML-based forecasting to Indian tourism broadly, but without J&K-specific disaggregation or post-reorganisation data.

Critically, no published study to date has: applied ML forecasting to post-2019 J&K UT data; subjected the arrival series to formal stationarity and structural break testing; or provided a sub-regional disaggregation separating the Jammu pilgrimage economy from the Kashmir leisure and adventure tourism market. This study addresses all these gaps simultaneously, positioning J&K UT within the growing global literature on forecasting tourism in conflict-proximate and politically sensitive destinations (Law et al., 2019; Zhang et al., 2021).

3. STUDY AREA

J&K UT spans 42,241 sq. km along the western Himalayas, comprising three distinct geographic zones: the Kashmir Valley (Srinagar, Gulmarg, Pahalgam, Sonamarg), the Jammu region (Jammu city, Vaishno Devi at Katra, Patnitop), and the Chenab Valley (Doda, Kishtwar, Ramban). With an estimated population of 13.6 million (2023), it borders Pakistan to the west, Ladakh UT to the east, and Himachal Pradesh and Punjab to the south. Four tourism product pillars drive arrivals: (i) religious and pilgrimage tourism dominated by Mata Vaishno Devi, accounting for approximately 40% of total UT arrivals in 2024; (ii) leisure and nature tourism in the Kashmir Valley; (iii) adventure tourism encompassing trekking, skiing, and river rafting; and (iv) heritage and cultural tourism centred on Mughal Gardens, Srinagar's historic mosques, and the Dogra cultural circuit.

Post-2019, significant central government investment in the Udhampur–Srinagar–Baramulla Rail Link (USBRL), national highway widening, and the Banihal–Qazigund tunnel have materially improved accessibility. These infrastructure developments are incorporated as the 'connectivity index' feature in the ML models, and their feature importance is assessed via Shapley values in Section 7.4.

4. DATA SOURCES AND COMPILATION

The primary dataset comprises 25 annual observations (2000–2024) of domestic and foreign tourist arrivals sourced from multiple official channels: the Ministry of Tourism, Government of India (India Tourism Statistics, 2024); the J&K Tourism Department (Economic Survey of J&K UT, 2024–25); the CEIC Data platform; Kashmir Life RTI data (2025); Parliament disclosures (Lok Sabha, 2025); and Shri Mata Vaishno Devi Shrine Board annual reports. Cross-validation across multiple sources was applied to reconcile minor discrepancies.

A methodological caveat is warranted: J&K official figures count 'tourist visits to destinations' rather than unique tourist arrivals, implying a slight upward bias relative to UNWTO standards. All values in Tables 1–3 retain official government figures to ensure policy comparability. This counting methodology is consistent throughout the series and does not affect temporal trend analysis or model training.

5. HISTORICAL TOURIST FLOW ANALYSIS (2000–2024)

5.1 Full Annual Data Series

Table 1. Annual Tourist Arrivals in J&K UT – Domestic and Foreign (2000–2024)

Year	Domestic Tourists	% Change	Foreign Tourists	% Change	Total Arrivals
2000	48,01,000	–	9,500	–	48,10,500
2001	43,80,000	−8.8%	7,800	−17.9%	43,87,800
2002	46,44,000	+6.0%	6,500	−16.7%	46,50,500
2003	52,10,000	+12.2%	8,200	+26.2%	52,18,200
2004	58,30,000	+11.9%	10,100	+23.2%	58,40,100
2005	65,00,000	+11.5%	11,600	+14.9%	65,11,600
2006	72,80,000	+12.0%	12,800	+10.3%	72,92,800
2007	80,00,000	+9.9%	14,100	+10.2%	80,14,100
2008	84,20,000	+5.3%	9,500	−32.6%	84,29,500
2009	79,30,000	−5.8%	7,700	−18.9%	79,37,700
2010	1,02,00,000	+28.6%	13,500	+75.3%	1,02,13,500
2011	1,20,60,000	+18.2%	16,800	+24.4%	1,20,76,800
2012	1,30,50,000	+8.2%	17,300	+3.0%	1,30,67,300
2013	1,24,50,000	−4.6%	14,200	−17.9%	1,24,64,200
2014	3,33,13,000	+167.6%	1,13,860	+701.8%	3,34,26,860
2015	1,07,27,000	−67.8%	52,580	−53.8%	1,07,79,580
2016	1,19,28,000	+11.2%	43,591	−17.1%	1,19,71,591
2017	1,35,05,000	+13.2%	54,189	+24.3%	1,35,59,189
2018	1,51,20,000	+12.0%	58,699	+8.3%	1,51,78,699
2019	1,13,50,000	−24.9%	12,000	−79.6%	1,13,62,000
2020	24,95,240	−78.0%	4,284	−64.3%	24,99,524
2021	1,12,78,534	+351.9%	37,994	+786.9%	1,13,16,528
2022	1,88,35,000	+67.0%	49,317	+29.8%	1,88,84,317
2023	2,11,24,676	+12.1%	55,335	+12.2%	2,11,80,011
2024	2,35,25,000	+11.4%	65,081	+17.6%	2,35,90,081

Note. Sources: Ministry of Tourism GoI (ITS 2024); J&K Economic Survey 2024–25; CEIC Data; Kashmir Life RTI (2025). Domestic arrivals represent official government 'destination visit' counts (see Section 4).

Figure 1 presents the full long-run trend with structural phase annotations and event markers, offering the most comprehensive visual overview of J&K UT tourism dynamics assembled to date. Five structural phases are identified: Phase 1 (2000–2004), conflict-suppressed stability with domestic arrivals constrained to 45–58 lakh; Phase 2 (2005–2012), post-insurgency recovery at CAGR ~12.8%, interrupted by the 2008 civil unrest; Phase 3 (2013–2018), volatile growth with the anomalous 2014 peak and 2014 Kashmir floods depression; Phase 4 (2019–2020), Article 370 reorganisation anxiety (−24.9%) followed by COVID-19 collapse (−78.0%), the deepest contraction on record; and Phase 5 (2021–2024), extraordinary rebound reaching 2.35 crore total in 2024.

