Superstore Marketing Presentation

By Bridget Bremenour, Zeynep Cetin, and Mustafa Mansour

Data and Background Information
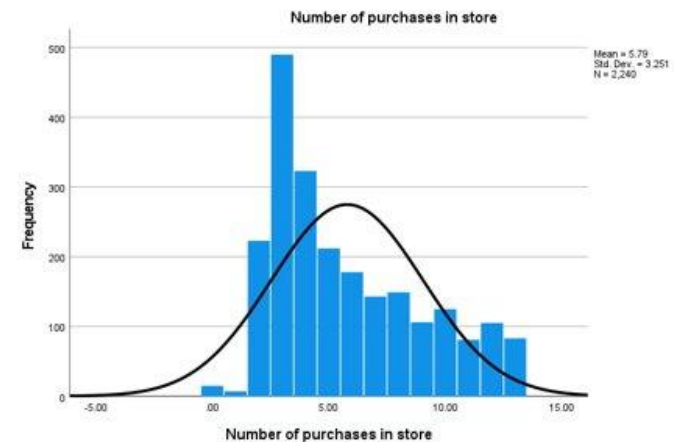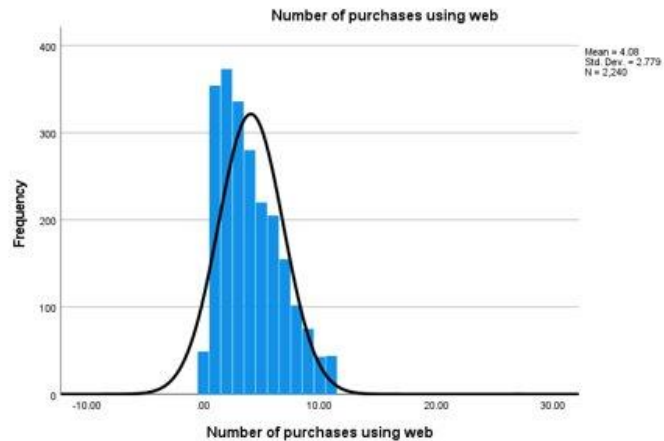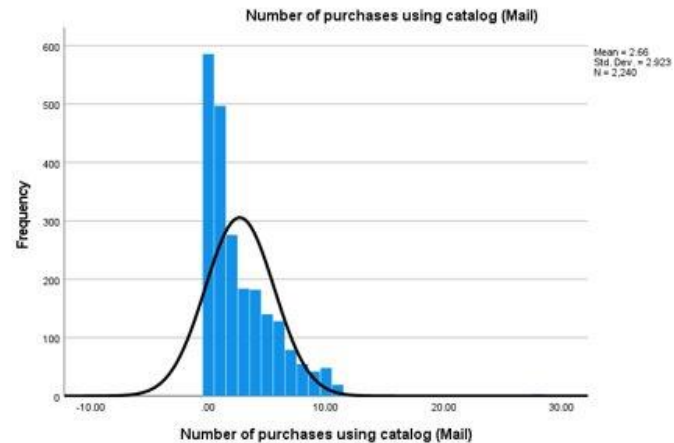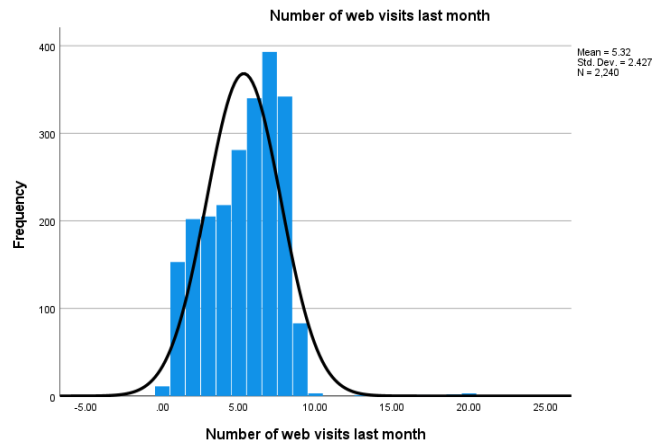
Context: A superstore is planning for the year-end sale. They want to launch a new offer - gold membership, that gives a 20% discount on all purchases, for only $499 which is $999 on other days. It will be valid only for existing customers and the campaign through phone calls is currently being planned for them.

Objective: Build the best predictive model to determine the likelihood of a customer purchasing the membership or not.
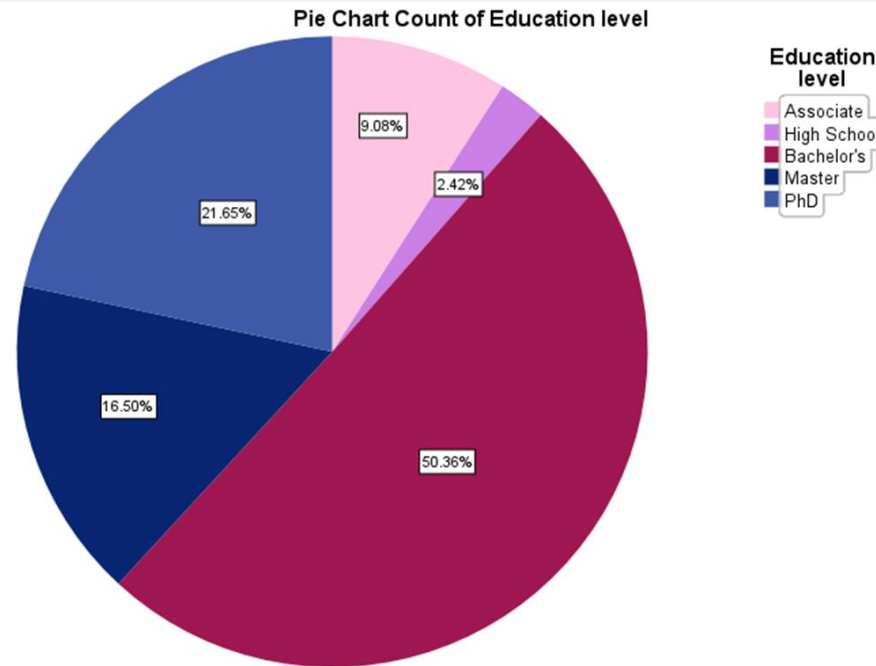
Additional Information: The dataset consists of 19 possible predictors and 2240 observations.

