Supply Velocity Client – Customer Profitability Prediction Analysis

Contents

Supply Velocity Client – Customer Profitability Prediction Analysis

Executive Summary

Supply Velocity Client had engaged our research team to help them identify factors that would enable them to predict or identify potential profitable customers. Our team was able to determine that five of the seven factors (independent variables) that were discussed in our initial meeting were good predictors of gross profit per customers. Key customer, customers that purchase on the web, customers that are located in coastal states, the number employees at a customer and the population of the customers' billing zip code were all shown, to improve the average profit per customer with all other factors held constant. Sunshine percent of the state and the median income of the zip code of the customer's billing address did not help in predicting gross profit per customer (Statistically not significant).

Problem Statement

The problem we were engaged to solve was to help predict which prospects are likely to be highly profitable customers for Supply Velocity Client. By understanding the attributes of a highly profitable customer, Supply Velocity Client can effectively target its marketing campaigns and resources. At Supply Velocity Client the marketing campaign is primarily conducted by mailing out printed catalogues to customers and to rented mailing lists. By targeting their marketing efforts, Supply Velocity Client may be able to reduce rental fees for mailing lists and reduce the numbers of catalogues mailed. In addition, Supply Velocity Client can use the predictions from this analysis to assign existing sales representatives to higher profit potential prospects. The goal was to build a model that would help Supply Velocity Client achieve greater expected gross profit (sales minus standard cost) and lower overall marketing expenses.

Performance & Predictor Variables

Our team met with the CEO and Director of Marketing to discuss the performance variable (also known as dependent variable) we want to predict and the variables that would help predict the performance variable (also known as independent variables). There were many viable candidates for performance variables; some of the variables we discussed include customer-loyalty measures, volume measures and profit measures. The team came to a mutual agreement along with Eric and Katherine that the best performance variable to help Supply Velocity Client improve profit was gross profit per customer. Gross profit was calculated as sales dollars minus standard cost per item sold.

Next we discussed the independent variables that could predict gross profit per customer. Some variables were suggested by our team, while other variables were suggested by the CEO and Director of Marketing. A key consideration for the selection of a variable was the availability of data, either within the systems at Landscapes Brands or in the public domain of the web. The independent variables selected are as follows:

1. **Key customer:** Key customer was a binary variable (1 = key customer, 0 = not). Key customers were determined exogenously by Supply Velocity Client. If a customer had $20,000 in sales in a single year or $5000 per year over three years they were designated a key customer.

2. **Whether the customer ordered product via Supply Velocity Client' web site or over the phone:** Web Y/N was a binary variable (1 = order via web site, 0 = order via phone).

3. **Whether the customer was located in a coastal state:** Supply Velocity Client manufactures outdoor furnishings that are highly corrosion resistant. Therefore, management wondered if customers in coastal states, where outdoor furnishings are more prone to corrosion were the more profitable customers.

4. **The sunshine percentage of the state the customer was located within:** Supply Velocity Client also a wondered the effect of sunshine, i.e. how sunny/rainy the location of the customer was on total profit. We used sunshine percent of the state as a proxy for sunniness of customers' locale.

5. **The number of employees of the customer:** Most of Supply Velocity Client' customers are privately held businesses. Supply Velocity Client paid a data firm to overlap the number of employees of these companies, from other data sources, onto their customer base. We used this as a proxy for customer size.

The marketing team at Supply Velocity Client wondered if the income and population affected customer profit. The hypothesis would be that companies ordering from areas having higher median incomes and companies ordering from more populated areas purchased more products. We were able to acquire median income and population by zip code from the 2000 census. This data was matched to customer billing zip code.

6. **The median income of the zip code that the customer was located within,**

7. **The population of the zip code that the customer was located within.**

Note: All of the customer location variables were based on customers' billing addresses for orders, rather than on the addresses to which orders were shipped.

These variables were used in single and multiple-variable regression models to predict gross profit per customer.

Data Summary

Our original database had over 74,000 customers, 50,505 of which had employee data. Due to time constraints and the statistical significance of customer employee count in our fitted model, we chose to conduct our analyses using the sample of customers with employee counts. Also to be noted is that one of the customers without an employee count was the armed forces. This one customer had total gross profit over 10 times more than the second highest total gross profit customer. We believe this customer was an outlier that significantly skewed the data, and that removal of this customer was appropriate.

