CIS 570 – BI Project – Part 2

Note: You should have completed the Data Mining Assignment before starting on this project. This project is similar to the mining assignment. If you are unsure about how to complete a particular step, refer back to the detailed instructions in that assignment.

1. Start Visual Studio 2017. Click File →Open→Analysis Services Database. In the Connect To Database dialog box, enter **buscissql\cisbi** in the Server textbox and select **your BI Project analysis services database (e.g., if your last name is Smith, your analysis services database will be called SmithBIP)** from the Database drop-down list. Click on the Browse button and select the folder where you would like the solution file to be stored (this can be your class folder for CIS570 – **i.e., "S: drive"**).
2. Create a new Data Source to DMProject (name of the database) on BUSCISSQL\CISBI.
3. Create a new Data Source View for the Data Source you created in Step 2. Add the Customers table to the data source view.
4. Do preliminary analysis on the Customers table data (if necessary, use a pivot table) to answer the following questions:
    a) How many cases (i.e., records or rows) are in the Customers table?
    b) What percentage of customers own homes, and what percentage rent?
    c) Identify any one variable that could be a good predictor of home ownership (i.e., own or rent). Provide evidence for your choice.
    d) What is the range (i.e., minimum and maximum) of values for Number of TVs?
    e) What is the mean (i.e., average) age?
5. Create a new Mining Structure. Keep the default of Microsoft Decision Trees for the mining technique. Select the Data Source View you created in Step 3. Select the Customers table as the Case table. Check Key for CustomerID. Check Input for all the other variables with the exception of HomeOwnership.   Check Predictable for HomeOwnership. Set the percentage for testing to 25%. Change the Mining Model Name to Decision Trees – Home Ownership. Check Allow drill through. Click Finish.
6. Select the Mining Models tab, and set the "HoldoutSeed" property to 15.
7. Process the Mining Model.
8. 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?   What is the rule for this node?
    b) Identify the node that has customers with the highest probability of renting. What is the probability? What is the rule for this node?
    c) What are the weakest and strongest predictors of Home Ownership?
    d) Complete the following table:

Predicted	Own (Actual)	Rent (Actual)
Own		
Rent		

    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?
    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 a) the ideal model, and b) the Decision Tree model?
    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, and b) the Decision Tree model?
    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 the population (i.e., 30,000 people) will receive your email solicitation?

9. Submit the answers for Steps 4 and 8 either directly or as a DOC/DOCX, PDF or TXT attachment.