Who Are My Best Customers?

Using SPSS to get greater value from your customer database

Table of contents

Introduction

Who are my best customers? If you manage sales, marketing, or customer service, you want an answer to this question. In fact, you want to know more about all your customers—from the best to the worst. That's because planning and implementing successful, cost-effective strategies for every customer segment is critical to increased business profits. This paper will focus on how a company might identify its best customers, but the same process could be used for other customer segments.

Knowledge about your best customers—their attitudes, purchase patterns, and demographic profiles—is the key to developing and implementing successful marketing and customer relationship management programs. Such knowledge helps you effectively target your promotional, advertising, and marketing campaigns, as well as develop up-sell and cross-sell programs and more long-term customer loyalty, retention, and rewards programs.

Coordinating these efforts is particularly important today, as marketing moves away from mass marketing and toward greater customization, with an emphasis on positioning particular products or services for specific types of customers. Reliable, detailed information about customer behavior, attitudes, and other characteristics offers a real competitive advantage and helps improve the return on investment for all your customer interactions. The insight gained from even the most elementary analysis of customer characteristics can have profound implications for your business.

This white paper demonstrates how you could analyze a customer database using SPSS. This integrated suite of products for statistical analysis and data management supports you throughout the data analysis process, whether you perform your analysis from a single desktop computer or across an extended network.

In this paper, the marketing database of 2,000 customers includes the following information:

- Date when customer first became a customer
- Purchase history by dollar value of orders
- Response to different offers
- Household income level
- Geographic classification
- Gender and other demographic variables

Our goal is to identify unique customer segments using the classic measures of recency, frequency, and monetary value (RFM) in order to identify a company's best customers. While this paper doesn't describe the process for RFM scoring, it does provide all the preliminary analysis needed to do so. We use various data analysis techniques to extract this information and suggest how it can be used to guide business decisions.

Initial exploration of customer data

We begin by exploring the different variables in our database to answer questions such as:

- Where do our customers live?
- What is an average customer's household income?
- How long have our customers been customers?
- How much money do our customers spend with us?

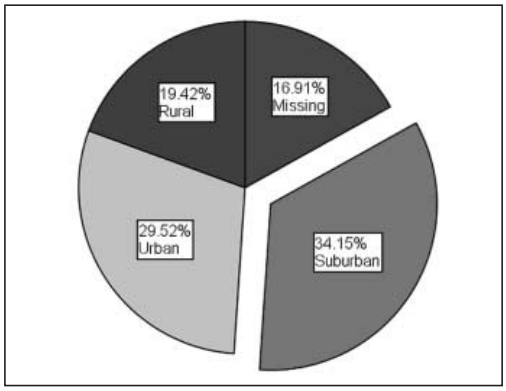
SPSS offers several methods to quickly obtain the answers to these questions. SPSS Frequencies and Descriptives procedures are very good at providing a first look at our data, and the results often suggest other kinds of analyses we might perform.

Where do our customers live?

Analyzing whether customers are urban, suburban, or rural can help us determine an optimal marketing mix. SPSS Frequencies provides a table of counts and percents by category along with a visual representation of the data in a bar, histogram, or pie chart. SPSS presents the results as a table and chart, complete with explanatory labels.

From the results shown in Table 1, we learn that the largest proportion of our customer base (34 percent) lives in a suburban area, and the smallest proportion (19 percent) lives in a rural area. We also see that 16.9 percent have no area listed.

SPSS automatically flags missing data for special treatment. It is useful to know when and why information is missing. For example, you might want to distinguish between data missing because they don't apply and data missing because they are unavailable. In Table 1, the "percent" column includes the missing data, but the "valid percent" column excludes it from the calculations. This provides a fast side-by-side comparison of how the missing data affect the results.



Number of Customers by Area

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Suburban	707	34.2	41.1	41.1
	Urban	611	29.5	35.5	76.6
	Rural	402	19.4	23.4	100.0
	Total	1720	83.1	100.0	
Missing	System	350	16.9		
Total		2070	100.0		

Chart 1 and Table 1. The SPSS table and chart, automatically created with complete labels, reveal that most customers (34 percent) live in the Suburban area.

What is an average customer's household income?

