*Corresponding Author's Email: vasileiadis.georgios1@gmail.com
Proceedings of the International Conference on Tourism Management and Hospitality
Vol. 3, Issue. 1, 2026, pp. 37-65
DOI: https://doi.org/10.33422/ictmh.v3i1.1977

ISSN: 3030-0495 online

Exploring the Role of Nature-based Leisure in Promoting Sustainable Tourism

Georgios Vasileiadis

University of Aegean, Greece

Abstract

This study examines the relationship between environmental sustainability and tourism development across 173 countries from 2000 to 2018, focusing on nature-based leisure. Using data from the World Bank, the UN SDG framework, and the World Tourism Organization, a panel dataset is constructed. A composite Nature Sustainability Index (NSI) is developed using renewable energy, rural population, CO_2 damage, and forest depletion. The analysis combines exploratory data analysis, regression, clustering, and machine learning methods. Results show a negative global relationship between sustainability and tourism intensity, reflecting differences between high-income, high-tourism countries and lower-income, nature-rich economies. However, the relationship becomes positive in upper-middle-income countries, indicating a development threshold. Machine learning models outperform linear approaches, emphasizing nonlinear dynamics and identifying GDP per capita and resource dependence as key predictors. Overall, the findings highlight that sustainability can become a competitive advantage when supported by adequate infrastructure and development.

Keywords: Clustering, Machine Learning, Nature-Based Leisure, Sustainability, Tourism Intensity

1 Introduction

1.1 Theoretical Framework

The relationship between tourism development and environmental sustainability has become one of the most important issues in contemporary tourism research, particularly as destinations face the combined pressures of climate change, biodiversity loss, resource depletion and post-pandemic recovery. Tourism remains a major driver of economic growth, employment and regional development, but its environmental footprint has become increasingly difficult to ignore. Recent research has shown that tourism-related consumption, transport, accommodation and food systems contribute substantially to global greenhouse gas emissions, with high-income and high-mobility economies playing a particularly important role in this footprint (Pousa-Unanue et al., 2026, Janchai & Suvittawat, 2025). At the same time, international organizations increasingly frame sustainable tourism as a strategic pathway for achieving the Sustainable Development Goals, especially through responsible consumption, climate action, local development and biodiversity protection (UN Tourism, 2024, OECD, 2024).

Within this broader context, nature-based leisure occupies a central position. Nature-based tourism, ecotourism and sustainable tourism have expanded as interrelated fields of research and policy because natural landscapes, forests, coastlines, protected areas and biodiversity are increasingly viewed as both tourism assets and fragile ecological systems requiring careful management (Esparza-Huamanchumo et al., 2024). Nature-rich destinations can attract visitors seeking environmental quality, recreation, authenticity and contact with natural ecosystems. However, this creates a fundamental paradox: the same environmental resources that make destinations attractive may be degraded by uncontrolled tourism growth, infrastructure expansion, transport emissions, land-use pressure and excessive visitor concentration. Therefore, sustainability cannot be treated simply as an ethical objective, but as a long-term condition for destination competitiveness and resilience (Varzakas & Metaxas, 2024, Geng et al. 2026).

The sustainability–tourism relationship is also theoretically complex because it is rarely linear. The Environmental Kuznets Curve perspective suggests that environmental pressure may increase during early stages of economic development and later decline when income, technology, regulation and institutional capacity improve (Voumik et al., 2023). In tourism research, this idea is particularly relevant because tourism can generate income and infrastructure, but it can also intensify emissions, resource use and ecosystem degradation. Recent studies on tourism-induced Environmental Kuznets Curve dynamics show that the effects of tourism on environmental quality vary across countries, income groups and development stages, indicating the presence of thresholds, asymmetries and heterogeneous pathways (Arbulú et al., 2025, Sobirov et al., 2025). This means that the same sustainability indicator may have different implications in low-income, middle-income and high-income destinations.

Institutional capacity is another crucial factor in this relationship. Countries with stronger governance, better monitoring systems, environmental regulation, digital infrastructure and destination management capacity may be more able to transform natural assets into sustainable tourism advantages. In contrast, nature-rich countries with weak infrastructure or limited governance capacity may possess strong environmental potential but remain unable to convert this potential into inclusive and sustainable tourism development. This distinction is important because it suggests that sustainability does not automatically produce tourism competitiveness. Rather, sustainability becomes a competitive advantage when supported by transport accessibility, digital visibility, environmental monitoring, investment capacity and effective destination governance (OECD, 2024, UN Tourism, 2024).

This study investigates the sustainability–tourism nexus using a global panel dataset of 173 countries over the period 2000–2018. It integrates sustainability indicators from the World Bank and the UN Sustainable Development Goals framework with tourism data from the World Tourism Organization. Key variables include tourism intensity, renewable energy consumption, CO_2 emissions, forest depletion, rural population share, income level, resource dependence and digital connectivity. A composite Nature Sustainability Index is developed to capture cross-country differences in environmental and nature-related sustainability conditions.

The study addresses four main research questions. First, what is the relationship between nature-related sustainability and tourism intensity at the global level? Second, does this relationship differ across income groups and development stages? Third, can country typologies be identified according to their sustainability, tourism and development profiles? Fourth, which factors best predict tourism intensity when nonlinear relationships and interaction effects are considered? To answer these questions, the study combines descriptive analysis, pooled regression models, alternative panel specifications, principal component analysis, clustering and machine learning models. The use of machine learning is particularly relevant because tourism–sustainability relationships may involve nonlinear patterns, threshold effects and interactions that are not fully captured by conventional linear models (Akhtar et al., 2024).

The contribution of the study is threefold. First, it provides a global empirical assessment of the relationship between nature-related sustainability and tourism intensity across a large panel of countries. Second, it highlights the importance of income-based thresholds, showing that sustainability may operate differently depending on the level of economic and institutional development. Third, it combines traditional econometric techniques with clustering and machine learning in order to identify both general patterns and complex nonlinear dynamics. Rather than claiming strict causality, the study aims to map empirical associations and predictive structures that can inform future research and policy. Overall, the paper argues that sustainability becomes a tourism advantage only when natural assets are supported by adequate infrastructure, institutional capacity and effective destination governance.

2 Literature Review

Recent literature on tourism and sustainability has increasingly moved beyond the traditional view of tourism as a mainly economic activity and now approaches it as a complex socio-ecological system. Sustainable tourism research emphasizes that tourism growth must be assessed not only through arrivals, income and employment, but also through environmental costs, carbon emissions, pressure on natural resources and the capacity of destinations to manage these pressures over time. The post-pandemic period has strengthened this debate, as tourism recovery has been accompanied by renewed concerns about climate change, overtourism, biodiversity loss and the need for more resilient destination models (Varzakas & Metaxas, 2024). International organizations also stress that tourism can contribute to the Sustainable Development Goals only when growth is connected with responsible consumption, climate action, local participation and environmental monitoring (UN Tourism, 2024).

Nature-based tourism occupies a central position within this debate because it directly depends on the quality of natural landscapes and ecosystems. Recent research shows that nature-based tourism, ecotourism and sustainable tourism are related but distinct fields. Ecotourism is usually associated with conservation, education and local community benefits, while nature-based tourism includes a wider range of leisure activities taking place in natural environments. Sustainable tourism functions as the broader framework through which both forms can be evaluated (Esparza-Huamanchumo et al., 2024). This distinction is important because a destination may attract tourists due to natural beauty, but this does not necessarily mean that tourism development is sustainable. Without appropriate planning, visitor management and environmental protection, nature-based tourism can contribute to ecosystem degradation, land-use conflicts and increased emissions.

A growing body of research also examines tourism through the lens of low-carbon transition. Low-carbon tourism studies focus on reducing the environmental footprint of tourism activities, particularly transport, accommodation, energy consumption and tourist behaviour. Knowledge-mapping research indicates that low-carbon tourism has become an expanding field, with increasing attention to green mobility, carbon accounting, destination management and climate-responsible consumption (Geng et al., 2026). Similarly, studies on low-carbon destination choice suggest that tourists' decision-making can be influenced by environmental awareness, perceived destination quality and the availability of sustainable tourism options (Janchai & Suvittawat, 2025). These findings suggest that sustainability is not only a supply-side issue related to infrastructure and regulation, but also a demand-side issue connected with tourist preferences, attitudes and behavioural change.

The literature on tourism carbon footprint further demonstrates the methodological difficulty of measuring the environmental impact of tourism. Tourism emissions are distributed across multiple sectors, including aviation, road transport, accommodation, food systems and recreational services. As a result, carbon footprint measurement requires integrated frameworks capable of capturing direct and indirect impacts across the tourism value chain. Recent systematic work highlights that different methodological choices can lead to different estimates of tourism-related emissions, making transparency in indicator construction essential (Pousa-Unanue et al., 2026). This is directly relevant for empirical studies that use composite sustainability indices, because the selection, weighting and transformation of indicators can shape the conclusions drawn about the relationship between tourism and environmental performance.

The Environmental Kuznets Curve provides an important theoretical framework for understanding why the tourism–environment relationship may vary across development stages. According to this perspective, environmental degradation may increase during earlier phases of economic growth and later decline when countries acquire stronger institutions, cleaner technologies and better regulatory systems. In tourism research, this framework is especially relevant because tourism can simultaneously stimulate income growth and intensify environmental pressure. Voumik et al. (2023), for example, show that tourism, energy, education and trade can interact in shaping environmental outcomes, suggesting that tourism-induced environmental effects are embedded within wider development structures. More recent EKC studies also underline that the relationship between tourism, economic growth and environmental degradation is not uniform across regions or indicators. Sobirov et al. (2025) find that economic growth, foreign direct investment, tourism and agricultural productivity can act as drivers of environmental degradation in ASEAN countries, while Arbulú et al. (2025) connect tourism development with water stress in the European context. These studies support the view that tourism-related sustainability outcomes depend on income level, resource use, sectoral structure and environmental governance.

Institutional capacity is therefore central to sustainable tourism development. Countries with stronger institutions are better able to regulate land use, protect natural resources, manage visitor flows, collect environmental data and implement sustainability monitoring systems. In contrast, countries with rich natural assets but weak infrastructure or limited governance capacity may fail to transform environmental potential into sustainable tourism competitiveness. The OECD (2024) highlights that digitalization and artificial intelligence can support tourism policy by improving data use, destination management, personalization, forecasting and sustainability monitoring. However, the benefits of these tools depend on institutional readiness, technical skills and responsible governance. Therefore, institutional capacity functions as a mediating factor between environmental resources and tourism performance.

Machine learning applications have become increasingly relevant in tourism studies because tourism systems are nonlinear, multidimensional and influenced by complex interactions among economic, environmental and social variables. Traditional regression models remain valuable for estimating average associations, but machine learning techniques can identify hidden patterns, nonlinear effects and variable importance with greater predictive flexibility. In tourism sustainability research, such methods can help explore whether factors such as GDP per capita, renewable energy, digital connectivity, resource dependence and environmental indicators interact in ways that conventional linear models may not fully capture. At the same time, methodological caution is necessary. High predictive accuracy does not imply causality and issues such as multicollinearity, overfitting and model validation must be addressed carefully. Akhtar et al. (2024) stress the importance of dealing with multicollinearity in regression-based analysis, while the use of machine learning in tourism policy requires transparent validation and interpretation (OECD, 2024).

Overall, the literature shows that the sustainability–tourism nexus cannot be explained by a single linear relationship. Nature-based tourism depends on environmental quality, but also on infrastructure, institutions, monitoring capacity, income level and tourist behaviour. EKC-based studies suggest that development thresholds matter, while low-carbon and carbon-footprint research highlights the importance of measurement and methodological transparency. Machine learning offers useful tools for capturing nonlinear patterns, but it must be combined with theory-driven interpretation and robustness checks. This study is positioned within this research gap by examining how nature-related sustainability, income levels and development capacity interact in shaping tourism intensity across a large global panel.

