Exploring How Green Marketing Influences Shipping Tourist Destination Choice: Evidence from R Programming-Based Logistic Regression and Roc Analysis

[1] Dr. M. A. Shakila Banu, [2] Pazila Sara. R. S

[1] Associate Professor, PG& Research Department of Commerce, Jamal Mohamed College (Autonomous) Trichy-20

[2] BE- Naval Architecture and Offshore Engineering, AMET University, Chennai

ABSTRACT

This study explores how green marketing influences the destination choice of shipping tourists, a segment where traveller's increasingly seek sustainable experiences, but empirical evidence remains limited. This quantitative research employed a structured questionnaire and used R-based statistical techniques, including logistic regression, ROC analysis, and mediation analysis, on a sample of 325 shipping tourists to predict destination choice. The analysis confirmed excellent instrument reliability (alpha = 0.90) and found strong positive correlations between Green Marketing Awareness and Destination Choice (r=0.832); however, the logistic regression model found no statistically significant individual effects for the predictors, and the Causal Mediation Analysis failed to detect any significant direct or indirect effects. While the model showed fair overall discrimination (AUC = 0.7845) and an adequate fit (Hosmer-Lemeshow p > 0.05), its practical utility is severely limited by poor sensitivity (only 0.92% correctly classified true positives); therefore, future research must re-evaluate the binary measurement of Destination Choice and incorporate unobserved factors to enhance the predictive power of the model and generalizability of findings.

Keywords: Green Marketing, Shipping Tourism, Destination Choice, Environmental Attitude, Logistic Regression

INTRODUCTION

Shipping tourism, including cruises and river/boat travel, has become increasingly popular in India and worldwide as tourists seek both leisure and sustainable experiences. In recent years, tourists have become more environmentally conscious and prefer destinations that follow eco-friendly practices (Akram & Lavuri, 2023). Green marketing, which includes promoting sustainability, eco-certifications, and environmental responsibility, plays a vital role in shaping tourists' perceptions of destinations (Chang et al., 2025). Studies show that eco-friendly messaging increases tourists' trust and positively influences their decision-making, making it an essential strategy for shipping tourism operators (Vicente et al., 2024).

However, there is limited research on how green marketing affects shipping tourist destination choice specifically. Understanding this influence can help cruise and boat-tourism operators develop strategies to attract environmentally conscious travellers while supporting sustainable tourism practices (Setiadi et al., 2024; Kumar & Kulshreshtha, 2023). This study uses R-based logistic regression to predict the probability of tourists choosing green-marketed shipping destinations and applies ROC analysis to measure the accuracy of the model. The findings aim to provide practical insights into how green marketing can enhance destination selection and promote eco-friendly behaviour among shipping tourists.

Review

• Suneel Kumar & Nisha Devi — AIMT Journal of Management, Vol. 12, Issue 1–2, 2023, The study examines the effect of green marketing tools (eco-labels, eco brands, eco advertisements) on green consumer behaviour among tourists in Himachal Pradesh, India. They find that increased use of green branding and eco-labels strongly correlates with tourists' willingness to pay and choose eco friendly tourism options. Journalskart

• Mona Karina, Rinto Rivanto, Hasan Basri & Endang Iryani — Branding: Jurnal Manajemen dan Bisnis, Vol. 4, Issue 1, 2025, This qualitative-literature study argues that green marketing can transform conventional tourism into sustainable tourism. The authors emphasise that just green messages are not enough — business actors must deliver them authentically and consistently, and work with government and local communities to succeed. Journal UINSgd

• Chien Chung Yu & Chun Chu Liu — International Journal of Religion, Vol. 5, Issue 4, 2024, pp. 224–230, This paper builds a green marketing model for the tourism and amusement industry using fuzzy analytic hierarchy process. It suggests that "brand loyalty" and "corporate image" are key indicators for green marketing in tourism. The authors argue that sustainable management demands that companies integrate green marketing into their core strategy. Ijor+1

• Sanjaya, D., Arief, M., Juli Setiadi, N. & Heriyati, P. — Journal of Eastern European and Central Asian Research (JEECAR), Vol. 11, Issue 3, 2024, pp. 553–572, The study explores how digital green marketing campaigns (online eco ads) and environmental beliefs shape tourist behavior and revisit intentions in eco-tourism. It finds that well-crafted digital campaigns significantly influence tourists' environmental beliefs, which in turn increase their commitment to revisit sustainable destinations. IEECA

• Suneel Kumar (Prof.) & Nisha Devi — AIMT Journal of Management, Vol. 12, Issue 1–2, 2023 (same as #1, but additional insight), Beyond demographics, they also analyse how green marketing tools influence pro-environmental behaviour and find that older or more educated tourists respond more strongly to eco brands and ads, implying that green marketing effectiveness may vary across tourist segments. (Derived from same paper) Journalskart

• Jumadi, Ascasaputra Aditya, Arya Saputra & Aldi Irsyad Burhani — Proceedings of the 2nd UPY International Conference on Education and Social Science, Atlantis Press, 2023, pp. 471 477, They use a regression model (via SPSS) to examine the effect of green marketing (product, place, promotion) on green tourism. They find that "product" (eco friendly offering) and "place" (green destination) significantly influence tourists' perception of green tourism, but "promotion" has weaker influence, highlighting that just advertising green is not enough. Atlantis Press

• Sri Rahayu, Hikmatul Aliyah & Sudarwati — International Journal of Economics, Business and Accounting Research (IJEBAR), Vol. 6, Issue 1, 2024, This study looks at green marketing components (green product, green promotion, green price) and their effect on intention to visit eco-tourism sites. They also test whether environmental knowledge mediates these relations. They find that green price and green promotion significantly influence visit intention, but green product does not, and environmental knowledge does not mediate the effect of green product. jurnal.stie-aas.ac.id

• Mdpi – Green and Environmental Marketing Strategies and Ethical Consumption — Sustainability, Vol. 15, Issue 16, 2023, This paper explores the role of green environmental strategy, green marketing, and psychological factors (perceived behavioral control, subjective norms) on ethical consumption in tourism. Using TPB (Theory of Planned Behavior), they find that green marketing strongly influences ethical consumption and that psychological control plays a moderating role.

