



University of
Massachusetts
Amherst

Modeling Tourism Advertising Conversion in the United States

Item Type	article;event
Authors	Park, Sangwon;Fesenmaier, Daniel R.
Download date	2026-03-17 09:29:55
Link to Item	https://hdl.handle.net/20.500.14394/48193

Modeling Tourism Advertising Conversion in the United States

Sangwon Park

National Laboratory for Tourism & eCommerce

School of Tourism & Hospitality Management

Temple University, Philadelphia, PA USA

Daniel R. Fesenmaier, Ph. D.

National Laboratory for Tourism & eCommerce

School of Tourism & Hospitality Management

Temple University, Philadelphia, PA USA

ABSTRACT

This study attempts to develop tourism advertising conversion modeling throughout the United States. Geographic and demographic data were taken into account for developing the model using multivariate technique such as logistic regression and CHAID. The results of this research showed that residence states are the most important factor affecting the conversion and interestingly, the information that is relevant with the focal destination is more important than number of brochure information travelers received. This research ends with discussion of theoretical and practical implications for destination marketing.

INTRODUCTION

Tourism researchers have explored/developed a variety of methods and models to assess the conversion rates of advertising campaigns and have investigated advertising effectiveness from many perspectives (Kaminski, Gordon, di Benedetto, and Schoenbachler 1995; Kim, Hwang and Fesenmaier 2005; Siegel and Ziff-Levine 1990; Stergiou and Airey 2003; Woodside 1990; Woodside and Ronkainen 1986). Importantly, the majority these studies evaluate the effectiveness of an advertising program for a specific destination and/or for a campaign season, and therefore the results of these studies lack generalizability. That is, the problem with most conversion studies, however, is that they are not designed with the goal of the comparative evaluation whereby certain aspects of the campaign are somehow controlled (e.g., through experimental or statistical manipulation). This paper reports the results of a study that focused on developing a general conversion model for tourism destinations located throughout the United States.

BACKGROUND

In order to achieve this goal, two multivariate techniques (logistic regression and

CHAID) were used to develop a family of models that can be used to predict conversion rates for tourism advertising throughout the United States. These approaches were chosen as they have various strengths and weakness for model construction within large data systems whereby the data are typically discrete values and do not have normal distributions. Indeed, recent studies suggest that they can be complementary whereby the results of one analysis can be used as supportive of the other.

Logistic Regression

Logistic regression model, for modeling response in data base marketing, is a standard approach for predicting a dichotomous dependent variable (i.e., visit or not visit) analyzed by a set of predictor variables (Sloane and Morgan 1996). The probability of logistic regression model with independent variables can be written by:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

Where p ($Y = \text{yes}$) represents the probability of the presence of visitation and mean the regression coefficients of the model. The natural logarithm of the ratio of $P(Y = 1)$ to $1 - P(Y=1)$ gives a linear model in X_i :

$$\begin{aligned} \text{LR}(x) = \ln \frac{P(Y = 1)}{1 - P(Y = 1)} \\ = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \end{aligned}$$

$\text{LR}(x)$ has the desirable properties of linear regression and the independent variables can be the combination of categorical and continuous variables (Kurt, Ture, and Kurum, 2008). However, the regression model estimates the relative odds of response.

CHAID

CHAID (Chi-square Automatic Interaction Detector) is a tree-based classification procedure which splits data into subsets that best describe a dependent variable (Kass, 1980). CHAID is derived from the technique of Automatic Interaction Detection (AID) that was originally developed by Sonquist and Morgan (1964). AID defines the combinations of predictor variables to find out homogeneous groups according to the dependent variable (Morgan and Sonquist, 1963) but has several limitations (Baron and Phillips, 1994). First, AID assumes that the dependent variable, or behavioral criterion, is measured at the interval level, though the actual behavior can be converted into dichotomous variable like ‘buying’ or ‘not buying’. Second, the predictor variables are divided into just two categories regardless of

the number of original categories. This is not a problem when the predictor variable composes of dichotomous variables such as gender; however, in case of other predictors consisting of more than two subsets such as education, income, and region, the breaking down process into the dichotomous split can lead to obscure differences in responses between categories. However, CHAID addresses the problems of AID by examining all possible cross-tabulations of the data and rejecting insignificant cross-tabulations. That is, CHAID partitions data into more than two subgroups and find out the most significant splits to reveal “best” split while AID utilizes the most explanatory splits that do not consider the type of predictors and the number of each category. This system allows researchers to focus on useful subsections.

