

1. Using the information in Exhibits 5 and 6, summarize – for the Director of Marketing – which customer groups should be targeted with this mailing.

EXHIBIT 5

CHAID Tree
Based on 2/3rds of customer records used for 'training' sample

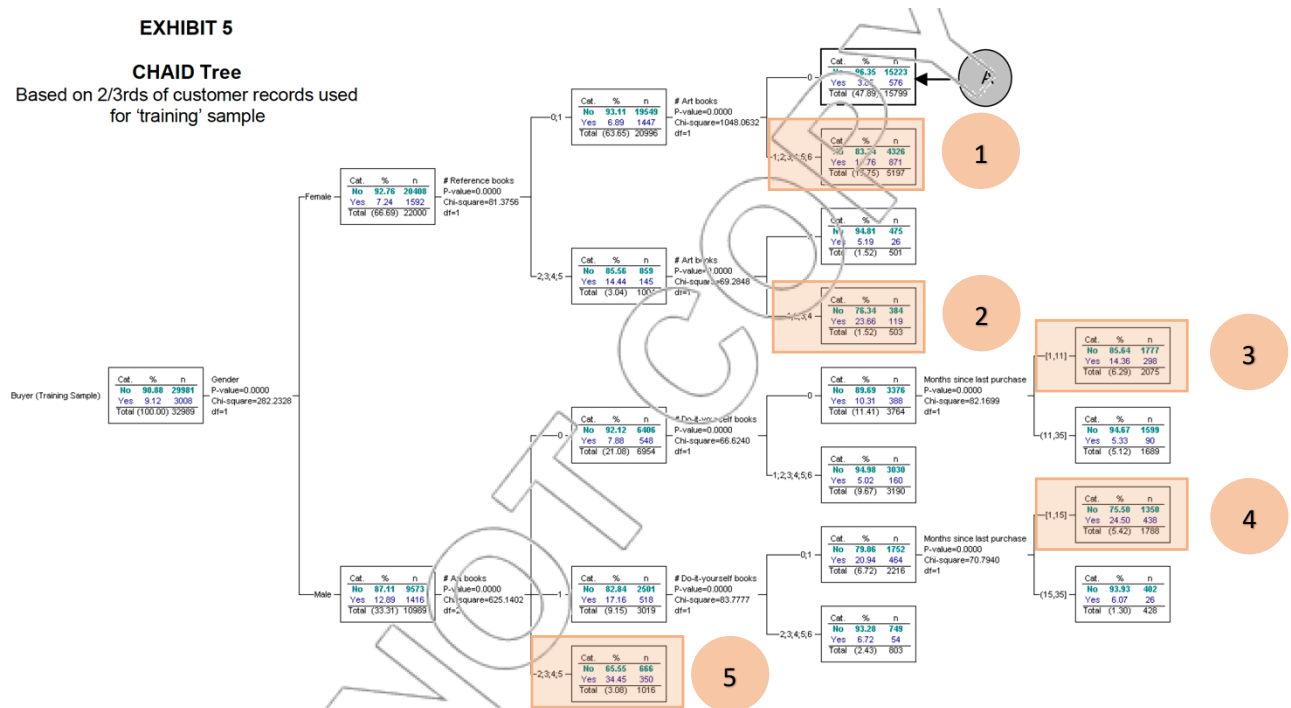
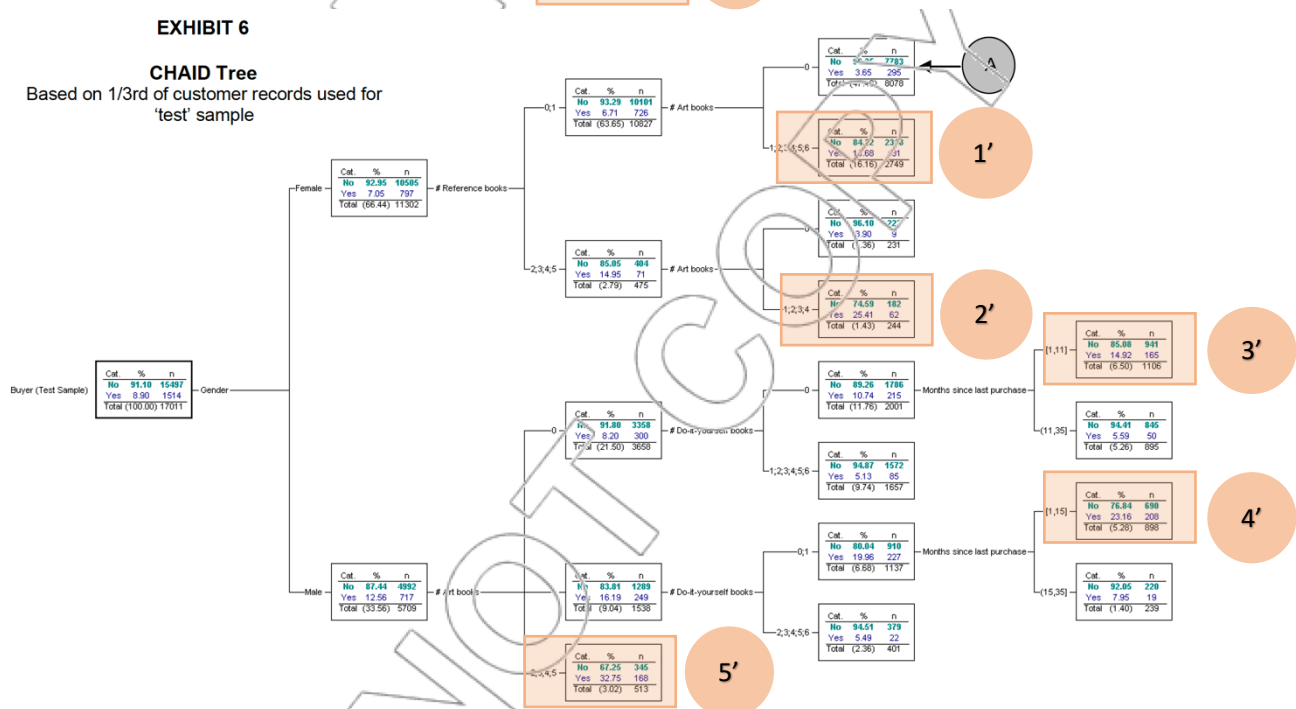