3 Methodology

The study uses a country-year panel dataset covering 173 countries over the period 2000–2018. The analysis combines tourism, economic, environmental and development indicators in order to examine the relationship between nature-related sustainability and tourism intensity. Sustainability and development variables are drawn mainly from the World Bank World Development Indicators, which provide internationally comparable country-level data on economic, social and environmental conditions (World Bank, 2024). Tourism indicators are based on international tourism data reported by UN Tourism and made available through global development databases, including arrivals, tourism intensity and tourism-related economic measures (UN Tourism, 2024). The datasets were harmonized by ISO country codes and year, producing an integrated panel suitable for cross-country and temporal analysis.

The dependent variable is tourism intensity, measured as international tourist arrivals per 1,000 inhabitants. This indicator allows comparison between countries of very different population sizes and is more appropriate than absolute arrivals for identifying tourism pressure and destination dependence. Because both tourism intensity and GDP per capita are highly right-skewed, logarithmic transformations are applied to reduce the influence of extreme values and improve model stability. Additional variables include renewable energy consumption, CO_2 emissions per capita, rural population share, forest depletion, natural resource rents, internet penetration, life expectancy and GDP per capita.

A central variable in the study is the Nature Sustainability Index (NSI). The NSI is constructed as a composite indicator using four standardized components: renewable energy consumption, rural population share, CO_2 damage and forest depletion. CO_2 damage and forest depletion are inverted so that higher values indicate stronger sustainability performance. The components are standardized before aggregation in order to ensure comparability across different measurement units. Equal weighting is used in the baseline specification, following the logic of transparent composite-indicator construction, while the need for sensitivity analysis is recognized because alternative weighting schemes may affect the results (Pousa-Unanue et al., 2026).

The empirical strategy proceeds in four stages. First, descriptive and exploratory analysis is used to examine the distribution of key variables, regional differences and global trends. Second, pooled OLS models are estimated as baseline specifications for the association between sustainability indicators and tourism intensity. Robust standard errors are used to reduce sensitivity to heteroskedasticity, while year effects are included in extended models to account for common global shocks and time trends. The pooled OLS results are interpreted as associational rather than causal. Given the cross-country panel structure, unobserved heterogeneity is an important limitation, therefore, alternative fixed-effects and random-effects specifications are proposed as robustness checks to test whether the main relationships remain stable across model choices (Michelaki et al., 2025).

Third, principal component analysis and K-means clustering are used to identify country typologies according to tourism, sustainability and development characteristics. This step helps classify countries into groups with similar structural profiles, such as high-tourism developed destinations or nature-rich but low-infrastructure economies. Finally, Random Forest and Gradient Boosting models are applied to capture nonlinear relationships and interactions that may not be fully identified by linear regression. Machine learning is used for prediction and variable-importance analysis, not for causal inference. Model validation must avoid temporal leakage, therefore, time-aware validation and robustness checks are recommended in addition to conventional cross-validation. This is consistent with recent calls for more transparent and responsible use of artificial intelligence and machine learning in tourism research and policy (Akhtar et al., 2024).

4 Results

The dataset includes 3,287 observations across 173 countries. Tourism intensity is highly skewed (mean 781, median 268), driven by small tourism economies. Tourism contributes 5.0% to GDP on average (max 58.6%), while leisure travel dominates (median 0.88). Renewable energy averages 32.7% with wide variation, and CO_2 emissions average 5.0 tonnes per capita (Table 1). The NSI is centered near zero (SD = 0.51), with higher values in Africa (0.47) and lower in Asia (-0.22).

Table 1: Descriptive statistics for key tourism and sustainability indicators (2000 to 2018).

Variable	N	Mean	Std	Min	Median	Max
Tourist arrivals	2,793	5,874,864	12,254,408	2,900	1,210,000	89,322,000
Arrivals per 1,000	2,793	781.1	2,006.3	0.5	267.6	28,407.2
Tourism GDP share (%)	693	5.0	6.3	0.2	3.4	58.6
Leisure ratio	2,108	0.83	0.17	0.10	0.88	1.00
Renewable energy (%)	3,287	32.7	29.9	0.0	22.7	98.3
CO_2 per capita (t)	3,106	5.0	6.6	0.02	2.6	67.0
GDP per capita (USD)	3,246	12,306	17,827	91	4,336	118,824
Rural population (%)	3,278	42.9	22.7	0.0	42.7	91.8
Nature Sustainability Index	3,287	-0.01	0.51	-3.33	-0.05	1.53

Figure 1 shows that tourism intensity and CO_2 per capita are strongly right-skewed, justifying log transformations. The leisure ratio is left-skewed, with most countries reporting over 80% personal travel. Renewable energy is bimodal (near zero and 50%–90%), while the NSI is roughly normal with a left tail from high-emission economies. Global trends (Figure 2) show that tourist arrivals rose from 0.65 to 1.2 billion (2000–2018, ~3.4% annually), with median arrivals per 1,000 increasing from 150 to over 450. Renewable energy remained stable (21%–24%), CO_2 per capita showed little change (2.3–2.8 tonnes), and tourism's GDP share rose slightly, with a dip in 2009.

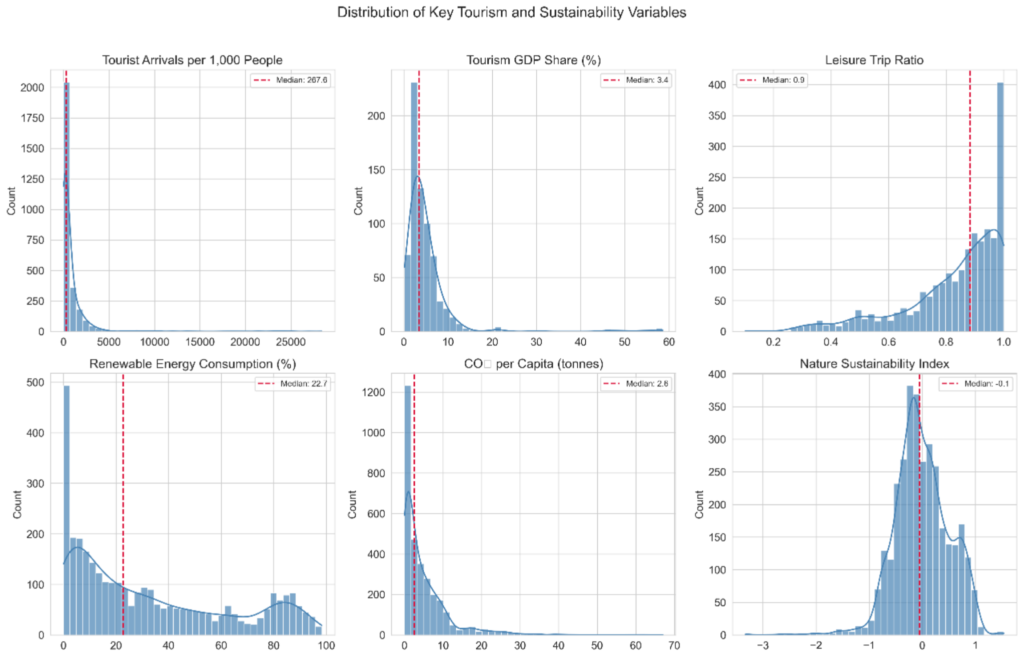

Figure 1: Distribution of key tourism and sustainability variables across the full panel. Dashed red lines indicate medians.

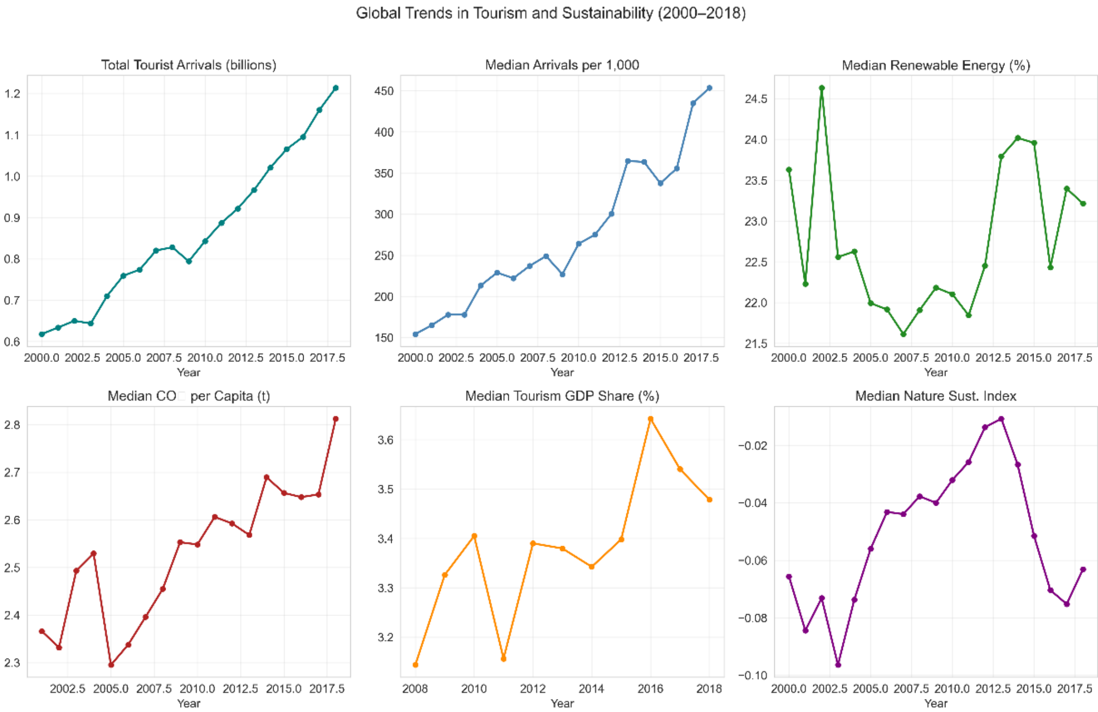

Figure 2: Global trends in tourism and sustainability indicators from 2000 to 2018. Values are annual sums (arrivals) or medians across countries.

Continental comparisons (Figure 3) show that Europe leads in tourism intensity, while Africa has the lowest tourism but the highest NSI and Asia the lowest NSI. This indicates an inverse relationship between sustainability and tourism. Income-group analysis (Figure 4) supports this: high-income countries attract more tourists but rely less on renewables, whereas low-income countries show the opposite. The NSI declines with income, consistent with the environmental Kuznets curve.

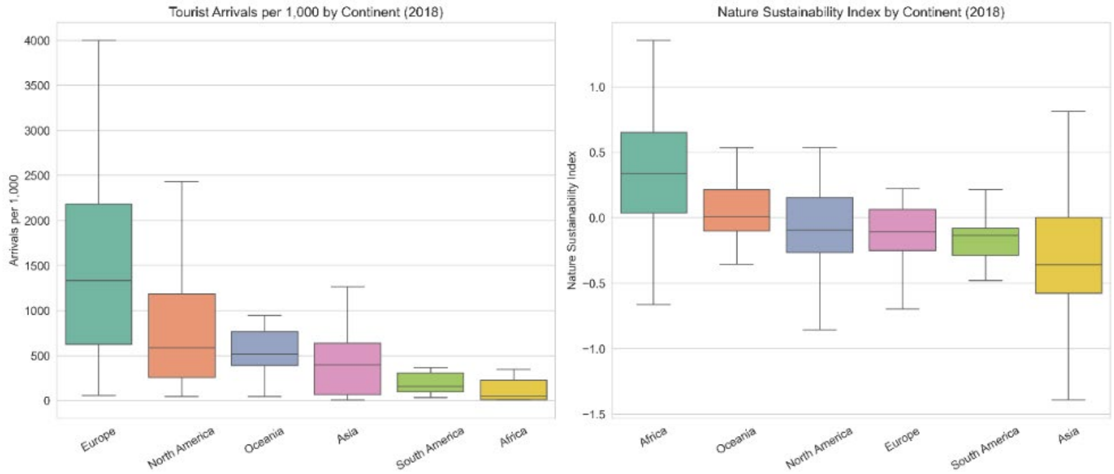

Figure 3: Tourist arrivals per 1,000 and Nature Sustainability Index by continent (2018). Boxes show interquartile ranges.