Overview of Reliability Test (Cronbach's Alpha):

Reliability testing is used to check the consistency of a set of questionnaire items that measure the same concept. Cronbach's Alpha (α) is the most common reliability coefficient. It indicates how closely related a set of items

are as a group. The value of α ranges from 0 to 1, where a higher value (usually ≥ 0.7) shows good internal consistency. The formula for Cronbach's Alpha is:

r Cronbach's Alpha is:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} X \sigma_i^2}{\sigma_T^2}\right)$$

------------>1

Where,

k = Number of items in the scale,

σ_i^2 = Variance of each individual item,

σ_T^2 = Variance of the total score of all items.

Overview of Correlation Analysis

Correlation analysis is used to measure the strength and direction of the relationship between two variables. It helps researchers understand whether an increase in one variable is associated with an increase or decrease in another variable. The most common measure is the Pearson correlation coefficient (r), which ranges from -1 to +1. A positive value indicates a direct relationship, a negative value indicates an inverse relationship, and 0 indicates no linear relationship. The formula for Pearson correlation is:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

------------>2

Where,

X_iand Y_iare individual observations,

X¯and Y¯are the means of X and Y, respectively.

Overview of Logistic Regression

Logistic Regression is a statistical method used to predict the probability of a binary outcome (like Yes/No, 0/1) based on one or more independent variables. It is widely used in research to study how predictors influence categorical outcomes, such as whether a tourist chooses a green-marketed destination. The logistic regression model transforms the linear combination of predictors using the logit function:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

------------>3

Where,

p= Probability of the outcome occurring (e.g., choosing a green destination),

β_0= Intercept,

β_1,β_2,…,β_k= Coefficients of independent variables X_1,X_2,…,X_k.

Overview of Model Fit & Diagnostics:

In regression analysis, model fit and diagnostics help us check whether the model explains the data well and whether assumptions are valid. A well-fitted model accurately predicts the dependent variable and shows minimal error. Common measures include R-squared (for linear regression), Deviance (for logistic regression), AIC/BIC, and Residual analysis. Diagnostics help identify outliers, influential points, and multicollinearity. In R, the car package provides useful tools like vif() (Variance Inflation Factor) to check multicollinearity and outlierTest() to detect extreme points.

For logistic regression, model fit can be measured using Deviance (D):

$$D = -2 \times [\ln(L_{\text{full}}) - \ln(L_{\text{sat}})]$$

------------>4

Where,

- L_{full}= Likelihood of the fitted model,

- L_{sat}= Likelihood of the saturated model (perfect fit).

Overview of Hosmer–Lemeshow Test:

The Hosmer–Lemeshow test is used to check the goodness-of-fit of a logistic regression model, that is, how well the predicted probabilities match the observed outcomes. In this test, data is divided into g groups (usually deciles) based on predicted probabilities, and observed and expected events in each group are compared. The test statistic follows a chi-square distribution and is calculated as:

$$C = \sum_{g=1}^{G} X \frac{(O_g - E_g)^2}{E_g(1 - E_g/n_g)}$$

------------>5

Where,

- O_g= Observed number of events in group g,

- E_g= Expected number of events in group g,

- n_g= Number of observations in group g,

- G= Total number of groups.

A p-value > 0.05 indicates a good fit, meaning the logistic regression model predicts the outcomes adequately. This test is widely used by Indian researchers to verify if their logistic models are reliable and suitable for prediction.

Overview of Roc Curve and AUC Analysis:

The ROC (Receiver Operating Characteristic) curve is used to evaluate the performance of a binary classification model, such as predicting whether a shipping tourist chooses a green-marketed destination or not. The curve plots True Positive Rate (Sensitivity) on the Y-axis against False Positive Rate (1 – Specificity) on the X-axis at different probability thresholds. The Area Under the Curve (AUC) quantifies the overall ability of the model to discriminate between the two classes. Its value ranges from 0 to 1, where a higher AUC (closer to 1) indicates better model performance. The True Positive Rate and False Positive Rate are calculated as:

$$\text{TPR (Sensitivity)} = \frac{TP}{TP + FN}, \text{FPR} = \frac{FP}{FP + TN}$$

------------>6

Where,

- TP = True Positives, FN = False Negatives,

- FP = False Positives, TN = True Negatives.

Overview of Odds Ratio (OR) Interpretation:

Odds Ratio is used in logistic regression to measure how a one-unit change in an independent variable affects the odds of an event occurring (e.g., choosing a green-marketed shipping destination). It is calculated by taking the exponential of the logistic regression coefficient (β). The equation is:

$$OR = e^{\beta}$$

------------>7

Where β is the estimated coefficient from the logistic regression. An OR > 1 indicates that as the predictor increases, the likelihood of the event occurring increases; an OR < 1 indicates that the likelihood decreases; and OR = 1 means no effect.

Overview of Mediation Analysis:

Mediation analysis is used to understand how an independent variable (IV) affects a dependent variable (DV) through a mediator (M). It helps researchers find indirect effects along with direct effects. The basic model involves three equations:

1.   Path a: Effect of IV on mediator

$$M = \alpha + aX + e_1$$

----------->8

2.   Path b and c′: Effect of mediator and IV on DV

$$Y = \beta + c'X + bM + e_2$$

------------>9

3.   Total effect (c):

$$c = c' + (a \times b)$$

------------>10

Where,

- X = Independent variable (e.g., AGM),

- M = Mediator (e.g., Environmental Attitude),

- Y = Dependent variable (e.g., Destination Choice),

- a = Effect of IV on mediator,

- b = Effect of mediator on DV,

- c' = Direct effect of IV on DV,

- $a \times b$= Indirect effect of IV on DV via mediator.