EXHIBIT 6

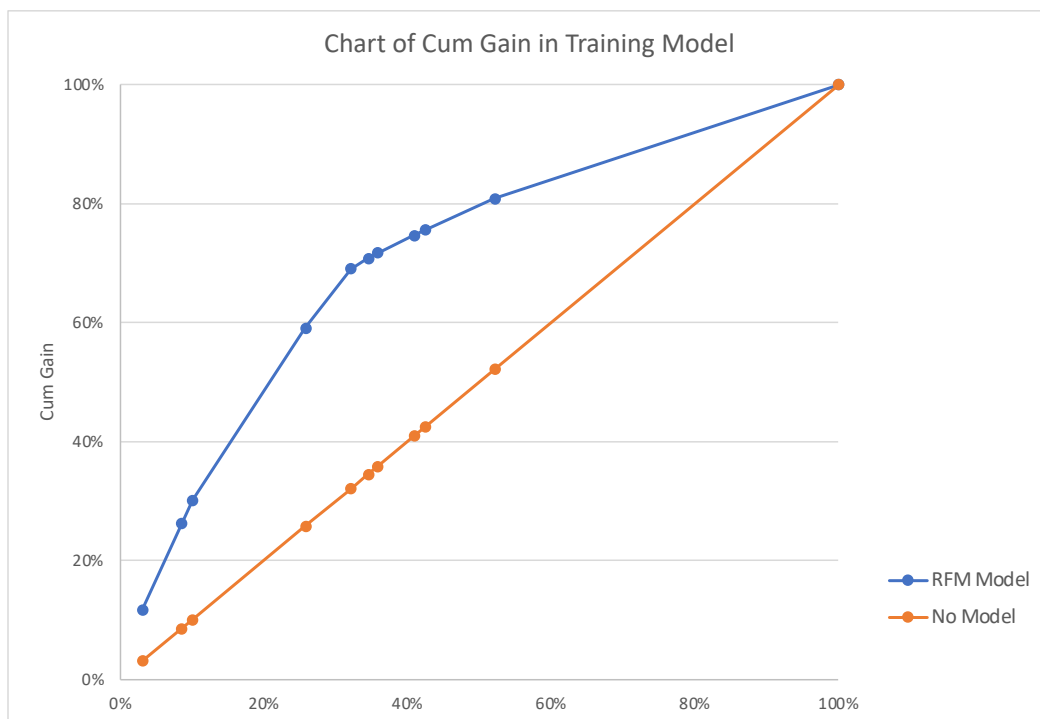
CHAID Tree
Based on 1/3rd of customer records used for 'test' sample