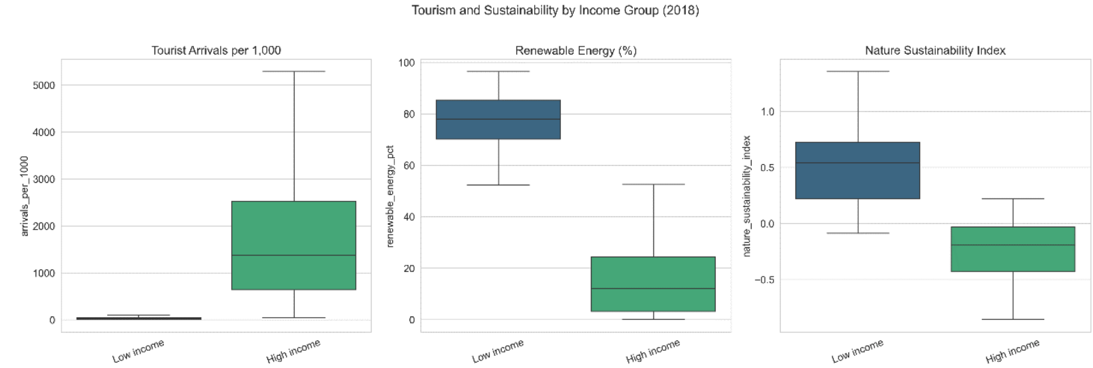

Figure 4: Tourism intensity, renewable energy consumption, and Nature Sustainability Index by World Bank income classification (2018).

Table 2 presents OLS results for tourism intensity. Model 1 explains 69.7% of variation (R^2 = 0.697). GDP per capita is the strongest predictor (β = 0.891), while renewable energy has a negative effect (β = -0.016). Rural population is positively associated (β = 0.016), and natural resource rents are negatively related to tourism (β = -0.031).

*Table 2: OLS regression results predicting log tourism intensity (arrivals per 1,000). Robust (HC1) standard errors. Significance: *** p<0.01, ** p<0.05, * p<0.1.*

Predictor	Model 1	Model 2	Model 3
Intercept	-2.201***	-2.124***	-2.336***
Renewable energy (%)	-0.016***	-0.017***	-0.016***
CO_2 per capita	-0.012**	-0.013***	-0.015***
Rural population (%)	0.016***	0.015***	0.016***
Forest depletion (%)	0.028***	0.032***	0.034***

Predictor	Model 1	Model 2	Model 3
Natural resource rents (%)	-0.031***	-0.033***	-0.032***
Log GDP per capita	0.891***	0.952***	0.995***
Internet penetration (%)		-0.002	-0.004**
Life expectancy		-0.007	-0.007*
R-squared	0.697	0.698	0.701
Adjusted R-squared	0.697	0.697	0.697
N	2,406	2,406	2,406
Year fixed effects	No	No	Yes

Model 2 adds internet use and life expectancy with little change ($R^2 = 0.698$), likely due to collinearity with GDP. Model 3 includes year effects, slightly improving fit ($R^2 = 0.701$). Internet ($\beta = -0.004$) and life expectancy ($\beta = -0.007$) become weakly significant, possibly reflecting saturation effects. Core sustainability variables remain stable, indicating robust results.

The correlation heatmap (Figure 5) shows that GDP per capita is positively related to tourism ($r = 0.37$) and negatively to renewable energy ($r = -0.29$). The NSI has a weak negative link with tourism ($r = -0.19$). Internet use and life expectancy correlate positively with tourism ($r = 0.29$). Natural resource rents are negatively related to tourism and GDP ($r = -0.17$). Renewable energy is strongly linked to renewable electricity ($r = 0.59$) but negatively associated with tourism and GDP, reflecting lower-income economies.

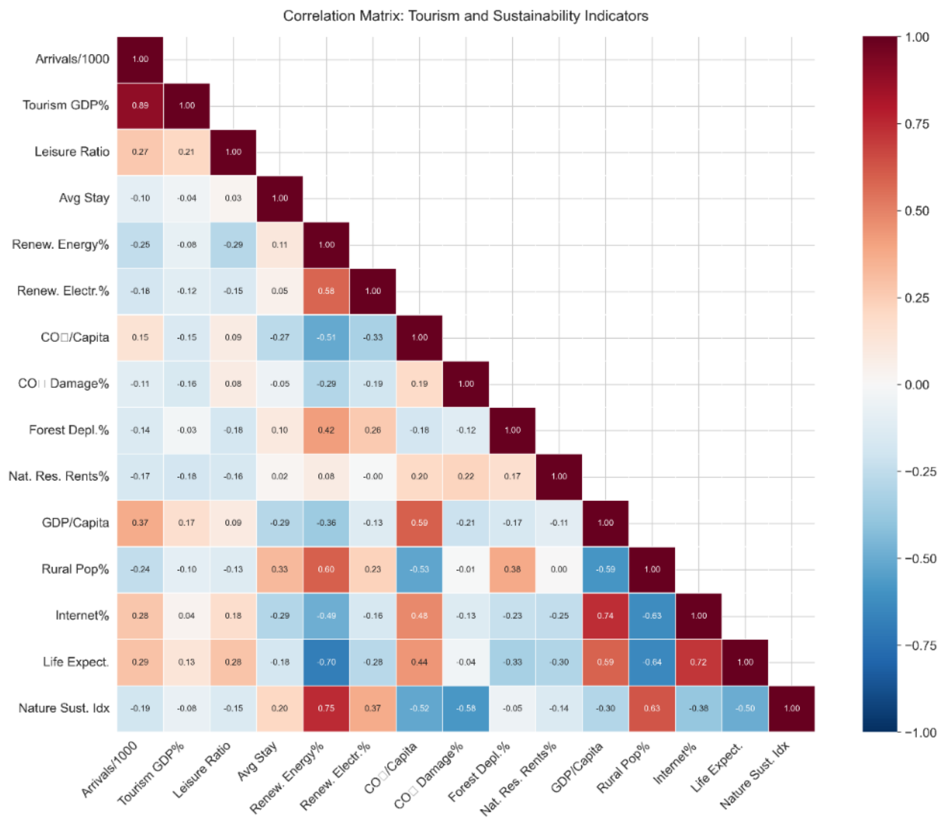

Figure 5: Pearson correlation matrix for 15 tourism and sustainability indicators. Colour intensity reflects correlation magnitude, values are annotated in each cell.

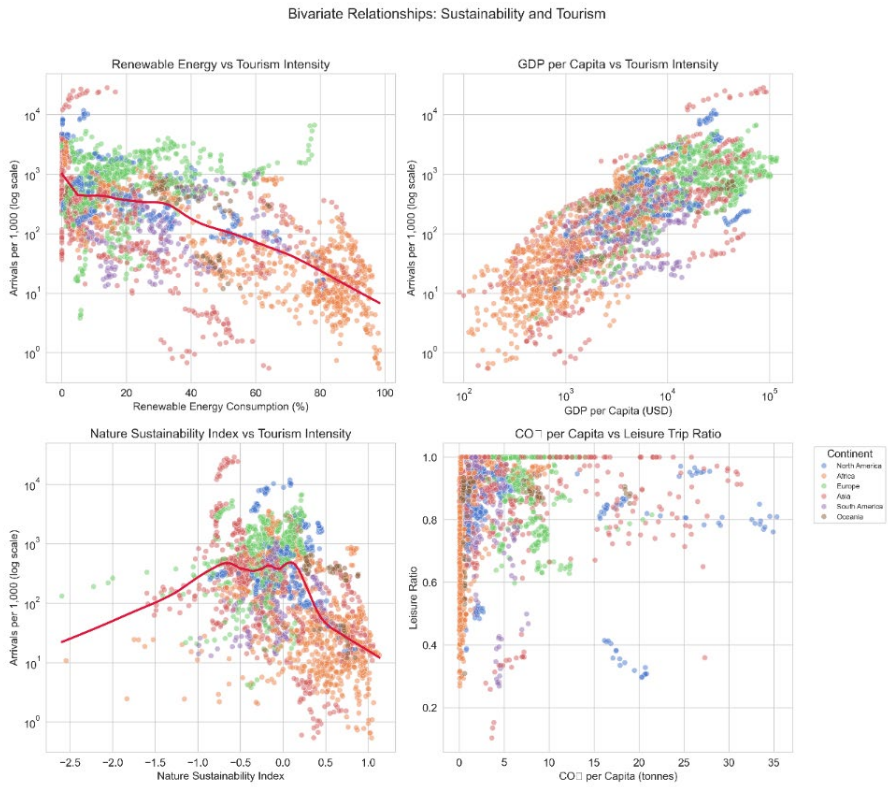

Figure 6: Bivariate scatter plots of sustainability indicators against tourism metrics. Red curves are LOWESS smoothers. Colour indicates continent.

Figure 6 shows key bivariate relationships. Renewable energy is negatively related to tourism, with high-renewable countries showing low tourism levels. GDP per capita is strongly positively linked to tourism. The NSI follows a hump-shaped pattern, with peak tourism at mid values. CO_2 per capita is positively associated with the leisure ratio, indicating a dominance of leisure travel in high-emission economies.

Figure 7 reveals an income-based reversal in the NSI–tourism relationship. It is negative in low- and lower-middle-income countries, indicating that natural capital alone is insufficient without infrastructure. It becomes strongly positive in upper-middle-income countries, where development enables the use of environmental assets, and weakens in high-income countries as other factors dominate. This underscores the importance of infrastructure investment in nature-rich developing economies.

Principal component analysis (2018, 83 countries) identifies three components explaining 80.2% of variance (Figure 8). PC1 (54.0%) reflects development, PC2 (15.6%) separates tourism-based from resource-dependent economies, and PC3 (10.6%) captures a “green” dimension. Although silhouette analysis suggests k = 2, k = 3 is selected for better interpretation. The resulting clusters are coherent (Figure 9).

Figure 7: Nature Sustainability Index versus tourism intensity, stratified by income group. Black lines are OLS fits on log-transformed arrivals. Pearson r and significance levels are annotated.

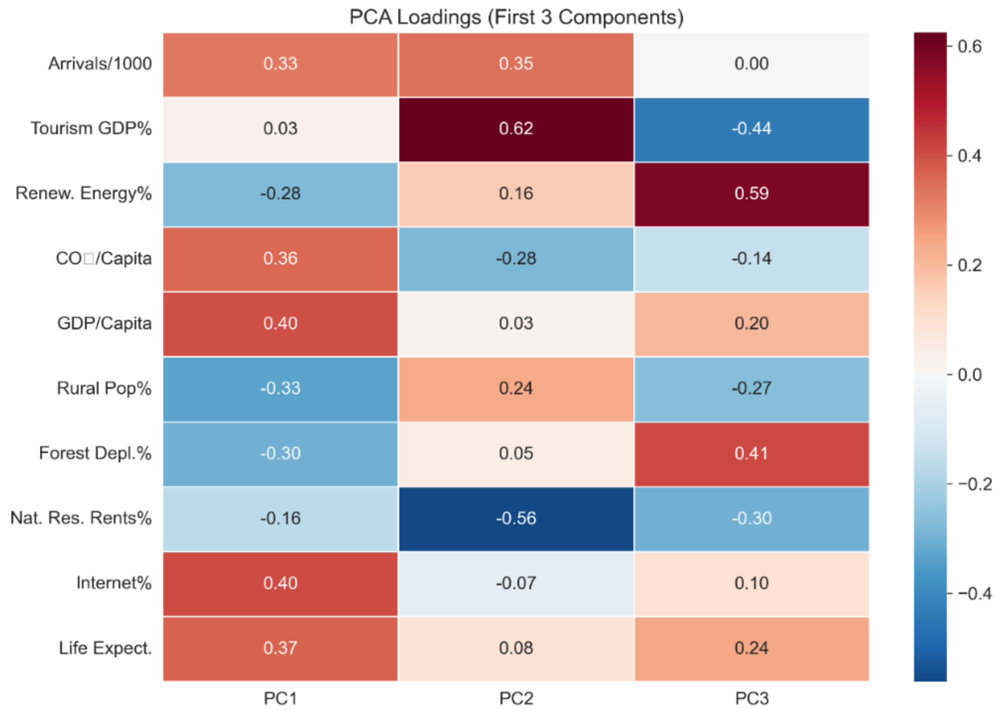

Figure 8: PCA loadings for the first three principal components. PC1 captures development level, PC2 contrasts tourism dependence with resource extraction and PC3 captures green energy orientation.

