

July 11, 2018

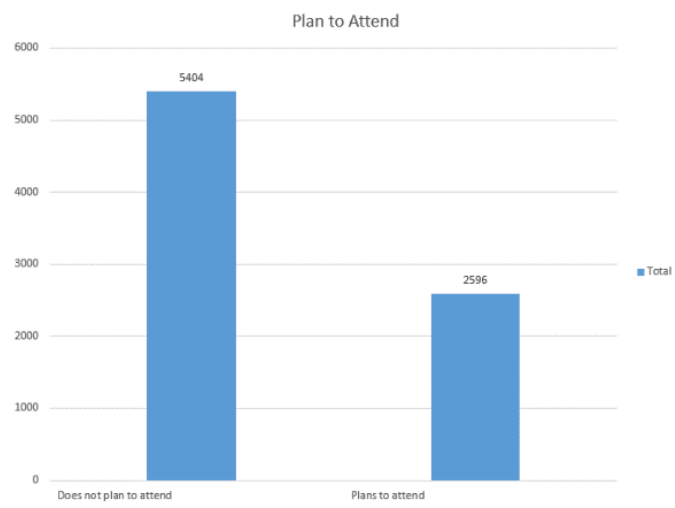
Course: CIS570 – Business Intelligence  
Name: Robert Palumbo  
Assignment: Final Exam  
Due Date: Wednesday, July 11@ 11:59pm

Re: Answers to Part A - SAS Visual Analytics Final Exam questions.

Question 1.

- a) How many students are in the sample data set? **8000 Students**
- b) What percentage of them have plans to attend college?  **$(2596 / 8000) = 32.45\%$**
- c) What percentage do not have plans to attend college?  **$(5404 / 8000) = 67.55\%$**

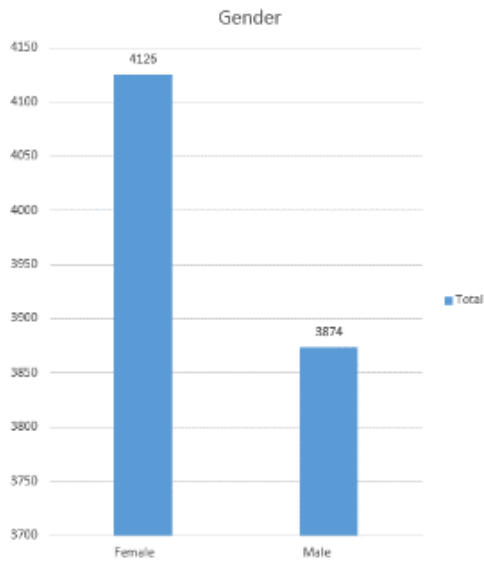
Support picture.



Question 2.

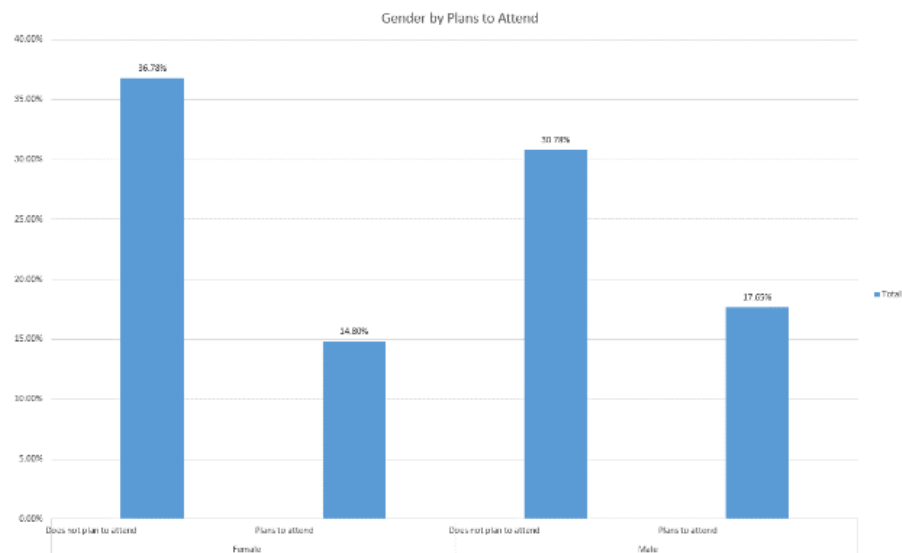
a) How many male students are in the sample? 3874