The dataset is obtained from Kaggle.com

Preliminary Analysis – 1

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Year of birth	2240	1893.00	1996.00	1968.8058	11.98407
Yearly income	2216	1730.00	666666.00	52247.2514	25173.07666
Number of kids at home	2240	.00	2.00	.4442	.53840
Number of teens at home	2240	.00	2.00	.5062	.54454
Months since customer joined	2240	99.00	134.00	115.8728	7.68590
Number of days since last purchase	2240	.00	99.00	49.1094	28.96245
Amount spent on wine purchases last 2 years	2240	.00	1493.00	303.9357	336.59739
Amount spent on fruit purchases last 2 years	2240	.00	199.00	26.3022	39.77343
Amount spent on meat product purchases last 2 years	2240	.00	1725.00	166.9500	225.71537
Amount spent on fish product purchases last 2 years	2240	.000	259.000	37.52545	54.628979
Amount spent on sweet product purchases last 2 years	2240	.00	263.00	27.0629	41.28050
Amount spent on gold product purchases last 2 years	2240	.00	362.00	44.0219	52.16744
Number of purchases made with discount	2240	.00	15.00	2.3250	1.93224
Number of purchases using web	2240	.00	27.00	4.0848	2.77871
Number of purchases using catalog (Mail)	2240	.00	28.00	2.6621	2.92310
Number of purchases in store	2240	.00	13.00	5.7902	3.25096
Number of web visits last month	2240	.00	20.00	5.3165	2.42665
Accepted the offer in the last campaign	2240	0	1	.15	.356
Valid N (listwise)	2216				

Number of web visits last month

Mean = 5.32
Std. Dev. = 2.427
N = 2,240

Number of purchases using catalog (Mail)

Mean = 2.66
Std. Dev. = 2.923
N = 2,240

Number of purchases using web

Mean = 4.08
Std. Dev. = 2.779
N = 2,240

Number of purchases in store

Mean = 5.79
Std. Dev. = 3.251
N = 2,240

4

Preliminary Analysis – 2

Pie Chart Count of Education level

The pie chart was built based on the education levels of the existing customers. We can see that more than half (50.36%) of the existing customers have at least a Bachelor's Degree, and the least amount of the customers have a high school degree (2.42%).

Preliminary Analysis - 3

Marital status * Accepted the offer in the last campaign Crosstabulation

Count

		Accepted the offer in the last campaign		
		No	Yes	Total
Marital status	Divorced	184	48	232
	Married	766	98	864
	Single	376	107	483
	Together	520	60	580
	Widow	58	19	77
Total		1904	332	2236

In this crosstab, we can see the marital status of the people who accepted the offer in the last campaign or not. 32.22% of the total amount of people who accepted the offer in the last campaign's marital status is single.

However, single people only make up 21.6% of the total customers in the campaign. Widowed people make up the lowest group of people who accepted the offer in the last campaign, with 5.72%, but 32.76% of the widowed people accepted the offer.

CART

▶ The best CART model was found by running several different models, and obtaining an equilibrium between a low risk, high sensitivity and specificity, and high index.

▶ After running several trees, the best model was achieved by setting the minimum cases in parent nodes to 80, and child nodes to 40, and changing the threshold for classifying successes from 50% to 35%. As our overall response rate contained only 15% of the successes.

Model Summary

|		
Specifications	Growing Method	CRT
	Dependent Variable	Accepted the offer in the last campaign
	Independent Variables	Year of birth, Amount spent on meat product purchases last 2 years, Number of purchases in store, Number of purchases using catalog (Mail), Number of web visits last month, Number of purchases using web, Amount spent on fish product purchases last 2 years, Number of purchases made with discount, Amount spent on fruit purchases last 2 years, Amount spent on gold product purchases last 2 years, Education level, Amount spent on sweet product purchases last 2 years, Amount spent on wine purchases last 2 years, Yearly income, Months since customer joined, Number of days since last purchase, Marital status, Number of teens at home, Number of kids at home
	Validation	Split Sample
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	80
	Minimum Cases in Child Node	40
Results	Independent Variables Included	Yearly income, Number of purchases made with discount, Amount spent on meat product purchases last 2 years, Amount spent on sweet product purchases last 2 years, Amount spent on fruit purchases last 2 years, Number of purchases using catalog (Mail), Number of web visits last month, Amount spent on wine purchases last 2 years, Number of days since last purchase, Number of purchases using web, Amount spent on gold product purchases last 2 years, Number of purchases in store, Amount spent on fish product purchases last 2 years, Education level, Months since customer joined, Year of birth, Number of kids at home, Marital status
	Number of Nodes	13
	Number of Terminal Nodes	7
	Depth	5

The Overall Tree & Nodes 0-4:

➢ The overall tree consists of 5 levels and 12 nodes (13 including the parent node).

➢ The best predictor for whether the customer accepted the offer in the last campaign or not is the yearly income. 40.2% of those who have a yearly income of 81930 or higher accepted the offer.

➢ For those who earned less than or equal to 81930 annually, 12.8% accepted the offer.

➢ Out of those who earned less than or equal to 81930, the best predictor was the number of days since last purchase. For those who had number of days since last purchase less than or equal to 15.5 days, 25.7% accepted the offer. For those who had more than 15.5 days, only 10.3% accepted the offer.
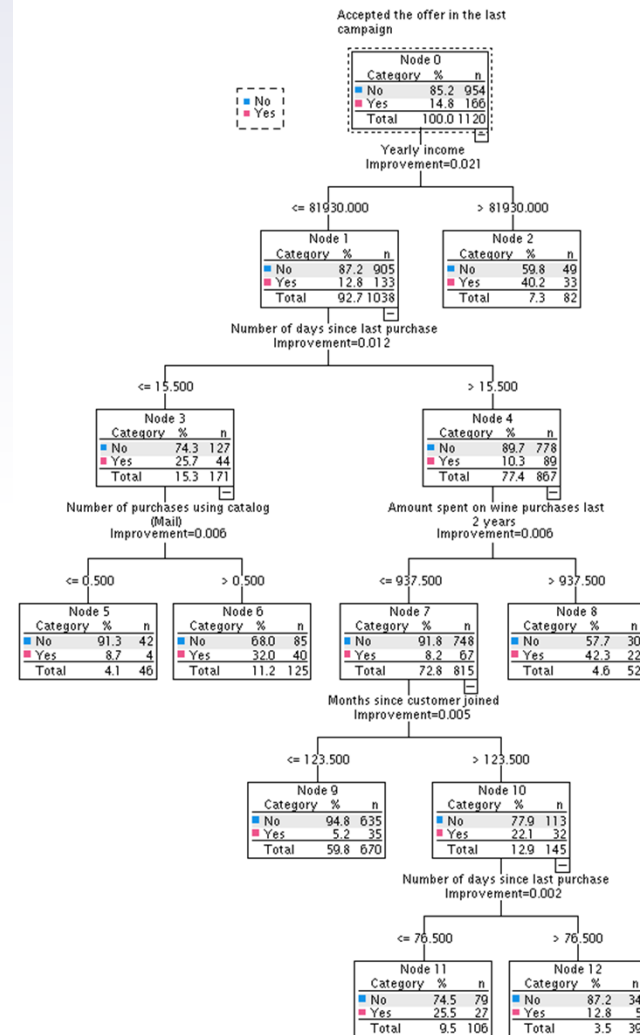
Nodes (5-8):

➢ Out of those whose last purchase was less than or equal to 15.5 days, the best predictor was the number of purchases using the catalog. For those who had a purchase of more than 1, (Since one can't make 0.5 purchases) 32% accepted the offer, compared to 8.7% of those who did not make any purchase using the catalog.