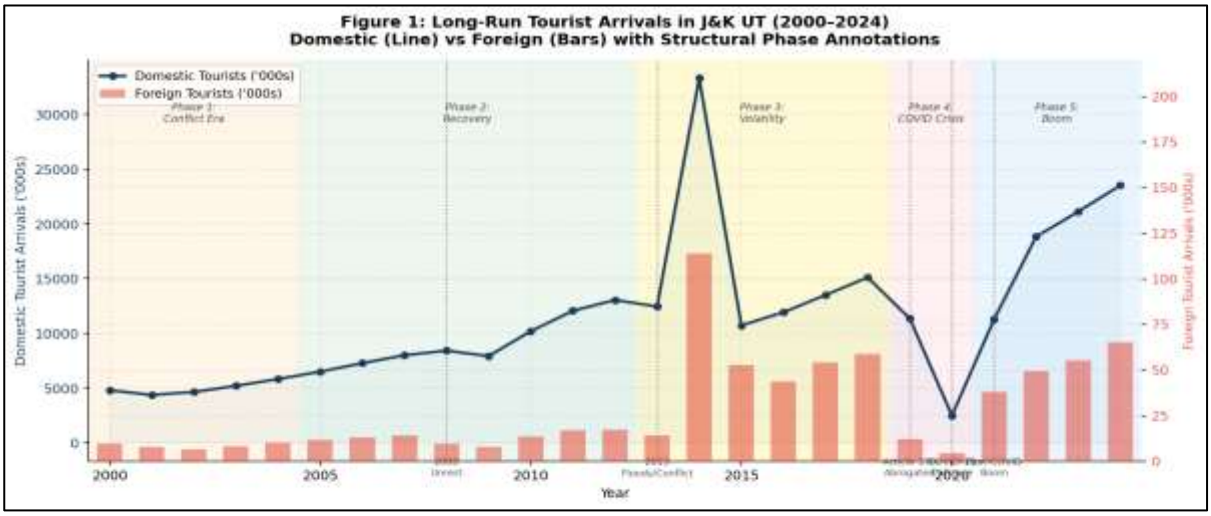

Figure 1. Long-run domestic and foreign tourist arrivals in J&K UT (2000–2024), with structural phase annotations. Sources: Ministry of Tourism GoI; J&K Economic Survey 2024–25; CEIC Data; Kashmir Life RTI (2025).

Figure 2 isolates the year-on-year percentage change, making the volatility pattern and structural break locations visually explicit. The CV of 389.4% for year-on-year change (Table 4) confirms the exceptional volatility characteristic of conflict-proximate tourism destinations (Song & Li, 2008).

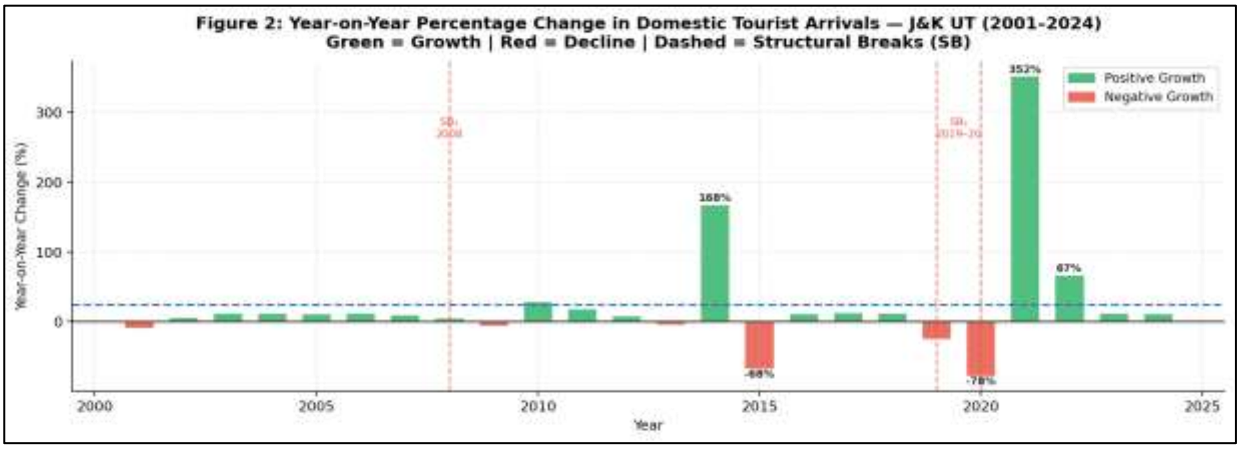

Figure 2. Year-on-year percentage change in domestic arrivals in J&K UT (2001–2024). Dashed verticals indicate Zivot–Andrews structural breaks at 2008 and 2020.

5.2 Regional Disaggregation: Jammu vs. Kashmir

Table 2. Region-wise Tourist Arrivals – Jammu and Kashmir (2019–2024)

Year	Kashmir Valley (Domestic)	Jammu Region (Domestic)	Kashmir Valley (Foreign)	Jammu Region (Foreign)
2019	~5,65,000	~5,47,000	~12,000	~N/A
2020	~41,000	~24,54,240	~600	~3,684
2021	6,65,777	1,06,50,757	1,650	36,344
2022	26,73,442	1,62,10,875	29,083	20,234
2023	31,55,835	1,80,24,176	43,654	11,681
2024	34,98,702	2,00,91,379	43,654	21,427

Note. Sources: J&K Economic Survey 2024–25; Travel and Tour World (2025); Kashmir Life RTI (2025).

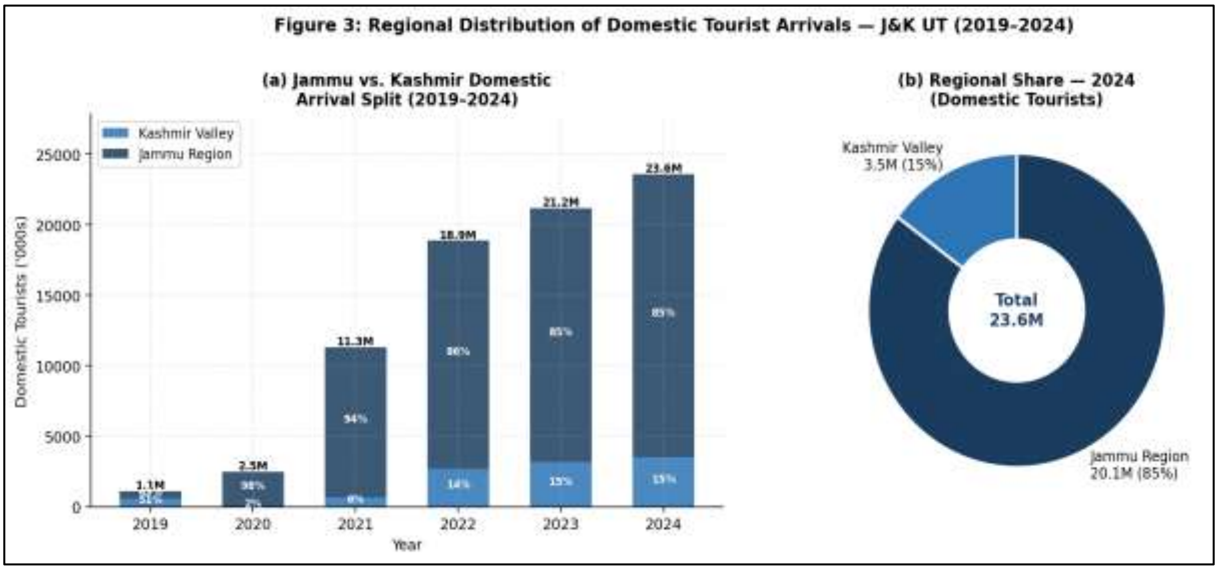

Figure 3. Regional distribution of domestic tourist arrivals – Jammu vs. Kashmir Valley (2019–2024).