Support picture.



b) What percentage of the female students have plans to attend college?  $(1184 / 8000) = 14.8\%$

Support picture.



Question 3.

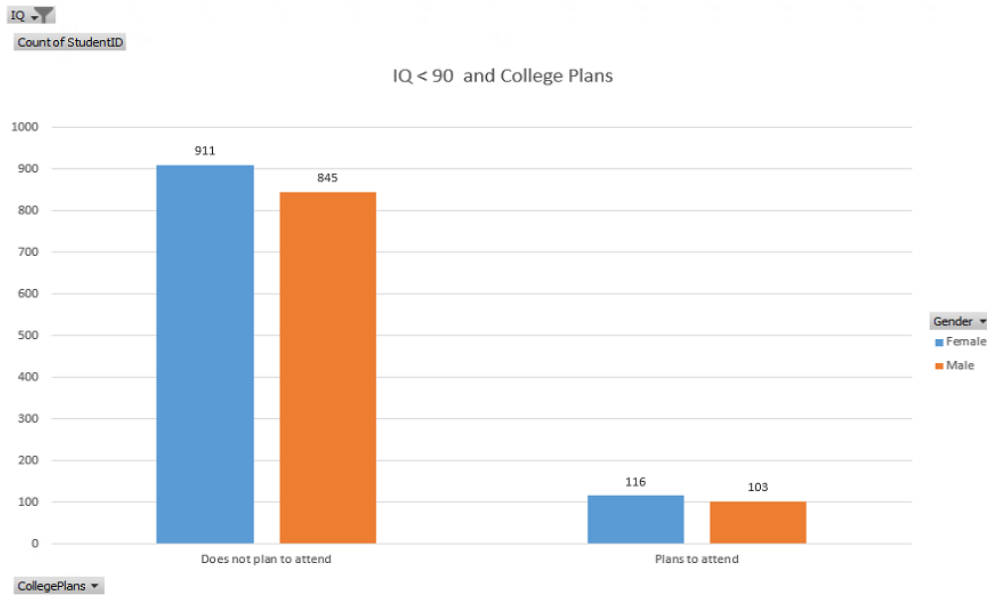
- a) What is the mean (average) IQ of the students in the sample dataset? 99.57

IQ	
Mean	99.57775
Standard Error	0.211572895
Median	100
Mode	107
Standard Deviation	18.92365499
Sample Variance	358.104718
Kurtosis	-0.432735465
Skewness	-0.003688643
Range	80
Minimum	60
Maximum	140
Sum	796622
Count	8000

- b) If a student's IQ is less than 90, is he/she more or less likely to have plans to attend college?

Student is **LESS** likely to have plans to attend college.

Support picture.



Question 4.