Below are some descriptive statistics on the variables used in this analysis.

- The average gross profit per customer = $1333
- The average orders per customer = 1.57
- There were 284 key customers in our data set of 50,505 customers
- 13,641 customers ordered via the web
- 26,331 customers were located in coastal states
- The least sunny state had sunshine percentage of 37.6%, the sunniest state had sunshine percentage of 84.5%, and the average was 60%
- The fewest employees per customer was 1, the most was 11,000, and the average was 51.9
- The lowest median income for a customer billing zip code of was $2499, the highest was $200,001 and the average was $47,714
- The lowest population for a customer billing zip code was 5 people, the highest was 143,987 and the average was 24,818

Supply Velocity Client' products last a very long time and they warrantee their products for seven years. Therefore, many customers will order one time and not need to replace these products for many years. To capture enough customers and repeat business we therefore used data that was cumulative for 2006 through 2011. In addition, we randomly set aside 10% of the customers to test the predictive power of our models.

Prediction Model Discussion

We created multiple-variable regression models to predict gross profit per customer using two approaches. The first approach, called the direct model predicted gross profit per customer directly from all of the independent variables identified for the study. Five of the seven independent variables proved to be statistically significant. Sunshine percentage and population of the customer's zip code were not statistically significant. Therefore, we removed these variables from the regression model, and used only the remaining five statistically significant independent variables in our model.

In the second approach we used two steps to predict gross profit per customer. First, we fit a model using the same seven independent variables to predict orders per customer, and then we fit a second model to predict gross profit per order. We calculated gross profit per customer by multiplying orders per customer by gross profit per order. All seven independent variables were statistically significant in both models.

Results – Direct Model

The direct gross profit per customer prediction model used five independent variables. The effects of these variables on gross profit per customer will be discussed briefly below. The R-square of this model was 0.0738. This means that the five independent variables predicted 7.38% of the observed variation in profit per customer. Therefore, 92.62% of the variation in gross profit per customer is due to other untested variables. While the predictive value of the model may seem low, we will show why it has good predictive power.

Direct Gross Profit Prediction Equation

Gross profit per customer = 425.66 + 19977(KeyCustomerYes) + 741.93(WebYes) +

332.83(CoastalStateYes) + 1.29(EmployeeCount) + 0.0074(ZipInc.)

The coefficients in the model were all positive. For the binary (yes-no) independent variables (key customer, web-yes, coastal state), when a customer had one of these attributes, all other variables held constant, the average profit of these customers will be higher by the value of the coefficient.

- A key customer, with all other variables held constant, will have an average gross profit $19,977 higher than non key customers
- Customers that use the web, with all other variables held constant, will have an average gross profit $741.94 higher than customers that call in orders
- Customers located in coastal states, with all other variables held constant, will have an average gross profit $332.83 higher than non coastal state customers

For the continuous independent variables (employee count, population of the zip code) an increase of 1 person (employee or resident), with all other variables held constant, will increase the average profit per customer by the value of the coefficient.

- For each customer employee, with all other variables held constant, a customer will on average have gross profit $1.29 higher
- For each $1000 of median income in a customer's zip code, with all other variables held constant, a customer in that zip code will on average have gross profit $7.4 higher

Because the average profit per customer is $1,332, even the median income coefficient of 0.0074 is economically meaningful. A company in a zip code with $10,000 higher median income, will, with all other variables held constant, have on average gross profit $74 higher. This is about a 5% increase in gross profit per customer for every $10,000 in median income of their zip code.

It should also be noted that the individual regressions of the five independent variables to gross profit per customer had correlation coefficients that were also statistically significant and had positive slopes. This is further verification that these independent variables are good predictors of gross profit per customer.

We tested the fitted models' predictive powers by using the estimating equation for the fitted model with the set-aside customer data. These set-aside customers were not used to create the prediction model. Using the values of the five independent variables for the 5,664 set-aside customers we calculated each customer's predicted gross profit. We then sorted these from highest to lowest and grouped them into tranches of 1,000 customers and calculated the average gross profit per customer for each tranche. Note, the last tranche was the bottom 564 customers. The results of the actual versus predicted gross profit per customer is shown below in Table 1 and Figure 1. It is apparent that the direct profit per customer prediction model did a good job, overall. Of the six tranches, four are in the correct order. Only the bottom two tranches should be reversed. In addition, the errors are relatively small.