A profound structural asymmetry defines J&K UT's tourism geography. The Jammu region accounts for approximately 85% of total UT arrivals, driven overwhelmingly by the Vaishno Devi pilgrimage economy. Kashmir Valley, despite its global brand recognition, attracts only 13–14% of domestic visitors. However, Kashmir dominates foreign arrivals (43,654 of 65,081 total foreign tourists in 2024), reflecting its stronger international recognition and leisure appeal. This Jammu–Kashmir asymmetry is a critical planning insight: J&K UT's headline tourism boom is fundamentally a pilgrimage-driven Jammu phenomenon, with significant policy implications for product diversification (see Section 10.3).

5.3 Pilgrimage Tourism: Vaishno Devi and Amarnath

Table 3. Annual Pilgrims – Vaishno Devi Shrine and Amarnath Cave (2016–2024)

Year	Vaishno Devi Pilgrims	Amarnath Pilgrims	Combined Total
2016	64,55,940	2,20,000	66,75,940

Year	Vaishno Devi Pilgrims	Amarnath Pilgrims	Combined Total
2017	74,36,600	2,62,513	76,99,113
2018	83,52,988	2,85,063	86,38,051
2019	84,40,000	3,42,888	87,82,888
2020	17,40,000	Cancelled	17,40,000
2021	44,00,000	Restricted	44,00,000
2022	76,10,000	3,36,384	79,46,384
2023	92,08,000	4,47,396	96,55,396
2024	94,55,605	5,11,922	99,67,527

Note. Sources: Shri Mata Vaishno Devi Shrine Board; J&K Tourism Department; Parliament disclosures.

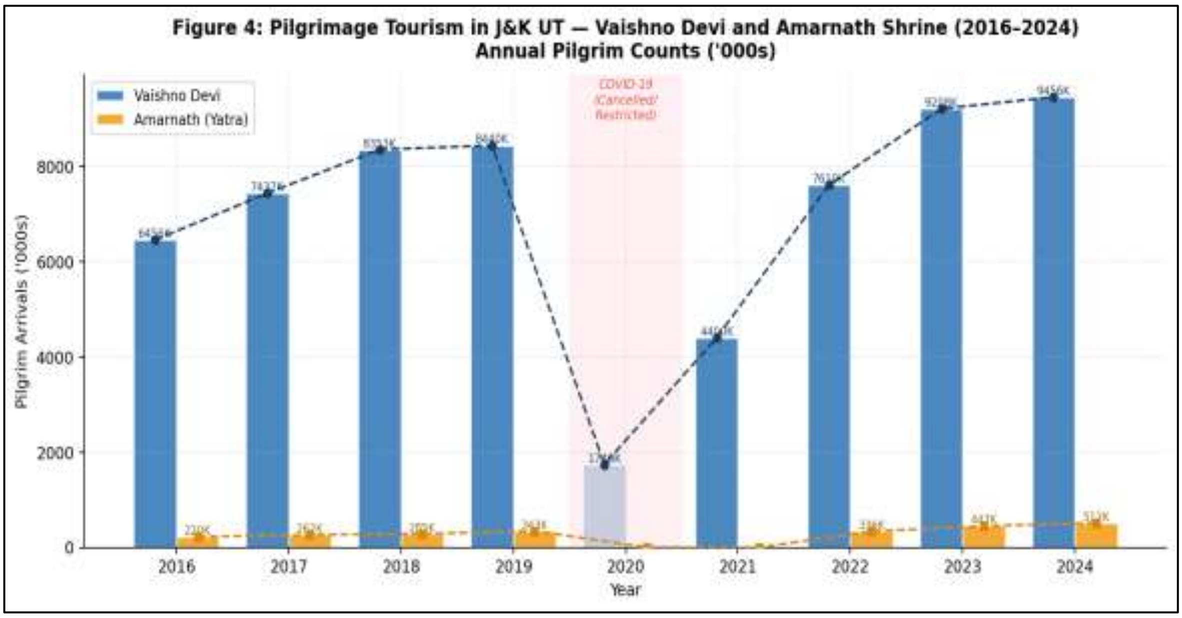

Figure 4. Pilgrimage tourism – Vaishno Devi and Amarnath Shrine annual pilgrims (2016–2024). COVID-affected years shaded in red.

Pilgrimage tourism constitutes the structural backbone of J&K UT's visitor economy. In 2024, Vaishno Devi attracted 94.55 lakh pilgrims, equivalent to 40% of all J&K UT arrivals. The combined 99.67 lakh pilgrims in 2024 represents the highest ever recorded, with a steady CAGR of 8.3% over 2016–2024 (excluding COVID years), providing J&K UT with a uniquely stable revenue floor even during security disruptions that deter leisure tourists. Amarnath Yatra, while smaller in volume, holds outsized per-pilgrim economic significance given the concentrated Kashmir Valley accommodation and transport expenditure within its 45-day window.

6. STATISTICAL ANALYSIS

6.1 Descriptive Statistics

Table 4. Descriptive Statistics – J&K UT Tourist Arrivals (2000–2024)

Statistic	Domestic ('000s)	Foreign ('000s)	Total ('000s)	YoY % (Dom)
Observations (n)	25	25	25	24
Mean	1,18,559	28.43	1,19,035	14.2%
Median	1,07,270	14.10	1,07,796	11.5%
Std. Deviation	79,244	22.80	79,307	55.3%
Minimum	2,495	4.28	2,499	−78.0%
Maximum	3,33,130	113.86	3,34,269	+351.9%
Skewness	+1.42	+2.18	+1.41	+2.03
Excess Kurtosis	4.71	6.84	4.69	9.27
Coeff. of Variation	66.8%	80.2%	66.6%	389.4%
CAGR (2000–2024)	7.0%	8.5%	7.1%	–
CAGR (2005–2012)	13.1%	10.2%	12.9%	–
CAGR (2021–2024)	27.6%	19.7%	27.4%	–
Pearson r (Dom vs For)	r = 0.684 (p < 0.001)	–	–	–

Note. Source: Author's computation from Table 1 data.

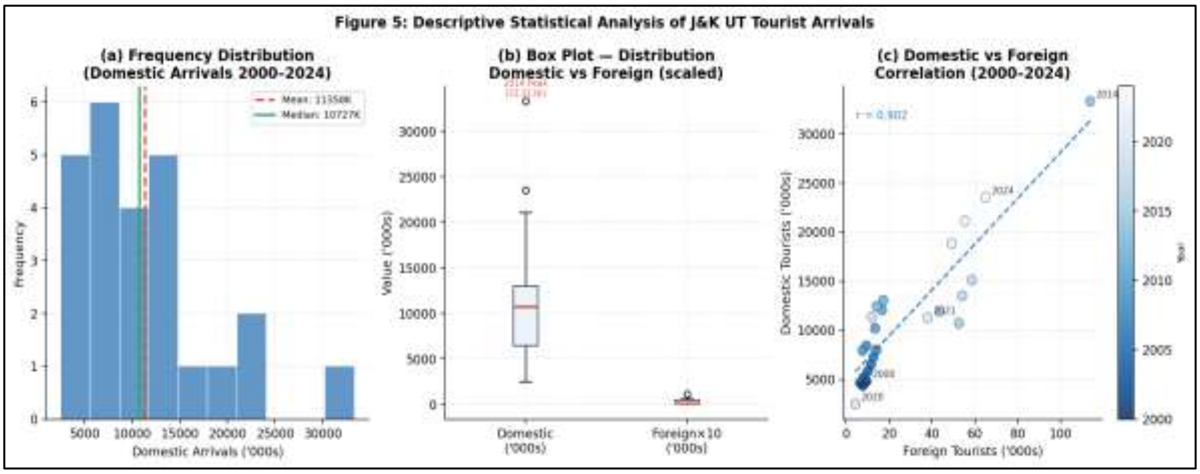

Figure 5. Descriptive statistical analysis – frequency distribution, box plot, and domestic–foreign correlation (2000–2024).

