Knowledge Wave International, inc

See your business like never before

www.knowledge-wave.com
info@knowledge-wave.com
800.329.9159

HOW TO BUILD A PREDICTIVE MODEL WITH CUSTOMER ANALYTICS.

A Knowledge Wave International Executive Guide – May 2010

Why do we try to predict what customers will do?

Most Voice of the Customer (VOC) programs strive to answer questions like …

- *What do our customers like?*
- *What do they not like?*
- *What is most important to them?*

Some VOC programs go to the next level and include both opinion data *and* operational / behavioral data. This expands the scope of understanding beyond what the customers are thinking and includes who the customers are and how they impact your business. This allows you to address the above questions in a more robust and targeted manner:

- *What do our most-profitable customers like?*
- *What do those customers who are close renewing their maintenance agreement not like?*
- *What is most important to customers in the markets most-prized by our competitors?*

The more narrowly you target your actions, the faster and easier it is to successfully drive improvements.

Applying basic significance analyses (e.g. t-test, u-test, ANOVA) to these differences enables you to gain a better understanding of your strengths and weaknesses accounting for the variability in scores from statistical sampling. This increases your confidence in taking action and helping to ensure that your investments are targeted at the most-profitable returns representing true differences in customer opinions.

But the real value in gathering customer feedback and tracking it to actual customer behavior lies in the ability to create a model that helps you predict future customer behavior. This whitepaper will provide an overview of the process and science around building a predictive model for your business.

Isn't this just so much magic?

Trying to predict the future of customer behavior is nothing new. Firms have always wanted to know how much inventory to stock, how many staff to schedule, and (not least importantly) how much revenue to expect. Using statistical analysis tools to make these predictions is old hat for most marketers. The laws of statistical sampling provide an almost miraculous ability to extrapolate the answers of a few to the actions of an entire market.

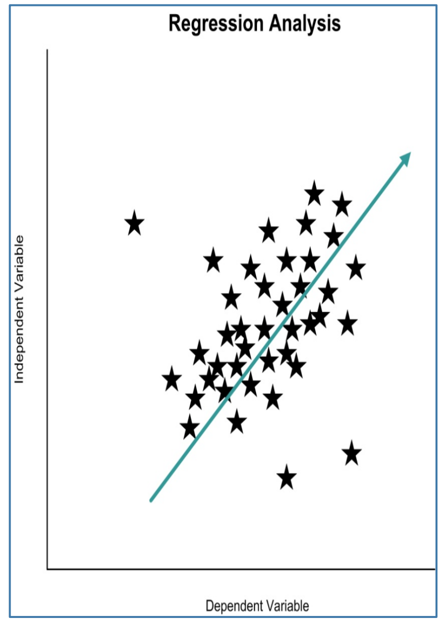
Regression analysis is the basis for these predictions and over the years more and more complicated versions of regression analysis have been developed in order to take into account the truly complex relationships between customers and companies. A detailed explanation statistical tools is outside the scope of this briefing but a brief explanation will help to set the framework.

Understanding regression analysis starts with defining two types of variables: Dependent and independent. The variable you are looking to predict is called the *dependent* variable because its predicted value *depends* on the values of the other variables. Those are called the *independent* variables.

An easy example is overall satisfaction with a product … let's use a new coat as an example.

There are a number of features of your new coat that contribute to how satisfied you are with it (the *dependent* variable in our example): Fabric, weight, color, style, how well it keeps rain out, whether other people have the same coat, price, how much you paid for shipping or how far you had to drive, and so on. All of these *independent* variables impact (to one extent or another) your overall satisfaction with the coat. We care because your overall satisfaction with the coat will impact future revenue: Will you buy other coats (either for yourself or others) from the same company? Will you tell friends and family how much you like the coat?

You start your regression analysis by observing data today and, for all intents and purposes, plotting them along a graph and trying to draw a line that best matches the fit of the data … for a simple regression with just one independent variable, the relationship is very easy to visualize (see the chart to the right).