There are several ways to gain a more detailed view of our customers. To obtain information about household income, for example, we examine basic summary statistics, such as the mean, minimum, and maximum values. Interval or continuous variables, such as income measured in dollars, are best first examined with descriptive statistics. The SPSS Descriptives procedure gives us a set of summary statistics. We learn from Table 2 that the average annual household income of the 2,000 customers in our database is approximately $61,000, and that the majority of customers have incomes between $50,000 and $72,000.

Descriptive Statistics of HH Income

	N	Minimum	Maximum	Mean	Std. Deviation
HH Income	1984	$38,552.00	$85,571.00	$61,386.39	$11,006.75
Valid N (listwise)	1984				

Table 2. The SPSS Descriptives procedure provides a quick summary showing that average household income is approximately $61,000.

Number of years as customer

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	169	8.2	8.5	8.5
	2	172	8.3	8.7	17.2
	3	147	7.1	7.4	24.6
	4	144	7.0	7.3	31.9
	5	132	6.4	6.7	38.5
	6	121	5.8	6.1	44.6
	7	145	7.0	7.3	51.9
	8	135	6.5	6.8	58.7
	9	112	5.4	5.6	64.4
	10	136	6.6	6.9	71.2
	11	117	5.7	5.9	77.1
	12	128	6.2	6.5	83.6
	13	105	5.1	5.3	88.9
	14	110	5.3	5.5	94.4
	15	111	5.4	5.6	100.0
	Total	1984	95.8	100.0	
Missing	System	86	4.2		
Total		2070	100.0		

Table 3. An SPSS Frequencies chart indicates that almost 52 percent of our customers have been customers for seven years or more.

How long have our customers been customers?

To determine how long our customers remain with us, we must manipulate a field in our database and then count the number of customers in each period. Since the database contains the date we entered the customer into the database, we first compute a new variable: length of time as a customer. By using one of the many time functions available in SPSS, we can easily transform the date into the length of time, in years, since we acquired the customer. After computing this new variable, we can request a frequency chart (Table 3) of the length of time a customer has been a customer. From Table 3, we learn that about 29 percent of our customers have been in the database for more than 10 years, and that just over half have been with us for seven years.

How much money do our customers spend with us?

Next, we can determine who are our best customers. "Best customers" are typically defined as the most profitable customers or the ones that spend the most money with our organization. To obtain the most accurate picture of customer lifetime value, we create a predictive model that uses data describing previous purchases and behavior to forecast future purchases. In this simplified example, we begin with the total value of the orders placed by each customer.

First, we create a new variable, total order value (in dollars), by summing the value of each order (Value1, Value2 and so on) in our database. Since total value is a continuous variable, a histogram is the most efficient way to graphically display the results. In a histogram, each bar represents a range of data.

From the histogram in Chart 2, we learn that the majority of customers spent $500 or less and that at higher dollar-value levels the number of customers making purchases steadily declines. The average amount spent by customers is $1,360. A very small number of customers spent in excess of $7,000.

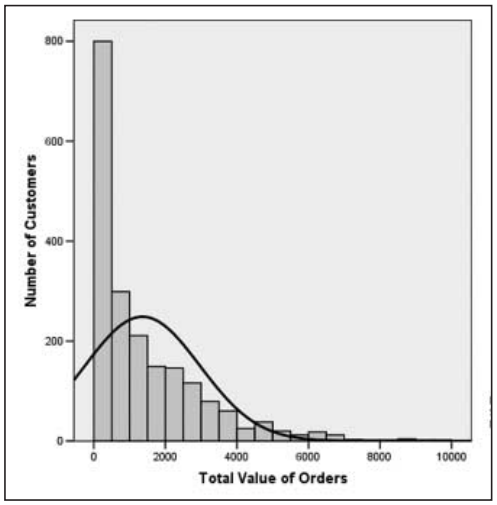


Chart 2. From the histogram, we see the majority of customers spent $500 or less and that at higher dollar-value levels the number of customers making purchases steadily declines.

So far, we know that a typical customer:

- Lives in a suburban area
- Has a household income of $61,000
- Has been a customer for seven years
- Spends $1,360 on our products and services

Further explorations of customer data

How do customers respond to different promotional offers?

Analyzing the results of specific marketing promotions is another important step toward understanding customers. Evaluating past efforts helps identify what worked and what did not, so you can duplicate your successes and learn from your failures.

