**Introduction**

The purpose of this assignment is to develop a supervised learning model that will be used 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. The model was developed using a 1,698 in-person survey conducted at various locations and attractions across the Wisconsin Dells area.

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 decision tree classification method was used to

**Results**

The EDA was facilitated 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
* Sex – does not show a strong correlation 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 –
* Region – there are imbalances showing correlations between certain regions that are more likely to ride the ducks.
* Children –
* Adults
* Nights – there is a strong correlation that visitors that are staying 2 or more nights are more likely to ride the ducks.

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

**Conclusions**

**Code**