The slope of the line basically represents the relationship between the two variables, discounted by how far the average data point falls from the line. Regressions involving multiple variables or relationships that are not linear are a little harder to describe but the basic concept is the same … they just involve many more lines and slopes and relationships. For multivariate regressions, the independent variables will be described with a *coefficient*, which indicates the strength of the relationship between that variable and the dependent variable.

For example, if you look at overall satisfaction with a meal in a restaurant (which contributes in part to whether or not you eat there again in the near future), the variables of interest will be things like ambiance, whether or not the menu options fit your taste, how friendly / professional / helpful the wait staff were, and so on.

Regression analysis may show you that the *coefficients* for these variables and their impact on overall satisfaction look something like this:

- (.3789) Ambiance
- (.1620) Wait time for a table
- (.0549) Healthy food options
- (.5629) Wine list
- (.2772) Server's attentiveness

From this, we would expect that changes in the wine list and ambiance would have a greater impact on overall satisfaction than having healthy food options or a shorter wait time.

What we really care about, however, is not just increasing the customer's stated level of satisfaction but improving the kinds of actual customer behavior that benefits our business. Does the customer actually come back to the restaurant in the next three months? Does the customer actually give the coupon for a free dessert to a friend? Do our actions actually improve our business by creating a loyal and engaged customer?

The good news is that most statistical software packages can handle large numbers of variables and large data sets. The bad news is that you may not be able to get some of this data even if you wanted to. It may not be in your system, you may not have thought to ask the customer, or it may not be reliable.

This speaks to a real problem that you have to consider when using the results of the regression analysis: Did you include all of the important variables? You may have noticed a glaring omission in the above list of variables for our restaurant: Price. With few exceptions, how much I pay for a restaurant experience has an impact on whether or not I return and how frequently.

There are potentially hundreds of other variables, which influence my perception of the value of the meal: Who was the sous-chef? Were the vegetables bought at the farmer's market or delivered in bulk? Was I there for an anniversary dinner or a first date?

Or consider the earlier example of a new coat. There may be absolutely nothing wrong with the coat but if I purchased a lighter-weight garment and the winter turns out to be heavier than expected, I may not use it as much as the manufacturer might hope. And that may impact overall satisfaction. Or maybe I buy a heavy winter coat but then by a turn-in-fortune get transferred to a new position in Florida or Texas and so have no need for a heavy overcoat. Again, non-use may impact satisfaction and both are outside of the company's control or even knowledge.

Of course, there are a number of other biases (applicable to any research or statistical project) and risks that you should consider and should be acknowledged and managed during the program design and implementation: Is the questionnaire designed correctly to provide valid and reliable feedback? Is the sample size unbiased? Is the operational data I am pulling from my systems accurate? Consistent? Timely?

That's why you will never look at your predictive model as a cold-hard number. Rather it will represent a range and the unknown and influencing factors should be called out explicitly.

Once you have your regression model, you then test it by applying the formula to a known data set and measuring how far the predicted values fall from the actual values. How well the model fits actual data determines its "goodness of fit" and is an indication of how confident you can be when taking action. We often use R-squared in our predictive models but other common tools are Root Mean Squared Error and Mean Absolute Error.

Two final cautions:

The customer's perceptions of your product and service are different for different types of customers, so be careful not build a regression model that is too broad. An easy example is cultural: Customers in Europe will have different expectations and respond to different situations … well, *differently* than customers in Asia or the United States. To continue the restaurant metaphor, a family with three small children will have different expectations about dining out than a couple on their anniversary or a group of high school friends re-uniting after 15 years.

Finally, it is important to remember that an analysis performed six months ago may (or may not) be applicable today. Or in another six months. You customers change, the market changes, your offerings change, technology changes, competitors change … the world is in a constant state of flux. Which is why it is important to continually update and test your predictive models. Again, the speed and power of today's statistical programs will help support this process fairly easily.

Once I have a model, what can I do with it?

GENERAL IMPLICATIONS

We all know why most companies focus primarily on their existing customer base: Earning repeat business or up-selling additional products and services is by far the easiest and most profitable sale for any company. Further, understanding the customer base is key to maintaining the firm's niche in the market it fought so hard to have earned in the first place. You know what incremental features you need to add and what revolutionary, unmet needs are out there among the people who already trust you.