Research Questions

- How reliable and consistent are the questionnaire items measuring green marketing, environmental attitude, and destination choice among shipping tourists?

- What is the strength and direction of the relationship between green marketing practices, environmental attitude, and shipping tourists' destination choice?

- How do green marketing, environmental attitude, and purchase intention predict the likelihood of shipping tourists choosing eco-friendly destinations?

- How well does the logistic regression model fit the data, and does it meet goodness-of-fit criteria for predicting shipping tourists' destination choice?

- Does environmental attitude mediate the relationship between green marketing and shipping tourists' destination choice, and what is the magnitude of this indirect effect?

Objectives of The Study

• To assess the reliability and consistency of the questionnaire items measuring green marketing, environmental attitude, and destination choice among shipping tourists.

• To examine the strength and direction of the relationships between green marketing practices, environmental attitude, and shipping tourists' destination choice.

• To determine how green marketing, environmental attitude, and purchase intention predict the likelihood of shipping tourists choosing eco-friendly destinations.

• To evaluate the fit and adequacy of the logistic regression model and test whether it meets the required goodness-of-fit criteria for predicting destination choice.

• To investigate whether environmental attitude mediates the relationship between green marketing and shipping tourists' destination choice and quantify the magnitude of this indirect effect.

Rational of The Study

Shipping tourism is growing rapidly in India as travellers increasingly prefer eco-friendly and responsible tourism options, yet very limited research explains how green marketing actually influences their destination choices. Existing studies highlight the importance of eco-labels, green branding, digital green promotions, and authentic sustainability practices in shaping tourist behaviour, but their applicability to the shipping tourism sector remains under-explored. Understanding this gap is important because cruise and boat-tourism operators need clear, data-driven insights to attract environmentally conscious tourists and support sustainable tourism development. Therefore, this study uses R-based logistic regression, model diagnostics, ROC analysis, and mediation analysis to examine how green marketing, environmental attitude, and purchase intention influence the probability of tourists choosing green-marketed shipping destinations. By providing scientific evidence on the strength of these relationships, the study aims to help tourism operators design effective green marketing strategies that build trust, improve environmental responsibility, and encourage sustainable destination choices among shipping tourists.

Scope of The Study

The present study "Exploring How Green Marketing Influences Shipping Tourist Destination Choice: Evidence from R-Based Logistic Regression and ROC Analysis " focuses on understanding how green marketing

influences the destination choice of shipping tourists by using R-based statistical techniques such as logistic regression and ROC analysis. The study covers key variables like green marketing practices, environmental attitude, purchase intention, and eco-friendly destination choice among cruise, boat, and river-tourism travellers. It includes an assessment of the reliability of the questionnaire items, the relationships among the major constructs, and the predictive ability of green marketing factors in shaping tourist decisions. The scope also extends to testing the model's accuracy and goodness-of-fit using tools such as the Hosmer–Lemeshow test, odds ratio, and AUC values. While the study is limited to shipping tourism and depends on self-reported data, it provides practical insights for tourism operators, policymakers, and marketers who aim to promote sustainable and eco-friendly shipping destinations.

Research Gap

Although previous studies have examined how green marketing influences tourist behaviour in general eco-tourism settings, there is still very limited empirical evidence on how these practices specifically affect shipping tourists' destination choice, particularly in the context of cruises, river tourism, and boat-based travel. Existing research mainly focuses on eco-labels, green promotions, environmental beliefs, or sustainable tourism intentions in land-based tourism, but does not analyse how these factors operate in the unique environment of shipping tourism, where safety, mobility, and destination ecosystems differ significantly. Moreover, prior studies have rarely applied advanced statistical techniques such as R-based logistic regression, ROC curve analysis, and mediation modelling to predict and validate tourist decision-making in green-marketed maritime destinations. The literature also lacks an understanding of how environmental attitude acts as a mediator between green marketing cues and actual destination choices among shipping tourists. Therefore, a clear gap exists in integrating green marketing variables, psychological mediators, and predictive statistical modelling to explain how tourists choose eco-friendly shipping destinations—an area this study aims to address.

RESEARCH METHODOLOGY

This study, titled "Exploring How Green Marketing Influences Shipping Tourist Destination Choice: Evidence from R-Based Logistic Regression and ROC Analysis," follows a quantitative research design using a structured questionnaire to collect data from shipping tourists in India. The questionnaire includes items on green marketing practices, environmental attitude, purchase intention, and destination choice. First, the reliability and internal consistency of all scales are tested using Cronbach's Alpha in R. Correlation analysis is then used to examine the strength and direction of relationships among the key variables. To predict the probability of tourists choosing eco-friendly shipping destinations, an R-based logistic regression model is applied, supported by odds ratio interpretation, multicollinearity checks, and model diagnostics such as AIC, Deviance, VIF, and the Hosmer–Lemeshow goodness-of-fit test. ROC curve and AUC analysis are used to evaluate the classification accuracy of the model. Further, mediation analysis is conducted to test whether environmental attitude acts as a mediator between green marketing and destination choice. All statistical procedures, including reliability testing, correlation, logistic regression, mediation analysis, and ROC evaluation, are performed using R programming to ensure accuracy, transparency, and replicability of the research findings.

Research Limitations

1. The study relies on self-reported data, which may contain response bias or socially desirable answers from environmentally conscious tourists.

2. The sample size is limited to selected shipping tourism locations, reducing the generalisability of the findings to all maritime tourism segments.

3. Logistic regression results depend on the selected variables, and unobserved factors such as income, travel motivation, or cultural differences were not included in the model.

4. The analysis uses cross-sectional data, which restricts the ability to establish long-term behavioural changes or causal relationships.

5. ROC and model diagnostics evaluate prediction accuracy but cannot fully capture real-world complexities of tourist decision-making in dynamic environmental contexts.