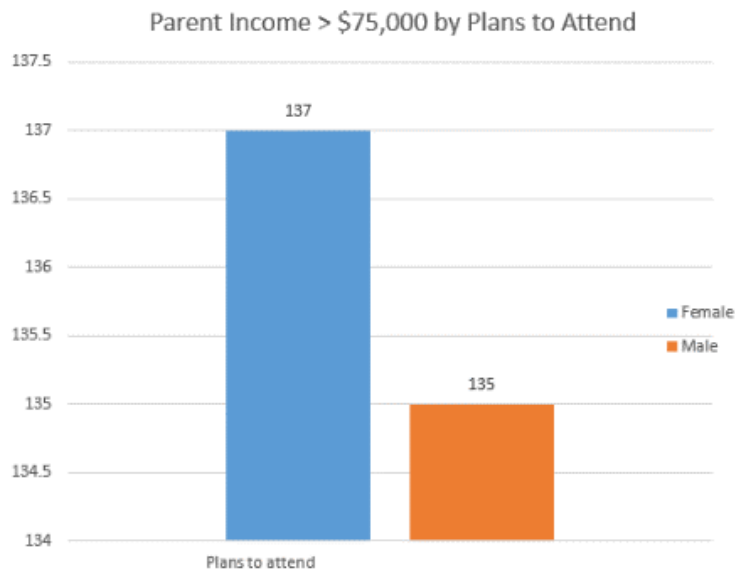
- a) What is the minimum parental income in the sample dataset? \$4500.00

Parent Income	
Mean	40584.0175
Standard Error	201.598371
Median	39330
Mode	31500
Standard Deviation	18031.50647
Sample Variance	325135225.5
Kurtosis	-0.469054067
Skewness	0.193386716
Range	77890
Minimum	4500
Maximum	82390
Sum	324672140
Count	8000

- b) If a student's parents have income greater than \$75,000, is he/she more or less likely to have plans to attend college? Offer support for your answer. (2 points)

In **ALL** cases, the student is **MORE LIKELY** to attend college with this parental income.

Support picture.



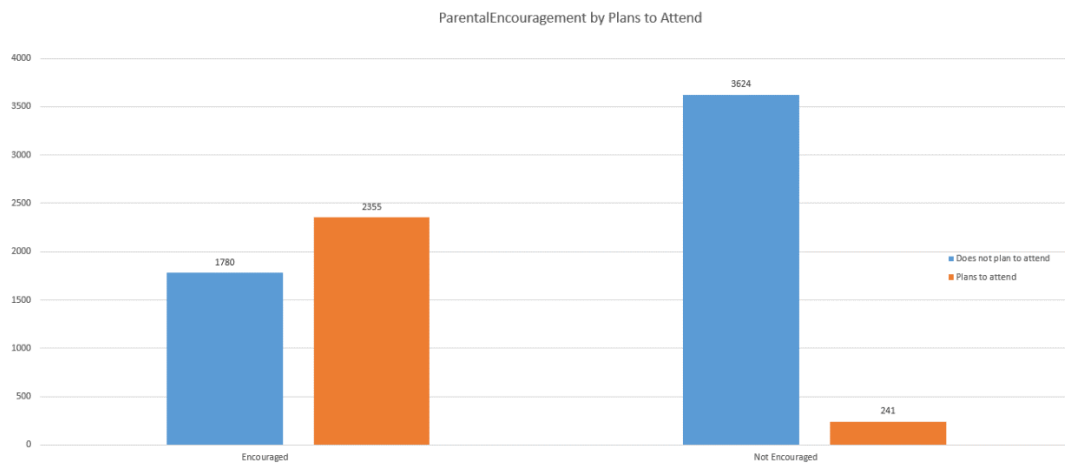
Question 5.

- a) Is Parental Encouragement likely to be a good predictor of a student's plans to attend college or not? Explain.

With reference to the following figure, in my opinion Parental encouragement is only a good predictor of a student's plans to NOT attend college. In this case 45.3% of students plan on NOT attending while 3.01% plan on attending when a parent does NOT encourage attendance. There is a decidedly heavy bias towards NOT attending college when a parent does not encourage it.

However, when parents do encourage students to attend college, 29.44% will plan ON attending whereas 22.25% will plan on NOT attending. This is a fairly close margin in this case. In my opinion here, a student is sort of on the fence as to whether they will plan on attending college or not if a parent encourages it, thus for this case parental encouragement does not make a good predictor.

Support picture.



