**Introduction**

The purpose of this assignment is to inform a marketing plan that will target future visitors to one of the Wisconsin Dells seven attractions referenced in the Harrington (2007) case study. This study will outline a specific set of recommendations to drive more visitors to “The Ducks” tour. A supervised learning model was developed using a 1,698 in-person survey conducted at various locations and attractions across the Wisconsin Dells area to predict those visitors who will take the Duck tour given a specific set of attributes.

The predictors and response variables are as follows (from Exhibit 1)

nnights = length of stay  
nadults = number of adults in party  
nchildren = number of children under 18 in the party  
planning = how far in advance the vacation was planned  
sex = sex of survey respondent  
age = age category  
education = highest level of education completed  
income = level of total household income  
region = zip code of region

An EDA is conducted to evaluate the predictive accuracy of the attributes of the variables to determine those that are appropriate for the model. A classification tree is then used to determine the attributes associated with visitors who are likely to take The Ducks tour.

**Results**

The first step of the study conducts an Exploratory Data Analysis (EDA) by calculating a distribution for each of the categorical variables to determine those variables that may be valid for predicting those visitors most likely to ride the duck tour.

* age – the age of survey respondent had minimal variation on those visitors that took the Duck tour and therefore this variable will not be used in the model.
* sex – the sex of the survey respondent does not show a significant difference in those that did or did not take the Ducks tour so it will not be used in the model.
* education – there are imbalance showing correlations between age groups that are more likely to ride the ducks.
* Income – the amount of household income had minimal variation across the categories on those groups that did and did not take the Ducks tour. This variable will not be used in the model
* region – there are imbalances showing correlations between certain regions that are more likely to ride the ducks. This variable will be used in the model.
* nchildren – there is variation across the distribution that the number of children influences those that took the Ducks tour so this variable will be used in the model.
* nadults – the number of the adults in the group that take the Ducks tour has minimal variation across the categories and therefore will not be used in the model.
* nnights – there is a strong correlation that visitors that are staying 2 or more nights are more likely to ride the ducks. This variable will be used in the model.
* planning – there is a correlation that the more days that the vacation is planned in advance the more likely visitors are to take the Ducks tour.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  | |

**Conclusions**

**Code**

**References**