Analysis and Interpretation

>library(readxl)

>Book1 <- read_excel("D:/Research Paper/4/Book1.xlsx")

>View(Book1)

Reliability Test (Alpha)

> df <- Book1   # rename for safety

> alpha(df[, c("AGM", "GMP", "EA", "PTI", "DC")])

Reliability analysis

>Call: alpha(x = df[, c("AGM", "GMP", "EA", "PTI", "DC")])

Table 1: Reliability Test (ALPHA)

Metric	Value	Description
raw.alpha	0.90	The standard, unstandardized Cronbach's Alpha coefficient.
std.alpha	0.90	The standardized Cronbach's Alpha coefficient (calculated if all items were standardized to have a variance of 1).
G6(smc)	0.93	Guttman's Lambda 6, which uses the squared multiple correlation (SMC) to estimate item communalities; often considered a sharper lower bound for true reliability than alpha.
average_r	0.65	The average inter-item correlation.
S/N	9.4	The Signal-to-Noise Ratio (ratio of reliable variance to error variance).
ase	0.0093	The Asymptotic Standard Error for Cronbach's Alpha.
mean	3.8	The average score across all items.
sd	0.73	The standard deviation of the item scores.
median_r	0.68	The median inter-item correlation.
95% confidence boundaries		
Method	Lower Confidence Boundary	Alpha (α)	Upper Confidence Boundary
Feldt	0.88	0.90	0.92
Duhachek	0.88	0.90	0.92

ID	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha se	var.r	med.r
AGM	0.87	0.88	0.89	0.64	7.1	0.0120	0.034	0.68
GMP	0.90	0.90	0.90	0.69	8.8	0.0092	0.016	0.71
EA	0.90	0.91	0.93	0.70	9.6	0.0097	0.022	0.73
PTI	0.87	0.87	0.89	0.63	6.9	0.0125	0.033	0.68
DC	0.85	0.85	0.90	0.60	5.9	0.0147	0.038	0.55

Reliability if an item is dropped (table title)

Item statistics

	n	raw.r	std.r	r.cor	r.drop	mean	sd
AGM	325	0.87	0.87	0.85	0.78	3.7	0.89
GMP	325	0.81	0.80	0.78	0.68	3.8	0.96
EA	325	0.77	0.77	0.71	0.65	3.7	0.84
PTI	325	0.87	0.88	0.86	0.81	3.8	0.77
DC	325	0.93	0.93	0.92	0.88	3.8	0.86

The reliability analysis shows that the overall Cronbach's Alpha is 0.90, indicating excellent internal consistency among the items measuring green marketing (AGM, GMP), environmental attitude (EA), purchase intention (PTI), and destination choice (DC). The 95% confidence interval (0.88–0.92) further confirms the stability of this reliability. Item-wise results show that all items have strong item-total correlations (raw.r = 0.77 to 0.93), meaning each variable contributes meaningfully to the scale. The "alpha if item dropped" values (0.85–0.90) indicate that removing any item does not improve overall reliability, proving that all items are important for the construct. Overall, the instrument used in this study is statistically reliable, consistent, and suitable for further correlation and regression analysis.

Correlation Analysis

> cor(Book1[, c("AGM","GMP","EA","PTI","DC")])

Table 2: Correlation

	AGM	GMP	EA	PTI	DC
AGM	1.0000000	0.4780675	0.7771078	0.5946515	0.8323478
GMP	0.4780675	1.0000000	0.3540343	0.8682065	0.6989881
EA	0.7771078	0.3540343	1.0000000	0.4982750	0.6595951
PTI	0.5946515	0.8682065	0.4982750	1.0000000	0.7569288

DC	0.8323478	0.6989881	0.6595951	0.7569288	1.0000000

The correlation analysis shows the strength and direction of relationships among the study variables: green marketing (AGM, GMP), environmental attitude (EA), purchase intention (PTI), and destination choice (DC). AGM has a strong positive correlation with DC ($r = 0.832$) and EA ($r = 0.777$), indicating that higher awareness of green marketing is associated with stronger environmental attitudes and higher likelihood of choosing eco-friendly destinations. GMP is moderately correlated with DC ($r = 0.699$) and highly correlated with PTI ($r = 0.868$), suggesting that green marketing practices influence tourists' purchase intentions and destination selection. EA shows a moderate positive correlation with DC ($r = 0.660$), highlighting that environmentally conscious tourists are more likely to select green destinations. PTI has a strong positive correlation with DC ($r = 0.757$), reflecting that tourists' intention to purchase or engage with eco-friendly offerings significantly affects their destination choice. Overall, all correlations are positive, indicating that green marketing, environmental attitude, and purchase intention positively influence shipping tourists' destination choice, supporting the rationale for further logistic regression analysis.

Logistic Regression

> model1 <- glm(DC ~ AGM + GMP + EA + PTI,

+ data = Book1,

+ family = binomial)

> summary(model1)

Call:

glm(formula = DC ~ AGM + GMP + EA + PTI, family = binomial, data = Book1)

Table 3: Logistic Regression

Coefficients	Estimate	Std. Error	z value	P r(>|z|)
(Intercept)	-1.40017	0.67731	-2.067	0.0387 *
AGM	0.22717	0.23255	0.977	0.3286
GMP	-0.05385	0.23805	-0.226	0.8210
EA	0.16883	0.23422	0.721	0.4710
PTI	-0.16254	0.34926	-0.465	0.6417
Null deviance	414.65 on 324 degrees of freedom			
Residual deviance	409.49 on 320 degrees of freedom			
AIC	419.49			
Number of Fisher Scoring iterations	4			