Use the data to build a) Decision-Tree and b) Neural Network models to predict whether a student has plans to attend college or not. Use gender, IQ, parental income and encouragement as inputs. Reserve 25% of the data for testing and set the “HoldoutSeed” property to 75.

Question 6.

- a) Based on the results of the Neural Network model, which input variables (e.g., IQ) and their associated values (e.g., between 85 and 100) have the strongest influence on a female student's plans to a) attend college and b) not attend college. What are the probabilities for the two outcomes? (3 points)

Refer to the figure below for verification. The input variables with the strongest influence for females to **ATTEND** college are:

- i) Parental Income
  - (1) Range: **52,812.45 – 82,390**
  - (2) Percentages
    - (a) Does not plan to attend: **39.69%**
    - (b) Plans to attend: **60.28%**
- ii) Parent Encouragement
  - (1) Value: **Encouraged**
  - (2) Percentages
    - (a) Does not plan to attend: **58.79%**
    - (b) Plans to attend: **41.19%**
- iii) IQ
  - (1) Range: **112.416 – 140.0**
  - (2) Percentages
    - (a) Does not plan to attend: **75.71%**
    - (b) Plans to attend: **24.27%**
- iv) Parent Income
  - (1) Range: **40,683.56 – 52,938.26**
  - (2) Percentages
    - (a) Does not plan to attend: **78.39%**
    - (b) Plans to attend: **21.58%**

The input variables with the strongest influence for females to **NOT ATTEND** college are:

- i) Parent Encouragement
  - (1) Value: **NOT Encouraged**
  - (2) Percentages
    - (a) Does not plan to attend: **96.21%**
    - (b) Plans to attend: **3.76%**
- ii) Parental Income
  - (1) Range: **4,500 – 28,428.86**
  - (2) Percentages

- (a) Does not plan to attend: **95.94%**
- (b) Plans to attend: **4.04%**

iii) IQ

- (1) Range: **60.0 – 86.947**
- (2) Percentages
  - (a) Does not plan to attend: **92.47%**
  - (b) Plans to attend: **7.51%**

iv) Parental Income

- (1) Range: **28,428.86 – 40,683.56**
- (2) Percentages
  - (a) Does not plan to attend: **90.49%**
  - (b) Plans to attend: **9.49%**

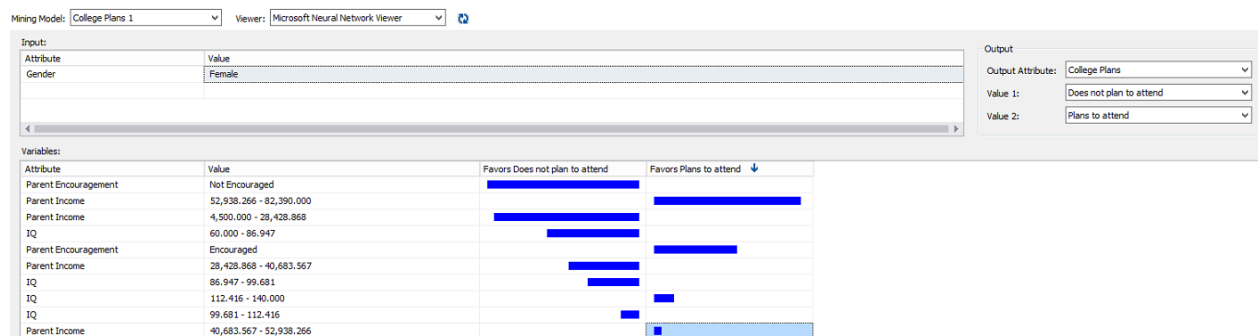
v) IQ

- (1) Range: **86.947 – 99.681**
- (2) Percentages
  - (a) Does not plan to attend: **88.18%**
  - (b) Plans to attend: **11.79%**

vi) IQ

- (1) Range: **99.681 – 112.416**
- (2) Percentages
  - (a) Does not plan to attend: **83.37%**
  - (b) Plans to attend: **16.61%**

Support picture.