➢ For those whose last purchase was more than 15.5 days ago, the best predictor was the amount spent on wine purchases in the last 2 years. Of those who spent more than 937.50 dollars, 42.3% of them accepted the offer, compared to 8.2% of those who spent less than or equal to 937.50 dollars.
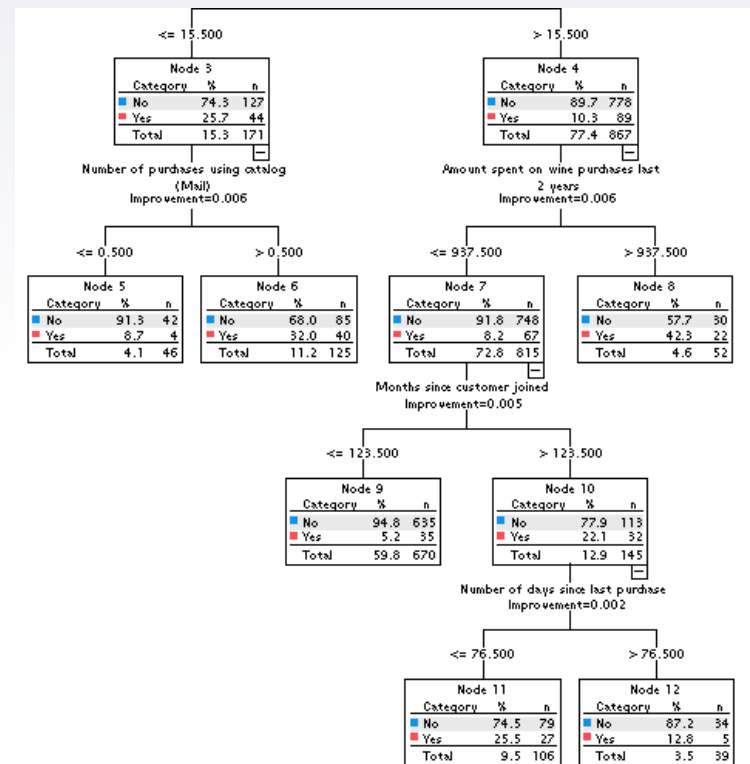
Nodes (9-12):

➢ Out of those who spent less than or equal to 937.50 on wine in the last two years, the best possible predictor was months since the last purchase. For those customers who joined the company more than 123 months ago, 22.1% accepted the offer compared to 5.2% of those who joined the company less than or equal to 123 months ago.

➢ For those who joined the company more than 123 months ago, the best predictor was number of days since last purchase. The customers whose last purchase was more than 76.5 days ago, 12.8% accepted the offer, compared to 25.5% of those whose last purchase was less than or equal to 76.5 days ago.
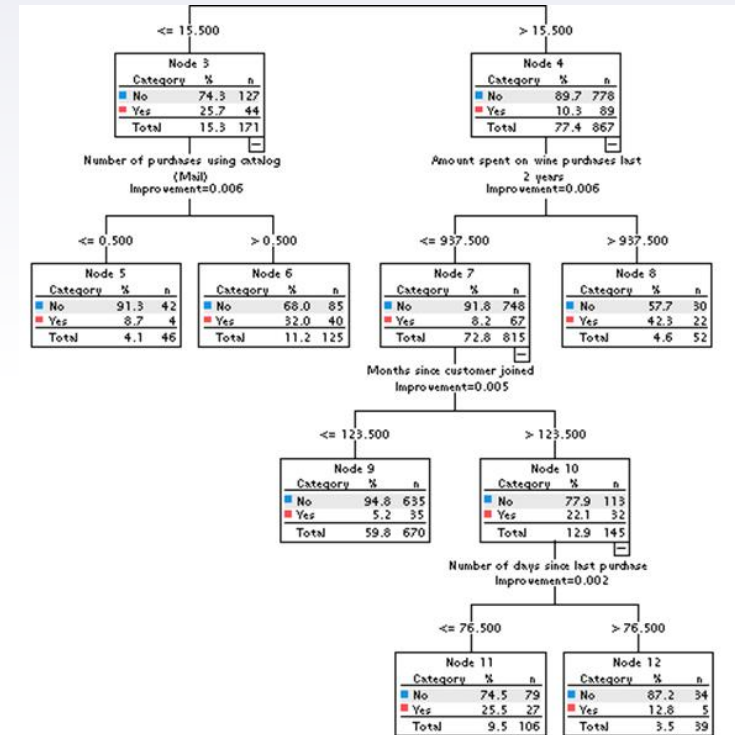
Gains Table and Gains Plots: Target Category Yes

➢ While the Node section of the gains table shows the percentage of the customers in the node compared to the total customers in the dataset, the response percentage was sorted by the target category, which is Yes.

➢ From the index, we could interpret that the individuals who fell in the node 2, they responded at a rate 171.5% higher compared to the overall (testing). The ones who fell under node 9 terminal node had the worst rate of accepting the offer with a rate of 4.5%.

➢ From the Gains Plots, we can see that there is a slight difference in the model benefit area between the testing and training samples, but it's a decent size.

➢ The top 50% of the file contained around 82% of all the customers who accepted the offer (testing sample).

Gains for Nodes

Sample	Node	Node N	Node Percent	Gain N	Gain Percent	Response	Index
Training	2	84	7.5%	42	25.3%	50.0%	336.1%
	6	124	11.1%	48	28.9%	38.7%	260.2%
	8	40	3.6%	15	9.0%	37.5%	252.1%
	11	85	7.6%	23	13.9%	27.1%	181.9%
	5	49	4.4%	4	2.4%	8.2%	54.9%
	12	48	4.3%	3	1.8%	6.3%	42.0%
	9	686	61.5%	31	18.7%	4.5%	30.4%
Test	2	82	7.3%	33	19.9%	40.2%	271.5%
	6	125	11.2%	40	24.1%	32.0%	215.9%
	8	52	4.6%	22	13.3%	42.3%	285.4%
	11	106	9.5%	27	16.3%	25.5%	171.9%
	5	46	4.1%	4	2.4%	8.7%	58.7%
	12	39	3.5%	5	3.0%	12.8%	86.5%
	9	670	59.8%	35	21.1%	5.2%	35.2%

Growing Method: CRT
Dependent Variable: Accepted the offer in the last campaign

Target Category:Yes

Training Sample

Test Sample

Sample
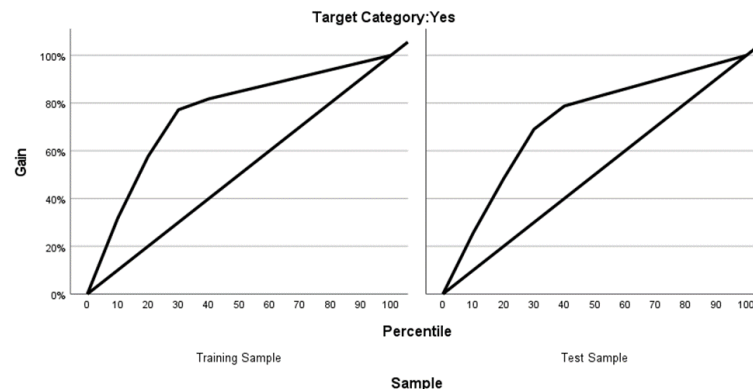
Growing Method:CRT
Dependent Variable:Accepted the offer in the last campaign

The Misclassification Table, Risk and ASE

➤ We predicted those who did not accept the offer 84.9% correctly (specificity) and we predicted those who accepted the offer correctly 63.3% after changing the classification threshold to 0.35.