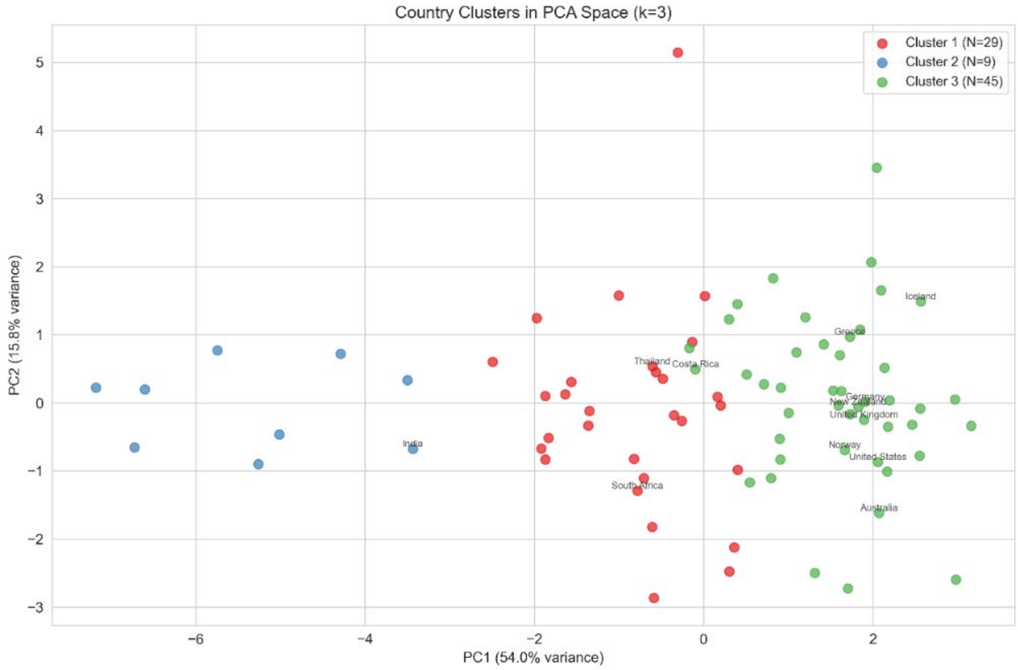

Figure 9: Country clusters projected onto the first two principal components. Cluster sizes are indicated in the legend.

Cluster 1 – "Emerging Middle" (29 countries): Upper-middle-income economies (e.g., Colombia, Egypt) with moderate tourism (≈400 per 1,000), low renewable use (18%), and growing but not dominant tourism sectors. **Cluster 2** – "Nature-Rich, Low-Infrastructure" (9 countries): Low-income countries (e.g., India, Rwanda) with high renewables (72%) and rural populations, but very low tourism (117 per 1,000) due to limited infrastructure. **Cluster 3** – "Developed, High-Tourism" (45 countries): High-income economies (e.g., Australia, France) with the highest tourism (1,648 per 1,000), strong connectivity, but lower renewable shares and higher emissions, where tourism is driven by infrastructure and competitiveness (Table 3).

Table 3: Mean values of clustering variables by cluster. Based on most recent available year per country (83 countries with complete data).

Indicator	Cluster 1 (Emerging)	Cluster 2 (Nature-Rich)	Cluster 3 (Developed)
N	29	9	45
Arrivals per 1,000	400	117	1,648
Tourism GDP (%)	5.8	3.4	5.2
Renewable energy (%)	18.4	71.9	21.6
CO_2 per capita (t)	4.2	0.5	8.8
GDP per capita (USD)	6,200	1,382	37,245
Rural population (%)	41.5	70.7	21.4
Forest depletion (%)	0.08	2.28	0.06
Internet (%)	59.4	18.6	85.2
Life expectancy (years)	72.9	62.6	79.7

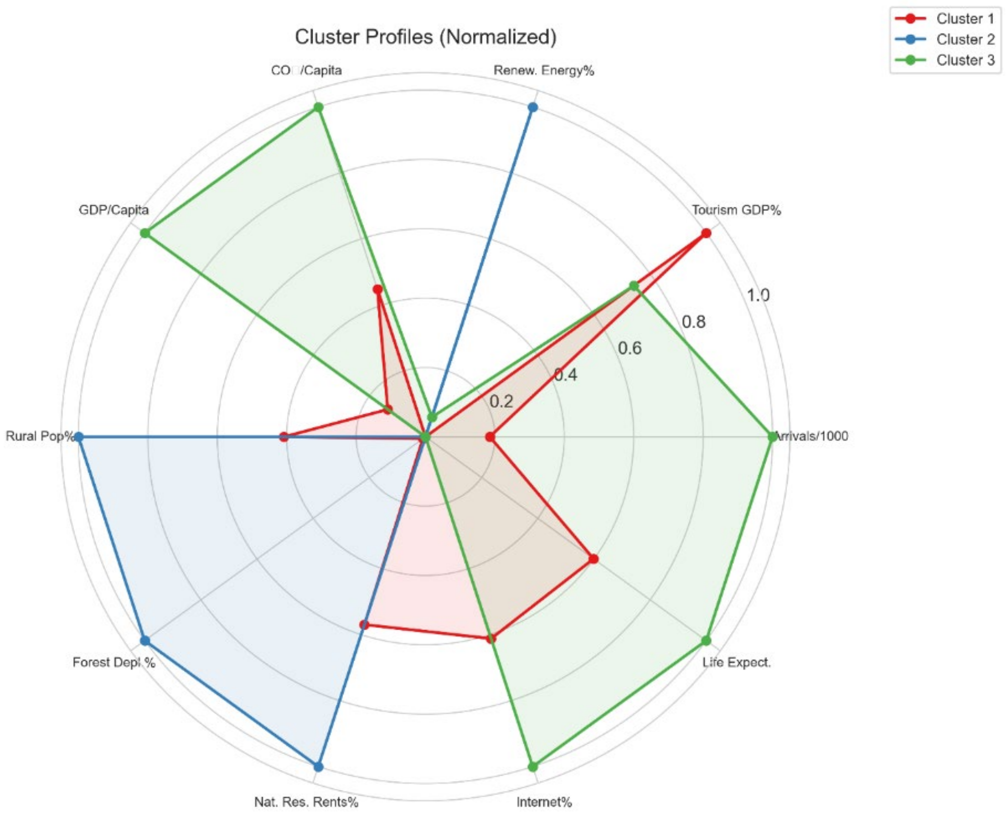

Figure 10: Normalised radar chart of cluster profiles. Each axis is scaled 0 to 1 within the variable's range across clusters.

The radar chart (Figure 10) visualises these profiles on a normalised 0 to 1 scale. Cluster 2 (blue) dominates on renewable energy, rural population and forest depletion but collapses on tourism, GDP, and internet. Cluster 3 (green) leads on arrivals, GDP, life expectancy and internet but lags on renewable energy and rural population. Cluster 1 (red) sits between the two on nearly every dimension.

To assess the relative importance of sustainability factors in predicting tourism intensity and to capture nonlinear relationships that OLS may miss, we trained Random Forest and Gradient Boosting regressors using 12 features. Table 4 summarises the 5-fold cross-validated performance.

Table 4: Cross-validated model performance (5-fold CV). MAE and RMSE are in units of arrivals per 1,000.

Model	R-squared (mean)	R-squared (std)	MAE	RMSE
Linear Regression	0.308	0.043	422.7	n/a
Random Forest	0.912	0.031	134.7	283.0
Gradient Boosting	0.947	0.015	111.0	216.4

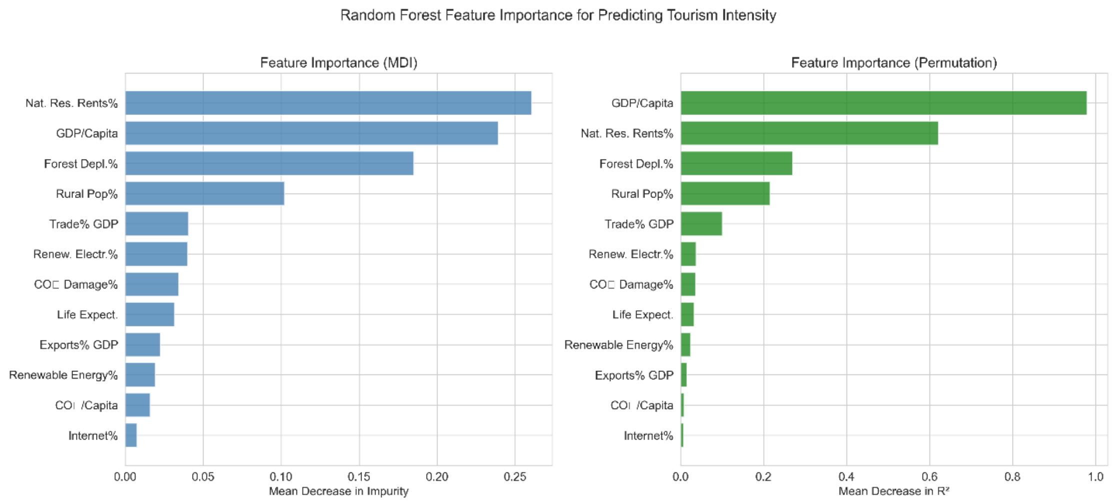

Figure 11: Feature importance from Random Forest, measured by mean decrease in impurity (left) and permutation importance (right).

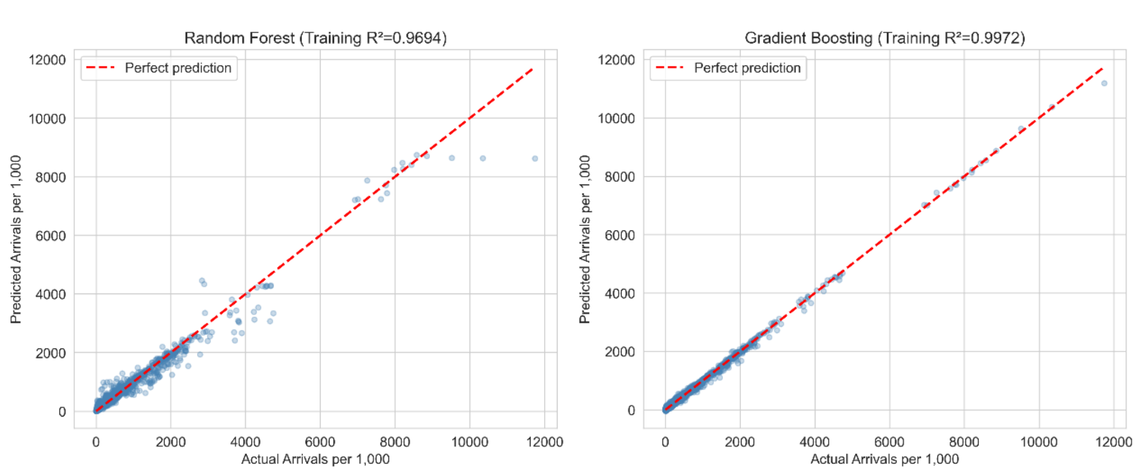

Figure 12: Actual versus predicted tourism intensity for Random Forest and Gradient Boosting. The dashed red line indicates perfect prediction.