But just surveying your customers isn't enough. Once you understand what your customer base says they will do, you compare that against what they actually do. For example: If 80% of new customers *say* they will buy from you again (by selecting 7 out of 7 on a standard Likert scale) in the next three months but only 50% actually follow-through, you now know four things:

The first is that you can take a shot at predicting revenue from repeat customers over the next three months. If 1,000 new customers this month say they will purchase from you again in the future, you can estimate that 500 actually will. If your average sale is $250, then you can predict $125,000 over three months from those repeat purchasers.

If you factor in the probable repurchasing of those who answer 6 and 5 on the 7-point scale, the number grows to $152,000:

Response	Actual additional purchase rates	Actual additional purchases	Average Purchase Price	Additional Income
7	50%	500	$250	$125,000
6	10%	100	$250	$25,000
5	1%	10	$250	$2,500

The second thing you know is how much more revenue you could expect to earn if you could increase the actual additional purchase rates by just a few percentage points. Maintaining the same hypothetical survey results nets $182,500 in additional revenue merely by growing th number of responders who select 7 / 7:

Response	Percentage of Customers	Actual additional purchases	Average Purchase Price	Additional Income
7	55%	550	$250	$137,500
6	15%	150	$250	$37,500
5	3%	30	$250	$7,500

The third thing you know (or, at least, can analyze) are the differences between customers who say they will purchase but do not and those that do. This understanding of what distracts the customer from your brand and message can help you adjust and adapt to convince the customer to purchase from you again. Are those customers more sensitive to price or timeliness? Do they place a higher premium on a feature that you haven't developed yet? Or, do they simply not need your products or services again? If that is the case, you can avoid wasting resources on unfruitful customer acquisition strategies.

Finally, you know the specifics about the customers who are answering 7 versus 6, 6 versus 5, and so on. This allows you to do build profiles of customers who are most likely to repurchase and those that are not. In effect answering the question: How do we move 6s up to 7s? This has implications for internal policies and procedures, external marketing messages and offers, and coaching of sales and service staff. This also allows you to treat each individual customer uniquely and craft a message to which he or she will be most receptive.

Of course, "Will you buy from us again?" is not the only question you will measure that has an impact on future sales. "Would you recommend us to a friend or colleague?" can also be a predictor of both the customer's referability and the customer's own willingness to spend money with your company in the future. And "Would you consider purchasing other products from us?" gives you a measure of how much share of wallet you can hope to earn from the customer.

Understanding that there is a direct, though discounted, line between the customer's stated intentions and actual behavior, the next step is to understand how you can improve the odds of the customer actually following through and investing more money with your firm.

SPECIFIC ISSUES

How you increase the percentage of customers who say they will purchase from you again is a matter of shaping the customer's experience to increase the perceived value your company provides. Value is a complicated notion and the quality of the product or service is only a part of the equation. Your staff, the cost, how closely the product or service meets the customer's actual needs, how easy it is to do business with you all factor into the emotional decision of how willing the customer is to do more business with your firm. These are the Key Performance Indicators (KPIs) you measure both through surveys *and* operational metrics.

To further complicate the matter, they will be different for different customer experiences (sales versus service, return versus order, cancellation versus renewal). You also have different types of customers, or customer profiles. This is where the "lots and lots of data" comes into play. Some customers may interact with you only through your web site. Others may order over the phone but use your on-line communities for support issues. Still others may call your support line on a weekly basis. You have customer demographics (including personal, psychographic, and level of sophistication with your offerings), products or services purchased, competitive profile, internal team or individual the customer works with, predicted value to your company … there are literally hundreds if not thousands of different variables you can consider in creating customer profiles.

You cannot focus on everything so you have to choose the drivers that are the most important to customer satisfaction and have the lowest level of satisfaction. There are a number of ways to determine the level of importance for any individual KPI but regression analysis linked to loyalty measures and actual customer behaviors is one of the most powerful.

One important profile to consider is where the customer is in the customer lifecycle. Did you just complete the sale? Has the customer been out of training for six months? Is the product past the end of its predicted life span? Is the customer's warranty up for renewal in 90 days? Building your models to allow for these different states of customer need enables you to design very narrow, highly-targeted plans for improving customer satisfaction.