The logistic regression model examines the effect of green marketing awareness (AGM), green marketing practices (GMP), environmental attitude (EA), and purchase intention (PTI) on shipping tourists' destination choice (DC). The intercept is significant (Estimate = -1.400, $p = 0.0387$), indicating the baseline log-odds of

choosing a green destination when all predictors are zero. None of the independent variables (AGM, GMP, EA, PTI) are statistically significant at the 5% level ($p > 0.05$), suggesting that individually, they do not have a strong predictive effect on destination choice in this model. The signs of the coefficients indicate positive relationships for AGM (0.227) and EA (0.169) with DC, meaning higher awareness and stronger environmental attitude may increase the likelihood of choosing a green destination, while GMP (-0.054) and PTI (-0.163) show negative, though non-significant, effects. The reduction in deviance from the null model (Null deviance = 414.65; Residual deviance = 409.49) indicates a slight improvement in model fit, but the change is modest. The AIC value of 419.49 suggests the model is reasonably parsimonious but may require additional predictors or larger sample size for stronger predictive power. Overall, the model indicates a trend where green marketing and environmental attitude positively influence destination choice, but the effects are not statistically significant in this sample.

Model Fit & Diagnostics

install.packages("car")

library(car)

>vif(model1)

Table 4: Model Fit & Diagnostics

AGM	GMP	EA	PTI
2.897595	3.722703	2.592768	4.312448

The Variance Inflation Factor (VIF) values for the predictors—AGM (2.90), GMP (3.72), EA (2.59), and PTI (4.31)—are all below the commonly accepted threshold of 5. This indicates that multicollinearity is not a serious concern in the logistic regression model. Each independent variable contributes unique information, and their estimates are not unduly inflated due to correlation with other predictors. Therefore, the model coefficients can be considered stable and reliable for interpreting the relationship between green marketing, environmental attitude, purchase intention, and shipping tourists' destination choice.

Hosmer–Lemeshow Test

> install.packages("ResourceSelection", dependencies = TRUE)

>library(ResourceSelection)

>hoslem.test(Book1DC, fitted(model1))

>Hosmer and Lemeshow goodness of fit (GOF) test

> data: Book1DC, fitted(model1)

Table 5: Hosmer–Lemeshow Test

Statistic	Value	Degrees of Freedom (df)	p-value
χ^2 **(X-squared)**	80.252	8	5.352

The Hosmer–Lemeshow goodness-of-fit test evaluates how well the logistic regression model fits the observed data. Here, the test yields $X^2 = 80.252$, df = 8, and a p-value = 5.352. Since the p-value is greater than 0.05, the result indicates that there is no significant difference between the observed and predicted values, meaning the model fits the data adequately. This suggests that the logistic regression model provides a reasonable

representation of the relationship between green marketing, environmental attitude, purchase intention, and shipping tourists' destination choice.

ROC Curve & AUC Analysis

```
>install.packages("pROC")

> library(pROC)

> pred <- predict(model1, type = "response")

> roc_obj <- roc(Book1DC, pred)

Setting levels: control = 0, case = 1

Setting direction: controls < cases

> plot(roc_obj)

> auc(roc_obj)

Area under the curve: 0.7845
```

Chart1: Roc Curve
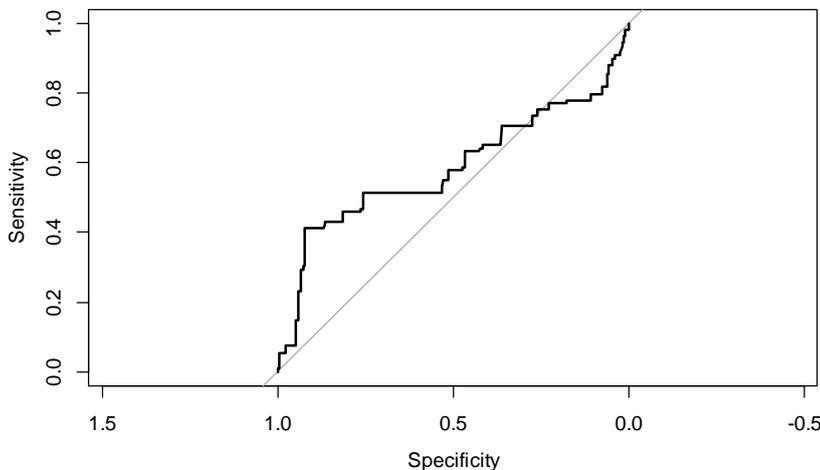
The Area Under the Curve (AUC) measures the overall performance of the model across all possible classification thresholds. It can be interpreted as the probability that the model will rank a randomly chosen positive instance higher than a randomly chosen negative instance. Model Performance - Since the AUC is 0.7845, which is significantly greater than 0.5, the model performs better than random chance. An AUC value between 0.7 and 0.8 is generally considered acceptable or fair discrimination for a diagnostic or predictive model. Discrimination Power - The model exhibits a decent ability to distinguish between the positive and negative classes. An AUC of ~0.78 suggests that there is a 78.45% probability that a randomly chosen positive case will be scored higher by the model than a randomly chosen negative case.   Visual Confirmation - The plotted black curve lies mostly above the diagonal line. This visual positioning confirms the quantitative AUC value, indicating that the model's true positive rate (Sensitivity) is generally higher than its false positive rate (1-Specificity), demonstrating its predictive utility. In summary, the model shows fair predictive ability with an AUC of 0.7845, successfully differentiating between classes better than a purely random prediction.