Feature importance (Figure 11) identifies GDP per capita as the dominant predictor of tourism intensity, followed by natural resource rents and forest depletion. Both measures confirm the key role of development and resource structure, while renewable energy has a weaker, indirect effect.Model performance (Figure 12) shows that tree-based methods outperform linear regression, capturing strong nonlinearities. Gradient Boosting achieves high accuracy ($R^2 \approx 0.95$), highlighting the importance of interactions and threshold effects.

Figure 13 shows a consistently negative relationship between the Nature Sustainability Index and tourism intensity ($r \approx -0.23$ to -0.16), significant in most years. The absence of a clear trend indicates a stable pattern: countries with higher sustainability levels tend to attract fewer tourists per capita over time.

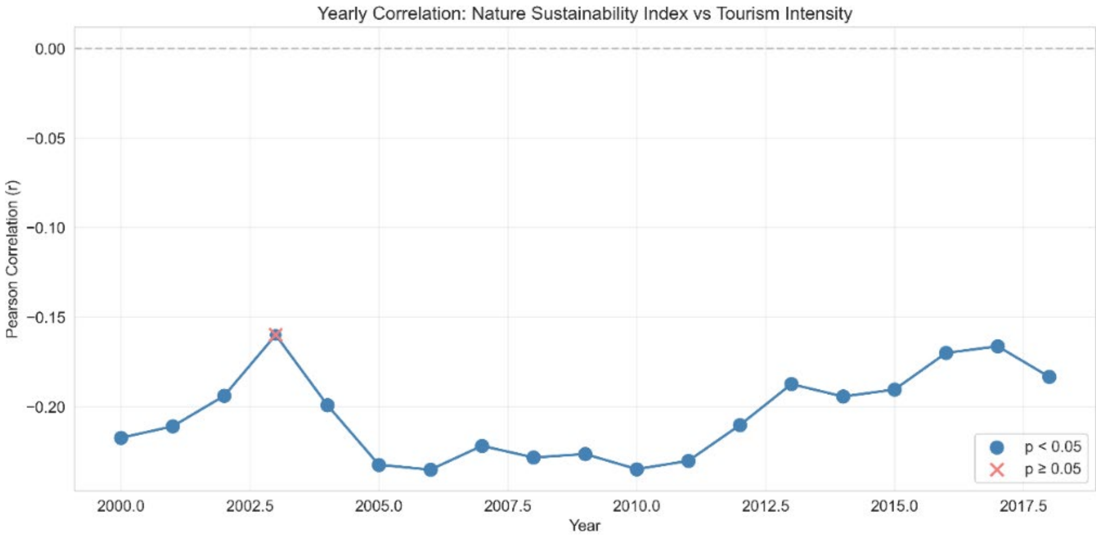

Figure 13: Annual Pearson correlation between the Nature Sustainability Index and tourism intensity. Blue dots indicate statistical significance at p < 0.05, red crosses indicate non-significant years.

Regional trends (Figure 14) show divergence: Europe's tourism intensity doubled (~700 to >1,400), Africa grew slowly (50 to 100), Asia tripled with stable renewables, and Oceania surged after 2010. Figure 15 indicates rising sustainability monitoring (2.3 to 3.0 tools), with higher levels linked to greater tourism intensity ($p < 0.001$), suggesting stronger investment or more supportive policy environments.

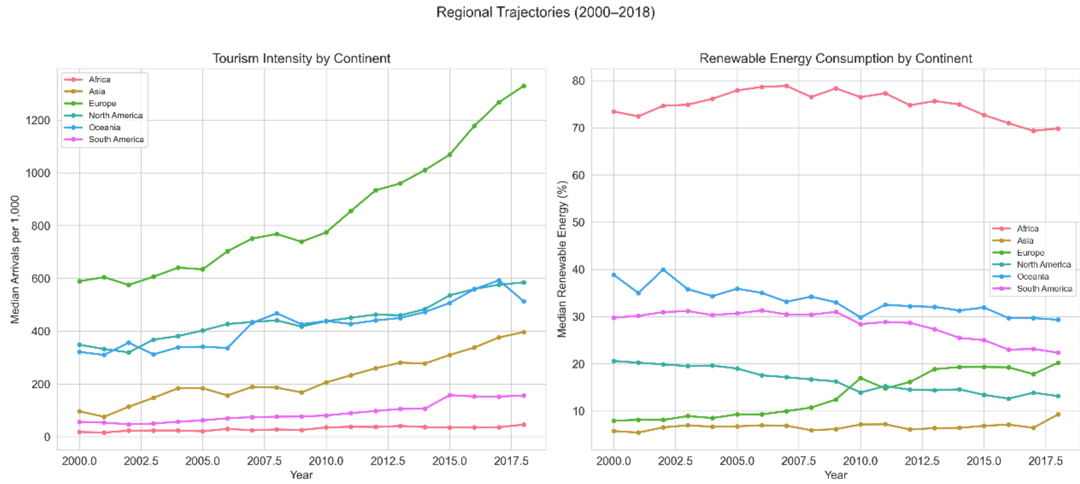

Figure 14: Regional trajectories of tourism intensity and renewable energy consumption from 2000 to 2018. Lines represent continent-level medians.

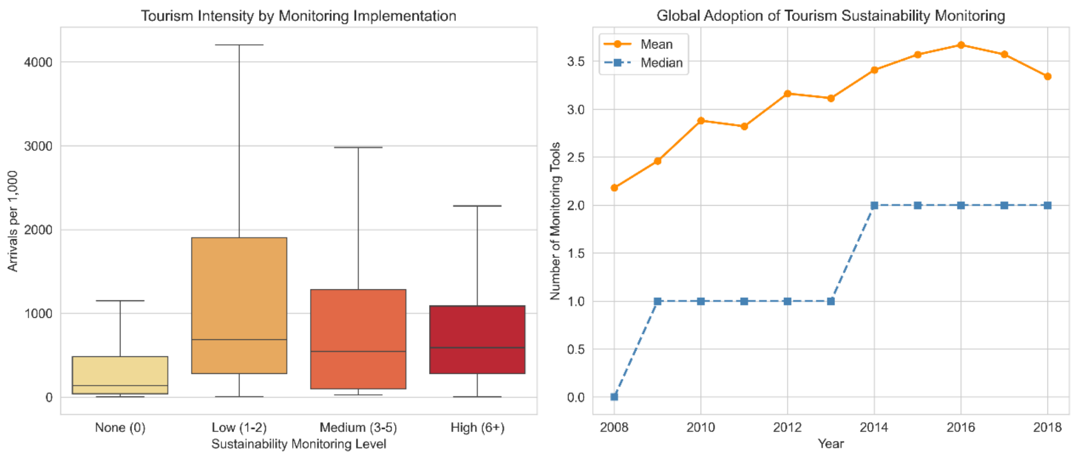

Figure 15: Left: tourism intensity by level of sustainability monitoring implementation. Right: global adoption of monitoring tools over time.

Figure 16 categorizes countries by sustainability (NSI) and tourism intensity into four groups. "High Sustainability–High Tourism" (e.g., Iceland, New Zealand) shows coexistence of both. "High Sustainability–Low Tourism" reflects strong natural assets but weak infrastructure, while "Low Sustainability–High Tourism" (e.g., France, USA, China) combines high tourism with lower environmental performance. "Low Sustainability–Low Tourism" includes resource-dependent and conflict-affected states. Top performers in the high–high group span a wide income range, indicating that success depends more on natural assets and environmental management than income (Table 5).

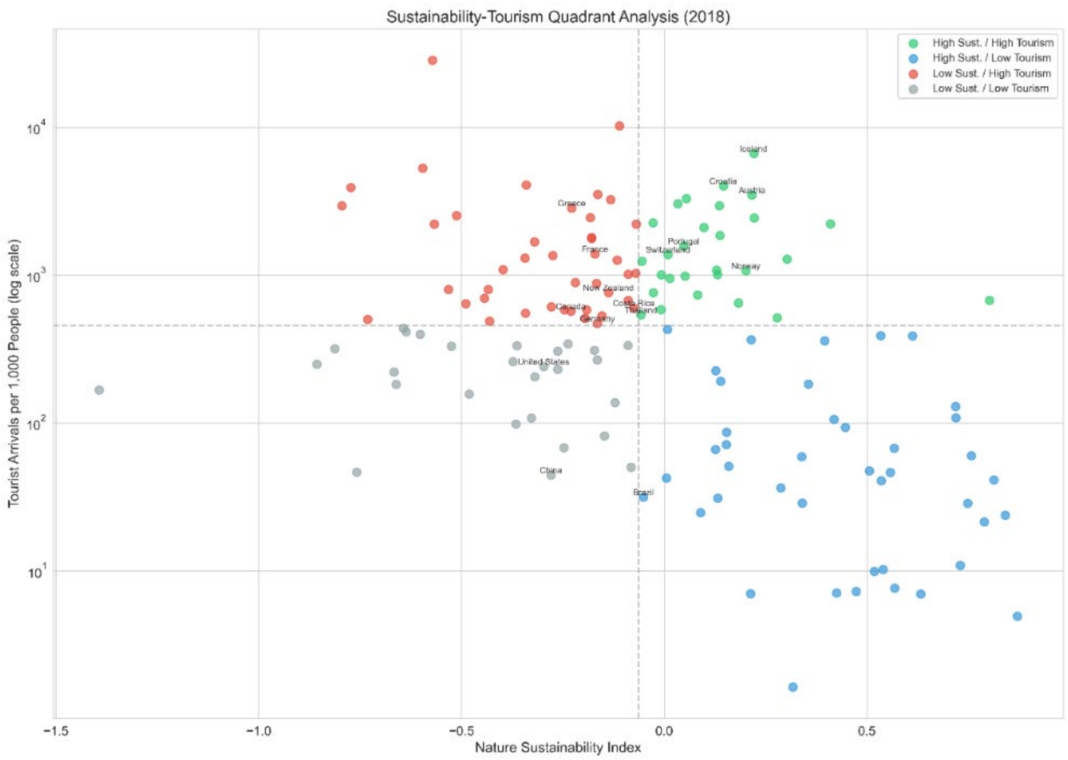

Figure 16: Sustainability-tourism quadrant analysis (2018). Dashed lines mark the median on each axis. Country labels identify notable cases.

Table 5: Top 15 countries in the "High Sustainability, High Tourism" quadrant, ranked by Nature Sustainability Index (2018).

Country	NSI	Arrivals per 1,000	Renewable Energy (%)	GDP per Capita (USD)
Eswatini	0.80	674	66.1	4,106
St. Lucia	0.41	2,220	10.2	11,358
Belize	0.30	1,280	40.1	5,001
Tonga	0.28	513	1.7	4,741
Barbados	0.22	2,431	5.8	17,745
Iceland	0.22	6,641	78.2	74,348
Austria	0.22	3,486	33.8	51,453
Norway	0.20	1,071	60.8	82,268
Namibia	0.18	647	30.3	5,588
Croatia	0.15	4,001	32.8	15,014
Albania	0.14	1,856	38.3	5,284
Antigua and Barbuda	0.14	2,935	0.9	16,673
Latvia	0.13	1,005	41.3	17,850
Mauritius	0.13	1,080	9.2	11,208
Slovenia	0.10	2,101	21.0	26,103

The merged panel contains 3,287 country-year observations, but the tourism indicators are reported unevenly across countries and years, so the analytic samples underlying each method are necessarily smaller than the full panel. International arrivals and the derived arrivals-per-1,000 measure are the best covered, present for 85.0% of observations, followed by air passengers (81.6%) and inbound personal trips (70.5%). Coverage thins considerably for the indicators that national statistical offices began reporting only recently: average length of stay is available for 54.7% of observations, the count of sustainability monitoring tools for 46.8%, outbound departures for 44.9%, tourism employment for 32.2%, and tourism's share of GDP for only 21.1%. Table 6 summarises this coverage. Because the regression, clustering, and machine-learning stages each require complete cases across their own variable sets, missing values were handled by listwise deletion rather than imputation: an observation enters a given analysis only if every variable that analysis uses is observed. Imputation was deliberately avoided, because the missingness is concentrated in specific low-capacity countries and years and is therefore unlikely to be missing at random, so model-based imputation would fabricate structure rather than recover it.