Question 7.

- a) Based on the results of the Neural Network model, which input variables (e.g., IQ) and their associated values (e.g., between 85 and 100) have the strongest influence on a male student's plans to a) attend college and not attend college. What are the probabilities for the two outcomes? (3 points)

Refer to the figure below for verification. The input variables with the strongest influence for males to **ATTEND** college are:

- i) Parent Encouragement
  - (1) Value: **Encouraged**
  - (2) Percentages
    - (a) Does not plan to attend: **48.23%**
    - (b) Plans to attend: **51.75%**
- ii) Parental Income
  - (1) Range: **52,938.26 – 82,390**
  - (2) Percentages
    - (a) Does not plan to attend: **50.34%**
    - (b) Plans to attend: **49.63%**
- iii) IQ
  - (1) Range: **112.416 – 140.0**
  - (2) Percentages
    - (a) Does not plan to attend: **68.57%**
    - (b) Plans to attend: **31.41%**
- iv) Parental Income
  - (1) Range: **40,683.56 – 52,938.26**
  - (2) Percentages
    - (a) Does not plan to attend: **69.59%**
    - (b) Plans to attend: **30.39%**
- v) IQ
  - (1) Range: **99.681– 112.416**
  - (2) Percentages
    - (a) Does not plan to attend: **72.89%**
    - (b) Plans to attend: **27.09%**
- vi) IQ
  - (1) Range: **86.947 – 99.681**
  - (2) Percentages
    - (a) Does not plan to attend: **76.83%**
    - (b) Plans to attend: **23.15%**
- vii) Parental Income
  - (1) Range: **38,428.86 – 40,683.56**
  - (2) Percentages
    - (a) Does not plan to attend: **79.54%**

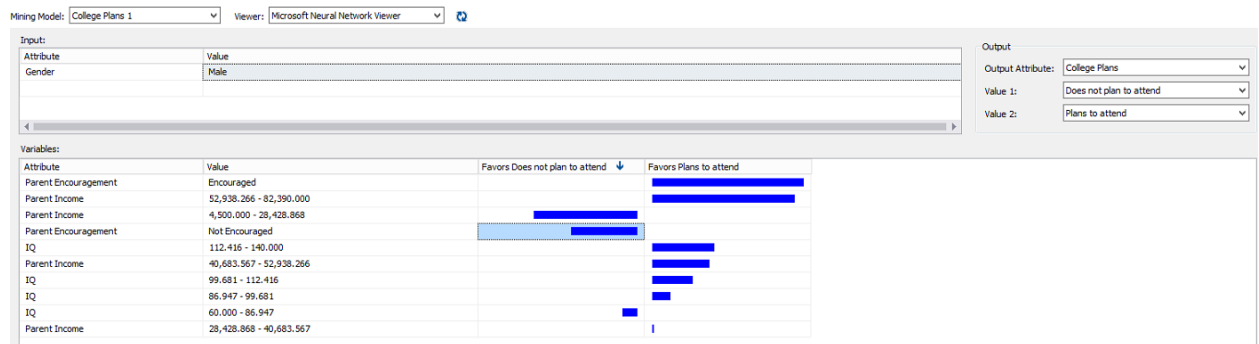


(b) Plans to attend: **20.44%**

The input variables with the strongest influence for males to **NOT ATTEND** college are:

- i) Parental Income
  - (1) Range: **4,500 – 28,428.86**
  - (2) Percentages
    - (a) Does not plan to attend: **91.47%**
    - (b) Plans to attend: **8.5%**
- ii) Parent Encouragement
  - (1) Value: **NOT Encouraged**
  - (2) Percentages
    - (a) Does not plan to attend: **88.29%**
    - (b) Plans to attend: **11.69%**
- iii) IQ
  - (1) Range: **60.0 – 86.947**
  - (2) Percentages
    - (a) Does not plan to attend: **82.15%**
    - (b) Plans to attend: **17.82%**

Support picture.