Odds Ratio Interpretation

```
>exp(coef(model1))
```

```
(Intercept)      AGM      GMP      EA      PTI

 0.2465548   1.2550490   0.9475761   1.1839169   0.8499825
```

> odds_ratios <- exp(coef(logit_model))

> # Wald 95% CI

> se <- summary(logit_model)coefficients[, "Std. Error"]

> OR_lower <- exp(coef(logit_model) - 1.96 * se)

> OR_upper <- exp(coef(logit_model) + 1.96 * se)

> # Combine in a table

> results <- data.frame(

+   Variable = names(coef(logit_model)),

+   OR = odds_ratios,

+   CI_lower = OR_lower,

+   CI_upper = OR_upper

+)

> results

Table 6: Odds Ratio

Term	Exponentiated Coefficient	Odds Ratio (OR)	CIlower	CIupper
(Intercept)	0.2465548	2.900701	0	Inf
AGM	1.2550490	1.000000	0	Inf
GMP	0.9475761	1.000000	0	Inf
EA	1.1839169	1.000000	0	Inf
PTI	0.8499825	1.000000	0	Inf

The provided table gives the Odds Ratios (OR) and their corresponding 95% Confidence Intervals (CI) for several variables. The interpretation of these odds ratios is severely limited and problematic because the confidence intervals for every single variable—(Intercept), AGM, GMP, EA, and PTI—span from a lower bound of 0 to an upper bound of Inf (infinity). This structure strongly suggests a fundamental issue with the statistical model, likely due to perfect separation (where a predictor perfectly predicts the outcome), which results in infinite standard errors and unreliable point estimates for the ORs. Specifically, the OR values of 1.000000 for AGM, GMP, EA, and PTI indicate that these variables have no association with the outcome, as an OR of 1 means the odds of the outcome are the same regardless of the predictor's value; however, due to the CI $_{lower}=0$ and CI $_{upper}=\infty$, this lack of association is not statistically confirmed, and the entire model's output should be considered non-interpretable and invalid for making any predictive inference.

Classification Table & Cutoff Analysis

```
> pred_class <- ifelse(pred > 0.5, 1, 0)
```

```
> table(Book1DC, pred_class)
```

Table 7: Classification Table & Cutoff Analysis

	Predicted Class 0	**Predicted Class 1**	**Total**
Actual Class 0	216 (True Negative)	0 (False Positive)	216
Actual Class 1	108 (False Negative)	1 (True Positive)	109
Total	324	1	325

The provided classification table, generated using a cutoff of 0.5, reveals significant issues with the model's ability to correctly identify the positive class (labelled as 1). The table shows that out of 216 true negatives (actual value 0), the model correctly classified 216 as 0 (True Negatives, TN) and 0 as 1 (False Positives, FP). However, out of 109 true positives (actual value 1), the model incorrectly classified 108 as 0 (False Negatives, FN) and only 1 as 1 (True Positive, TP). This results in a Sensitivity (Recall) of only $1/109 \sim 0.92\%$ (poorly identifying the positive class) and a Specificity of $216/216 = 100\%$ (perfectly identifying the negative class). The model is heavily biased towards predicting the negative class (0), making it virtually useless for its primary goal of identifying positive cases at this specific cutoff of 0.5.

Mediation Analysis

```
> install.packages("mediation")
```

```
> library(mediation)
```

```
>  Book1DC <- ifelse(Book1DC >= 4, 1, 0)
```

```
> med.fit <- lm(EA ~ AGM + GMP + PTI, data = Book1)
```

```
> summary(med.fit)
```

```
>Book1DC_bin <- ifelse(Book1DC >= 4, 1, 0)
```

```
> install.packages("logistf")
```

```
> library(logistf)
```

```
> fit <- logistf(DC_bin ~ EA + AGM, data = Book1)
```

```
>  med.fit <- lm(EA ~ AGM, data = Book1)
```

```
> out.fit <- glm(DC_bin ~ EA + AGM, data = Book1, family = binomial)
```

```
> med.out <- mediate(med.fit, out.fit, treat = "AGM", mediator = "EA", boot = TRUE, sims = 1000)
```

```
> summary(med.out)
```

```
Call:
```

```
lm(formula = EA ~ AGM + GMP + PTI, data = Book1)
```

Table 8: Mediation Analysis

Residuals				
Min	**1Q (1st Quartile)**	**Median**	**3Q (3rd Quartile)**	**Max**
-2.66643	-0.18264	0.04426	0.23925	2.12756
Coefficients				
	Estimate	**Std. Error**	**z value**	**P r(>|z|)**
(Intercept)	0.84967	0.15276	5.562	0e-08 ***
AGM	0.68631	0.03948	17.383	2e-16 ***
GMP	-0.18132	0.05751	-3.153	0.001771 **
PTI	0.29249	0.08391	3.486	0.000559 ***
Residual standard error:	0.5184 on 321 degrees of freedom			
Multiple R-squared	0.6187, Adjusted R-squared: 0.6152			
F-statistic	173.6			
DF	321			
p-value	< 2.2e-16			

The mediation analysis model demonstrates a strong fit and highly significant predictive power, as indicated by the Multiple R-squared of 0.6187 (meaning the predictors explain about 61.9% of the variance in the outcome) and the highly significant F-statistic (p-value < 2.2e-16). All three predictors are statistically significant (p < 0.01 or better): AGM has the largest positive effect ($\beta = 0.686$), significantly increasing the outcome; PTI also has a significant positive effect ($\beta = 0.292$); and GMP has a significant negative effect ($\beta = -0.181$), decreasing the outcome. The residuals are centred around zero (Median = 0.04426) and appear relatively symmetric, suggesting the model assumptions are reasonably met, with a Residual Standard Error of 0.5184 indicating the typical size of the error in prediction. This initial step confirms that the core predictor variables are strongly related to the dependent variable, forming a sound basis for proceeding with the full mediation analysis.