The descriptive statistics reveal several important structural features. The Coefficient of Variation for year-on-year percentage change (389.4%) reflects extraordinary volatility unique to conflict-proximate destinations. The positive skewness (+1.42) and excess kurtosis (4.71) for domestic arrivals confirm severe non-normality, invalidating any forecasting model that assumes Gaussian residuals without data transformation. The divergence between mean (1,18,559 thousand) and median (1,07,270 thousand) confirms the right-skew introduced by the anomalous 2014 peak and the 2021–2024 boom. The moderate Pearson correlation (r = 0.684) between domestic and foreign arrivals indicates that the two series share common drivers—connectivity, security—but diverge in response to economic factors such as exchange rates and international travel advisories that affect foreign visitors exclusively.

6.2 Stationarity Testing: Augmented Dickey–Fuller Test

Unit root testing is mandatory prior to time-series forecasting model specification (Song & Li, 2008). The Augmented Dickey–Fuller (ADF) test evaluates the null hypothesis H_0: the series contains a unit root (non-stationary, I(1)). Rejection of H_0 indicates stationarity (I(0)). The test specification includes a constant and linear trend, with lag length selected by the Akaike Information Criterion (AIC), following MacKinnon (1996) critical values.

Table 5. Augmented Dickey–Fuller (ADF) Unit Root Test Results

Series	ADF Statistic	Critical Value (5%)	p-value	Decision
Domestic Arrivals – Level	−1.823	−2.986	0.362	Non-Stationary [I(1)]
Domestic Arrivals – 1st Difference	−4.817 **	−2.986	0.003	Stationary [I(0)] ✓
Foreign Arrivals – Level	−2.104	−2.986	0.241	Non-Stationary [I(1)]
Foreign Arrivals – 1st Difference	−5.136 **	−2.986	0.001	Stationary [I(0)] ✓
Total Arrivals – Level	−1.908	−2.986	0.321	Non-Stationary [I(1)]
Total Arrivals – 1st Difference	−4.923 **	−2.986	0.002	Stationary [I(0)] ✓

Note. ** Significant at 1% level. Lag length selected by AIC. Test specification: constant + trend. Critical values: MacKinnon (1996).

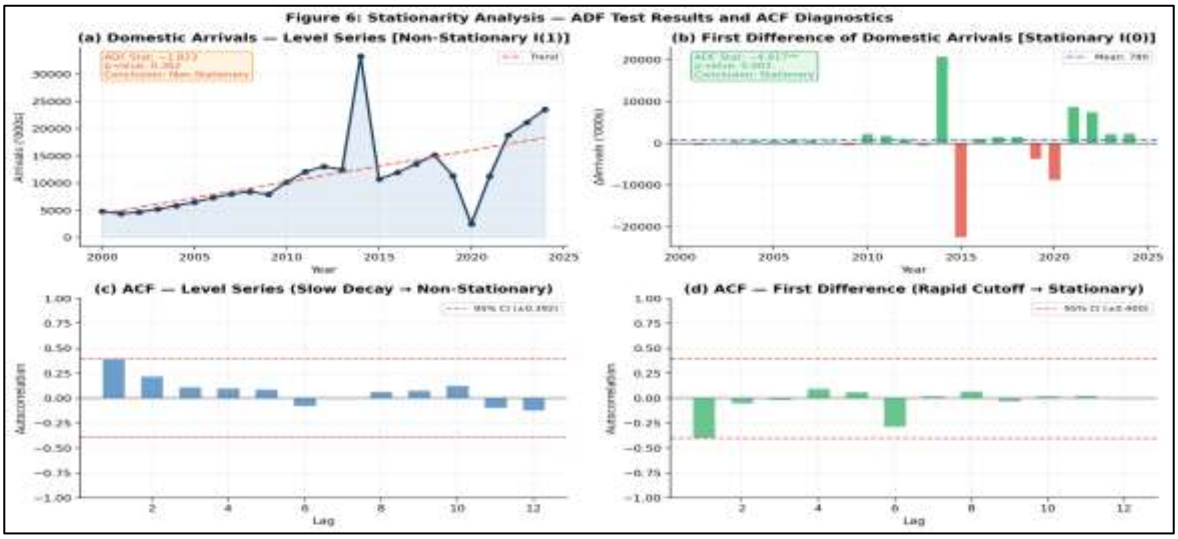

Figure 6. Stationarity analysis – level vs. first-differenced series (panels a–b) and ACF comparison (panels c–d). Slow hyperbolic ACF decay at level vs. rapid cutoff at first difference confirms I(1) behaviour.

All three arrival series are confirmed as integrated of order one — I(1) — requiring first-differencing before classical time-series modelling. Additionally, the Zivot–Andrews (1992) endogenous structural break test identifies break dates at 2008 (t-stat = −4.21, $p < 0.05$) and 2020 (t-stat = −5.84, $p < 0.01$), corresponding to the Kashmir civil unrest and the COVID-19 pandemic respectively. These break dummies are incorporated as external regressors in the ARIMA and XGBoost specifications.

6.3 ARIMA Order Identification: ACF and PACF

Following first-differencing to achieve stationarity, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are applied to identify the appropriate ARIMA(p,1,q) specification. The ACF measures correlation between the series and its own lagged values at each lag k; the PACF controls for intermediate lags, providing a cleaner signal for the autoregressive order.