➤ The Average Squared Error for the adjusted classification threshold is higher than the original .50, However, the ASE for both testing and training samples are the very close for both cases, which tells us that the model does not have any issues with overfitting.

➤ 14.9% of the training dataset is classified incorrectly, compared to 14.8% of testing. Since they are almost identical, this tells us there are no issues with overfitting.

Predicted Probability * Accepted the offer in the last campaign Crosstabulation[a]

| | | | Accepted the offer in the last campaign | | |
			No	Yes	Total
Predicted Probability	.00	Count	807	61	868
		% within Accepted the offer in the last campaign	84.9%	36.7%	77.8%
	1.00	Count	143	105	248
		% within Accepted the offer in the last campaign	15.1%	63.3%	22.2%
Total		Count	950	166	1116
		% within Accepted the offer in the last campaign	100.0%	100.0%	100.0%

a. Sample Assignment = Training Sample   (0 stands for No 1 stands for Yes)

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
SQUEadjusted	1116	.00	.39	.0818	.13994
SQUEoriginal	1116	.00	.25	.0478	.07773
Valid N (listwise)	1116				

a. Sample Assignment = Training Sample

Risk

Sample	Estimate	Std. Error
Training	.149	.011
Test	.148	.011

Growing Method: CRT
Dependent Variable: Accepted the offer in the last campaign

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
SQUEadjusted	1120	.00	.39	.0869	.14277
SQUEoriginal	1120	.00	.25	.0501	.07750
Valid N (listwise)	1120				

a. Sample Assignment = Testing Sample

CHAID and Exhaustive CHAID

CHAID

- Number of nodes: 14

- Terminal nodes: 9

- I chose the standard depth of 3, I did run several other models with 4 and 5 levels, but they did not improve anything significantly

- There were originally 19 predictor variables, but the tree only included 5 total.

Model Summary

|		
Specifications	Growing Method	CHAID
	Dependent Variable	Accepted the offer in the last campaign
	Independent Variables	Year of birth, Education level, Marital status, Yearly income, Number of kids at home, Number of teens at home, Months since customer joined, Number of days since last purchase, Amount spent on wine purchases last 2 years, Amount spent on fruit purchases last 2 years, Amount spent on meat product purchases last 2 years, Amount spent on fish product purchases last 2 years, Amount spent on sweet product purchases last 2 years, Amount spent on gold product purchases last 2 years, Number of purchases made with discount, Number of purchases using web, Number of purchases using catalog (Mail), Number of purchases in store, Number of web visits last month
	Validation	Split Sample
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	Amount spent on meat product purchases last 2 years, Number of days since last purchase, Number of web visits last month, Number of teens at home, Months since customer joined
	Number of Nodes	14
	Number of Terminal Nodes	9
	Depth	3

The Overall Tree:

➢ This is the overall CHAID model with 3 levels, the first split was based on amount spent on meat produces purchased in the last 2 years.

➢ It is a relatively small uncomplicated tree, and there is only one split resulting from the first level of the tree.

➢ The second split is based on number of days since last purchase

➢ Three of the terminal nodes come in the first level and there are a few reasons the model stopped at those nodes. It is mostly likely that splitting again would not have met the minimum case requirements for at least one of the child nodes which is 50. Nodes 1, 2, and 4 all have just above 100 cases and so the next split would have had to have been very close to 50/50 to meet child node case requirements

➢ This is just the tree for the training data set (50% of overall data 1130 cases)
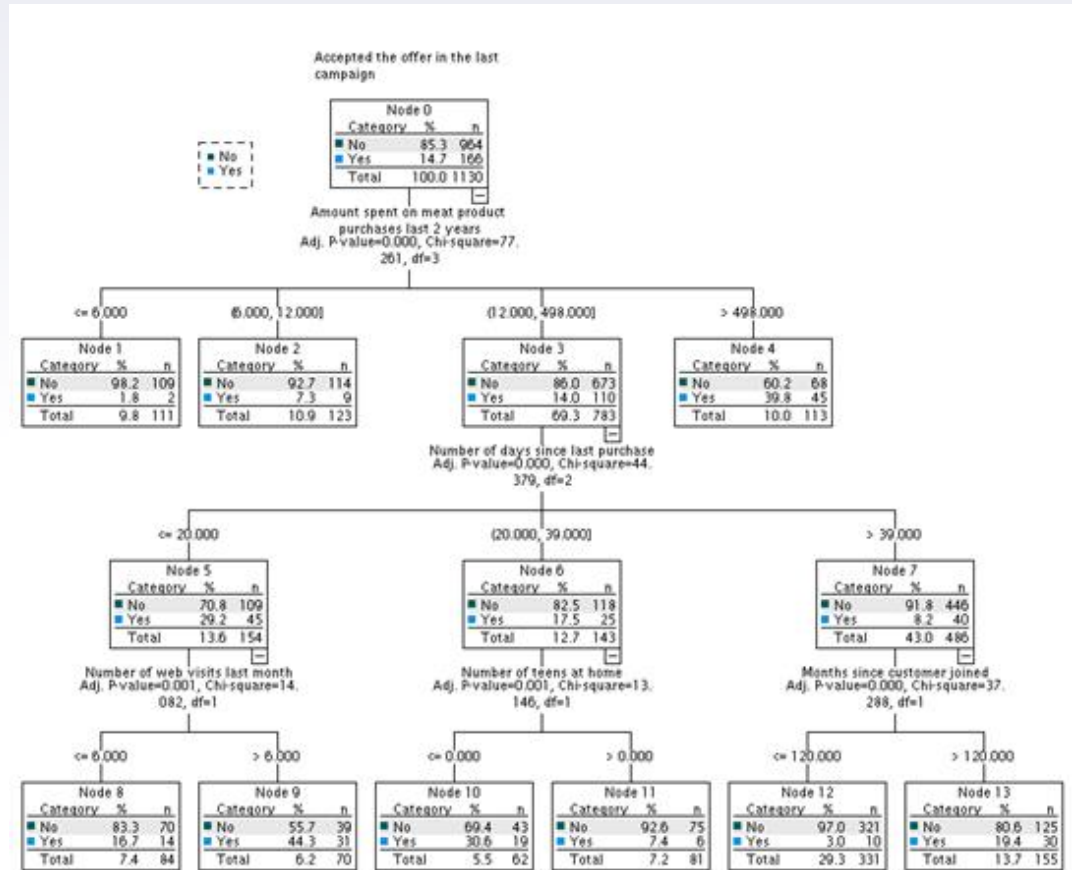
Level 1 and 2:

➤ This second level all stems from node 3
➤ Node 3 has a total of 783 cases and all of those customers purchased between $12 and $498 on meat products at this superstore
➤ Of the 783 cases 673 of those customers were predicted to not accept the offer in the last campaign
➤ The level two split is determined by number of days since last purchase, resulting in 3 child nodes none of which are terminal nodes.
➤ None of the nodes on the second level are particularly good at predicting a customer who accepts the offer from the sales campaign unfortunately
➤ Node 5 does have almost 30% of its case total predicted to have accepted the offer
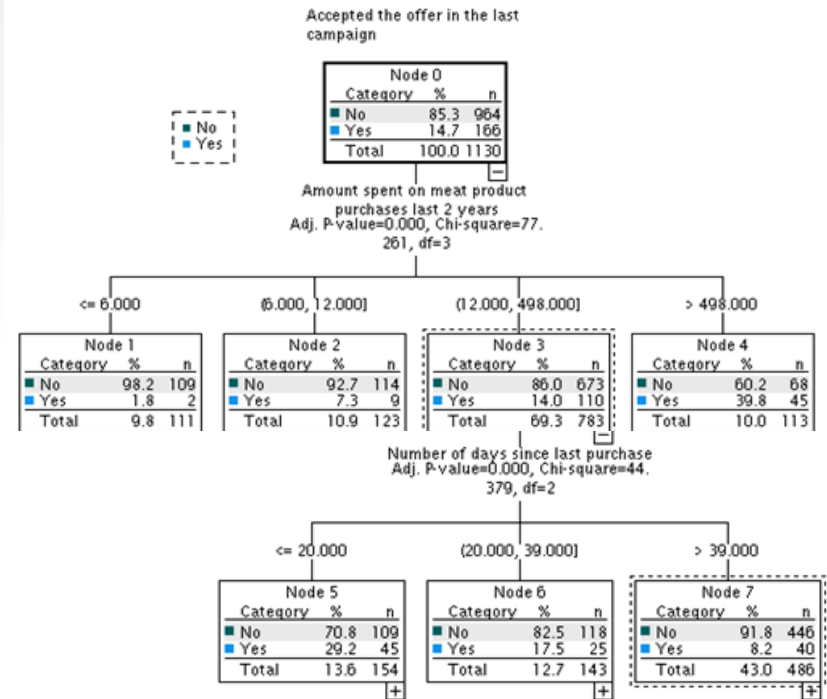
Level 3:

- ➤ The first circle is terminal nodes 8 and 9, node 8 contains 84 total cases and 83.3% of those cases are predicted not to response to the offer.
- ➤ Node 9 contains 70 customers and 55.7% are predicted not to response to the offer.
- ➤ The second circle is nodes 10 and 11, node 10 contains 62 customers and 69.4% are predicted not to response. Node 11 contains 81 cases and 92.6% are predicted not to response.
- ➤ The last two terminal nodes were both fairly good at predicting who wouldn't accept an ad offer, with 12 being close to 100% predicted to not accept.
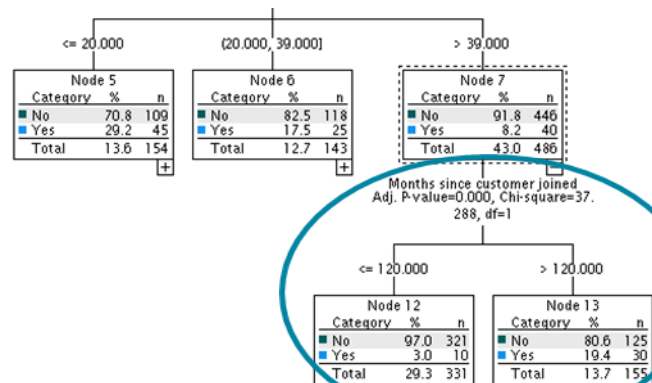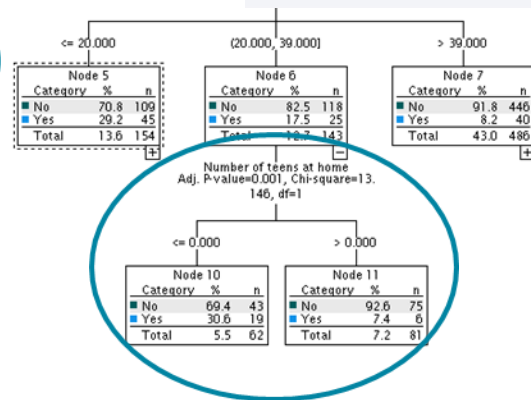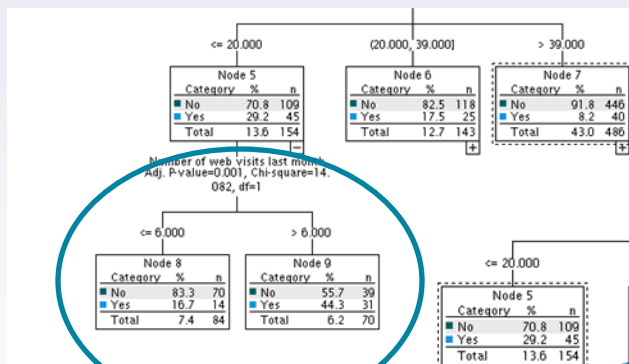
Risk, Misclassification, ASE:

➢ Risk was 14.7% for the training data and 15% for the testing set. It is good to see that those values are pretty close to each other.

➢ Default for classification is 50/50 so we changed that threshold, we lowered it to 0.35 in order to capture a more accurate picture of the data and with that we got a specificity of 88.9% and sensitivity of 45.8%.

➢ The average squared error was 0.1053

Risk

Sample	Estimate	Std. Error
Training	.147	.011
Test	.150	.011

Classification

Sample	Observed	Predicted No	Predicted Yes	Percent Correct
Training	No	964	0	100.0%
	Yes	166	0	0.0%
	Overall Percentage	100.0%	0.0%	85.3%
Test	No	940	0	100.0%
	Yes	166	0	0.0%
	Overall Percentage	100.0%	0.0%	85.0%

Accepted the offer in the last campaign * misclass_0.35 Crosstabulation[a]

			misclass_0.35 no	misclass_0.35 yes	Total
Accepted the offer in the last campaign	No	Count	857	107	964
		% within Accepted the offer in the last campaign	88.9%	11.1%	100.0%
	Yes	Count	90	76	166
		% within Accepted the offer in the last campaign	54.2%	45.8%	100.0%
Total		Count	947	183	1130
		% within Accepted the offer in the last campaign	83.8%	16.2%	100.0%

a. S...