Here, we want to answer two questions:

- How many people responded to each of our four offers?
- What is the average amount spent in response to our different promotions?

To do so, we run SPSS Frequencies on each offer response and SPSS Descriptives on the order value for the four offers.

Offer 1 response

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not respond	1110	53.6	55.5	55.5
	Responded	890	43.0	44.5	100.0
	Total	2000	96.6	100.0	
Missing	System	70	3.4		
Total		2070	100.0		

Table 4. Almost 45 percent, or 890 of the people in the customer database, responded to Offer 1.

In Table 4, we see that 890 customers, or almost 45 percent of the customer database, responded to Offer 1. Similar analysis for the other offers, shown in table 5, shows a 39 percent response to Offer 2, a 37.4 percent response to Offer 3, and a 17.4 percent response to Offer 4.

Descriptive Statistics of Offers

	N	Minimum	Maximum	Mean	Std. Deviation
Order value (offer 1)	2000	$.00	$3,280.23	$376.64	$602.32
Order value (offer 2)	2000	$.00	$3,477.31	$312.36	$569.28
Order value (offer 3)	2000	$.00	$3,377.18	$293.98	$536.49
Order value (offer 4)	2000	$.00	$3,339.58	$375.63	$595.73
Valid N (listwise)	2000				

Table 5. The analysis of purchase history reveals that the average value for Offer 3, $294, is lower than the average value for the other offers.

This raises a new question: Were there unique characteristics in Offer 1 that made it more effective? In other words, is this result significant? The information on purchase history (Table 5) reveals that the average value for Offer 3, $294, is lower than the other offers. Whether this difference is significant can be quickly and easily determined, by using SPSS.

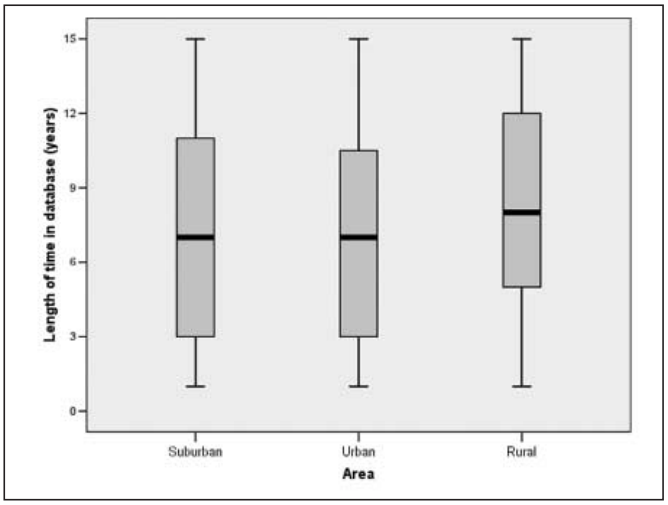


Chart 3. The boxplot displays both the mean and distribution of the data. It is easy to see that customers in rural areas have been customers longer, on average, than those in other areas.

Does customer retention vary by area?

To explore our first question, we generate a powerful statistical chart, the boxplot. This displays both the mean and the distribution of the data. From the boxplot in Chart 3, we learn that customers in rural areas have been customers longer, on average, than those in other areas.

Comparison of Means provides summary statistics for a joint distribution. Table 6 shows the same information as the boxplot in table format. It reveals that while the overall average length of time in the database is 7.49 years, customers in rural areas have remained customers longer, on average, than those in suburban or urban areas.

Is this a significant finding? Statistical significance tells us if the differences we see are random or if they are sufficiently large to justify further consideration. If the differences are statistically significant, this suggests the potential influence of some non-random factor. When statistical significance exists, it is a strong indication for further exploration.

The ANOVA report in Table 7, created in SPSS with just a few mouseclicks, shows that the differences between area and length of time as customers are statistically significant. Convention holds that the Pearson Chi-square statistic should be less than .05 for the exhibited differences to be statistically significant (at the 95 percent confidence level). Since the significance for these results is .000, or less than .05, we can conclude the differences in means are likely significant: The overall distribution of average customer retention and area is probably not due to random causes, but to something else.

Report

Length of time in database (years)

Area	Mean	N	Std. Deviation
Suburban	7.38	703	4.398
Urban	6.96	607	4.266
Rural	8.48	398	4.306
Total	7.49	1708	4.366

Table 6. This Comparison of Means report shows that while the overall average length of time in the database is 7.49 years, customers in rural areas have remained customers longer, on average, than those in suburban or urban areas.