METHODOLOGY

The data used in this study were obtained from two sources. First, geographic and demographic data was obtained in order to establish a comparative basis for modeling conversion rates and were used as the independent variables in the model. This information was obtained when the person contacted a tourism office through a website (either the DMO website or that of the advertiser). Specifically, the information obtained about each inquirer was: (1) Residence state - measured by the state in which the inquirer is currently living; (2) Target market – defined as whether or not (0/1) the inquirer resides within or outside of the target market of the advertising campaign; (3) InfoRequest – defined as the number of brochures the inquirer had requested about a variety of the travel destinations (maximum of 8 destinations); (4) Match - measured by whether or not (0/1) the information that the inquirer requested matched the focal destination; and, (5) PRIZM - defined as being a member of one of 67 demographic groups as defined by Claritas, Inc.; this variable was assigned based upon the 5 digit zipcode of inquirer’s residence.

Second, an online survey that was sent to 196,200 persons who had requested information about 20 different tourism destination marketing organizations throughout the United States from May to August, 2008. The survey included a number of questions related to travel to a specific destination; in particular, the respondent was asked if he/she visited the destination since receiving/obtaining travel information; the response (yes/no) is the dependent variable for this study. The survey process used a three-step process: (1) the initial invitation was sent out along with the URL of the survey; (2) four days later, a reminder was delivered to those who had not completed the survey; and, (3) the final request for participation was sent out to those who had not completed the survey one week later. An Amazon.com gift card valued at \$100 was provided to one winner for each destination as an incentive of survey participation. The survey effort resulted in 11,546 usable responses out of 134,256 collected ones, reflecting a response rate of 8.6%

Data analysis was conducted in three stages where first, the data obtained from the conversion survey was extracted from the survey results and matched with the inquiry data provided by the advertising firm. Second, logistic regression was used to estimate the predictive power of the variables without the effect of interaction. In next step, CHAID was employed to assess the main effects generated from logistic regression and to assess the interaction effects.

RESULTS

As can be seen Table 1, Model 1 variables appears that the model provide reasonably accurate and reliable estimates of conversion (Nagelkerke $R^2 = .45$ and the prediction rate = 76.8%). The results indicate that the state of residency of the respondent with respect to the destination state is an important predictive variable. For example, living in Arizona negatively relates to travel to destination 1 while living in Massachusetts and New Hampshire positively related to travel. Also, the results indicate that the extent to which respondents requested travel information (InfoRequest) was a significant predictor for a number of state advertising programs; interestingly, the result showed significant negative impact of information search. In addition, the positive estimate for 1st MATCH (the travel information being asked first brochure “matches” the focal destination) indicates that the order of requesting travel information reveals one destination preference. Finally, the Prizm Segments were found to have substantial and differing impact on the conversion.

The results from the CHAID analysis verified the importance of residence state in conversion with 68% of accuracy rate through 10 folds cross-validation (see Figure 1); for example, those living in Massachusetts, New Hampshire and Vermont are more likely to have converted. The next meaningful variable is InfoRequest (number of brochure) whereby the travelers who received brochures (93.18%) at least 1 time show a much higher rate of conversion than those who did not receive any information (76.19%). However, in the case of the group who live in New Jersey, Pennsylvania, Florida, Texas, Indiana, Utah, Wisconsin, Arizona, and Alabama, travelers who received information over 2 times show relatively lower rate of convert than those who received information below 2 times. This result suggests the importance of relevant information rather than number of information requested.