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
ASE_1	1130	.00	.96	.1053	.20138
Valid N (listwise)	1130				

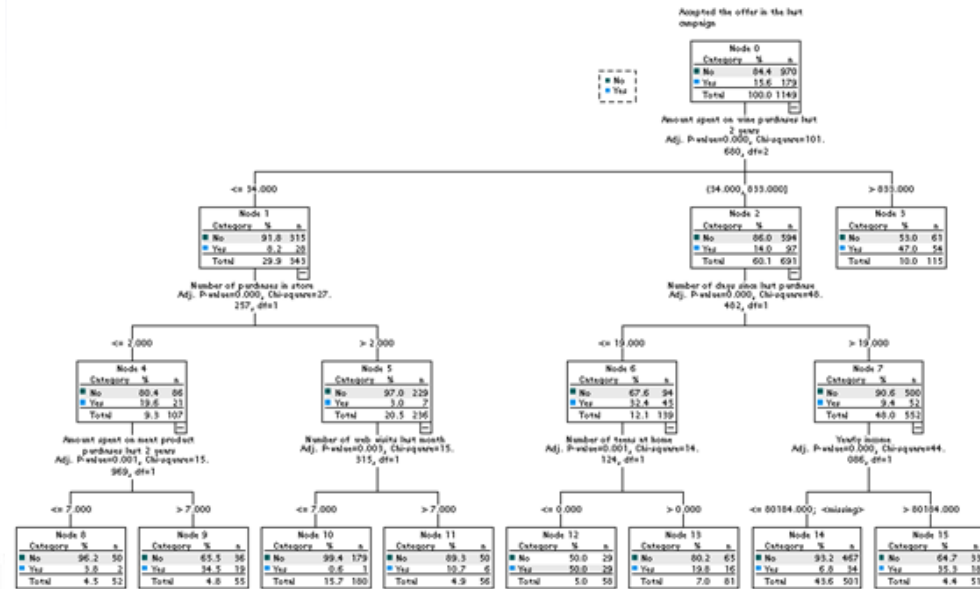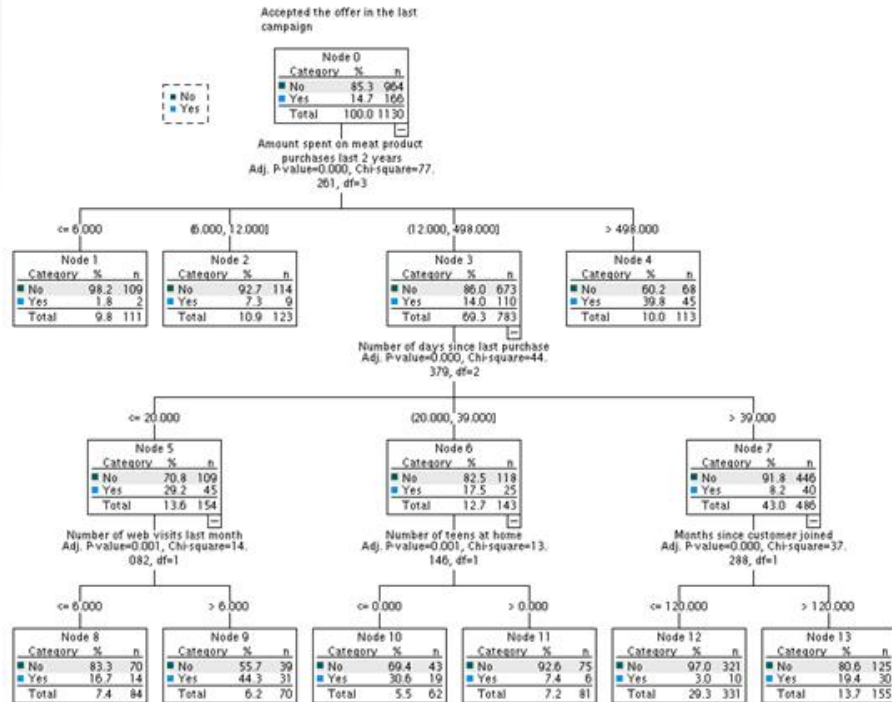a. Sample Assignment = Training Sample

Exhaustive CHAID

➢ This is just model summary for the exhaustive CHAID model

➢ The difference between CHAID and exhaustive CHAID is exhaustive CHAID takes merging one step further and keeps merging until there are only two categories within each predictor variable. So, after the first level each parent node only results in two child nodes

➢ A few more nodes than CHAID but the same amount of terminal nodes
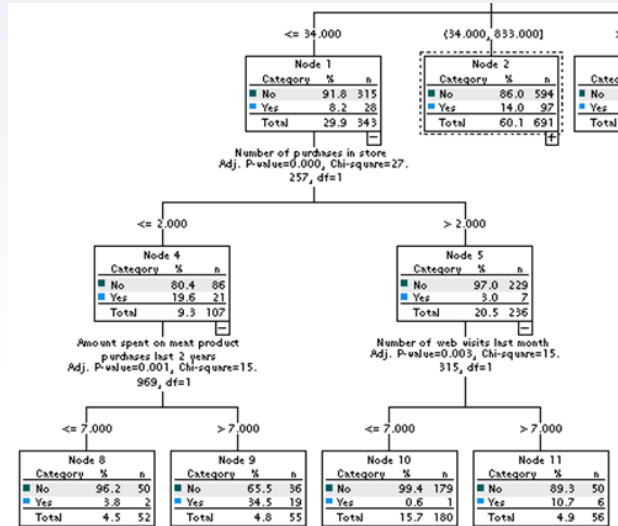
Model Summary

Specifications	Growing Method	EXHAUSTIVE CHAID
	Dependent Variable	Accepted the offer in the last campaign
	Independent Variables	Year of birth, Education level, Marital status, Yearly income, Number of kids at home, Number of teens at home, Months since customer joined, Number of days since last purchase, Amount spent on wine purchases last 2 years, Amount spent on fruit purchases last 2 years, Amount spent on meat product purchases last 2 years, Amount spent on fish product purchases last 2 years, Amount spent on sweet product purchases last 2 years, Amount spent on gold product purchases last 2 years, Number of purchases made with discount, Number of purchases using web, Number of purchases using catalog (Mail), Number of purchases in store, Number of web visits last month
	Validation	Split Sample
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	Amount spent on wine purchases last 2 years, Number of purchases in store, Amount spent on meat product purchases last 2 years, Number of web visits last month, Number of days since last purchase, Number of teens at home, Yearly income
	Number of Nodes	16
	Number of Terminal Nodes	9
	Depth	3

Comparing CHAID to Exhaustive CHAID:

➤ My CHAID (left) and exhaustive CHAID (right)

Levels 2 and 3:

> This is the center of the exhaustive CHAID tree, the right node on the first level is a terminal node and was pretty good at predicting customers who would accept the offer.