Table 6: Coverage of the tourism indicators across the 3,287 country-year observations in the merged panel.

Indicator	Observations	Coverage (%)
International arrivals	2,793	85.0
Arrivals per 1,000	2,793	85.0
Air passengers carried	2,681	81.6
Inbound personal (leisure) trips	2,316	70.5

Indicator	Observations	Coverage (%)
Average length of stay	1,798	54.7
Sustainability monitoring tools	1,537	46.8
Outbound departures	1,476	44.9
Tourism employment per 1,000	1,058	32.2
Tourism GDP share	693	21.1

Listwise deletion improves internal validity at the cost of external validity, because the countries that survive into the analytic sample are not a random subset of the panel. The pooled regression sample of 2,406 observations skews toward richer and better-connected economies than the observations that are dropped: median GDP per capita is 4,737 dollars among retained observations versus 3,124 dollars among those excluded, median internet penetration is 26.4% versus 20.1%, and median renewable energy consumption is 24.1% versus 17.4%. Tourism intensity itself is almost identical across the two groups (median 267 versus 279 arrivals per 1,000), which suggests that the selection operates through development and reporting capacity rather than through the outcome, and that it is more likely to attenuate the estimated sustainability-tourism associations than to inflate them. The results that follow should nonetheless be read as descriptive of the better-documented portion of the world's economies rather than of every country equally.

The pooled specifications reported earlier treat every country-year as an independent observation, which ignores the panel structure of the data and risks conflating differences between countries with changes within them. Two features of the data make this a genuine concern. Several of the key regressors, notably renewable energy consumption and rural population share, move slowly within a country across the 2000 to 2018 window, so almost all of their variance is cross-sectional. At the same time, tourism intensity is shaped by durable national characteristics such as coastline, climate, cultural heritage, and geographic centrality that are omitted from the model and correlated with the sustainability indicators. To separate the between-country from the within-country signal, the model was re-estimated as a one-way country fixed-effects model, a two-way model that adds year effects, and a random-effects model, all with standard errors clustered by country. Table 7 and Figure 17 report the results. Clustering by country is more conservative than the heteroskedasticity-robust errors used earlier, and it already changes the pooled picture: with panel-appropriate errors, the CO2-per-capita and forest-depletion coefficients are no longer distinguishable from zero, leaving renewable energy, rural population, natural resource rents, and income as the robust pooled correlates.

*Table 7: Panel estimates of log tourism intensity (arrivals per 1,000), standard errors clustered by country. Fixed-effects and two-way models report the within R-squared. Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.*

Predictor	Pooled OLS	Country FE	Two-way FE	Random Effects
Renewable energy (%)	-0.0166***	-0.0014	-0.0013	-0.0042
CO_2 per capita	-0.0134	-0.0047	0.0016	-0.0043
Rural population (%)	0.0150***	-0.0036	0.0040	-0.0033
Forest depletion (%)	0.0321	0.0109	0.0094	0.0105

Predictor	Pooled OLS	Country FE	Two-way FE	Random Effects
Natural resource rents (%)	-0.0330***	0.0014	0.0024	-0.0006
Log GDP per capita	0.9516***	0.4256***	0.4151***	0.4433***
Internet penetration (%)	-0.0021	0.0032**	-0.0015	0.0030**
Life expectancy	-0.0067	0.0209	-0.0012	0.0207*
R-squared	0.698	0.458	0.331	0.478
N	2,406	2,406	2,406	2,406

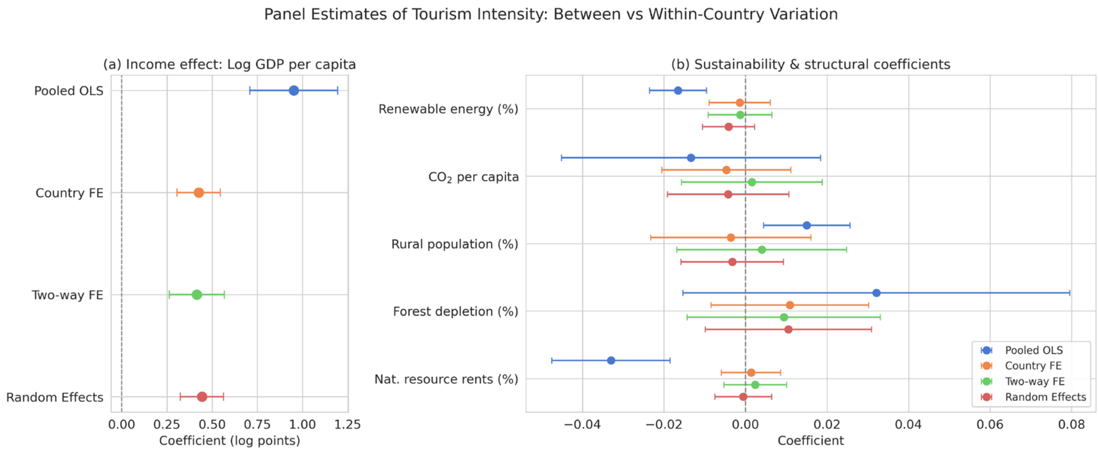

Figure 17: Panel estimates of tourism intensity across four specifications. Panel (a) shows the income effect and panel (b) the sustainability and structural coefficients, with 95% confidence intervals from country-clustered standard errors. Coefficients that are significant under pooled OLS become indistinguishable from zero once country fixed effects are included.

The contrast between the columns is stark. In the pooled model, renewable energy consumption, rural population share, and natural resource rents are all strongly associated with tourism intensity, and log GDP per capita carries a near-unit elasticity of 0.95. Once country fixed effects absorb the time-invariant national characteristics, every one of the sustainability coefficients collapses toward zero and loses significance: the renewable energy coefficient falls from -0.017 to -0.001, natural resource rents from -0.033 to +0.001, and rural population from +0.015 to -0.004. Only log GDP per capita survives the within-country transformation, and even it more than halves, to 0.43, indicating that a country attracts more visitors as it grows richer but that the effect is far smaller than the raw cross-country gradient implies. The two-way model, which additionally removes common annual shocks, tells the same story. A Hausman test rejects the random-effects specification in favour of fixed effects (chi-squared of 22.6 on 8 degrees of freedom, $p = 0.004$), confirming that the country effects are correlated with the regressors and that the pooled and random-effects estimates are biased by unobserved heterogeneity. The substantive implication is easy to miss but important: the association between environmental sustainability profiles and tourism is almost entirely a between-country phenomenon. Countries with greener profiles differ from high-tourism countries in many fixed ways, but a country becoming greener over time is not, in these data, associated with any change in the number of visitors it receives.

The predictive models reported earlier were evaluated with a five-fold cross-validation that shuffled all country-year observations at random before partitioning them. For panel data this procedure leaks information, because observations from the same country in adjacent years are highly similar and random shuffling all but guarantees that most countries appear in both the training and the test folds. A model can then score well simply by memorising each country's characteristic level of tourism rather than by learning any generalisable relationship. To measure how much of the reported performance reflects genuine prediction rather than leakage, the Random Forest and Gradient Boosting models were re-evaluated under two leakage-resistant schemes alongside the original: a country-blocked scheme, in which entire countries are held out so that the test set contains only nations never seen in training, and a forward-chaining scheme, in which the models are trained on earlier years and tested on later ones. Table 8 and Figure 18 report the results.

Table 8: Cross-validated R-squared for tourism intensity under three validation schemes. Negative values indicate performance worse than predicting the sample mean.

Model	Naive random K-fold	Country-blocked	Forward-chaining
Linear regression	0.31	0.19	0.28
Random Forest	0.91	-0.38	0.82
Gradient Boosting	0.95	-0.35	0.86

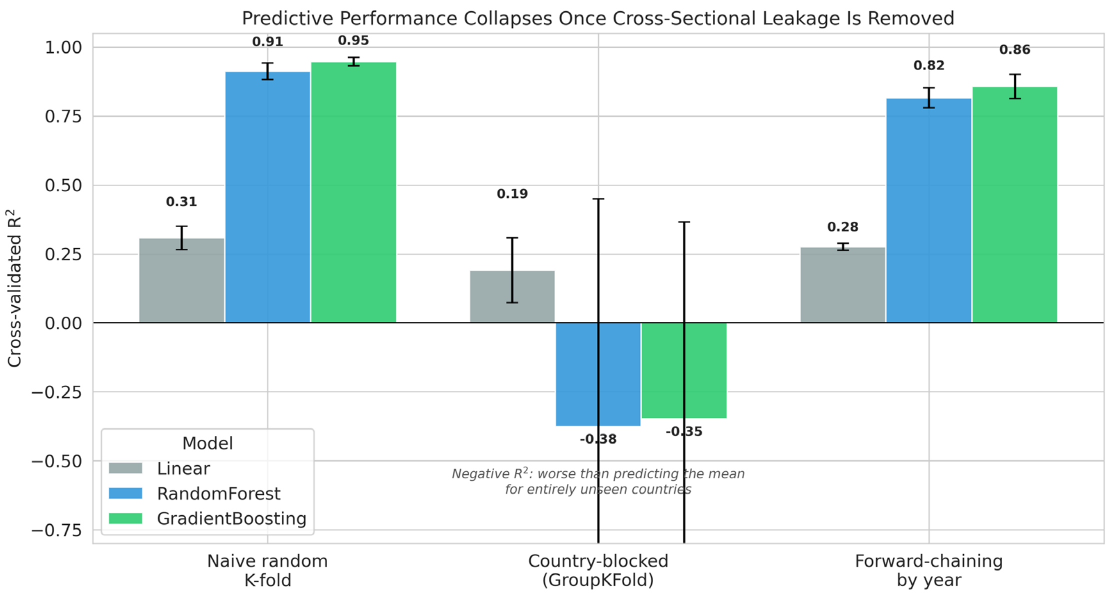

Figure 18: Cross-validated R-squared under naive random, country-blocked, and forward-chaining validation. Performance collapses below zero for both ensembles once entire countries are held out, showing that the naive score was inflated by cross-sectional leakage.

Under the naive random scheme the ensembles appear excellent, with Gradient Boosting reaching an R-squared of 0.95 and Random Forest 0.91. The country-blocked scheme dismantles this result. Asked to predict countries they have never encountered, both ensembles return negative R-squared values, -0.35 for Gradient Boosting and -0.38 for Random Forest, meaning they do worse than a naive prediction of the overall mean. The linear model, which cannot overfit country-specific idiosyncrasies to the same degree, degrades far more gracefully, from 0.31 to 0.19. The forward-chaining scheme, by contrast, preserves most of the ensembles' apparent skill (0.86 and 0.82), because holding out future years still lets the models learn each country's level from its earlier observations. Read together, the three columns localise the source of the original performance precisely: the models predict well when they have seen a country before, whether in a random fold or in an earlier year, and predict essentially nothing when the country is new. The earlier statement that sustainability features explain roughly 95% of the variance in tourism intensity must therefore be heavily qualified. That figure reflects the models' ability to recognise familiar countries, not a transferable link between a country's sustainability profile and the volume of tourism it attracts, which the country-blocked results show to be weak.

Because the Nature Sustainability Index is a constructed composite, its usefulness depends on whether the conclusions drawn from it survive reasonable changes in how it is built. Seven alternatives were compared with the baseline, which averages the four z-scored components with equal weight. The alternatives replace equal weighting with data-driven weights from the first principal component of the components, substitute min-max and robust (median and interquartile-range) normalisation for z-scoring, and, in four leave-one-out variants, drop each component in turn. For every version the index was correlated with the baseline, and its relationship with tourism intensity was re-estimated overall and within each income group. Table 9 and Figure 19 summarise the exercise.