1000 Customer Tranche	Actual Average Decile Profit per Customer	Predicted Average Decile Profit per Customer	Error
1 - 1000	$2,339	$2,519	7.73%
1001 - 2000	$1,211	$1,489	22.97%
2001 - 3000	$1,165	$1,180	1.30%
3001 - 4000	$1,015	$1,020	0.52%
4001 - 5000	$896	$818	8.70%
5001 - 5664	$917	$661	27.88%

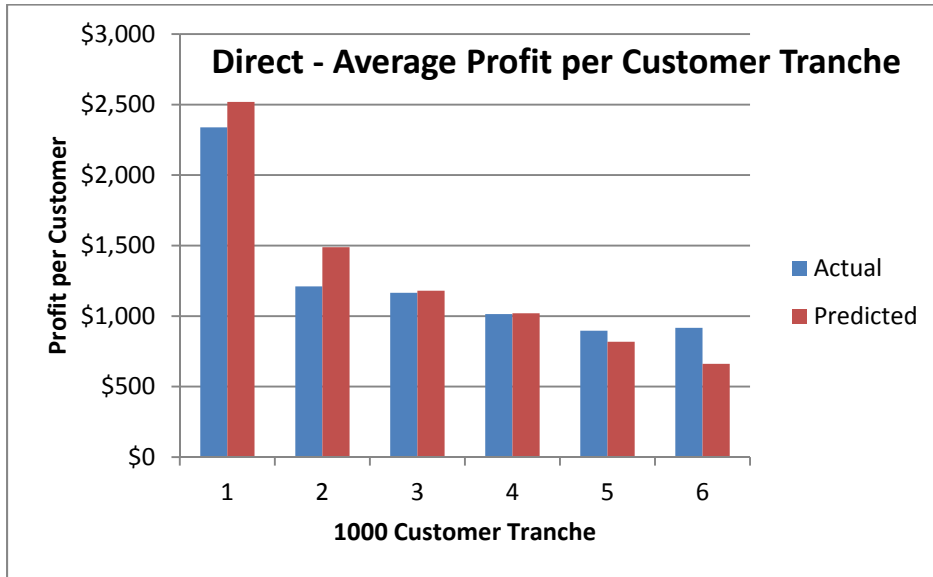**Table 1: Actual versus Predicted Average (over the 1000 customer tranche) Gross Profit per Customer for the Direct Model**

Figure 1: Actual versus Predicted Average (over the 1000 customer tranche) Gross Profit per Customer for the Direct Model

Results – Two Stage Model

Our second approach used two steps to calculate gross profit per customer. The two stage model's overall predictive power was very similar to the direct model. The graph of actual versus predicted gross profit for the 1,000 customer tranches looks very similar. The R-squares for the orders and profit per order multiple variable regression models were slightly lower for these models than the direct model. Therefore we will focus our conclusions on the direct model.

1000 Customer Tranche	Actual Average Decile Profit per Customer	Predicted Average Decile Profit per Customer	Error
1 - 1000	$2,374	$2,601	9.57%
1001 - 2000	$1,304	$1,159	11.13%
2001 - 3000	$1,065	$991	6.94%
3001 - 4000	$1,079	$892	17.34%
4001 - 5000	$830	$746	10.07%
5001 - 5664	$878	$643	26.70%