Every company will have a different set of drivers and profiles, depending on the market in general and the firm's differentiation strategy in particular. And the value of the early-stage predictions may be less important for longer-cycle purchases (consider purchasing a tractor versus an iPod versus your weekly groceries). But it is important to understand the points at which the level of engagement begin to fall … that is the ideal time to step-in and fight to keep the customer enthused about your offering.

How about a couple real-life examples?

Assume that you have two questions in your transactional service satisfaction survey that you have identified to be strongly related: "I waited a reasonable amount of time on hold" and "I am likely to make another purchase from Company in the next three months". Assume further that the average scores on these two questions are (on a 10-point scale) 7.50 and 6.85 respectively.

By using regression analysis, you can determine the coefficient describing the relationship between these two variables: .78. In other words, an increase in satisfaction with hold time of 1 point will lead to an increase for intent to purchase again of .78 points.

Using this coefficient, you can predict that a move from 7.50 to 8.50 for the independent variable ("I waited a reasonable amount of time on hold") would correspond to a move of 6.85 to 7.63 for the dependent variable ("I am likely to make another purchase from Company in the next three months").

Assume further that you have been monitoring actual customer purchase behavior and identified that there is a 35% likelihood that those customers who say they will make a purchase in the next three months actually do. By moving the score of "I am likely to make another purchase from Company in the next three months" from 6.85 to 7.63, you increase the number of probable purchasers from 35 / 100 to approximately 42 / 100 or 20%.

If your average monthly income is $10,000,000, then, a 20% increase in sales (or $2,000,000) is not insignificant.

However, you have to look at the investment required to increase satisfaction with hold time by 1 point in order to determine the probable ROI of such a change.

In other words: Is it worth the investment?

To demonstrate the point, consider two scenarios around current actual average hold time:

In the first scenario, 90% of callers are on-hold for three minutes or less. Based on your analysis of the relationship between actual hold time and customer satisfaction with hold time, you would have to move that number to 90% of callers being on hold for less than 90 seconds. If you calculate the number of agents you would have to add, train, and deploy; infrastructure you would have to build; and additional monitoring systems you would have to implement, you could be looking at an investment of $35,000,000 over 18 months before you would even start to see an impact on increased revenue. A three-year turn around for ROI in today's fast-based, highly competitive market may be a hard sell.

In scenario 2: 80% of your customers are on hold for less than seven minutes. The investment and time to improve that metric to 80% for less than 5 minutes might be only $2,000,000 over six months, making the ROI much more attractive.

The reality is that it is never just one driver which impacts customer satisfaction, loyalty, and, by extension, downstream revenue. The good news is that predictive analytics can easily take into account hundreds of variables quickly and easily. The better news is that more refined analysis of the data can help you to target more actionable change programs. In one study KWI conducted, the organization first targeted decreased hold times for VIP customers (identified as having the company's platinum rewards card) and eventually refined the program even further to minimize hold time for VIP customers who had made a purchase in the past week, since those customers were most likely calling about a problem with the order.

Two very small investments in infrastructure and process led to very dramatic improvements in customer satisfaction among the company's most valuable clientele and, eventually, share of wallet the company was earning from those customers.

About Knowledge Wave International

YOU'VE NEVER SEEN YOUR BUSINESS LIKE THIS BEFORE!

Founded in 1999, KWI ("*kee-wee*") provides innovative customer loyalty and employee engagement programs to clients in almost every market and industry. Our more than 250 clients drive business growth off of feedback programs in more than 35 languages, across all customer touchpoints. Our proprietary feedback platform provides the basis of a scientifically rigorous and effective research framework, where the philosophy is hard data and individual results rather than generic theory. Customer advocacy is different for every business and even for different types of customers within a business … KWI helps you to understand and take advantage of the unique needs and expectations of your different customers quickly, effectively, and quantitatively.

To learn more about KWI, please visit us at http://www.knowledge-wave.com or call 1.800.329.9159 or email info@knowledge-wave.com.