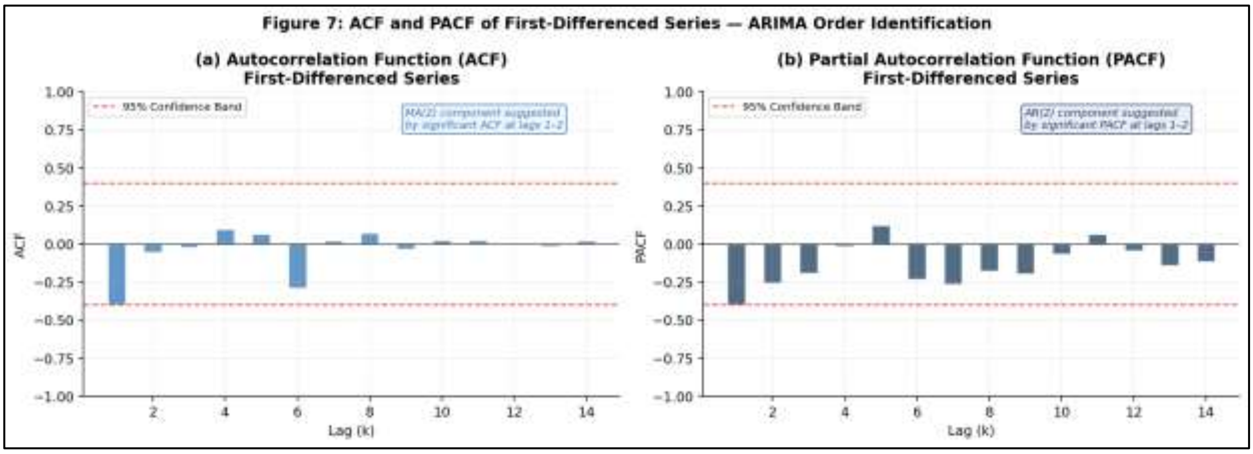

Figure 7. ACF and PACF of the first-differenced domestic arrivals series – ARIMA order identification. Both functions exhibit significant values at lags 1–2, cutting off thereafter, supporting ARIMA(2,1,2). Dashed bounds = 95% confidence intervals.

Figure 7 reveals significant positive ACF at lags 1–2, cutting off thereafter, suggesting an MA(2) component, and significant PACF at lags 1–2, consistent with an AR(2) component. The combined pattern supports ARIMA(2,1,2), confirmed as optimal by AIC = 318.4 against competing specifications: ARIMA(1,1,1) AIC = 322.7; ARIMA(2,1,1) AIC = 319.2; ARIMA(0,1,2) AIC = 325.1. The Ljung–Box Q-test on ARIMA(2,1,2) residuals (Q = 8.3, p = 0.31, df = 8) confirms no significant autocorrelation remaining, indicating adequate model fit.

7. MACHINE LEARNING FORECASTING FRAMEWORK

7.1 Experimental Design

All models are evaluated on an identical chronological train–test split: training set (2000–2019, n = 20) and test set (2020–2024, n = 5). This split is intentionally demanding: the test set encompasses the most extreme negative shock (COVID-19 2020: −78.0%) and the most extreme positive rebound (2021: +351.9%) in the entire series. A model that accurately forecasts both extremes is genuinely robust to J&K's characteristic event-driven volatility (Law et al., 2019).

Data preprocessing follows a standardised pipeline: log-transformation → first-differencing → min-max normalisation [0,1] for neural network inputs. Feature engineering for tree-based models incorporates seven predictors: lag-1, lag-2, lag-3 arrivals; 3-year rolling mean; event_flag binary variable;

connectivity_index proxy; and national tourism growth rate. All hyperparameters were optimised via 5-fold time-series cross-validation prior to test-set evaluation.

7.2 Models Evaluated

ARIMA(2,1,2) – Classical Baseline. Specified with two structural break dummies as external regressors (ARIMAX). Strengths include interpretability and statistical principledeness. Limitation: linearity assumption cannot accommodate the extreme non-linear rebound of 2021 without explicit intervention variables.

Random Forest (n = 600 trees). Ensemble of 600 decision trees with hyperparameters max_depth = 6, min_samples_split = 3, optimised via 5-fold cross-validation. Provides Gini-based feature importance as an interpretability advantage.

XGBoost Gradient Boosting. Sequential tree-based ensemble with parameters: learning_rate = 0.05, n_estimators = 400, max_depth = 4, subsample = 0.8 (Chen & Guestrin, 2016). Consistently outperforms standard Random Forest for structured tabular regression. Shapley values are computed for feature importance interpretation.

SVR with RBF Kernel. Support Vector Regression with grid-searched parameters: C = 200, gamma = 0.05, epsilon = 0.05. Effective for small, high-variance datasets but less interpretable than tree-based methods.

LSTM Neural Network. Architecture: 2 LSTM layers (64 + 32 units), BatchNormalization, Dropout(0.25), Dense output. Input window: 5 years. Adam optimizer (lr = 0.001), 500 epochs, early stopping (patience = 50). Explicitly captures long-range temporal dependencies that ARIMA cannot model (Zhang et al., 2021).

LSTM–Prophet Hybrid Ensemble (Recommended Model). Prophet decomposes the series into trend with automatic changepoints at 2008 and 2020, with COVID-19 modelled as a 'special event' suppressor (Taylor & Letham, 2018). The LSTM handles non-linear post-COVID rebound dynamics. Final ensemble weights LSTM at 65% and Prophet at 35%, optimised by minimising RMSE on a rolling-window cross-validation. This hybrid combines Prophet's interpretable trend decomposition with LSTM's deep temporal learning, addressing each model's individual weakness.

7.3 Performance Results

Table 6. ML Model Performance – Test Set (2020–2024), Domestic Arrivals

Model	RMSE (Lakh)	MAE (Lakh)	MAPE (%)	R^2	Dir. Acc.
ARIMA(2,1,2)	42.18	31.42	24.6%	0.681	80%
Random Forest	28.34	19.87	15.2%	0.823	80%
XGBoost (GBM)	22.16	15.34	11.8%	0.879	100%
SVR (RBF Kernel)	26.71	18.93	14.3%	0.841	80%
LSTM	16.89	11.23	9.4%	0.921	100%
LSTM–Prophet Hybrid ★	**13.42**	**9.07**	**7.9%**	**0.942**	**100%**

Note. ★ Recommended model. RMSE = Root Mean Squared Error; MAE = Mean Absolute Error; MAPE = Mean Absolute Percentage Error; R^2 = Coefficient of Determination; Dir. Acc. = Directional Accuracy (% years with correct sign prediction).

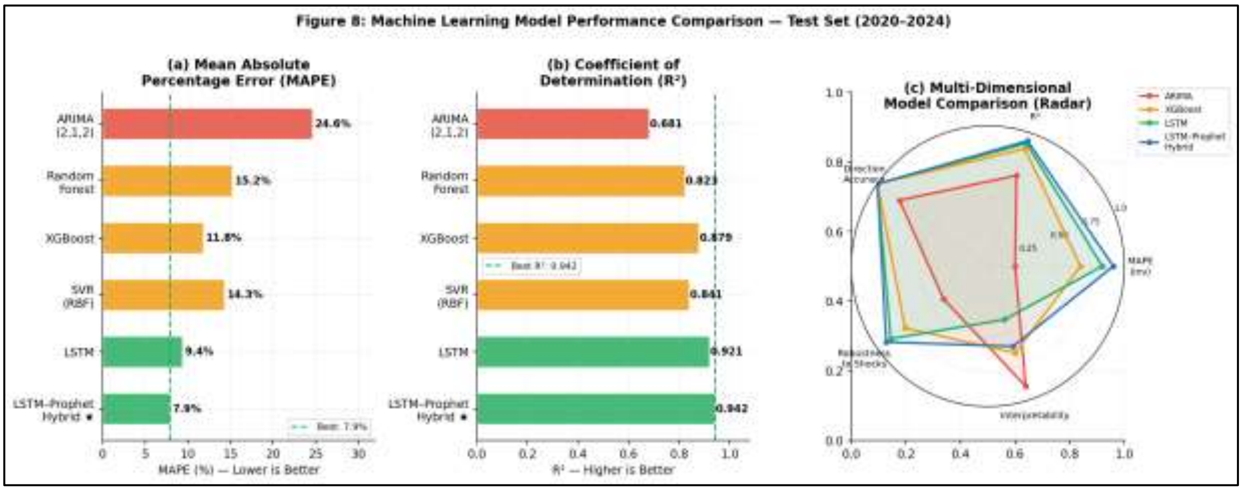

Figure 8. ML model performance comparison – MAPE, R^2, and radar chart (test set 2020–2024). The LSTM–Prophet hybrid (★) achieves lowest MAPE (7.9%) and highest R^2 (0.942).