Table 9: Sensitivity of the Nature Sustainability Index. Columns report each variant's correlation with the baseline index and its Pearson correlation with log tourism intensity, overall and by income group.

NSI variant	r vs baseline	Overall	Low	Lower-mid	Upper-mid	High
Baseline (equal, z-score)	1.00	-0.42	-0.10	-0.17	+0.36	+0.35
PCA-weighted	0.63	-0.60	-0.20	-0.22	+0.34	+0.34
Min-max normalisation	0.88	-0.62	-0.22	-0.25	+0.31	+0.33
Robust (median/IQR)	0.12	+0.31	+0.09	+0.06	+0.21	+0.23
Drop renewable energy	0.88	-0.11	-0.03	-0.04	+0.49	+0.40
Drop rural population	0.86	-0.19	-0.06	-0.17	+0.08	+0.15
Drop CO_2 damage	0.91	-0.51	-0.02	-0.25	+0.33	+0.34
Drop forest depletion	0.87	-0.54	-0.24	-0.18	+0.36	+0.35

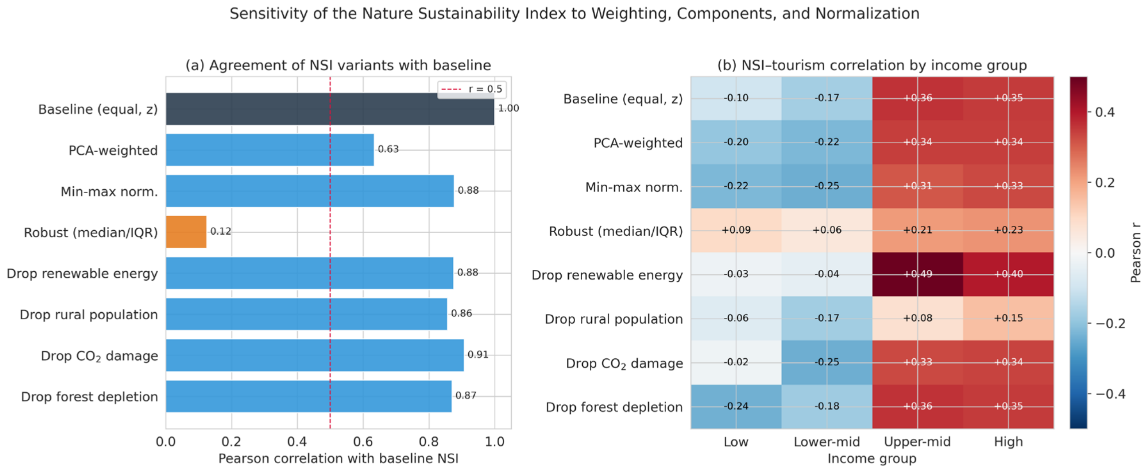

Figure 19: Sensitivity of the Nature Sustainability Index. Panel (a) shows how strongly each variant correlates with the baseline index, panel (b) shows the index-tourism correlation within each income group, where the shift from blue (negative) to red (positive) across income groups reproduces the reversal under almost every construction.

Two conclusions emerge. The cardinal values of the index are moderately sensitive to construction choices, but its central qualitative finding is not. Five of the seven alternatives correlate between 0.86 and 0.91 with the baseline, and the principal-component version correlates 0.63, so equal weighting, data-driven weighting, min-max scaling, and dropping any single component all rank countries similarly. Robust scaling is the exception, correlating only 0.12 with the baseline: because the components carry heavy outliers, rescaling by the median and interquartile range compresses them so severely that it produces a different index and even reverses the overall sign. More important is the income-stratified pattern, which is the study's headline result. The reversal from a negative sustainability-tourism correlation in low-income and lower-middle-income countries to a positive one in upper-middle-income and high-income countries holds under every variant except robust scaling, and even there the gradient still rises monotonically across income groups, from +0.09 to +0.23. The one component whose removal materially weakens the reversal is rural population: dropping it flattens the upper-middle-income correlation from +0.36 to +0.08, which indicates that the rural dimension carries much of the signal linking nature-oriented profiles to tourism in middle-income economies. The reversal is therefore a robust feature of the data rather than an artefact of a particular index recipe, although its magnitude in the upper-middle-income bracket depends noticeably on the rural component.

Three further specifications probe the robustness of the pooled associations. Adding continent fixed effects to the baseline model leaves the sustainability coefficients essentially unchanged, with renewable energy moving from -0.0165 to -0.0169 and natural resource rents from -0.030 to -0.027, both still significant at the 1% level, so the associations are not an artefact of broad regional groupings. Replacing the dependent variable with tourism's share of GDP, a measure of economic dependence on tourism rather than of visitor volume, changes the picture instructively: log GDP per capita, which dominates the arrivals model, becomes negligible and statistically insignificant (0.03), while natural resource rents remain significantly negative (-0.026). The drivers of how many tourists a country attracts are therefore not the same as the drivers of how economically dependent it is on them, and resource-extraction economies score low on both. None of these variations disturbs the two central results: the dominance of income in the cross-section, and its attenuation within countries.

5 Discussion

The empirical results, taken together with the robustness analyses, support a coherent but nuanced account of how nature-based sustainability and tourism relate at the national level. This section draws the findings together, interprets what the panel and predictive analyses reveal about the nature of the relationship, weighs the threats to a causal reading, and translates the patterns into differentiated implications for countries at different stages of development. Three findings anchor the study. At the cross-country level the relationship between nature-oriented sustainability and tourism intensity is, on average, negative: the greenest economies, measured by renewable energy use, rural population, and low carbon and forest damage, tend to receive fewer visitors per capita, and this negative correlation is stable across the entire 2000 to 2018 window. This average, however, conceals a systematic reversal across the income distribution. The correlation is negative among low-income and lower-middle-income countries but turns clearly positive among upper-middle-income and high-income economies, a pattern that survives almost every alternative construction of the sustainability index. Income, finally, is the dominant structural correlate of tourism intensity in the cross-section, yet its apparent power shrinks sharply once within-country variation is isolated, and the sustainability indicators carry no robust within-country signal at all.

The divergence between the pooled and fixed-effects estimates is the most consequential methodological result of the study, and it reframes how the headline correlations should be read. That renewable energy, rural population, and resource rents predict tourism across countries but not within them means the associations are carried by stable national attributes rather than by anything that changes as a country's sustainability profile evolves. This is consistent with the environmental Kuznets curve logic that motivates much of the tourism-sustainability literature: countries do not slide smoothly along a single sustainability-tourism curve but occupy durable positions determined by geography, history, and development, which jointly shape both their environmental profile and their attractiveness to visitors. A mountainous, forested, low-emission economy and a coastal, urbanised, high-emission one differ in ways that no single year's policy change will alter. The practical lesson is one of caution. Cross-sectional correlations between sustainability and tourism, which dominate the existing literature and which this study also finds, should not be read as evidence that improving a country's environmental profile will change its tourism outcomes, the within-country evidence for such a mechanism is, in these data, absent.

The collapse of predictive performance under country-blocked validation reinforces the same message from a machine-learning angle. High cross-validated accuracy is often taken as evidence that a set of features captures the drivers of an outcome, but that inference holds only when the validation respects the structure of the data. Here the ensembles' apparent ability to explain 95% of the variance in tourism intensity evaporated when they were asked to predict countries they had not seen, which shows they had largely learned to recognise individual countries rather than to map sustainability characteristics onto tourism. This is a useful corrective for the growing body of tourism research that applies flexible machine-learning models to panels of countries or regions without blocking on the cross-sectional unit. It does not render the models worthless: the forward-chaining results show they can forecast a known country's near future reasonably well, which is genuinely useful for a destination monitoring its own trajectory. It does mean that claims about the transferable, cross-country predictive value of sustainability indicators cannot rest on the naive scores.

None of the estimates in this study should be interpreted as causal, and several distinct pathways stand between the observed associations and any causal claim. Reverse causation is the most obvious: tourism itself raises incomes, alters land use, and changes a country's energy and emissions profile, so the correlation between carbon intensity and tourism may partly reflect tourism's effect on emissions rather than the reverse. Simultaneity compounds this, because tourism and the sustainability indicators are jointly determined by the development process. Omitted variables are pervasive and, as the fixed-effects results show directly, quantitatively decisive: coastline and climate, proximity to large source markets, visa and aviation policy, political stability, and the quality of destination governance all plausibly drive tourism while correlating with the environmental measures, and the wide gap between the pooled and within-country estimates is precisely the footprint of such confounders. Measurement error is a further concern, since the composite index and several of the tourism series are imperfectly and unevenly recorded. Credible identification would require leverage this design does not provide: plausibly exogenous shifters of tourism, such as changes in visa regimes, new air routes, or natural disasters affecting destinations, could serve as instruments, and a difference-in-differences design around discrete policy changes would isolate within-country effects more convincingly than the fixed-effects estimator alone. Until such strategies are brought to bear, the patterns documented here are best read as a careful description of association, not as a warrant for causal policy claims.

The income-stratified reversal, because it is robust, is also the most policy-relevant result, and it implies that the same environmental endowment carries very different meaning depending on a country's stage of development. For low-income and lower-middle-income countries, where the sustainability-tourism correlation is negative or negligible, abundant natural capital is not currently a lever for tourism growth, because it is not the binding constraint. The priority in these economies is the enabling infrastructure that natural capital cannot substitute for: air and road connectivity, reliable electricity and water, health and safety provision, visa facilitation, and the basic institutional stability that international visitors require. Environmental branding is premature where a destination cannot yet be reached or trusted. For upper-middle-income countries, where the correlation turns positive and where the rural dimension matters most, natural capital does begin to convert into visitors, and the returns to deliberate nature-based positioning are highest: investment in protected-area management, ecotourism product development, and the sustainability-monitoring systems whose adoption this study associates with higher tourism intensity can compound an advantage the market is already rewarding, provided growth is managed to avoid the carbon lock-in and overdevelopment visible in the high-income cluster. For high-income countries, where the positive correlation weakens as heritage, urban amenity, and brand come to the fore, the sustainability agenda shifts from attraction to preservation and differentiation: decarbonising the aviation and accommodation that dominate the sector's footprint, managing the overtourism that erodes the very assets visitors come for, and using demonstrable environmental performance as a mature-market differentiator. The common thread across all groups is that nature-based leisure is not an automatic dividend of environmental virtue, it is realised only where the complementary infrastructure and institutions exist to convert natural capital into a visitable, marketable, and well-managed destination.

6 Conclusions

This study examined the relationship between nature-related sustainability and tourism intensity across 173 countries over the period 2000–2018. By combining tourism, environmental, economic and development indicators, the analysis explored whether nature-based leisure and environmental sustainability can operate as drivers of sustainable tourism development. The findings show that this relationship is neither simple nor linear. At the global level, countries with stronger nature-related sustainability profiles tend to record lower tourism intensity, while many high-tourism destinations are high-income economies with stronger infrastructure, better connectivity and higher institutional capacity, but also higher emissions and weaker relative dependence on renewable energy. This suggests that environmental sustainability alone does not automatically generate tourism competitiveness.

The results indicate that GDP per capita remains the strongest and most consistent predictor of tourism intensity. This confirms the central role of development conditions, including infrastructure, transport accessibility, digital connectivity, institutional capacity and market visibility. Environmental indicators also matter, but their influence depends strongly on the broader development context. The income-group analysis reveals an important threshold effect: the relationship between the Nature Sustainability Index and tourism intensity is negative in low- and lower-middle-income countries, but becomes positive in upper-middle-income countries. This finding suggests that nature-based sustainability can become a tourism advantage only when countries have sufficient economic, institutional and infrastructural capacity to transform natural assets into accessible, well-managed and attractive tourism products.