ANOVA

Length of time in database (years)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	569.499	2	284.749	15.188	.000
Within Groups	31965.164	1705	18.748		
Total	32534.663	1707			

Table 7. The ANOVA report shows that the differences we see are statistically significant, a strong indication for further exploration.

Examples of possible causes include:

- Our first office was opened in a rural area
- There is more need for the product in one area than in another
- A certain product feature was introduced successfully in one area

Other causes may exist and bear investigation. This is why it is also important to know your business, in order to marshal data to support your intuition.

Did customer response to Offer 1 vary by area?

Next, we continue our analysis of offer response. SPSS provides an easy way to graphically present information on all four offers, using a clustered bar chart. Chart 4 provides a summary of response patterns by area. We see that customers in urban areas tend to under-order relative to the other two, particularly the rural. This is a finding we could not have guessed by looking at the frequency distribution of area, which showed us that the rural areas contained the fewer people.

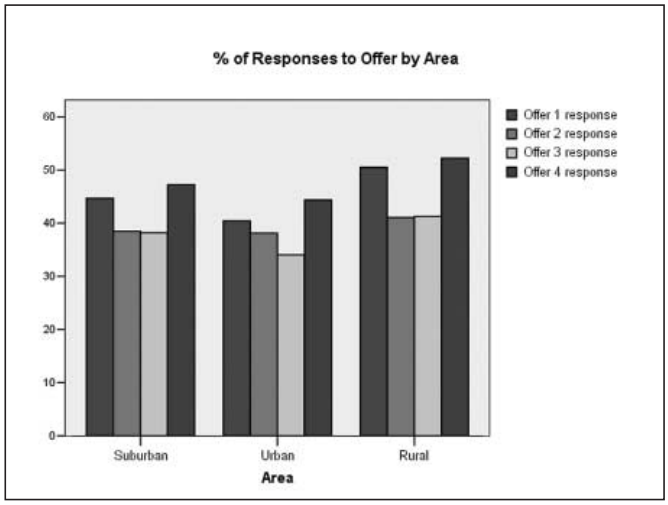


Chart 4. The SPSS clustered bar chart provides a quick and clear way to present response patterns by area.

To find out if this is significant, we can further explore the results of individual offers by area. To answer the question "How did people in each area respond to Offer 1?" we perform an SPSS crosstab on Offer 1 and area. Table 8 shows 41.3 percent of the people who responded to Offer 1 were from suburban areas. While only 26.5 percent of the people who responded to Offer 1 were from rural areas, over half (50.5 percent) of the rural customers responded to the offer.

To understand if area determines the likelihood of response to Offer 1, we compare the percentages in the "% of area" rows and find that 45 percent of people from suburban areas responded to this offer, and that 40 percent of people in urban areas responded. Based on this information, we conclude rural areas are good areas for an offer such as Offer 1. However, while it appears the percentages are different, that is insufficient reason to start duplicating Offer 1 in rural areas. First, we must determine if these percentages are statistically significant. Here, the Chi-square statistic indicates if statistical significance exists.

Area * Offer 1 response Crosstabulation

			Offer 1 response		
			Did not respond	Responded	Total
Area	Suburban	Count	391	316	707
		% within Area	55.3%	44.7%	100.0%
		% within Offer 1 response	41.0%	41.3%	41.1%
	Urban	Count	364	247	611
		% within Area	59.6%	40.4%	100.0%
		% within Offer 1 response	38.2%	32.2%	35.5%
	Rural	Count	199	203	402
		% within Area	49.5%	50.5%	100.0%
		% within Offer 1 response	20.9%	26.5%	23.4%
Total		Count	954	766	1720
		% within Area	55.5%	44.5%	100.0%
		% within Offer 1 response	100.0%	100.0%	100.0%

Table 8. While only 26.5 percent of the people who responded to Offer 1 were from rural areas, over half (50.5 percent) of rural customers responded to the offer.

Table 9 contains Chi-square information for the area and Offer 1. In this case, the Chi-square is .007 and, therefore, is significant. There could be a specific, identifiable reason that made Offer 1 more successful in rural areas. Perhaps the copy spoke more directly to their needs, or the media type was better matched to attract and keep their attention.