> One of the nodes worth note on this side is node 12 which was predicted a 50/50 split, half of those people were predicted to accept the sales offer from the campaign.
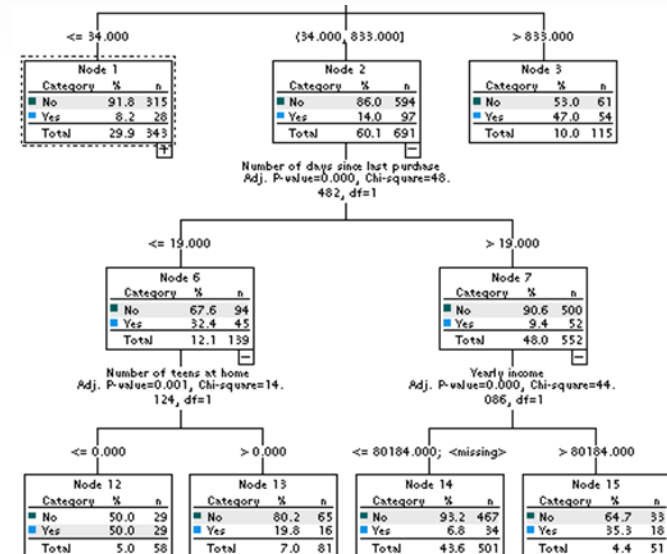

> The left side of the tree, and as characteristic of exhaustive CHAID there are only two resulting child nodes after each split

> 3 of the 4 terminal nodes on this side are really good at predicting customers who wouldn't accept the offer. The second level also has nodes fairly good at predicting not accepting the offer

Risk, Misclassification, ASE:

➤ Risk was 15.6% for the training data and 15.7% for the testing set. Again, very very close so no indication of overfitting.

➤ With the threshold lowered to 0.35 the specificity is 85.5% and the sensitivity was 49% which is a little better than the CHAID model.

➤ The average squared error was 0.1095, which is close the CHAID so not much better or worse in that perspective.

Risk

Sample	Estimate	Std. Error
Training	.156	.011
Test	.157	.011

Growing Method: EXHAUSTIVE

Classification

Sample	Observed	Predicted No	Predicted Yes	Percent Correct
Training	No	941	29	97.0%
	Yes	150	29	16.2%
	Overall Percentage	95.0%	5.0%	84.4%
Test	No	894	40	95.7%
	Yes	131	22	14.4%
	Overall Percentage	94.3%	5.7%	84.3%

Growing Method: EXHAUSTIVE CHAID

Accepted the offer in the last campaign * misclass2_0.35 Crosstabulation[a]

			misclass2_0.35 no	misclass2_0.35 yes	Total
Accepted the offer in the last campaign	No	Count	824	140	964
		% within Accepted the offer in the last campaign	85.5%	14.5%	100.0%
	Yes	Count	84	82	166
		% within Accepted the offer in the last campaign	50.6%	49.4%	100.0%
Total		Count	908	222	1130
		% within Accepted the offer in the last campaign	80.4%	19.6%	100.0%

a. Sa

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
ASE_2	1130	.00	.99	.1095	.20934
Valid N (listwise)	1130				

a. Sample Assignment = Training Sample

Logistic Regression

Correlations

	Amount spent on wine purchases last 2 years	Amount spent on fruit purchases last 2 years	Amount spent on meat product purchases last 2 years	Amount spent on fish product purchases last 2 years	Amount spent on sweet product purchases last 2 years	Amount spent on gold product purchases last 2 years		Yearly income	Number of kids at home
Year of birth									
Yearly income									
Number of kids at home								--	
Number of teens at home								-.429**	--
Months since customer joined								.020	-.036
Number of days since last purchase								-.018	-.056**
Amount spent on wine purchases last 2 years	--							-.004	.008
Amount spent on fruit purchases last 2 years	.390**	--						.579**	-.496**
Amount spent on meat product purchases last 2 years	.563**	.543**	--					.430**	-.373**
Amount spent on fish product purchases last 2 years	.401**	.594**	.569**	--				.585**	-.438**
Amount spent on sweet product purchases last 2 years	.387**	.567**	.525**	.582**	--			.439**	-.388**
Amount spent on gold product purchases last 2 years	.389**	.391**	.351**	.418**	.371**	--		.441**	-.371**
Number of purchases made with discount	.011	-.131**	-.122**	-.139**	-.119**	.050*		.325	-.349**
Number of purchases using web	.543**	.298**	.295**	.296**	.349**	.424**		-.083	.223**
Number of purchases using catalog (Mail)	.636**	.487**	.725**	.533**	.491**	.435**		.388	-.361**
Number of purchases in store	.642**	.462**	.480**	.461**	.449**	.383**		.589	-.503**
Number of web visits last month	-.321**	-.417**	-.539**	-.444**		-.248**		.529	-.500**
Accepted the offer in the last campaign	.247**	.124**	.238**	.110**	.11	.137**		.122**	-.553**

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

				Yearly income	Number of kids at home
Number of days since last purchase	-.020	-.004	.008		
Amount spent on wine purchases last 2 years	-.159**	.579**	-.496**		
Amount spent on fruit purchases last 2 years	-.019	.430**	-.373**		
Amount spent on meat product purchases last 2 years	-.030	.585**	-.438**		
Amount spent on fish product purchases last 2 years	-.043*	.439**	-.388**		
Amount spent on sweet product purchases last 2 years	-.019	.441**	-.371**		
Amount spent on gold product purchases last 2 years	-.065*	.325**	-.349**		
Number of purchases made with discount	-.061**	-.083*	.223**		
Number of purchases using web	-.146**	.388**	-.361**		
Number of purchases using catalog (Mail)	-.123**	.589**	-.503**		
Number of purchases in store	-.129**	.529**	-.500**		
Number of web visits last month	.122**	-.553**	.449**		
Accepted the offer in the last campaign	.019	.133**	-.079**		

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

			Number of purchases using catalog (Mail)	Number of purchases in store	Number of web visits last month	Accepted the offer in the last campaign
Number of purchases using catalog (Mail)			--			
Number of purchases in store			.519**	--		
Number of web visits last month			-.519**	-.429**	--	
Accepted the offer in the last campaign			.220**	.039	-.003	--

Marital status * Accepted the offer in the last campaign Crosstabulation

			Accepted the offer in the last campaign		Total
			No	Yes	
Marital status	Single	Count	618	174	792
		% within Marital status	78.0%	22.0%	100.0%
	With someone	Count	1286	158	1444
		% within Marital status	89.1%	10.9%	100.0%
Total		Count	1904	332	2236
		% within Marital status	85.2%	14.8%	100.0%