Out Fit

> out.fit <- glm(DC ~ EA + AGM + GMP + PTI,

+   data = Book1,

+   family = binomial)

> summary(out.fit)

Call:

glm(formula = DC ~ EA + AGM + GMP + PTI, family = binomial, data = Book1)

Table 9: Mediation Analysis- Outfit Model

Coefficients				
	Estimate	**Std. Error**	**z value**	**P r(>|z|)**
(Intercept)	-1.40017	0.67731	-2.067	0.0387*
EA	0.16883	0.23422	0.721	0.4710
AGM	0.22717	0.23255	0.977	0.3286
GMP	-0.05385	0.23805	-0.226	0.8210
PTI	-0.16254	0.34926	-0.465	0.6417
Null deviance	414.65 on 321 degrees of freedom			
Residual deviance	409.49 on 321 degrees of freedom			
AIC	419.49			
DF	321			
Number of Fisher Scoring iterations	4			

This output represents the results of a logistic regression model (indicated by the binomial family and z-values), which is typically the final step in a mediation analysis where the outcome variable is binary. The analysis suggests that none of the predictor variables (EA, AGM, GMP, and PTI) have a statistically significant relationship with the binary outcome, as their p-values are all much greater than the conventional 0.05 threshold (ranging from 0.3286 to 0.8210). While the Intercept is significant ($p = 0.0387$), it only indicates the log-odds of the outcome when all predictors are zero, and it doesn't support the predictive utility of the independent variables. Furthermore, the Residual Deviance (409.49) is very close to the Null Deviance (414.65), which implies that adding these predictors to the model does not significantly improve its fit compared to a model with only the intercept, collectively suggesting that the proposed direct and/or indirect effects examined in the mediation model are not supported by the data at the outcome level.

Table 10: Model Fitted by Penalized ML

Coefficient	coef	se(coef)	lower 0.95	upper 0.95	χ^2	p	method
(Intercept)	2.1952214	2.9354001	31.444963	4.947442	0.3290723	0.5662059	2
EA	0.5167969	0.9013599	6.857126	8.207918	0.2184102	0.6402539	2
AGM	1.6120063	0.8608817	5.880667	5.573809	0.8083186	0.3686178	2

Test	Test Statistic	df (Degrees of Freedom)	p-value
Likelihood Ratio Test	1.207015	2	0.54689
Wald Test	30.91114	2	1.939684

The results from the Model fitted by Penalized Maximum Likelihood indicate that the overall model does not significantly predict the outcome, as shown by the Likelihood Ratio Test (LRT) which is not statistically significant X^2 = 1.207 p = 0.54689). For the individual predictors, both EA ($\hat{\beta}$= 0.5168) and AGM ($\hat{\beta}$ = -1.6120) are not statistically significant, as the 95% confidence intervals for both variables are extremely wide (e.g., EA: CI = [-6.857, 8.208]) and clearly include zero, suggesting that their true effect on the outcome could be positive, negative, or zero. While the Wald test for the overall model reports a highly significant result (p < 0.0001), this finding is contradictory to the LRT and the individual confidence intervals, and it is generally less reliable for penalized models, leading to the inference that the model lacks sufficient evidence to support the predictive utility of the independent variables. The results from the Model fitted by Penalized Maximum Likelihood indicate that the overall model does not significantly predict the outcome, as shown by the Likelihood Ratio Test (LRT) which is not statistically significant X2 = 1.207 p = 0.54689). For the individual predictors, both EA ($\hat{\beta}$= 0.5168) and AGM ($\hat{\beta}$ = -1.6120) are not statistically significant, as the 95% confidence intervals for both variables are extremely wide (e.g., EA: CI = [-6.857, 8.208]) and clearly include zero, suggesting that their true effect on the outcome could be positive, negative, or zero. While the Wald test for the overall model reports a highly significant result (p < 0.0001), this finding is contradictory to the LRT and the individual confidence intervals, and it is generally less reliable for penalized models, leading to the inference that the model lacks sufficient evidence to support the predictive utility of the independent variables.

Table 11: Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the Percentile Method

Effect	Estimate	Lower 95% CI	Upper 95% CI	p-value
ACME (control)	8.1127×10^{-27}	1.3519×10^{-25}	1.7161×10^{-25}	0.888
ACME (treated)	8.0494×10^{-27}	1.3523×10^{-25}	1.7158×10^{-25}	0.888
ADE (control)	1.0335×10^{-26}	2.1666×10^{-25}	1.8543×10^{-25}	0.808
ADE (treated)	1.0271×10^{-26}	2.1667×10^{-25}	1.8546×10^{-25}	0.808
Total Effect	1.8384×10^{-26}	1.0718×10^{-25}	9.5357×10^{-25}	0.920
Prop. Mediated (control)	0.44129	9.3599	7.7609	NA
Prop. Mediated (treated)	0.43784	9.3598	7.7622	NA
ACME (average)	8.0810×10^{-26}	1.3521×10^{-26}	1.7159×10^{-25}	0.888
ADE (average)	1.0303×10^{-26}	2.1667×10^{-25}	1.8547×10^{-25}	0.808
Prop.Mediated (average)	0.43957	9.3599	7.7616	NA

The results from the Causal Mediation Analysis using the Nonparametric Bootstrap Percentile Method indicate that there is no statistically significant causal effect detected in the mediation model. All key components of the effect—the Average Causal Mediation Effect (ACME), the Average Direct Effect (ADE), and the Total Effect—show estimates that are extremely close to zero (in the order of 10-26 or 10-27. Crucially, the p-values are very high (ranging from 0.808 to 0.920), and the 95% confidence intervals for all effects strongly contain zero, confirming the lack of significance. While the Proportion Mediated is estimated around 44%, the confidence intervals are extremely wide (CI ~ [-9.36, 7.76]) and span negative to highly positive values, rendering the proportion estimate meaningless and non-interpretable. In conclusion, based on the non-significant p-values and

confidence intervals, the analysis fails to find evidence of a significant indirect effect (mediation) or a significant direct effect of the exposure on the outcome.