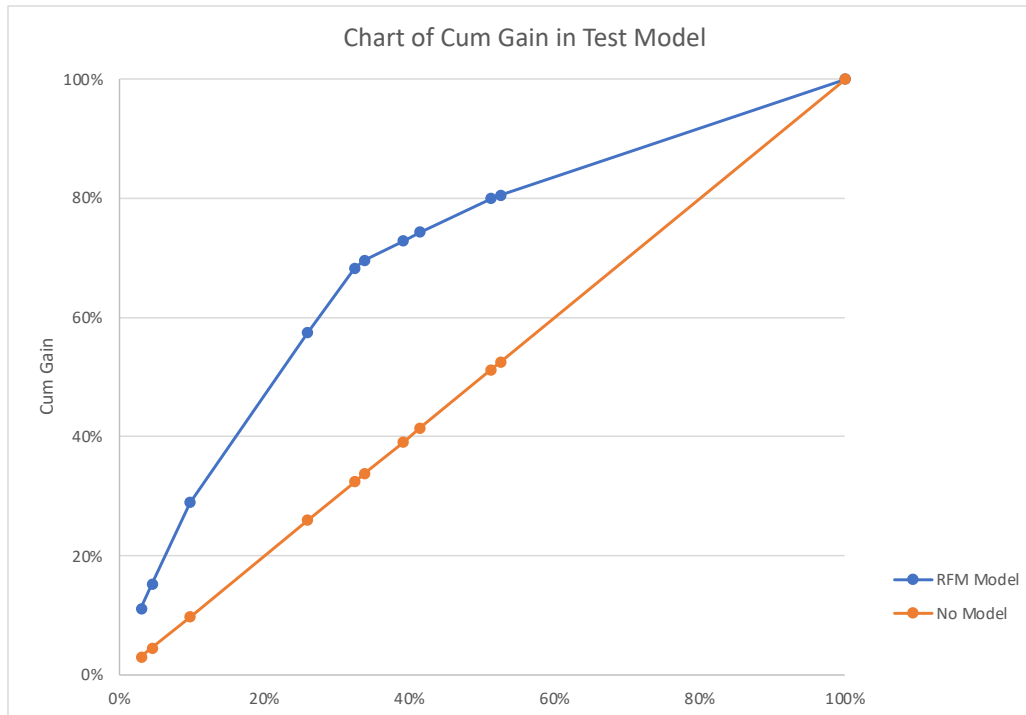


In Exhibits 5 and 6, decision tree model allows us making more effective mail campaign that could be rely on nodes where includes the specific customer segment. Nodes covered by colorful rectangular represent particular customer groups with a high response rate in Exhibit 5.

1. The 1st group are female customers who bought 0 or 1 reference books and more than 1 art books. The response rate is 16.76%.
2. The 2nd group are female customers who bought 2 or more reference books and purchased 1 or more art books. The response rate is 23.66%.
3. The 3rd group are male customers who bought neither art book nor Do-it-yourself book and purchased a book before 1 to 11 months. The response rate is 14.36%.
4. The 4th group are male customers who bought 1 art book and 0 or 1 Do-it-yourself book and purchased a book before 1 to 15 months. The response rate is 24.50%.
5. The 5th group are male customers who bought 2 or more art books. The response rate is 34.45%.

2. Use the information in Exhibit 7 to make a cumulative gains chart for both the training and test samples. Does the tree appear 'overfit'? Why or why not?





Basically, the training model performed well on 'new' observations. We get similar results from both training and test models. So, the decision tree does not appear the overfit in this scenario.

3. Using the same costs as before (\$18 selling price, \$9 wholesale price, \$3 shipping and \$0.50 mailing costs), estimate what the gross profit (in dollars and as a % of gross sales) as well as the return on marketing would be if the "The Art History of Florence" offer were only mailed to those predicted by the CHAID tree results to be good prospects for this offer.

Total # customer = 15586

Total # targeted customer = 3110

Total Mail cost = \$.5 * 15586 = \$7793

Sale Profit = (\$18 - \$9 - \$3) * 3110 = \$18660

Gross Profit = \$18660 - \$7793 = \$10876

Return on marketing = Gross Profit/ Total Mail Cost = \$10876/ \$7793 * 100% = 139%

4. Compare and contrast the results and insights from using RFM, logistic regression and CHAID decision tree analysis for targeting buyers for BookBinders offers.

Model Name	Gross Profit	ROI
Decision Tree	\$7,418	139%
Logistic Regression	\$12,157	156%
RFM Model	\$7,919	70%

Compared with ROI values across different models, logistic regression gets the most ideal outcome that the gross profit is the highest, \$12,157 and the value of the return on marketing is 156%. Between decision tree and RFM model acquiring similar gross profit, decision tree has a profitable performance with the ROI of 139%.

Although the logistic regression model is the better one, some weaknesses exist, that is the outcome would be influenced by outliers. For example, some customers purchased much far more than others.

Decision Tree, however, cannot be influenced by outliers. And it is easy to apply. What should be noticed is that this model is prone to split customers aggressively, which means we would get a lot of terminal nodes, customer groups, resulting to an overfit situation easily.

Overall, these two models could be tested with other data to identify if it is robust to allow us making marketing campaign plans.

Conducting RFM model is an easy way when we want a quick success. It groups customers from three dimensions, recency, frequency and monetary. Within the model, results could be disturbed among dimensions. A customer who has high purchasing frequency but spends less is combined into the group with lower response rate.