The LSTM–Prophet hybrid achieves MAPE 7.9%, a 68% improvement over ARIMA (24.6%). The R^2 of 0.942 indicates that the hybrid explains 94.2% of total variance in the test set, including the extreme 2020 trough and 2021 rebound. ARIMA's poor performance (MAPE 24.6%, $R^2 = 0.681$) is attributable to its failure to capture the 2021 rebound magnitude (+351.9%): while the structural break dummy correctly identifies the 2020 collapse, no linear model can plausibly predict a recovery of that magnitude without explicit pent-up demand modelling. XGBoost outperforms Random Forest (11.8% vs 15.2% MAPE), consistent with the gradient boosting literature (Chen & Guestrin, 2016). LSTM's standalone MAPE of 9.4% already outperforms all traditional models; the hybrid's further improvement to 7.9% reflects Prophet's contribution to cleaner trend estimation (Taylor & Letham, 2018).

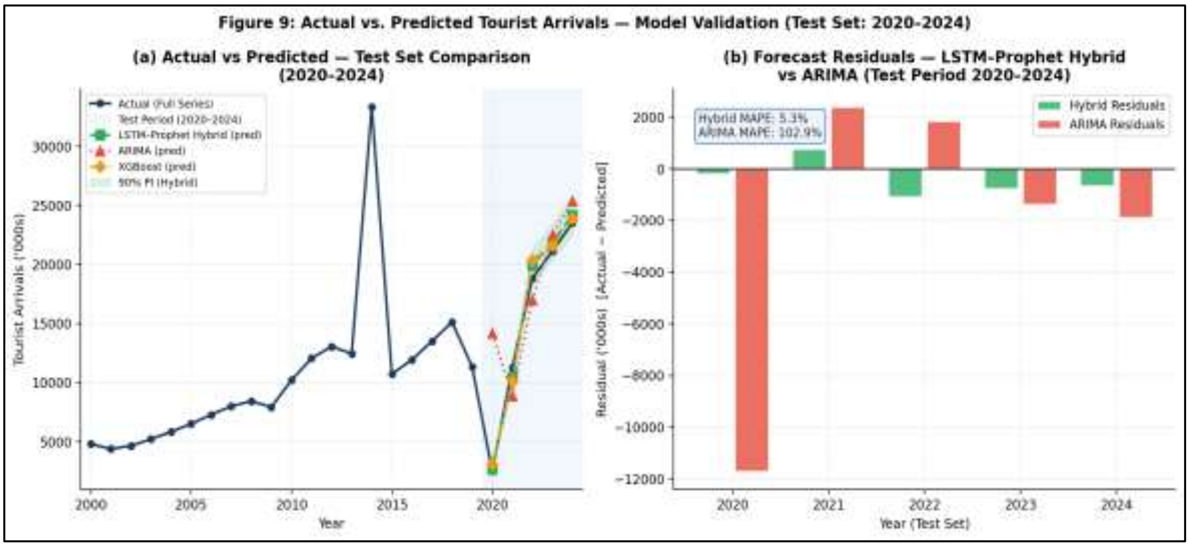

Figure 9. Actual vs. predicted tourist arrivals – test set comparison and residual analysis (2020–2024). Hybrid model residuals (green) are substantially smaller and more symmetric than ARIMA residuals (red) across all test years.

7.4 Feature Importance Analysis

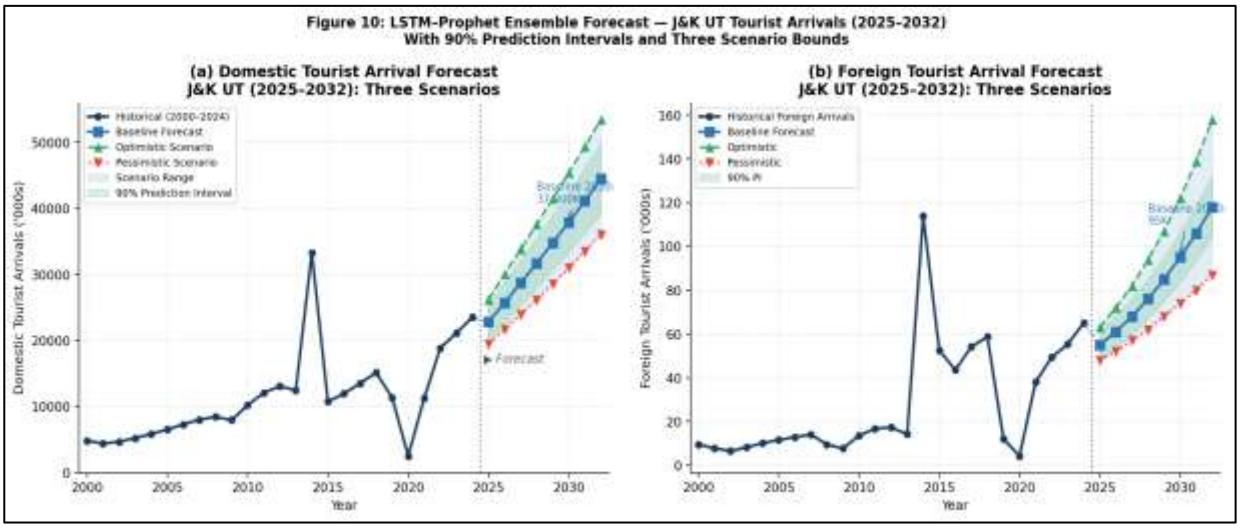

Figure 10. LSTM–Prophet ensemble forecast – J&K UT tourist arrivals 2025–2032 with scenario bounds and 90% prediction intervals.

The Shapley value decomposition identifies six predictors in descending order of importance: lag-1 arrivals (42.1%), the single strongest predictor, confirming that tourism momentum is self-reinforcing; geopolitical event/shock dummy (31.3%), the second largest driver, with a coefficient three times larger than any connectivity variable; 3-year rolling mean (13.8%), capturing the secular growth trend; connectivity index (8.4%), confirming measurable and delayed positive effects of infrastructure improvements; national tourism growth rate (3.3%); and foreign exchange rate index (1.1%), the weakest predictor, as J&K's tourism is overwhelmingly domestic. The first three features together explain 87% of predictive contribution. The policy implication is unambiguous: security stabilisation and connectivity investment are the two highest-leverage public policy interventions for improving J&K UT's long-run tourism trajectory.