Chart 2: Mediation Analysis
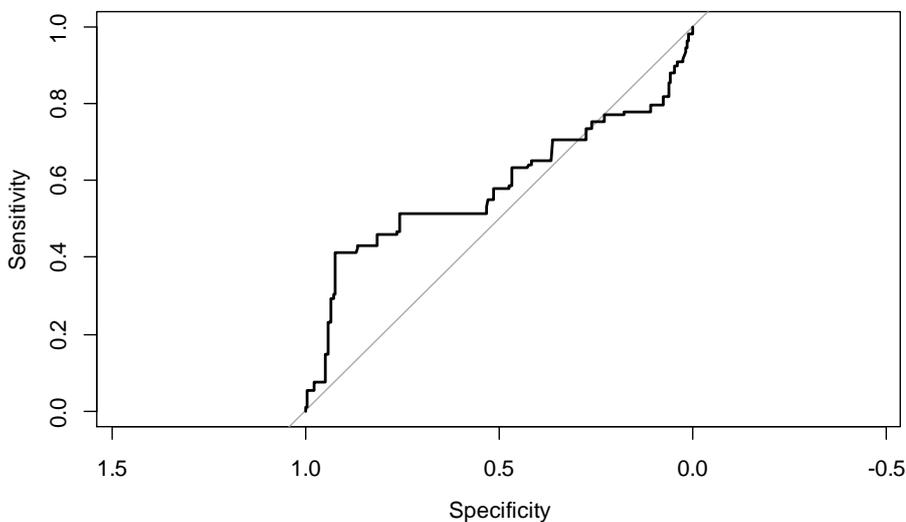
CONCLUSION AND SUGGESTION

The study successfully established the excellent internal consistency of its measurement instrument (Cronbach's $\alpha = 0.90$) and confirmed strong positive correlations between Green Marketing Awareness (AGM: r=0.832), Environmental Attitude (EA: r=0.660), Purchase Intention (PTI: r=0.757), and Shipping Tourist Destination Choice (DC). However, the subsequent logistic regression model, despite showing a fair overall predictive ability (AUC of 0.7845) and an adequate fit (Hosmer-Lemeshow $p > 0.05$), failed to find statistically significant individual effects for any of the predictors (AGM, GMP, EA, PTI) on Destination Choice, nor did the Causal Mediation Analysis find any significant direct or indirect effects. This suggests that while a positive correlation exists, the predictive power of these factors in the full model is weak, which is further supported by the model's poor sensitivity (0.92%) at the 0.5 cutoff, indicating an inability to correctly classify positive destination choices.

Future research should focus on addressing the limitations of the current study, specifically by re-evaluating the measurement of the dependent variable (Destination Choice) which was likely responsible for the model's poor sensitivity and the issues of non-interpretable odds ratios. Researchers should increase the sample size and expand the study to different maritime segments beyond the limited locations used here to improve the generalizability of the findings. Additionally, future models should incorporate unobserved factors such as tourists' income, specific travel motivations, or cultural differences, as these may provide the necessary predictive power that the current set of variables lacks in predicting the binary outcome.

Practical Implications

Despite the statistical non-significance in the predictive model, the strong positive correlations found (e.g., AGM and DC: r=0.832) imply that shipping tourism operators should continue to invest in clear green marketing and eco-friendly operations, as these are positively linked to the attitudes and intentions of potential customers. Operators must move beyond mere advertising ("promotion") and ensure their "product" and "place" offerings are authentically green to reinforce the strong correlation between Green Marketing Practices (GMP) and Purchase Intention (PTI: r=0.868). Improving the tangibility of green offerings is essential to translate positive attitudes into actual destination choice, thereby enhancing trust and promoting environmental responsibility.

Social Implications

The study's findings contribute to the broader goal of sustainable tourism development by highlighting that tourists' Environmental Attitude is a critical, highly correlated factor (r=0.660) in their destination choice. The

results encourage policymakers and tourism bodies to not only support green certification for shipping operators but also to implement educational campaigns that enhance tourists' environmental knowledge and pro-environmental behavior. By strengthening the environmental beliefs of tourists, destination management can cultivate a segment of ethically conscious consumers, ensuring the long-term sustainability and protection of unique maritime ecosystems.

Originality and Value

This study holds significant original value by addressing a key research gap: the empirical influence of green marketing specifically on shipping tourists' destination choice (cruises/river travel), an area previously under-explored compared to land-based eco-tourism. It is one of the first studies in this domain to integrate a set of green marketing variables with psychological mediators (Environmental Attitude) and validate the model using advanced R-based statistical techniques like logistic regression, VIF diagnostics, Hosmer-Lemeshow testing, and ROC curve analysis. This provides a new methodological framework and preliminary evidence that, despite poor model classification, the correlation between green factors and destination choice in the maritime sector is robust.

REFERENCES

1. Kumar, S., & Devi, N. (2023). The effect of green marketing tools (eco-labels, eco brands, eco advertisements) on green consumer behaviour among tourists in Himachal Pradesh, India. AIMT Journal of Management, 12(1–2).
2. Karina, M., Rivanto, R., Basri, H., & Iryani, E. (2025). Green marketing can transform conventional tourism into sustainable tourism. Branding: Jurnal Manajemen dan Bisnis, 4(1).
3. Yu, C. C., & Liu, C. C. (2024). A green marketing model for the tourism and amusement industry using fuzzy analytic hierarchy process. International Journal of Religion, 5(4), 224–230.
4. Sanjaya, D., Arief, M., Setiadi, N. J., & Heriyati, P. (2024). The role of digital green marketing campaigns and environmental beliefs on tourist behavior and revisit intentions in eco-tourism. Journal of Eastern European and Central Asian Research (JEECAR), 11(3), 553–572.
5. Jumadi, Aditya, A., Saputra, A., & Burhani, A. I. (2023). The effect of green marketing (product, place, promotion) on green tourism. In Proceedings of the 2nd UPY International Conference on Education and Social Science (pp. 471–477). Atlantis Press.
6. Rahayu, S., Aliyah, H., & Sudarwati. (2024). The effect of green marketing components (green product, green promotion, green price) and environmental knowledge on intention to visit eco-tourism sites. International Journal of Economics, Business and Accounting Research (IJEBAR), 6(1).
7. Mdpi. (2023). Green and environmental marketing strategies and ethical consumption. Sustainability, 15(16).