July 1, 2018

Course: CIS570 – Business Intelligence

Name: Robert Palumbo

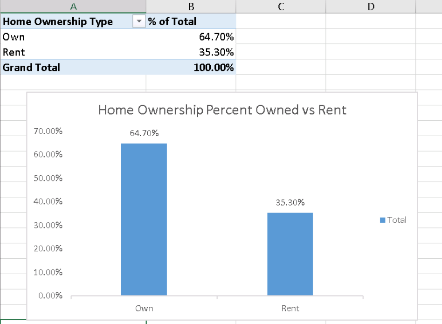
Assignment: BI Project Part2 - Submission Answers

Due Date: Sunday, July 1 @ 11:59pm

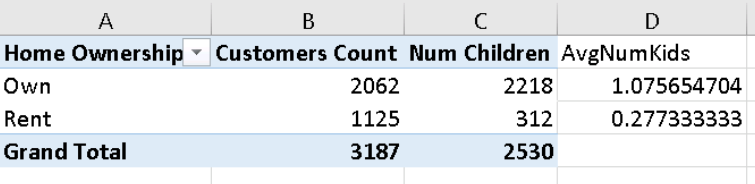
1. Do preliminary analysis on the Customers table data (if necessary, use a pivot table) to answer the following questions:
   1. How many cases (i.e., records or rows) are in the Customers table? **3187** **rows**
   2. What percentage of customers own homes, and what percentage rent?

**Owned: 2062 / 3187 = 64.7**

**Rented: 1125 / 3187 = 35.3**



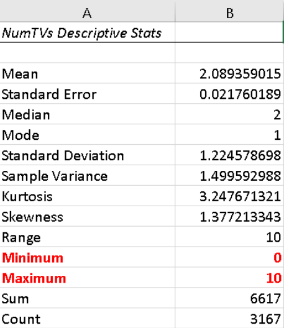
* 1. Identify any one variable that could be a good predictor of home ownership (i.e., own or rent). Provide evidence for your choice.



* 1. What is the range (i.e., minimum and maximum) of values for Number of TVs?

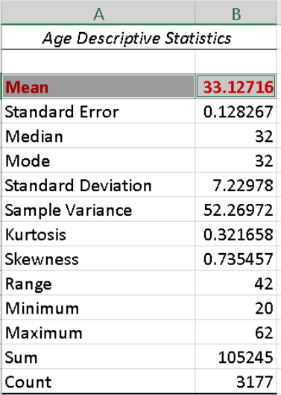
**Minimum TVs: 0**

**Maximum TVs: 10**



* 1. What is the mean (i.e., average) age? **Average age is: 33.12716.**

**NOTE: I removed the customer rows that did NOT have any age specified which would skew the average. This resulted in a new total of 3177 rows.**

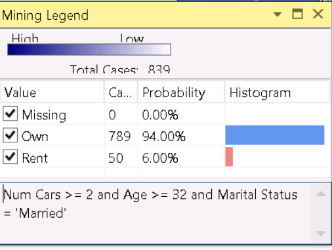


1. Use the Mining Model Viewer and Mining Accuracy Chart to answer the following questions (Note: Set the Slider to show all 5 levels):

a) Identify the node that has customers with the highest probability of owning homes.

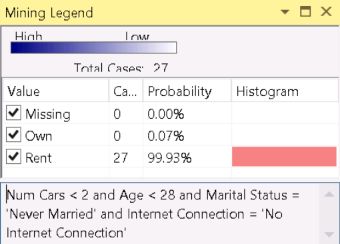
What is the probability? 94%

What is the rule for this node? (Num Cars >= 2 and Age >=32 and Marital Status = ‘Married’)



b) Identify the node that has customers with the highest probability of renting.

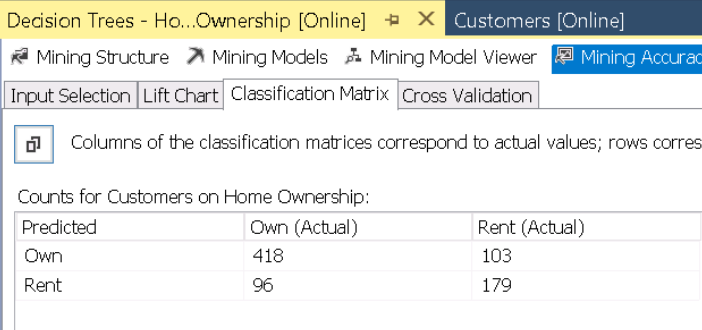
What is the probability? 99.93%  
 What is the rule for this node? Num Cars < 2 and Age < 28 and Marital Status = ‘Never Married’ and Internet Connection = ‘No Internet Connection’



c) What are the weakest and strongest predictors of Home Ownership?

Weakest Predictor? Weakest indicator is: Internet Connection  
 Strongest Predictor? Strongest indicator is: Num Cars

d) Complete the following table:



e) Based on the values in the table above, what percentage of the total numbers of cases in the test data, does the Decision Tree model predict correctly?

Total Test Cases: 2391 (from Root Node=2391 Own=1548 Rent=843)

True Positives: 418

True Negatives: 179

Total Correct: 597

Percent Correct: (597 / 2391) = 24.968%

XRLP: 597 / 796 = 75% [this is using the total of all matrix values in the table]

f) Set the Predict Value in the Input Selection Tab (within the Mining Accuracy Tab) to “Rent”. At 30% of the overall population.

What percentage of the target population is correctly predicted by:

1. the ideal model? 84.68%
2. the Decision Tree model? 57.8%

g) Set the Predict Value in the Input Selection Tab (within the Mining Accuracy Tab) to “Own”. At 65% of the overall population.

What percentage of the target population is correctly predicted by:

a) the ideal model? 100%

b) the Decision Tree model? 81.128%

h) Ram Realty will pay you $3 for each emailed solicitation that reaches a renter. Assume there are 30,000 people on an email list, the fixed cost for the email campaign is $2,500 and it costs $1 per email.

If your objective is to maximize profit, to how many people should you email the solicitation?

What will be your profit?

**Bonus Question:** If you followed the model’s recommendation, how many renters in thepopulation (i.e., 30,000 people) will receive your email solicitation?