DISCUSSION

It is crucial for DMOs to understand who actually “convert” from their advertising in order to design effective advertising/marketing programs. This research attempted to develop a predictive model of tourism conversion behavior of Americans to several US destinations. The results show a relatively high level of predictability where the key variables include the state of residence, the amount of information people requested, relevance of information, and

demographic characteristics of the household. From a practical perspective, this research is important in that the model will: (1) Enable DMOs to be better target advertising strategies and therefore, more effectively allocate their tourism advertising budget. That is, since DMOs know who visitors are based on target market, residence state, how many and what kinds of information they requested and demographic beforehand, they can target those markets that are more likely to visit the destination; (2) Enable DMOs to benchmark the response of various states and demographic groups in order to evaluate the potential for future investments; and, (3) Enable the advertising firm to evaluate the potential of their “customers” to visit various destinations throughout the United States, which in turn, enable them to develop more highly targeted communication programs.

Table 1. Summary of Logit Analysis for Destination 1

	Model1		Model2	
	Parameter Estimates	<i>P</i>	Parameter Estimates	<i>P</i>
Constructs				
Target market	.25	.35	.28	.30
Residency State				
AZ	-2.77	.04	-4.34	.02
CA			-3.49	.04
MA	2.33	.01		
NH	2.31	.03		
InfoRequest	-.10	.02	-.075	.18
Match				
1 st Match	1.35	.00	1.29	.00
Demo. Prism				
DD	2.05	.04	1.79	.05
SCE			2.25	.04
N	808		808	
Model χ^2	289.54		313.18	
Nagelkerke R^2	.48		.52	
DF	111		122	
Sig.	.000		.000	
Overall correct %	76.8		76.9	

Note: Only significant results are represented.

The significant PRISM groups are: SCE = Second City Elite, DD = Domestic Duos



Figure 1. The Result of CHAID Analysis for Destination 1.

REFERENCES

- Baron, S., and D. Phillips. (1994). "Attitude survey data reduction using CHAID: An example in shopping centre market research." *Journal of Marketing Management*, 10, 75-88.
- Kaminski, P. F., G. L. Gordon, A. di Benedetto, and D. D. Schoenbachler. (1995). "Destination advertising: Assessing effectiveness with the Split-Run technique." *Journal of Travel and Tourism Marketing*, 4(2), 1-21.
- Kass, G. V. (1980). "An exploratory technique for investigating large quantities of categorical data" *Applied Statistics*, 29(2), 119-127.
- Kim, D-Y, Y-H.Hwang, and D. R. Fesenmaier (2005). "Modeling Tourism Advertising Effectiveness." *Journal of Travel Research*, 44, 42-49.
- Kurt, I., Ture, M., and A. T. Kurum. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Systems with Applications*, 34(1), 366-374.
- Morgan, J. N., and J. A. Sonquist. (1963). "Problems in the analysis of survey data and a proposal." *Journal of the American Statistical Association*, 58, 415-434
- Siegel, W. and W. Ziff-Levine (1990). "Evaluating tourism advertising campaigns: conversion vs. advertising tracking studies." *Journal of Travel Research*, 28(3), 51-55.
- Sloane, D., and S. P. Morgan. (1996) An introduction to categorical data analysis. *Annual Review of Sociology*, 22, 351- 375.
- Stergiou, D., and D. Airey (2003). "Inquiry conversion and tourism website effectiveness: Assumptions, problems and potential." *Tourism and Hospitality Research*, 4(4), 355-366.
- Woodside, A. G. (1990). "Measuring advertising effectiveness in destination marketing strategies." *Journal of Travel Research*, 29(2), 3-8.
- Woodside, A. G., and I. A. Ronkainen (1986). "How serious is non-response bias in advertising conversion research." *Journal of Travel Research*, 26, 41-48.