Question 8.

- a) Based on the results of the Decision Tree model, identify the node (with a minimum of 100 cases) that has students with the highest probability of not planning to attend college? What is the probability? What is the rule for this node? (3 points)

Node: **Parent Income < 43445**

Probability: **97.18%**

Node Rule: **Parent Encouragement = 'Not Encouraged' and IQ < 100 and Parent Income < 43445**

Question 9.

- a) Based on the results of the Decision Tree model, identify the node that has students with the highest probability of planning to attend college? What is the probability? What is the rule for this node? (3 points)

Node: **Parent Income >= 67591**

Probability: **99.97%**

Node Rule: **Parent Encouragement = 'Encouraged' and Parent Income >= 67591 and < 74601**

Question 10.

- a) Based on the results of the Decision Tree model, which inputs are the weakest and strongest predictors of “College Plans”? (1 point)

Weakest Predictor: **IQ**

Strongest Predictor: **Parent Encouragement**

Question 11.

- a) Complete the following table for the a) Decision Tree and b) Neural Network models. (2 points)

i) **Decision Tree**

Mining Structure Mining Models Mining Model Viewer Mining Accuracy Chart Mining Model Prediction

Input Selection Lift Chart Classification Matrix Cross Validation

Columns of the classification matrices correspond to actual values; rows correspond to predicted values

Counts for College Plans on College Plans:

Predicted	Does not plan to attend (Actual)	Plans to attend (Actual)
Does not plan to attend	1231	191
Plans to attend	145	433

ii) **Neural Network**

Mining Structure Mining Models Mining Model Viewer Mining Accuracy Chart Mining Model Prediction

Input Selection Lift Chart Classification Matrix Cross Validation

Columns of the classification matrices correspond to actual values; rows correspond to predicted values

Counts for College Plans 1 on College Plans:

Predicted	Does not plan to attend (Actual)	Plans to attend (Actual)
Does not plan to attend	1217	166
Plans to attend	159	458

Question 12.

- a) Based on the values in the classification tables, which model has a lower percentage of cases predicted *incorrectly*? What is the percentage? (3 points)

Decision Tree:  $(191 + 145) / 2000 = 16.80\%$

Neural Network:  $(166 + 159) / 2000 = 16.25\%$

**The Neural Network with 16.25% predicted incorrectly is the lowest.**

Question 13.

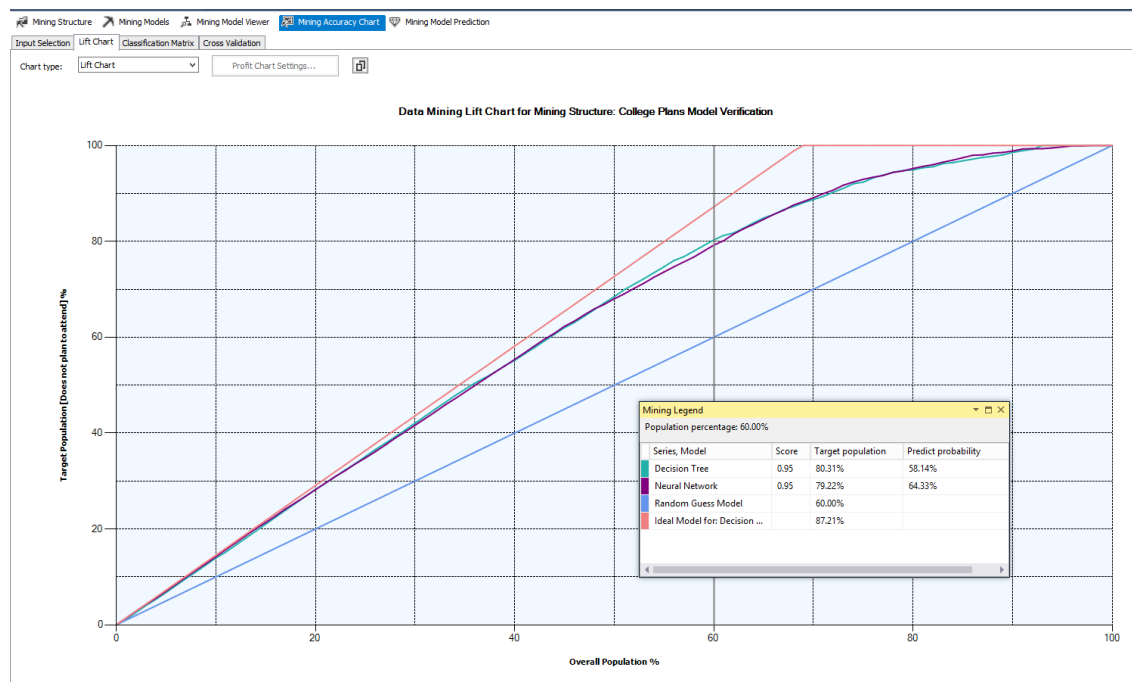
- a) For the Predict Value, “Does not plan to attend”, at 60% of the overall population, what percentage of the target population is correctly predicted by a) the ideal model, b) the Decision Tree model, and c) the Neural Network model? (3 points)