By identifying what made the campaign successful in rural areas, we can leverage that knowledge in future offers to this area. We also may choose to explore other relationships that underlie area.

How much have customers spent?

Another way to look at purchase history is to assess total amount spent, rather than just the money spent on individual orders. Perhaps a relationship between total money spent and area will reveal some insights.

A one-way ANOVA provides specific information about the significance of the differences in average values that you may see.

The first thing that one-way ANOVA provides is a table of Descriptive Statistics. Table 10 shows that the average total amount spent in response to each of the four offers by area varies widely. In urban areas, the average amount spent was $1,206; in suburban areas, $1,391 was the average spent; while in rural areas, the average spent was over $1,600.

The report also shows that the average difference exhibited between the spending levels in the suburban and the rural areas is not statistically significant. On the other hand, it shows that the difference between the rural and urban areas is significant.

You can use this information to further explore how and why these areas differ and develop targeted marketing plans to leverage the differences. For example, a different marketing and sales mix, different offer, or special bundle of products and services may work better in the urban areas. The marketing programs in rural areas should be repeated there for continued success.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	9.970[a]	2	.007
Likelihood Ratio	9.964	2	.007
Linear-by-Linear Association	1.999	1	.157
N of Valid Cases	1720		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 179.03.

Table 9. A Chi-square of .007 for the area and Offer 1 indicates that the differences between areas are significant.

Descriptives

Total Orders

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Suburban	707	$1,391.70	$1,640.44	$61.70	$1,270.57	$1,512.82	$.00	$9517.75
Urban	611	$1,206.01	$1,442.51	$58.36	$1,091.40	$1,320.61	$.00	$8402.60
Rural	402	$1,618.27	$1,772.04	$88.38	$1,444.52	$1,792.02	$.00	$8996.70
Total	1720	$1,378.69	$1,612.52	$38.88	$1,302.43	$1,454.95	$.00	$9517.75

Table 10. The average amount spent by customers in response to our four offers was $1,600, but this varies by region.

How much will customers spend?

Predictive models are powerful tools to help target prospects and optimize marketing resources. They help answer questions such as “How much will customers spend, given their income level?”

In many statistical studies, the goal is to establish a relationship, expressed as an equation, for predicting typical values of one variable given the value of another. SPSS offers several procedures for establishing relationships and defining predictive models. These procedures include scatterplots and correlations, linear and logistic regression analysis, and classification trees. With the step-by-step instructions and help features built into the SPSS product family, you can perform these procedures successfully, even if you aren’t a statistician.

Chart 5 shows the shape of the relationship between these two variables. The scatterplot is the correct chart to display the joint distribution of two continuous or interval variables. The correlation coefficient of 60.8 percent, displayed in Table 11, indicates a strong relationship between household income and total money spent. Regression analysis further defines the relationship with a model, as shown in Table 12 and Chart 6. This relationship shows that as household income increases, the total money spent on our products increases. We could use this finding to improve sales forecasts and the effectiveness of our marketing efforts.

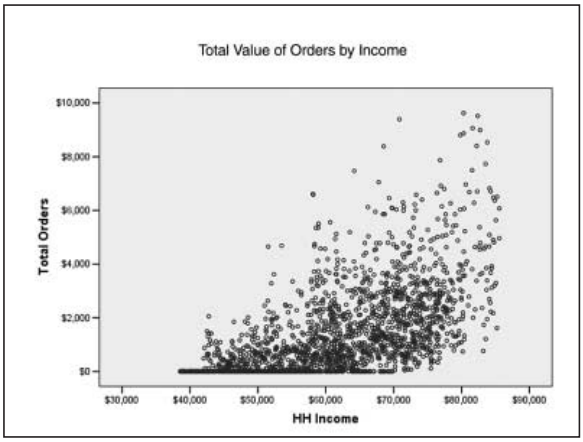


Chart 5. The scatterplot shows the shape of the relationship between these two variables.

Which groups are most likely to respond to an offer?

With the SPSS Classification Trees™ module, we can identify unique segments within our database. SPSS Classification Trees uses one of four established tree-growing algorithms to accomplish this. These algorithms include CHAID, exhaustive CHAID, classification and regression trees (C&RT), and QUEST. We can then use this analysis to more accurately forecast sales and improve our marketing efforts.