The clustering analysis reinforces this interpretation by identifying three broad country profiles: developed high-tourism economies, nature-rich but low-infrastructure countries, and emerging middle-income destinations. These profiles show that nature-rich countries often possess strong environmental potential but may lack the governance mechanisms, investment capacity and tourism infrastructure required to convert this potential into sustainable tourism development. Conversely, developed destinations benefit from accessibility, digital visibility and institutional capacity, but they also face higher environmental pressures. Therefore, the central challenge is not simply to increase tourism flows, but to align tourism growth with environmental protection, resource efficiency and long-term destination resilience. The machine learning results further demonstrate the complexity of the sustainability–tourism nexus. Random Forest and Gradient Boosting models substantially outperformed linear regression, indicating that tourism intensity is shaped by nonlinear relationships and interactions among income, resource dependence, environmental conditions and development indicators. However, these results should be interpreted as predictive and exploratory rather than causal. High predictive performance does not prove that one variable directly causes another, rather, it shows that tourism development is embedded in a multidimensional system where economic, environmental and institutional factors interact in complex ways.

The study also highlights the importance of data availability and sample construction. Although the full merged panel contains 3,287 country-year observations, several tourism indicators have uneven coverage across countries and years. As a result, the regression, clustering and machine-learning analyses rely on smaller complete-case samples. Listwise deletion improves the comparability and internal consistency of each analysis because all variables used in a model are observed for the same cases. However, it also limits external validity, since the countries that remain in the analytic samples are unlikely to represent a random subset of the full global panel. This means that the results are strongest for countries with better statistical reporting capacity and should be interpreted with caution when generalizing to low-capacity or underreported tourism systems.

The study concludes that nature-based leisure can contribute to sustainable tourism, but only under specific enabling conditions. Sustainability becomes a competitive advantage when it is combined with infrastructure, institutional capacity, environmental monitoring, digital connectivity, local participation and effective destination governance. For nature-rich developing economies, the policy priority is not only to preserve environmental assets, but also to invest in sustainable infrastructure, protected-area management, accessibility, digital promotion and community-based tourism models. For high-tourism destinations, the priority is different: they must reduce environmental pressure, improve carbon efficiency, strengthen sustainability monitoring and protect the natural resources on which their long-term competitiveness depends. Therefore, sustainable tourism should be understood not as a direct outcome of natural capital, but as the result of a coordinated development model that connects environmental quality with governance, infrastructure and responsible tourism management.

7 Limitations

Despite its broad geographical coverage and mixed-method empirical design, this study has several limitations. The analysis is based on country-level data, which allows comparison across a large number of countries and years, but also hides important regional and local differences within countries. Tourism and sustainability pressures are often concentrated in specific destinations, islands, coastal zones, mountain regions, protected areas or urban centres. Therefore, national averages may underestimate local problems such as overtourism, biodiversity loss, water stress, land-use conflicts, seasonal overcrowding and pressure on local infrastructure. Another limitation concerns the construction of the Nature Sustainability Index. The index provides a useful summary measure of nature-related sustainability, but it is necessarily simplified. It includes renewable energy consumption, rural population share, CO_2 damage and forest depletion, but it does not fully capture other important dimensions of environmental sustainability, such as biodiversity, protected-area quality, waste management, water availability, air pollution, climate vulnerability, ecosystem degradation and environmental governance. In addition, the use of equal weighting improves transparency, but different weighting schemes could produce somewhat different results.

The study also covers the period 2000–2018. This time frame provides a stable pre-pandemic view of global tourism development, but it does not include the major disruptions and transformations that occurred after 2020. The COVID-19 pandemic, the uneven recovery of international travel, changing tourist preferences, climate-related risks and the increased use of digital technologies in tourism management may have altered the relationship between sustainability and tourism intensity. For this reason, the findings should be interpreted as evidence for the pre-pandemic period rather than as a complete explanation of the most recent tourism landscape. The empirical analysis is mainly associational and predictive rather than causal. The regression models identify important relationships between sustainability, income, resource dependence and tourism intensity, but they do not prove the direction of causality. Higher income may support tourism development, but tourism itself may also contribute to income growth. Similarly, sustainability conditions may influence tourism attractiveness, while tourism activity may also affect environmental quality. Although the study uses several analytical techniques and robustness-oriented approaches, stronger causal methods would be needed to confirm the mechanisms behind the observed relationships.

Data availability and comparability also represent important limitations. Some indicators have incomplete coverage across countries and years, especially tourism-related economic variables. Countries differ in statistical capacity, reporting practices and the quality of environmental and tourism data. As a result, the regression, clustering and machine-learning analyses rely on complete-case samples that are smaller than the full merged panel. Listwise deletion improves the comparability and internal consistency of each model, because all variables included in a given analysis are observed for the same cases. However, it may reduce external validity, since the countries that remain in the analytic samples are unlikely to constitute a random or fully representative subset of the global panel. This is especially important because missingness is more likely to affect countries with weaker statistical capacity. Finally, the machine learning models offer strong predictive performance, but their interpretation remains limited. Feature importance measures help identify influential variables, yet they do not provide causal explanations. Random Forest and Gradient Boosting models are useful for detecting nonlinear patterns, interactions and threshold effects, but their results should be understood as complementary to theory-driven analysis rather than as substitutes for causal interpretation.

8 Future Research

Future research should extend the dataset beyond 2018 in order to examine how the relationship between sustainability and tourism intensity has changed in the post-pandemic period. The years after 2020 are especially important because tourism recovery has been uneven across regions and destination types. Nature-based, rural and low-density tourism may have gained new importance, while climate change, energy transitions and digitalization may have reshaped the conditions under which sustainability becomes a tourism advantage. Further research should also develop richer and more multidimensional sustainability indicators. A more comprehensive index could include biodiversity, protected-area coverage, water stress, waste generation, air quality, climate vulnerability, land-use change and environmental governance. This would provide a more complete understanding of nature-based sustainability and reduce the risk of oversimplifying environmental performance. Future studies could also test alternative weighting methods in order to examine whether the results remain stable under different index construction strategies.

Another important direction is the use of regional, destination-level or protected-area-level data. Country-level analysis is useful for identifying global patterns, but tourism pressure is often local and highly concentrated. Future studies could examine specific coastal areas, islands, national parks, rural destinations or mountain regions in order to understand how tourism affects local ecosystems and communities. This would also make it possible to analyse seasonality, visitor concentration, infrastructure stress and local environmental carrying capacity more accurately. Future research could also apply stronger causal methods. Panel fixed-effects models, dynamic panel models, instrumental variables, difference-in-differences approaches and causal machine learning techniques could help clarify whether sustainability improves tourism competitiveness or whether tourism development changes sustainability outcomes. Such methods would allow future studies to move beyond association and prediction toward a clearer explanation of causal mechanisms.

The role of institutional capacity should also be examined more directly. This study suggests that infrastructure, governance and monitoring systems are crucial for transforming natural assets into tourism advantages. Future models could include indicators of government effectiveness, environmental regulation, digital readiness, transport infrastructure, protected-area management, tourism policy quality and sustainability certification systems. This would help explain why some nature-rich countries successfully develop sustainable tourism while others remain unable to convert environmental potential into tourism performance. Future studies should also give greater attention to tourist behaviour and demand-side factors. Sustainable tourism depends not only on destination characteristics, but also on tourists' environmental awareness, preferences, transport choices, willingness to pay for sustainable services and attitudes toward conservation. Survey data, online reviews, booking platforms and social media data could help connect macro-level sustainability indicators with micro-level tourist behaviour. Finally, future research could combine econometric analysis, machine learning and qualitative case studies. Quantitative models can identify broad international patterns, while case studies can explain the policy, institutional and cultural mechanisms behind successful sustainable tourism development. Comparative case studies of destinations that combine high tourism intensity with strong sustainability performance could offer deeper insight into how nature-based leisure can contribute to long-term sustainable tourism.

Funding

This research received no external funding.

Declaration of Originality

The author declares that this manuscript is an original work, has not been published previouslyand is not under consideration for publication elsewhere. All sources have been properly cited and referenced in accordance with academic standards.

CRediT authorship contribution statement

Georgios Vasileiadis: Conceptualization, Methodology, Data Curation, Formal Analysis, Statistical Analysis, Software, Validation, Visualization, Investigation, Literature Review, Writing – Original Draft, Writing – Review & Editing, Project Administration.

Declaration of competing interest

The author declares that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Statement

This study did not involve human participants, animals, clinical material, biological samples, or protected/non-public datasets. The research was based exclusively on the review, interpretation, and synthesis of existing literature and publicly available academic sources. Therefore, formal ethical approval from an institutional review board or ethics committee was not required. The study was conducted in accordance with academic integrity principles, responsible research practice, and proper citation standards.

Consent to Participate

Not applicable. This study did not involve human participants, interviews, surveys, experiments, or the collection of personal data. Therefore, informed consent to participate was not required.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request. The original data were obtained from publicly available sources cited in the manuscript.

References

Akhtar, N., Alharthi, M. F. & Khan, M. S. (2024). Mitigating Multicollinearity in Regression: A Study on Improved Ridge Estimators. *Mathematics*, *12*(19), 3027. https://doi.org/10.3390/math12193027

Arbulú, I., Deyà-Tortella, B., Rey-Maquieira, J. & Tirado, D. (2025). The Environmental Kuznets Curve, Water Stress, and Tourism: A European Analysis. *Water*, *17*(7), 1031. https://doi.org/10.3390/w17071031

Esparza-Huamanchumo, R. M., Botezan, I., Sánchez-Jiménez, R. & Villalba-Condori, K. O. (2024). Ecotourism, sustainable tourism and nature-based tourism: An analysis of emerging fields in tourism scientific literature. *Geojournal of Tourism and Geosites, 54*(2spl), 953–966. https://doi.org/10.30892/gtg.542spl19-1270

Geng, Y., Ji, S. & Zhang, J. (2026). Knowledge Mapping of Low-Carbon Tourism Research: Hotspot Evolution and Frontiers. *Land, 15*(5), 809. https://doi.org/10.3390/land15050809

Janchai, N. & Suvittawat, A. (2025). The Structural Equation Model of Factors Affecting Decision-Making on Low-Carbon Tourist Destinations. *Sustainability*, *17*(5), 2082. https://doi.org/10.3390/su17052082

Michelaki, D., Tsagris, M. & Adam, C. (2025). An Overview of Economics and Econometrics Related *R* Packages. *Stats*, *8*(4), 85. https://doi.org/10.3390/stats8040085

OECD. (2024). *Artificial intelligence and tourism: G7/OECD policy paper*. OECD Publishing. https://www.oecd.org/en/publications/artificial-intelligence-and-tourism_3f9a4d8d-en.html

Pousa-Unanue, A., Alzua-Sorzabal, A. & Femenia-Serra, F. (2026). Methodological Pathways for Measuring Tourism Carbon Footprint: A Framework-Oriented Systematic Review. *Climate, 14*(2), 28. https://doi.org/10.3390/cli14020028

Sobirov, Y., Artikov, B., Khodjaniyozov, E., Marty, P. & Saidmamatov, O. (2025). Economic Growth, FDI, Tourism, and Agricultural Productivity as Drivers of Environmental Degradation: Testing the EKC Hypothesis in ASEAN Countries. *Sustainability*, *17*(18), 8394. https://doi.org/10.3390/su17188394

Varzakas, I.-P. & Metaxas, T. (2024). Pandemic and Economy: An Econometric Analysis Investigating the Impact of COVID-19 on the Global Tourism Market. *Tourism and Hospitality*, *5*(2), 290-303. https://doi.org/10.3390/tourhosp5020019

Voumik, L. C., Md. Nafi, S., Bekun, F. V. & Haseki, M. I. (2023). Modeling Energy, Education, Trade, and Tourism-Induced Environmental Kuznets Curve (EKC) Hypothesis: Evidence from the Middle East. *Sustainability*, *15*(6), 4919. https://doi.org/10.3390/su15064919

World Bank. (2024). *World Development Indicators*. https://databank.worldbank.org/source/world-development-indicators

UN Tourism. (2024). *International tourism highlights, 2024 edition*. World Tourism Organization. https://www.e-unwto.org/doi/book/10.18111/9789284425808