i) Ideal Model: 87.21%

ii) Decision Tree: 80.31%

iii) Neural Network: 87.21%

Support picture.

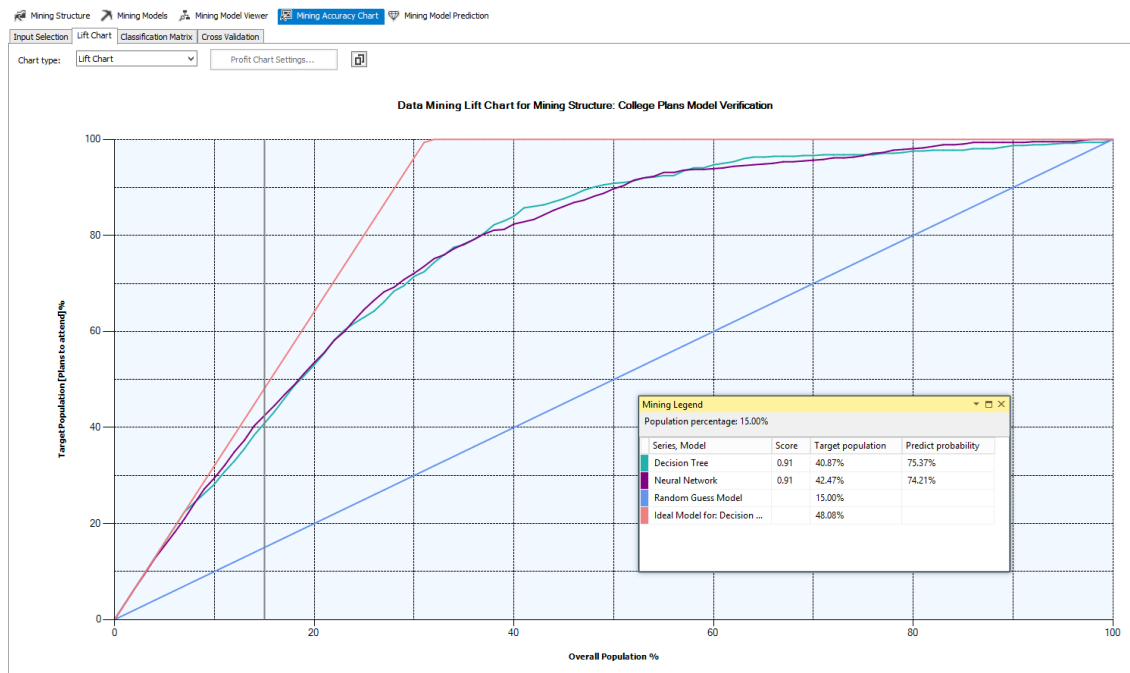


Question 14.