The analysis begins by our placing information about area, product class category, and categorized income into a model to find out which customers with certain specific characteristics are most likely to respond to Offer 1. SPSS Classification Trees automatically builds a tree diagram of the results, as in Chart 7.

Correlations

		HH Income	Total Orders
HH Income	Pearson Correlation	1	.608**
	Sig. (2-tailed)		.000
	N	1984	1984
Total Orders	Pearson Correlation	.608**	1
	Sig. (2-tailed)	.000	
	N	1984	2000

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

Table 11. The correlation coefficient shows a strong relationship of 60.8 percent, revealing that as household income increases, the total amount spent on our products increases.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-4075.349	162.252		-25.117	.000
	HH Income	.089	.003	.608	34.093	.000

a. Dependent Variable: Total Orders

Table 12. A linear regression defines the relationship between household income and the amount customers spend.

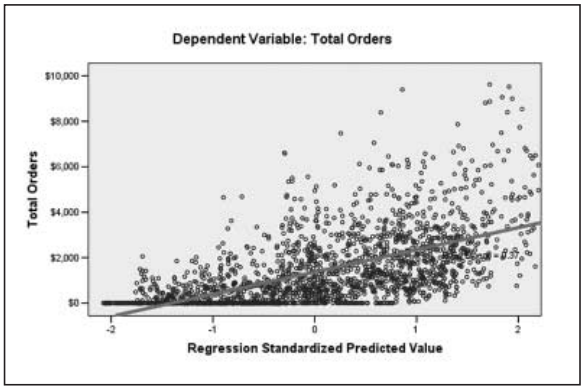


Chart 6. This chart shows that the more money customers earn, the more they spend on our products.

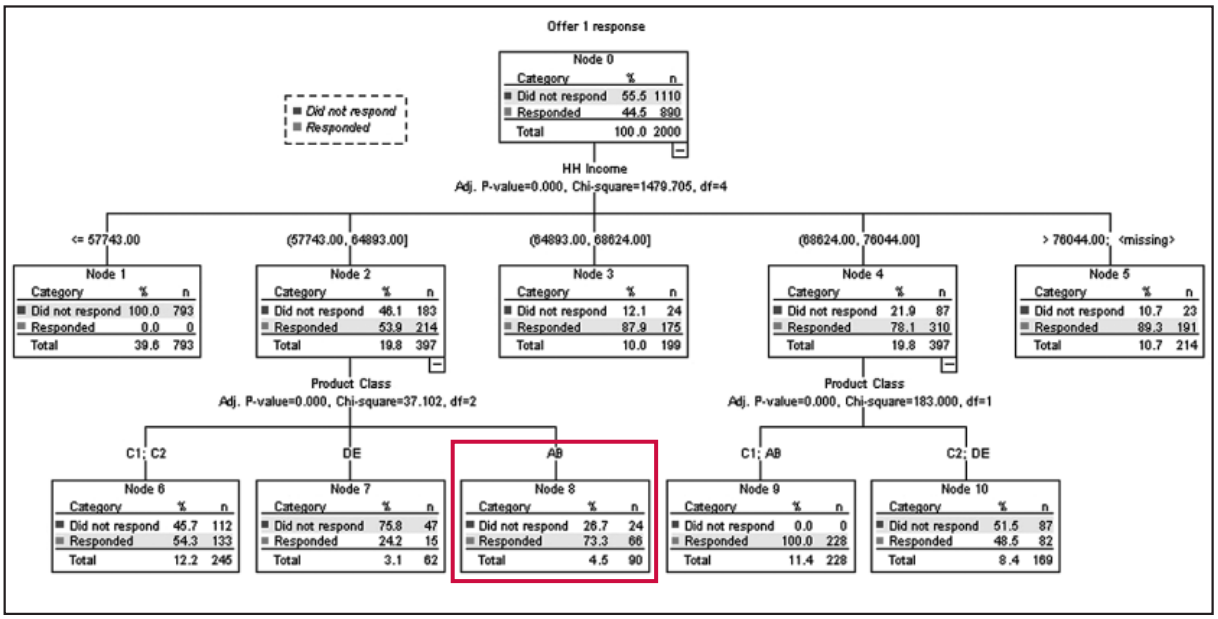


Chart 7. SPSS Classification Trees presents a model showing that customers with certain combinations of characteristics are most likely to respond to Offer 1.