		Accepted the offer in the last campaign		Total
		No	Yes	
	Count	181	22	203
	% within Education level	89.2%	10.8%	100.0%
	Count	52	2	54
	% within Education level	96.3%	3.7%	100.0%
	Count	975	151	1126
	% within Education level	86.6%	13.4%	100.0%
	Count	312	57	369

			Accepted the offer in the campaign	
			No	Yes
Marital status	Divorced	Count	184	
		% within Marital status	79.3%	
	Married	Count	766	
		% within Marital status	88.7%	
	Single	Count	376	
		% within Marital status	77.8%	
	Together	Count	520	
		% within Marital status	89.7%	
	Widow	Count	58	
		% within Marital status	75.3%	
Total		Count	1904	
		% within Marital status	85.2%	

Education level * Accepted the offer in the last campaign Crosstabulation

			Accepted the offer in the last campaign		Total
			No	Yes	
Education level	High School	Count	52	2	54
		% within Education level	96.3%	3.7%	100.0%
	BA/AS	Count	1156	173	1329
		% within Education level	87.0%	13.0%	100.0%
	Post-grad	Count	696	157	853
		% within Education level	81.6%	18.4%	100.0%
Total		Count	1904	332	2236
		% within Education level	85.2%	14.8%	100.0%

Model results

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical Variables Codings

		Frequency	Parameter coding (1)	Parameter coding (2)
Education level	BA/AS	666	1.000	.000
	High Sch	30	.000	1.000
	Post-Grad	425	.000	.000
Marital status	Single	397	1.000	
	With Someone	724	.000	

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	875.699[a]	.044	.078
2	822.585[b]	.088	.157
3	794.242[a]	.111	.197
4	767.728[c]	.132	.234
5	750.762[c]	.145	.258
6	734.103[c]	.157	.280
7	713.197[c]	.173	.308
8	704.838[c]	.179	.319
9	698.926[c]	.183	.326
10	689.851[c]	.190	.338

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

c. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Model results

Classification Table[a]

| | Observed | | Predicted | | |
| | | | Accepted the offer in the last campaign | | Percentage Correct |
			No	Yes	
Step 1	Accepted the offer in the last campaign	No	938	21	97.8
		Yes	145	17	10.5
	Overall Percentage				85.2
Step 2	Accepted the offer in the last campaign	No	912	47	95.1
		Yes	135	27	16.7
	Overall Percentage				83.8
Step 3	Accepted the offer in the last campaign	No	912	47	95.1
		Yes	121	41	25.3
	Overall Percentage				85.0
Step 4	Accepted the offer in the last campaign	No	906	53	94.5
		Yes	113	49	30.2
	Overall Percentage				85.2
Step 5	Accepted the offer in the last campaign	No	901	58	94.0
		Yes		49	30.2
	Overall Percentage				84.7
				53	94.5
				56	34.6
					85.8
				57	94.1
				61	37.7
					85.9
Step 8	Accepted the offer in the last campaign	No	900	59	93.8
		Yes	96	66	40.7
	Overall Percentage				86.2
Step 9	Accepted the offer in the last campaign	No	896	63	93.4
		Yes	95	67	41.4
	Overall Percentage				85.9
Step 10	Accepted the offer in the last campaign	No	893	66	93.1
		Yes	94	68	42.0
	Overall Percentage				85.7

a. The cut value is .350

Step 10	Accepted the offer in the last campaign	No	893	66	93.1
		Yes	94	68	42.0
	Overall Percentage				85.7

a. The cut value is .350

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 10	Education level			7.193	2	.027			
	Education level(1)	-.432	.211	4.199	1	.040	.649	.429	.981
	Education level(2)	-2.113	1.061	3.966	1	.046	.121	.015	.967
	Marital status(1)	1.092	.200	29.729	1	<.001	2.981	2.013	4.415
	Number of teens at home	-1.208	.215	31.497	1	<.001	.299	.196	.456
	Months since customer joined	.050	.014	13.959	1	<.001	1.052	1.024	1.080
	Number of days since last purchase	-.030	.004	60.350	1	<.001	.970	.963	.978
	Amount spent on wine purchases last 2 years	.001	.000	14.996	1	<.001	1.001	1.001	1.002
	Amount spent on gold product purchases last 2 years	.005	.002	8.187	1	.004	1.005	1.002	1.009
	Number of purchases using web	.179	.046	15.097	1	<.001	1.196	1.093	1.308
	Number of purchases using catalog (Mail)	.071	.031	5.168	1	.023	1.073	1.010	1.141
	Number of purchases in store	-.226	.042	29.536	1	<.001	.798	.735	.866
	Constant	-6.573	1.576	17.399	1	<.001	.001		

▸ $logit(\hat{\pi}) = -6.573 - 0.432\,(BA/AS\ \ indicator) - 2.113\,(High\ School\ indicator) + 1.092\,(single\ indicator) - 1.208\,(\#\ teens) + 0.050\,(Months\ since\ customer\ joined) - 0.030\,(Number\ of\ days\ since\ last\ purchase) + 0.001\,(\$Wine) + 0.005\,(\$Gold) + 0.179\,(\#\ of\ web\ purchases) + 0.071\,(\#\ of\ mail\ purchases) - 0.226\,(\#\ of\ store\ purchases)$

Comparing training to testing

▶ Training:

    ▷ ASE

Descriptive Statistics

	N	Mean
Square_err	1130	.1425
Valid N (listwise)	1130	

    ▷ Sens. & spec.

Classification Table[a]

			Predicted		
			Accepted the offer in the last campaign		Percentage Correct
	Observed		No	Yes	
Step 10	Accepted the offer in the last campaign	No	893	66	93.1
		Yes	94	68	42.0
	Overall Percentage				85.7

a. The cut value is .350

▶ Testing:

    ▷ ASE

Descriptive Statistics

	N	Mean
Square_err	1106	.1456
Valid N (listwise)	1106	

    ▷ Sens. & spec.

Accepted the offer in the last campaign * Predicted_value Crosstabulation

			Predicted_value		Total
			No	Yes	
Accepted the offer in the last campaign	No	Count	870	66	936
			92.9%	7.1%	100.0%
	Yes	Count	95	75	170
			55.9%	44.1%	100.0%

Comparing Models

- **CHAID**
  - Sensitivity: 45.8%
  - Specificity: 88.9%
  - ASE: 0.1053
  - Risk:14.7%

- **Exhaustive CHAID**
  - Sensitivity: 49.4%
  - Specificity: 85.5%
  - ASE: 0.1095
  - Risk:15.6%

- **CART**
  - Sensitivity: 63.3%
  - Specificity: 84.9%
  - ASE: 0.0869
  - Risk:14.8%

- **Logistic regression**
  - Sensitivity:44.1%
  - Specificity:92.9%
  - ASE:0.1456
  - Risk: 14.6%

CART was determined to be our best predictive model