- a) For the Predict Value, “Plans to attend”, at 15% of the overall population, what percentage of the target population is correctly predicted by a) the ideal model, b) the Decision Tree model, and c) the Neural Network model? (3 points)

- i) Ideal Model: 48.08 %
- ii) Decision Tree: 40.87%
- iii) Neural Network: 42.47%

Support picture.



Question 15.

Ram U will pay you \$10 for each mailed brochure that reaches a high school graduate who “Plans to attend college”. Assume that the population of high school graduates is 12,000, the fixed cost for mailing the brochures is \$2,000 and it costs \$2 to mail each brochure. How many students in this population will have plans to attend college? If your objective is to maximize profit, a) which model’s (i.e., Decision Tree or Neural Network) recommendation will you follow, b) to how many students should you mail the brochures, and c) what will be your profit? (4 points)

How many students in this population will have plans to attend college?

If your objective is to maximize profit:

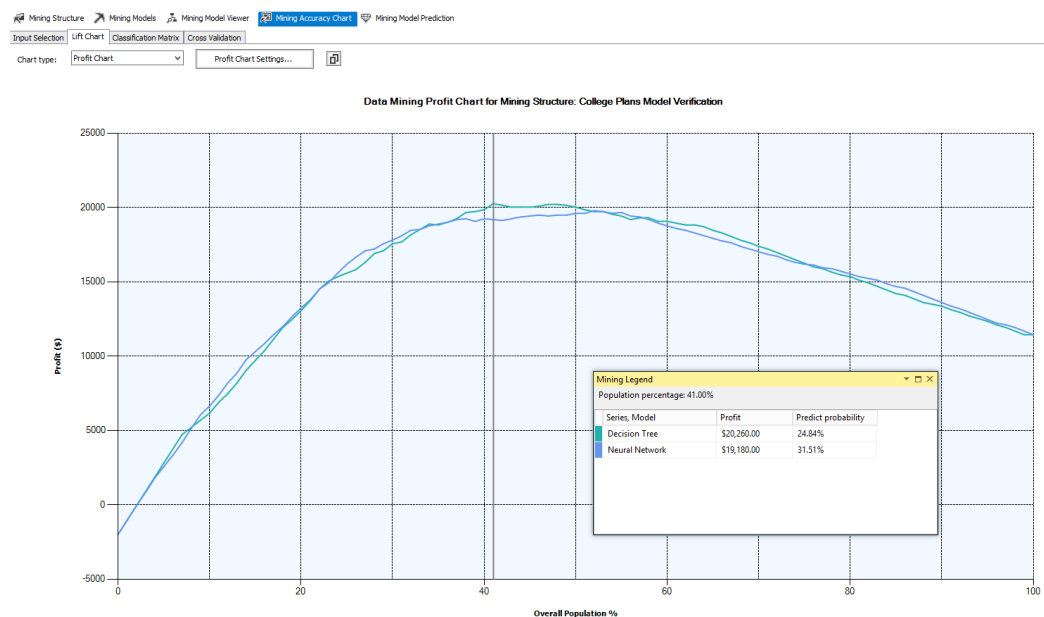
a) Which model’s (i.e., Decision Tree or Neural Network) recommendation will you follow?

Profit is maximized following the Decision Tree Model

Population %: 41%

Profit: \$20,260.00

Support picture.



b) To how many students should you mail the brochures?

Maximum profit is reached at 41% of the population of 12,000 students.

Thus, the brochure should be mailed to:  $0.41 * 12,000 = 4,920$  students.

c) What will be your profit?

Profit would be:  $4,920$  (students) \*  $\$10$  (profit per student) =  $\$49,200.00$