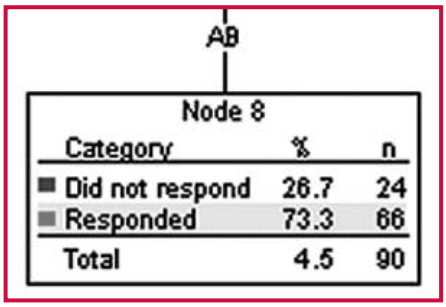


Chart 8. A detailed view of one node of the classification tree shows that customers with income in category 3 who purchase products in category 1 are 73 percent more likely to respond to a particular promotion.

Income is found to be the highest predictor, which corresponds to the earlier regression findings. In this case, SPSS Classification Trees goes beyond the regression example to explore further interactions.

In Chart 8, the details of the next level of branches reveal that if income is in category 3 ($57,750 to $65,000) and if product class is category 1, there is a 73 percent response rate. SPSS Classification Trees identifies the unique segments within the database, so you can leverage the combinations of characteristics to get the best results from your marketing programs.

We find from this analysis that households with income of $57,750 to $65,000 who purchased from product class category 1 are more likely to respond to Offer 1. Analysis using additional variables may lead to additional findings.

Taking action

Through the analyses described here, SPSS enabled us to quickly analyze our data in order to learn some important things about our typical customers: They tend to be longer-term customers, from suburban areas, are likely to have higher-than-average incomes, and have not responded well, on the whole, to Offer 3.

In addition, using powerful SPSS predictive modeling and segmentation techniques to identify relationships, we developed a model that describes the relationship between income and total money spent to help predict future sales. We also identified unique customer segments by their likelihood to respond to Offer 1.

By comparing multiple characteristics and groups, SPSS helped us learn more about underlying patterns: Not only was Offer 3 the least lucrative for us, it was particularly unproductive in urban areas, areas that tended to respond less well to our offers than the other two areas did. The fact that customers in urban areas had the lowest average income helps explain their relatively low response to our offers. By identifying such groups of customers, we can better target marketing and customer retention programs.

For instance, because higher-income households show greater revenue potential, we might offer them additional products and services, or develop customer retention programs that help keep them as happy, long-term customers. Alternatively, we might find that while customers in urban areas did not in general respond well to our offers, women of a particular income level in that area did, suggesting that it might be appropriate to target them in a certain type of campaign.

As a result of the analyses we conducted, we might make the following plans:

- Build a new customer retention program for our best customers, those defined as higher-income, long-time customers in the suburban area who purchase from product class 1
- Develop and test a new bundle of products and services to better target the needs of the lower-income urban area customers and prospects
- Repeat sales development of the rural area in the urban and suburban areas to build long-time customers
- Duplicate Offer 1 to prospects in rural areas
- Match the funds of future marketing campaigns to the predicted segment profitability (based initially on household income)

This paper describes just a few of the ways that you can use analytics to better understand your customers. By seeing your customers from a number of different perspectives, you can plan more effective programs and systematically measure results. In this way, you'll build stronger relationships with the customers you value most and decrease the costs of serving less valuable customer segments.

Other products offered by SPSS enable you to anticipate change in your customers' preferences and behavior. Predictive analytic solutions help you be proactive in planning your business strategies and provide a strong competitive advantage in any industry.

For the purposes of this paper, however, we have shown that the SPSS product family provides a host of analytic options, available in a single, integrated product suite. Even if you're not a statistician, you can apply this information to market more effectively, retain your most valuable customers, and increase the profitability of your business.

About SPSS Inc.

SPSS Inc. [NASDAQ: SPSS] is the world's leading provider of predictive analytics software and services. The company's predictive analytics technology connects data to effective action by drawing reliable conclusions about current conditions and future events. More than 250,000 commercial, academic, and public sector organizations rely on SPSS technology to help increase revenue, reduce costs, improve processes, and detect and prevent fraud. Founded in 1968, SPSS is headquartered in Chicago, Illinois. To learn more, please visit **www.spss.com**. For SPSS office locations and telephone numbers, go to **www.spss.com/worldwide.**

To learn more, please visit www.spss.com. For SPSS office locations and telephone numbers, go to www.spss.com/worldwide.