Table 2: Actual versus Predicted Average (over the 1000 customer tranche) Gross Profit per Customer for the Two Stage Model
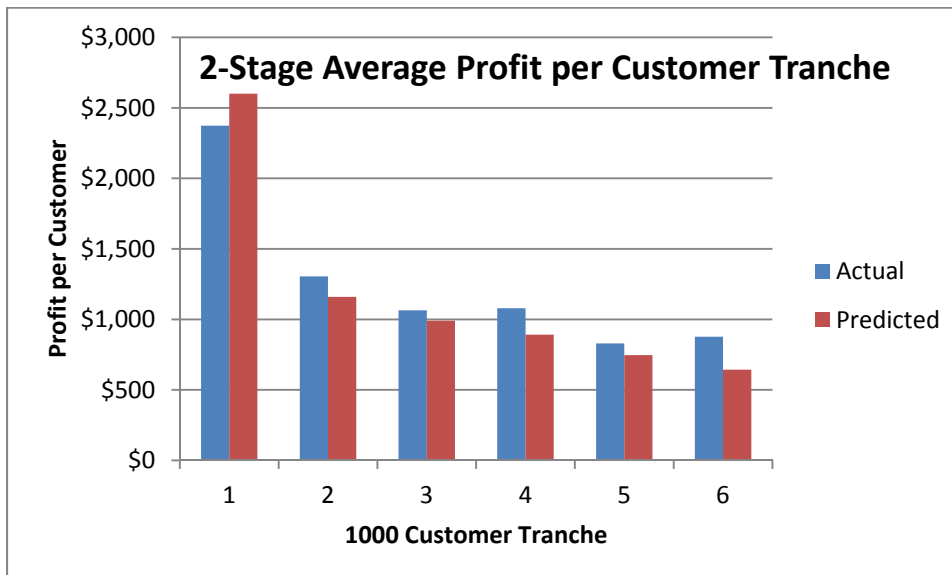

Figure 2: Actual versus Predicted Average (over the 1000 customer tranche) Gross Profit per Customer for the Two Stage Model

Conclusion

This analysis began with Supply Velocity Client asking if the seven independent variables could predict profitable customers. This information would be used to rent direct mail lists and choose customers and prospects to receive a mailed catalogue; and to determine which customers and prospects may get assigned to a senior sales representative. Because of the statistical significance of the prediction model, the significance of the five independent variables and the test of the model versus actual profit per customer, it is our recommendation that Supply Velocity Client should use these five variables (key customer, customers that purchase on the web, customers that are located in coastal states, greater number employees at a customer and greater median income of the customers' billing zip code) to select where to expend marketing and sales resources. Each variable, with all other variables held constant, predicts a more profitable customer.

Possible Next Steps

Due to time constraints our team gathered geographic data in the most expeditious manner possible. Our population and income data was taken from the 2000 US census. The 2010 US census data was not readily accessible at the time of the study. We believe it would be beneficial to revisit this analysis using the latest US census data once it is generally available in a downloadable dataset. The designation of "coastal" and sunshine percent was based on the customer billing state. Sunshine percent was averaged from national oceanic administration data for cities within these states. Some customers had multiple billing addresses. We manually had to aggregate the data for these customers into the state/zip that appeared to have the greatest activity. We believe further work to find the appropriate geographic area to use for these variables, would improve the reliability of the model. Choices include zip code, county, metropolitan statistical area (MSA) and state. Further, several larger customers shipped to multiple

states, some of which were geographically dispersed from the billing state.   We believe further work to assess the viability of shipping state and zip code as predictors is appropriate.

As stated above the direct profit per customer prediction model predicted only 7.38% of the variation of profit per customer. This indicates there are other variables that account for profit per customer. Additional discussion, using a cause and effect diagram, could discover additional potential independent variables. If data was readily available for these variables they could be added to the prediction model to determine if the variables improve the model's predictive power.

There is an important independent variable that was not included in this analysis, which likely helps predict gross profit per customer. That variable is the number of catalogues mailed to each customer. Given more time we would recommend adding this variable to the prediction model.

Some of the independent variables were correlated with each other. This may have caused hidden nonlinearities and may require further investigation.

Finally, we find the results for Web-Yes intriguing. Customers that buy on the web versus call in orders are on average, and with all other variables held constant, more profitable. We wondered if Supply Velocity Client has suggested selling hints on their web site that drive this behavior. Since web-ordering customers are, all other variables held constant, more profitable, perhaps we can use suggested selling to increase sales and profit even more. In addition, could the company incent its customers to buy on the web, where the firm may have an improved opportunity to sell its customers more products using cross-selling techniques? Perhaps a coupon would provide that incentive.   We recommend that such activities, if pursued, should include a study component to better understand the possible causal elements that lead to positive customer response and improved profitability.