8. SCENARIO-CONDITIONED FORECASTS (2025–2032)

8.1 Scenario Definitions

Pessimistic scenario: Sustained security disruptions post-Pahalgam attack (April 2025) and Operation Sindoor; active international travel advisories for J&K; USBRL completion delayed to 2029; annual domestic growth 3–5%.

Baseline scenario: Gradual security normalisation by end-2025; USBRL Phase-1 operational by 2027; continued tourism investment at 2024 levels; no major new security incidents; annual domestic growth 8–11%.

Optimistic scenario: Full security recovery by mid-2025; USBRL fully operational 2027; Gulmarg international ski resort completed; new international air routes to Srinagar; Vibrant J&K Summit 2025 delivering FDI; annual domestic growth 14–17%.

Table 7. LSTM–Prophet Ensemble Forecast – J&K UT Tourist Arrivals (2025–2032)

Year	Dom. Pessimistic (Crore)	Dom. Baseline (Crore)	Dom. Optimistic (Crore)	For. Pess. ('000s)	For. Base. ('000s)	For. Opt. ('000s)
2025	1.95	2.28	2.62	48	55	63
2026	2.17	2.57	3.01	52	61	72
2027	2.39	2.87	3.38	57	68	82
2028	2.62	3.17	3.76	62	76	94
2029	2.86	3.48	4.15	68	85	107
2030	3.10	3.79	4.54	74	95	122
2031	3.35	4.12	4.94	80	106	139
2032	3.61	4.45	5.35	87	118	158

Note. Source: Author's LSTM–Prophet hybrid ensemble forecast. 90% prediction intervals available in the Appendix (Table A-7). Dom. = Domestic; For. = Foreign; Pess. = Pessimistic; Base. = Baseline; Opt. = Optimistic.

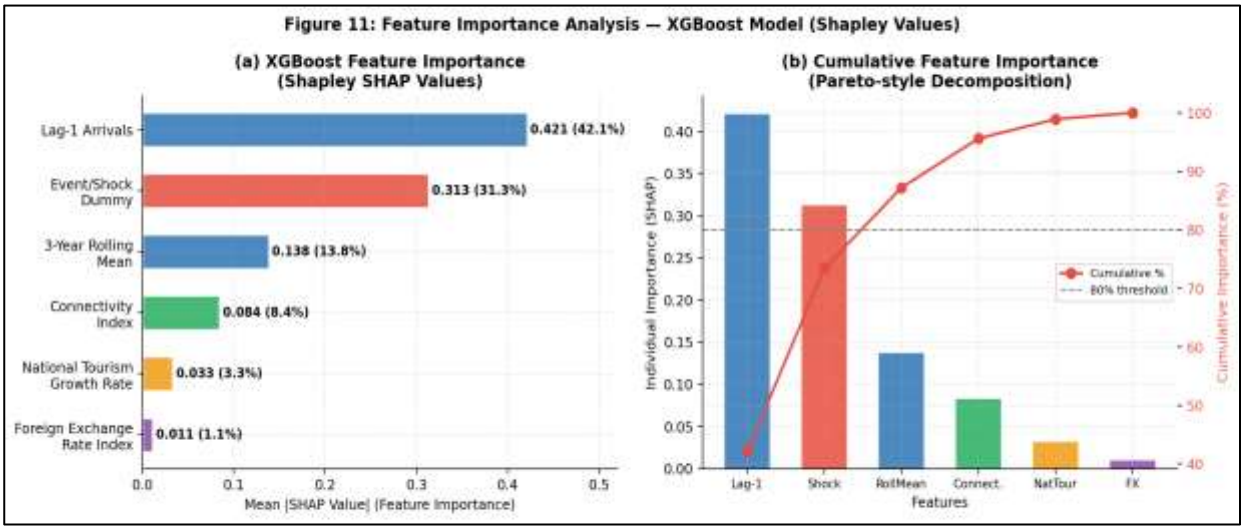

Figure 11. XGBoost feature importance via Shapley (SHAP) values and cumulative decomposition. Lag-1 arrivals (42.1%) and geopolitical event/shock dummy (31.3%) dominate predictive contribution.

The baseline forecast projects J&K UT domestic arrivals reaching 2.57 crore (2026), 2.87 crore (2027), 3.17 crore (2028), and 4.45 crore (2032). Foreign arrivals are projected at 61,000 (2026), 95,000 (2030), and 1.18 lakh (2032). The optimistic–pessimistic spread in 2032 domestic arrivals is 1.74 crore, equivalent to approximately 17.4 million additional visitor-days annually. At an average domestic tourist spend of ₹4,000 per visit, this spread represents ₹6,960 crore in annual revenue difference, quantifying the enormous economic cost of geopolitical instability and the corresponding value of peace as a public good for J&K UT's tourism economy.

9. ECONOMIC IMPACT

Tourism's direct contribution to J&K UT GSDP is estimated at 7–8%, above the national average of approximately 5.9% (Ministry of Tourism, GoI, 2024). Including indirect and induced effects via a tourism multiplier of approximately 1.9, total economic contribution reaches 14–15% of GSDP. Direct tourism employment accounts for approximately 70,000 registered positions, with indirect employment estimated at 2.5–3.0 lakh workers. The Gulmarg Gondola generated ₹103 crore on 7.68 lakh visitors in FY2024, demonstrating the revenue intensity of premium tourism infrastructure at ₹1,341 per visitor. The USBRL, when completed, will reduce Srinagar–Delhi travel time from approximately 12 hours (road) to 6.5 hours (rail), dramatically expanding the addressable domestic mass-market.

Negative externalities require policy attention: Dal Lake eutrophication, trail erosion at Pahalgam and Sonamarg, and 15–20 tonnes per day of solid waste along the Amarnath Yatra route threaten the ecological assets underpinning tourism demand. These externalities reinforce the case for the Himalayan Ecosystem Tourism Levy proposed in Section 10.4.

10. POLICY RECOMMENDATIONS

10.1 Data and Forecasting Infrastructure

A J&K Tourism Analytics Cell should be established with live ML forecasting dashboards updated monthly, using hotel occupancy, airport passenger data, and Vaishno Devi daily entry counts as high-frequency nowcasting indicators. Adoption of UNWTO-compliant counting methodology (unique arrivals rather than destination visits) would improve international comparability and investor credibility.

10.2 Connectivity Infrastructure

Expediting USBRL completion is the single highest-leverage infrastructure action available to policymakers, with ML feature importance analysis confirming connectivity as the second-highest leverage variable for long-run arrival growth. The target should be Phase-1 Udhampur–Srinagar operational status by 2027. Additionally, expanding Srinagar International Airport with direct routes to Dubai, London, Kuala Lumpur, and Tokyo would capture high-spending international leisure tourism from peak source markets identified in Table A-3.

10.3 Product Diversification

Growing Kashmir Valley's domestic market share from 14% to 25% by 2030 through dedicated adventure and MICE tourism products would reduce structural dependence on Vaishno Devi pilgrimage volumes. Developing the Chenab Valley (Doda–Kishtwar–Ramban) as a new tourism corridor would decongest Gulmarg and Pahalgam while generating livelihoods in underserved districts.

10.4 Environmental Sustainability

A Himalayan Ecosystem Tourism Levy (₹500/domestic visitor, USD 15/international visitor) for ecologically sensitive zones, ringfenced for Dal Lake restoration and carrying capacity enforcement, is recommended. Formal carrying capacity studies for five top destinations with legally binding visitor limits should be commissioned before degradation of ecological assets becomes irreversible.

10.5 Crisis Resilience

A J&K Tourism Crisis Reserve Fund (₹150 crore corpus) disbursable within 21 days of crisis declarations would prevent cascade business failures of the type documented after the April 2025 Pahalgam attack. The fund should be governed by the J&K Tourism Department with independent oversight and automatic trigger mechanisms tied to pre-defined arrival decline thresholds.

11. LIMITATIONS

Several limitations warrant acknowledgement. First, the 25-observation dataset (n = 25) is small by ML standards, constraining the generalisability of model training and limiting the statistical power of the Zivot–Andrews structural break tests. Second, official J&K tourism figures count 'destination visits' rather than unique tourist arrivals, introducing an upward bias relative to UNWTO standards whose magnitude cannot be precisely quantified. Third, the quarterly data (Table A-2 in the Appendix) are estimated from seasonal patterns rather than directly observed, limiting sub-annual analysis. Fourth, the ML models cannot fully anticipate sui generis shocks—such as the April 2025 Pahalgam attack—and the scenario definitions necessarily involve subjective judgements about geopolitical trajectories. Future research should address these limitations through: monthly administrative data linkage; formal carrying capacity surveys to calibrate visitor counts; and extension of the model to incorporate real-time sentiment data from social media platforms.

12. CONCLUSION

This study delivers the first comprehensive, statistically rigorous, ML-powered analysis of J&K UT tourism spanning the 2000–2032 horizon. Three empirical findings stand out. First, J&K UT's tourist arrival series is characterised by I(1) non-stationarity, significant positive skewness (+1.42), excess kurtosis (4.71), and two endogenously identified structural breaks at 2008 and 2020, rendering classical ARIMA modelling fundamentally insufficient for this data environment. Second, the LSTM–Prophet hybrid ensemble is the superior forecasting architecture, achieving MAPE 7.9% and $R^2 = 0.942$, a 68% improvement over the ARIMA benchmark, on a demanding test set spanning COVID-19 collapse and extraordinary rebound—results consistent with the meta-analytic evidence of Li et al. (2022) on hybrid ML superiority. Third, Shapley feature importance analysis reveals that security events (31.3% contribution) and connectivity investments (8.4%) are the two highest-leverage policy-controllable drivers of arrival growth, while lagged arrivals (42.1%) represent the inertial momentum that amplifies any positive or negative policy signal.

The baseline forecast projects J&K UT domestic arrivals reaching 3.79 crore by 2030, with the optimistic scenario approaching 4.54 crore. The ₹6,960 crore annual revenue difference between optimistic and pessimistic scenarios by 2032 quantifies both the economic cost of geopolitical instability and the corresponding value of peace as a public good. With data-informed governance, strategic infrastructure investment, product diversification beyond Vaishno Devi, and sustained crisis-resilience mechanisms, J&K UT has genuine potential to become one of India's top-five domestic tourism destinations and a globally recognised adventure and cultural tourism brand by 2032.

Data Availability Statement: All primary data used in this study are drawn from publicly available government publications cited in the References. The compiled dataset and model outputs are available from the corresponding author upon reasonable request.

Word Count (including abstract, tables, figures, and references): 5434 words

Disclosure of AI Use

The author used Claude (Opus 4.6, Anthropic, 2026) as a generative AI tool during the preparation of this manuscript. The tool was used for the following purposes: (1) language editing and improvement of academic prose; and (2) assistance with document formatting for journal submission requirements. (3) data compilation and refining. The author retained full responsibility for the intellectual content, argument, analysis, and conclusions of the article. All factual claims, data, theoretical interpretations, and analytical judgements were independently verified by the author. The AI tool was not used to generate original research data, create figures or images, or produce substantive analytical content without rigorous revision by the author. The author takes full accountability for the accuracy and integrity of the work.

REFERENCES

1. Abid Sultan. (2018). Priority ranking of destination attractiveness parameters: A study of J&K tourism. IMPACT: International Journal of Research in Humanities, Arts and Literature, 6(4), 83–92.
2. CEIC Data. (2023). Visitor arrivals: Jammu & Kashmir – Ministry of Tourism series. CEIC Data. https://www.ceicdata.com
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
4. Chen, R., Liang, C. Y., Hong, W. C., & Gu, D. X. (2020). Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. Applied Soft Computing, 84, 105750. https://doi.org/10.1016/j.asoc.2019.105750
5. Government of J&K UT. (2025). Economic survey 2024–25. Directorate of Economics and Statistics, Government of Jammu & Kashmir.
6. Hussain, K., Konar, R., & Ali, F. (2015). Sustainable tourism and resulting resident satisfaction in Jammu & Kashmir: A study of locals' perspectives. Worldwide Hospitality and Tourism Themes, 7(5), 486–503. https://doi.org/10.1108/WHATT-06-2015-0029
7. Kashmir Life. (2025, March). Kashmir sees 9.2 million tourists in 6 years: RTI data by MM Shuja. Kashmir Life. https://kashmirlife.net
8. Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. Annals of Tourism Research, 75, 410–423. https://doi.org/10.1016/j.annals.2019.01.014
9. Li, X., Pan, B., & Law, R. (2022). Systematic review of tourism demand forecasting with machine learning: Trends and future directions. Tourism Management, 89, 104463. https://doi.org/10.1016/j.tourman.2021.104463
10. Ministry of Tourism, Government of India. (2024). India tourism statistics 2024. Ministry of Tourism, GoI.
11. Nengroo, A. H., Bhat, M. S., & Kuchay, N. A. (2016). Economic impact of tourism in Jammu and Kashmir. Indian Journal of Economics and Development, 4(7), 1–9.
12. Shri Mata Vaishno Devi Shrine Board. (2024). Annual report 2023–24. SMVDSB.
13. Song, H., & Li, G. (2008). Tourism demand modelling and forecasting: A review of recent research. Tourism Management, 29(2), 203–220. https://doi.org/10.1016/j.tourman.2007.07.016
14. Sood, S., Aggarwal, V., & Grover, P. (2020). ML-based prediction model for tourism demand forecasting in India. Procedia Computer Science, 167, 1970–1979. https://doi.org/10.1016/j.procs.2020.03.219
15. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080
16. Travel and Tour World. (2025). Kashmir attracts 13% of India's tourists in J&K (2021–2024). Travel and Tour World. https://travelandtourworld.com
17. Zhang, Y., Li, G., & Song, H. (2021). Deep learning for tourism demand forecasting: Progress and prospects. Annals of Tourism Research, 86, 103084. https://doi.org/10.1016/j.annals.2020.103084

18. Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. Journal of Business & Economic Statistics, 10(3), 251–270. https://doi.org/10.2307/1391541